



Evaluation of proposed modes of binding of (2*S*)-2-[4-[(3*S*)-1-acetimidoyl-3-pyrrolidinyl]oxy]phenyl]-3-(7-amidino-2-naphthyl)propanoic acid hydrochloride and some analogs to Factor Xa using a comparative molecular field analysis

Roy J. Vaz*, Larry R. McLean & John T. Pelton

Hoechst Marion Roussel Inc., 2110 E. Galbraith Road, Cincinnati, OH 45215, U.S.A.

Received 11 April 1997; Accepted 11 September 1997

Key words: CoMFA, DX-9065a, factor Xa, systematic search

Summary

The binding mode of (2*S*)-2-[4-[(3*S*)-1-acetimidoyl-3-pyrrolidinyl]oxy]phenyl]-3-(7-amidino-2-naphthyl)propanoic acid hydrochloride (DX-9065a, **4**) to Factor Xa is examined using inhibition data for a series of analogs that have a hydrophobic group as well as basic or dibasic functionality. Comparative molecular field analysis is utilized on a series of DX-9065a analogs in a series of proposed alternative binding modes. A quantitative measure is provided that distinguishes between the proposed binding modes that describes 'how well' the binding mode explains the structure–activity relationship or the best 3D QSAR agrees with the crystallographically determined binding mode. The best model is in agreement with recently available data [Brandstetter et al., *J. Biol. Chem.*, 271 (1996) 29988]. The highest statistical correlation occurs with the second basic group accommodated in the vicinity of Glu⁹⁷ and a hydrophobic group accommodated in the pocket defined by Phe¹⁷⁴, Tyr⁹⁹ and Trp²¹⁵. Also, the best model arises when the conformation of the Glu⁹⁷ side chain is modified such that an H-bond interaction is maintained with the inhibitor if possible. The model also shows a tightening of the S1 pocket as is shown in the recent data described above.

Introduction

General background

The serine proteases represent a large class of closely related enzymes that employ a catalytic triad of Asp, His and Ser residues. They play a critical role in a variety of biologically significant processes, including hemostasis [1,2]. This intricately controlled system of enzymatic reactions and interfacial processes maintains blood in a fluid state under physiological conditions and allows rapid clotting in response to injury [4]. Recently, attention has been focused on the penultimate enzyme in the blood coagulation cascade [3], Factor Xa. This serine protease generates thrombin by limited proteolysis of prothrombin in a membrane associated complex with Factor Va and calcium. Its position within the coagulation cascade points to a

regulatory role for the enzyme by which the levels of active thrombin are established. Unlike thrombin itself, Factor Xa does not appear to play a direct role in the hemostatic mechanisms of clot formation and platelet aggregation [4].

Factor Xa has a γ -carboxyglutamic acid domain responsible for calcium and phospholipid binding and two epidermal growth factor (EGF)-like domains, the second of which has been suggested to mediate Factor VIIIa/Va binding. In the crystal structure of Factor Xa [5], Arg⁴³⁹ forms a salt bridge interaction with Asp¹⁸⁹ in the active site of a neighboring molecule. This interaction stabilizes the crystal to such an extent that complexes of Factor Xa with small-molecule inhibitors are difficult to crystallize. For this reason, an alternate method of determining the mode of binding of the inhibitors to Factor Xa was required. In this paper we report a molecular modeling approach em-

* To whom correspondence should be addressed.



ploying X-ray crystal data of Factor Xa and related serine proteases.

Both small-molecule and peptide inhibitors of Factor Xa have been identified. For the most part, high-affinity inhibitors directly block the active site of the enzyme. For example, the anticoagulant-antimetastatic leech protein ghlanten blocks two inhibitory domains in Factor Xa. The N-terminal domain of the inhibitor binds to the active site of the enzyme and the C-terminal domain interferes with the heparin binding region that may mediate interactions with cell-surface proteoglycans [6]. Small-molecule inhibitors have been based either on peptide aldehydes related to D-Phe-Pro-Arg-chloromethyl ketone (PPACK) or arylbenzamides.

A number of binding modes have been proposed for series of arylbenzamide inhibitors first described by Nagahara et al. [10]. However, no quantitative measure of the agreement between a set of proposed binding modes and the activity data has been reported. Three binding conformations have been proposed for the Daichi inhibitors [7–10]. The most recent is that proposed by Stubbs et al. [9]. They report X-ray crystallographic work on the complexes of two of the Daichi Factor Xa inhibitors, DX-9065a and BX-5633 with bovine β -trypsin, which is similar structurally to Factor Xa.

The key features of the Stubbs' model are (i) a salt bridge interaction between the amidino groups of the inhibitors and the Asp¹⁸⁹ of the bovine β -trypsin; (ii) a marginal edge-on interaction between the benzyl spacer of the inhibitors and the Tyr⁹⁹ ring, although this probably takes the form of van der Waals interactions; and (iii) an interaction between the distal positive charge from the pyrrolidinyl-acetimido group with the electronegative cavity formed near the vicinity of the backbone of Glu⁹⁷ and residues Thr⁹⁸ and Ile¹⁷⁵.

A second model by Lin and Johnson [8] proposes a cation– π interaction that stabilizes the distal positive charge of the pyrrolidinyl-acetimido group and the pocket formed by Trp²¹⁵, Phe¹⁷⁴ and Tyr⁹⁹. This model is similar to that proposed by Stubbs et al. [9]. Finally, while the model proposed by Katakura et al. [7] has a similar basic interaction with the Asp¹⁸⁹, the propionic group points towards Gly²¹⁹ and the acetimido group is closer to the carboxylic group in the side chain of Glu⁹⁷. That model has the benzene ring parallel to the planes of the benzene ring of Tyr⁹⁹ and Phe¹⁷⁴.

Only a limited number of compounds have been modeled in Factor Xa. In an effort to better define the mode of interaction as a first step in improving specificity, while retaining high affinity, we have employed a combination of molecular modeling and quantitative structure–activity relationships in three dimensions (3D QSARs). Previously [11], a statistically significant model was obtained for Rhinovirus14 inhibitors using Comparative Molecular Field Analysis (CoMFA). In this study, the alignment of the inhibitors was obtained from crystallographic data of these inhibitors. Also, alignments [12,13] derived from crystallographic information have produced statistically significant CoMFAs. The aim here is to determine if the binding mode which has the most statistically significant 3D QSAR model agrees with the experimentally determined binding mode. The experimentally determined binding mode for this study has just recently become available and hence the validation can be done. The basis set for this study consists of arylbenzamide inhibitors (Table 1) [7,10]. Inhibitors were docked into the crystal structures of the enzyme, the conformations were optimized together with a regional portion of the enzyme, and CoMFA was performed. This approach has allowed us to identify features that can be exploited in the design of new Factor Xa inhibitors with increased affinity and specificity.

Generally, the electrostatic fields have not been found to contribute significantly to CoMFA analysis or have been difficult to interpret [14]. However, the series of inhibitors in the current study have electrostatic coefficients that contribute substantially to the 3D QSAR and can be easily interpreted.

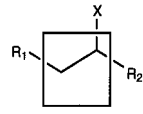
Theoretical

CoMFA [14,15] has become an increasingly common method for 3D QSAR studies. CoMFA analysis is based on the premise that the interaction of an inhibitor with an enzyme may be defined operationally in terms of a sampling of the steric and electrostatic fields surrounding a set of suitable inhibitors.

The method is primarily a way of 'translating' 3D information into 1D information and then back into 3D information. The technique can be described as consisting of the following steps:

(1) Establishing the conformation for each molecule in a series of analogs for which the property of interest is known for every molecule in the set. This property is usually some biological activity but could

Table 1. List of DX-9065a analogs used in the study



Primary
X= H, COOH or COOEt

Structure	R ₁ group	Primary	R ₂ group	IC ₅₀ (FXa) (μM)	IC ₅₀ (thrombin) (μM)
1	5-Am-2-BF	Ethane	5-Am-2-BF	0.1	5
4 (DX-9065a)	7-Am-2-NP	(S)-Propanoic acid	Ac-Pyr	0.07	>2000
1	5-Am-2-BF	Ethane	Pyr	0.94	100
14a	5-Am-2-BF	Ethyl propanoate	Pyr	1	21
14b	6-Am-2-BF	Ethyl propanoate	Pyr	1.7	36
14c	6-Am-2-NP	Ethyl propanoate	Pyr	4.2	38
14d	7-Am-2-NP	Ethyl propanoate	Pyr	0.31	>600
15a	5-Am-2-BF	Propanoic acid	Pyr	1.5	230
15b	6-Am-2-BF	Propanoic acid	Pyr	1.7	>1000
15c	6-Am-2-NP	Propanoic acid	Pyr	4.1	>100
15d	7-Am-2-NP	Propanoic acid	Pyr	0.2	>1600
16	5-Am-2-BF	Ethane	Ac-Pyr	0.64	4
17a	5-Am-2-BF	Propanoic acid	Ac-Pyr	0.5	15
17b	6-Am-2-BF	Propanoic acid	Ac-Pyr	0.84	>400
17c	6-Am-2-NP	Propanoic acid	Ac-Pyr	0.36	22
17d	7-Am-2-NP	(R)-Propanoic acid	Ac-Pyr	0.5	>1000
25	5-Am-2-BF	Ethyl propanoate	Pyr	52	20
26	5-Am-2-BF	Propanoic acid	Pyr	45	9
BX-5633	7-Am-2-NP	(S)-Propanoic acid	Ac-Pip	0.013	>400
7a	7-Am-2-NP	Propanoic acid	Phc	4.2	>1000
7b	7-Am-2-NP	Propanoic acid	Phn	7.3	>1000
7c	7-Am-2-NP	Propanoic acid	Pho	35	>1000
7d	7-Am-2-NP	Propanoic acid	Phe	26	>1000

5-Am-2-BF: 5-amidino-2-benzofuranyl; 6-Am-2-BF: 6-amidino-2-benzofuranyl; 6-Am-2-NP: 6-amidino-2-naphthyl; 7-Am-2-NP: 7-amidino-2-naphthyl; Pyr: [4-((3S)-3-pyrrolidinyl)oxy]phenyl; Ac-Pyr: [4-((3S)-1-acetimidoyl-3-pyrrolidinyl)oxy]phenyl; Ac-Pip: [4-((-1-acetimidoyl-4-piperidinyl)oxy]phenyl; Phc: [4-cyclopentyl]phenyl; Phn: [4-2-arninoethyl]phenyl; Pho: [4-methoxy]phenyl; Phe: [4-hydroxy]phenyl; All structures have the R₂ group at the 2 position of the propanoic acid except 25 and 26, which have the R₁ group at the 2 position.

correspond to any property for which a QSAR is to be established.

(2) Superimposing the molecules in 3D space.

(3) Defining a region rectangular in shape, encompassing all the superimposed structures in a 1 or 2 Å grid. Probes are placed at every grid point and the energy of interaction of the probe with each of the molecules in the region is calculated. The most common probes are an sp³ carbon without any charge and a proton with a +1 charge. The energy of interaction is given by:

$$E_{\text{interaction}} = \sum (E_{\text{van der Waals}} + E_{\text{Coulombic}}). \quad (1)$$

The sum is over all atom pairs, one of the atom pairs being the probe atom, the second the atoms in the test molecule within a certain radius of the probe (usually 8 Å). The van der Waals terms take the form

$$E_{\text{van der Waals}} = A/r^{12} - B/r^6, \quad (2)$$

where A and B are functions of the two atoms that are considered. This is the only term that contributes to the energy of the sp³ carbon probe interacting with the

molecule. The Coulombic term accounts for the proton probe interaction with the molecules:

$$E_{\text{Coulombic}} = q_1 q_2 / \epsilon r \quad (3)$$

where q_1 is the charge on the probe (+1), q_2 is the charge on an atom in the molecule, ϵ is the dielectric constant for the media in which the atoms are considered, and r is the distance between the two atoms.

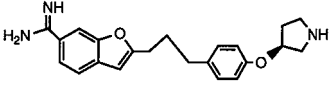
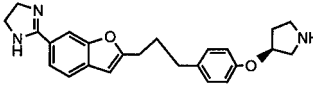
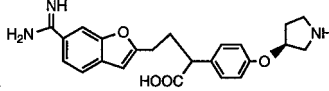
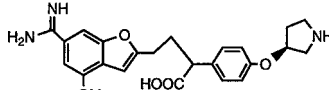
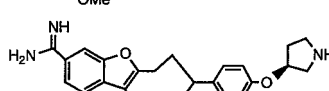
(4) Establishing a table with the rows corresponding to each molecule and columns corresponding to the interaction energies of the probe at each grid point. The first column will hold the value of the property discussed in (1) above.

(5) Performing a multivariate statistical analysis, such as partial least squares (PLS), correlating the probe energies and the property of interest. A stress test such as cross-validation is used for determining the optimal number of factors or components. The correlation is then repeated with the optimal number of components, if the stress test proves statistically valid. The validity [16] is based on the cross-validated r^2 or q^2 (>0.4) as well as a low number of optimal components. (Cross-validated r^2 or $q^2 = 1 - \text{SSE}/\text{SSK}$ where $\text{SSE} = \sum_y (Y_{\text{PRED}} - Y_{\text{ACTUAL}})^2 = \text{PRESS}$ and $\text{SSD} = \sum_y (Y_{\text{ACTUAL}} - Y_{\text{MEAN}})^2$, standard error of prediction = $[\text{PRESS}/(n - c - 1)]^{1/2}$.)

(6) The end result is a set of QSAR coefficients at every grid point, corresponding to the columns of the interaction energy table. Normally the QSAR coefficients themselves cannot be compared since each column has a different standard deviation value. Hence, for purposes of direct comparison, they are normalized by multiplying the coefficients by the standard deviation of the column from which they are derived. In a CoMFA study, a histogram of the QSAR coefficient multiplied by the standard deviation of the corresponding column is generated. A contour plot is then generated for values at the extremities of the histogram. These contours are termed 'coefficient contours'. Modifying the molecules to either increase steric interactions and/or charge interactions in these contour regions would enhance the activity. Agreement with the current data set is checked by placing molecules with high and low activities in the alignment used in the CoMFA to see if they have sections in the positive and/or negative regions, respectively.

Docked conformations are usually scored [17–19] by taking into account potentials including electrostatics. Here CoMFA was measured for the 23 inhibitors

Table 2. Structures that were utilized to 'anchor' structure 4, in a salt bridge interaction with Asp¹⁸⁹ in factor Xa

Inhibitor	IC ₅₀ (μM)
	0.94
	280
	1.0
	430
	22

in Table 1, with an alignment generated from different binding conformations and used to score the docked conformations.

Experimental

Crystal and other structures

The crystal structures of thrombin (pdb code: 1bbr), trypsin (pdb code: 1gbr) and Factor Xa (pdb code: 1hcg) were obtained from the Brookhaven Databank [20]. The numbering of residues follows the PDB convention of using the chymotrypsin equivalents. The Factor Xa coordinates were read into SYBYL6.22 [21], hydrogens were added and charges from AMBER [22] were loaded. The water molecules were deleted and the whole molecule was subjected to a minimization using the Tripos52 force field [23] to optimize the placement of the hydrogen atoms. In this optimization, all non-hydrogen atoms were kept fixed. This is the Factor Xa structure that is used, except where noted below. The thrombin crystal structure was used directly. The guanidinobenzoyl moiety of the bovine β -trypsin structure was deleted, hydrogens were added, AMBER [22] charges were calculated and the structure was optimized keeping all non-hydrogen atoms fixed. The three crystal structures were overlaid using the backbone of the structurally

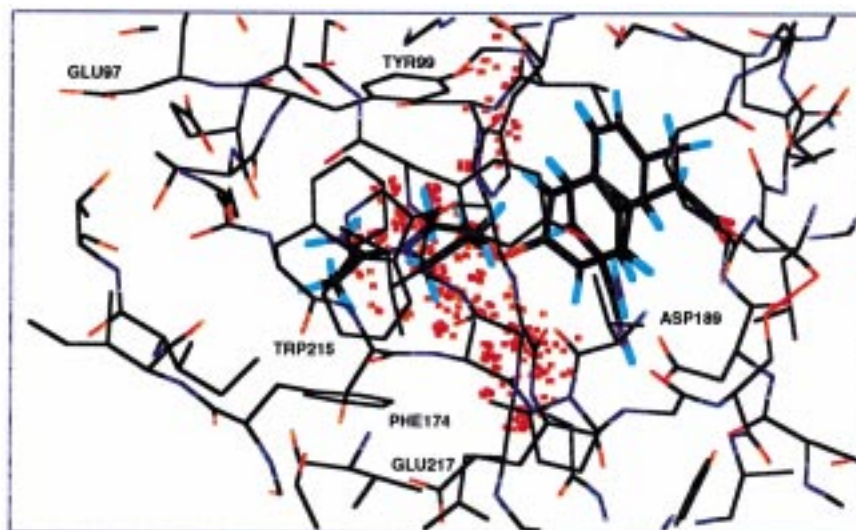


Figure 1. A trace of the carbon of the acetimidoyl groups across all conformations obtained from a conformational search performed on DX-9065a in the Factor Xa enzyme.

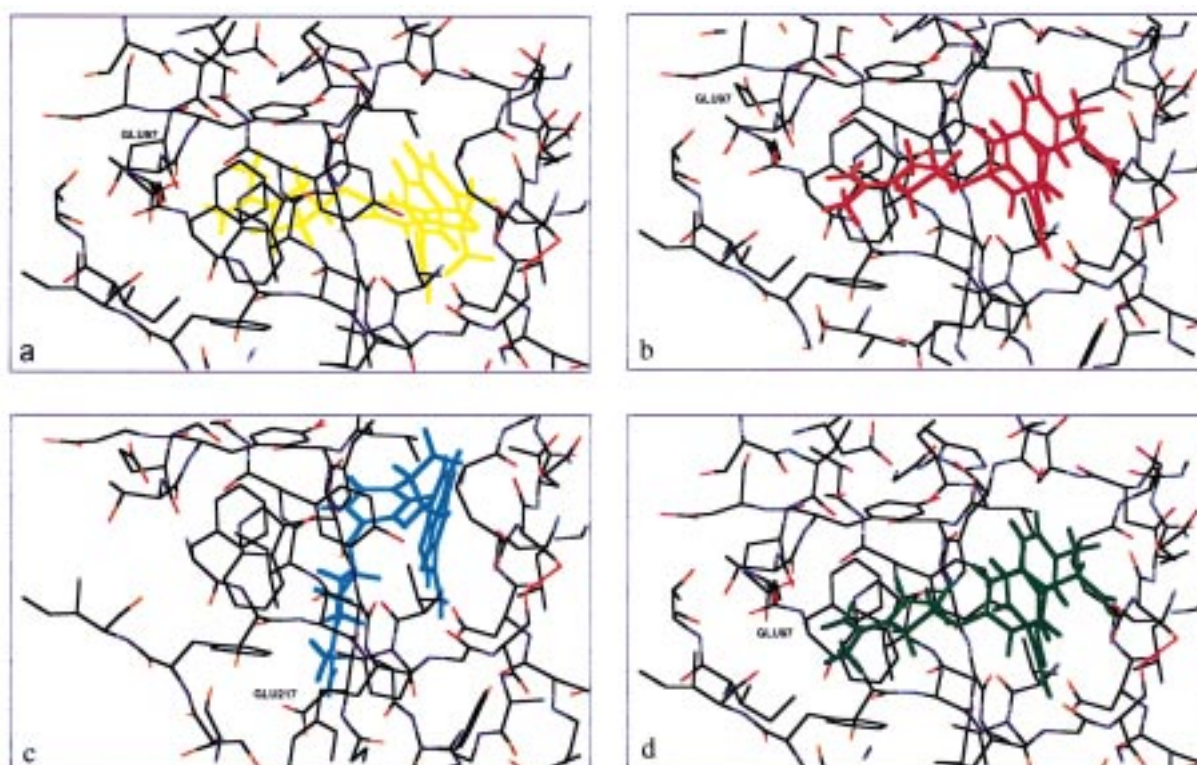


Figure 2. Structure 4 shown in the proposed binding mode represented by (a) E97S; (b) E97B; (c) E217; and (d) E97BS (same as E97B but the Glu⁹⁷ side chain is altered for contact).

Table 3. Hits from the CCD that have a salt bridge interaction between a amidino attached to a phenyl and a carboxyl group together with the dihedral angle between the phenyl and amidino groups

Ref. code in CCD	Angle between planes of amidino and phenyl
CONYAF10	20.88
FOYMAH	36.01
GADMUT	33.46
KABHIE	26.01

conserved regions as defined by the aligned sequences of Greer [24,25]. The molecules in Table 1 were subjected to a single point calculation using the AM1 Hamiltonian in MOPAC5 [26] to obtain electrostatic point charges for the atoms. When present, it was assumed that the amidino and acetimidoyl groups were positively charged and the carboxylic group of the propanoic acid moiety was negatively charged.

Model building

Based on a published model [27], the amidino group was assumed to form a salt bridge with the carboxylic group in the side chain of Asp¹⁸⁹ of Factor Xa. Four analogs listed in Table 2 showed that by enclosing the amidino group in a ring with no opportunity for salt bridge formation decreased the activity. Also, substitution on the naphthyl or benzofuran ring by a methoxy or methyl group decreased the activity, implying that the amidino moiety was in a tight pocket forming a salt bridge. This was confirmed by a recent publication [28].

To determine a good starting geometry for this salt bridge interaction, a search of the Cambridge Crystallographic Database (CCD) [29] was performed for a benzamidino group forming a salt bridge with a carboxylic function. Several crystal structures that met the requirement were obtained and the torsion angles between the amidino group and the benzene ring in these structures are listed in Table 3. One of the hits (ref. code: GADMUT) was superimposed with the carboxylic group of Asp¹⁸⁹ of the Factor Xa structure. Of the four hits, GADMUT resembled the proposed salt bridge interaction the most. The amidino group of structure 4 was then superimposed onto the amidino group of GADMUT. This orientation of the naphthyl amidino group of structure 4 in the Factor Xa study was the starting orientation in all optimizations.

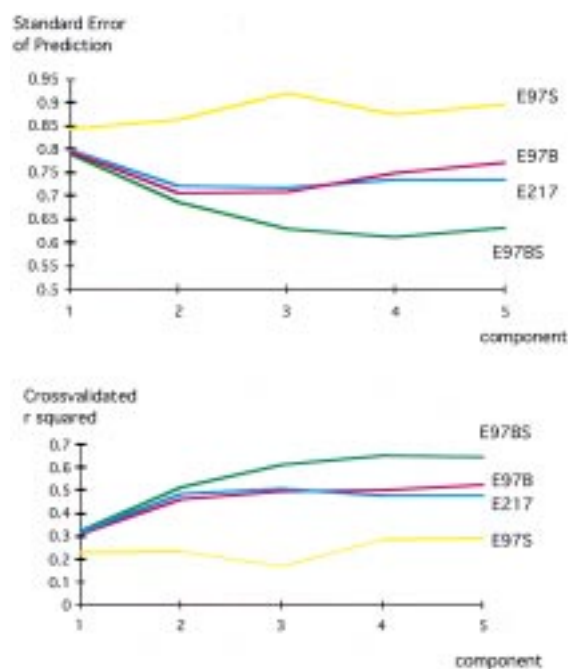


Figure 3. A plot of the cross-validated r^2 or q^2 as well as the standard error of prediction versus the number of components for the four models. The color scheme matches the color scheme followed in Figure 2 showing the models.

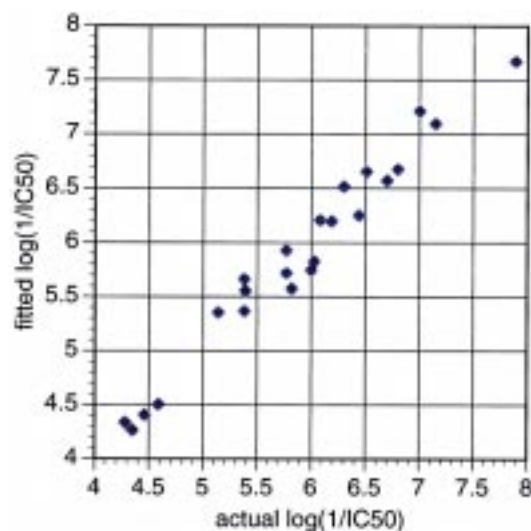


Figure 4. The fitted versus experimental $\log(1/IC_{50})$ from the E97BS model.

Structure **4** was placed into the active site of Factor Xa in the orientation described and the remaining bonds in the molecule were subjected to a systematic search [30–32] including the bonds of the acetimidopyrrolidinyl ring. Only those conformations where the nitrogen atom of the acetimidoyl group fell within 10 Å of any atom of the enzyme were retained. A total of 2079 conformations were obtained. A trace of the acetimidoyl carbon tracked in the enzyme for these conformations is displayed in Figure 1. Based on the clustering of the two dihedral angles represented by C2–C3 of the propionic group and the external bond of the pyrrolidinyl ring made with the oxygen atom, three sets of conformational families were obtained. Two of the families are represented by the binding modes proposed previously [7–9] for structure **4** as shown in Figures 2a and b. A third family is represented by an entirely new conformation (Figure 2c) in which the acetimidoyl group is placed in the vicinity of Glu²¹⁷. The models corresponding to these conformations will be designated as E97S, E97B and E217 respectively, defined by the interaction of the acetimidoyl nitrogen interacting with the Glu⁹⁷ side chain, backbone or the Glu²¹⁷ side chain. One of the negative features of the E217 conformational model is that Arg²²² is also in the vicinity of the acetimidoyl group. The distance from the acetimidoyl nitrogen to the side chain carboxyl carbon of Glu²¹⁷ is 6.07 Å. The distance from the same nitrogen to the carbon of the guanidino group in Arg²²² is 7.31 Å. In thrombin, Arg²²¹, which is equivalent to Arg²²² in Factor Xa, forms a salt bridge with Glu¹⁴⁶ in thrombin. In the crystal structure of Factor Xa, electron density corresponding to the Glu¹⁴⁶ residue is absent possibly as a result of proteolytic cleavage [5]. This could alter the position of the Arg²²² side chain. A positive feature of the E97B model is that the carboxylate group interacts with the side chain of Arg¹⁴³.

Two slightly different forms of the enzyme were used. For the models E217 and E97B, the enzyme as described earlier was employed. For E97S, the following changes were made. The Glu⁹⁷ angles χ_1 , χ_2 and χ_3 were changed from -47.61° , -76.44° and 175.53° to -167.27° , 55.03° and -96.77° , respectively. This allows the carboxyl group in the side chain of Glu⁹⁷ to interact with the acetimidoyl group of structure **4**. A fourth model corresponds to the E97B conformational model in the enzyme with the conformation of the Glu⁹⁷ sidechain modified. This model will be referred to as the E97BS model.

Two sequential optimizations were performed for the four models (E97B, E97S, E217 and E97BS). The first optimization held the non-hydrogen atoms of the enzyme fixed. The second optimization employed the structure obtained after the first optimization. The whole structure was re-optimized keeping only those residues fixed that were outside a 7 Å radius from the inhibitor. This optimization was terminated when the root mean square (rms) gradient of the energy was less than 0.05 kcal/(mol Å). The calculations were performed using the Tripos52 force field [23] and the Powell optimization routine with AM1 charges on the inhibitors derived from a single point calculation using MOPAC5 [26] and AMBER [22] charges on the enzyme. Once structure **4** was optimized corresponding to all four models described above, the molecule was removed from the enzyme and edited locally (using the editing tools in SYBYL6.22) to another molecule in Table 1. This new inhibitor was re-inserted into the enzyme retaining the same orientation for the amidino group. The chiral center in the propanoic group was assumed to be (*S*) when unspecified. The optimizations were then repeated. In this way the conformations of all 23 molecules of Table 1 were optimized in each of the four models (E97S, E97B, E217 and E97BS). With the optimizations completed, the four sets of molecules were removed from the enzyme and subjected to four sets of CoMFAs.

CoMFA

All four CoMFA models used the same region definition which extended from $(-24.23, -13.06, -5.01)$ to $(5.58, 8.89, 23.89)$ Å. All four models had a grid spacing of 2 Å. Also, those columns that had a minimum σ of less than 2 kcal were dropped. The scaling used was block standardization of each field without rescaling of the individual columns corresponding to individual lattice points. The probes used were the same in all four cases: an sp^3 hybridized carbon (C.3 atom type in SYBYL6.22) and a positive charge. PLS analysis with a maximum of 5 components and cross-validation of 23 groups was run on all three models. Once the cross-validation analyses were obtained, PLS analysis was performed on all three models without cross-validation. The model with the best statistical significance was E97BS with the number of optimal components equal to 4. E97S had the least statistical significance with the optimal number of components equal to 5. The optimal number of components for E97B was 3 while E217 had 5 components. The CoMFA for the E97BS models

was repeated with a region having a grid spacing of 1 Å instead of 2 Å in order to check sensitivity as well as obtain coefficient plots with a finer spacing. The statistical results for all four models are listed in Table 4.

Results and discussion

Of the models evaluated, the E97BS was the most significant with 4 components. The statistical parameters used were the standard error of prediction, the cross-validated r^2 or q^2 and the standard r^2 values for the four models. A plot of q^2 versus the number of components for the four models is shown in Figure 3. A plot of the fitted versus experimental $\log(1/IC_{50})$ for the E97BS model is shown in Figure 4.

Histograms of the values of the coefficient multiplied by the standard deviation for both electrostatic and steric fields of the E97BS model are shown in Figures 5a and b together with the contour plots of these values, colored similar to the histograms. These plots generally reflect the binding pocket. In Figure 5a, green indicates regions in which increasing bulk is expected to favor binding while red indicates that increasing steric bulk would decrease binding. In Figure 5b, cyan reflects regions where negative charge would enhance binding and red reflects regions where positive charge would enhance binding. The red region in Figure 5b includes the acetimidoyl group and the cyan region includes the propionic group.

The three models proposed by Stubbs et al. [9], Lin and Johnson [8] and Katakura et al. [7], have elements which resemble the model E97BS. The similarities with the model described by Stubbs et al. [9] include the pyrrolidiny ring of DX-9065a being stacked between Phe¹⁷⁴ and Tyr⁹⁹. Another similarity was the plane of the benzene ring being parallel with the plane of the indole ring of Trp²¹⁵. In the work of Stubbs et al., the crystallography was performed on BX-5633 and DX-9065a complexed with β -trypsin. The differences in the two enzymes, Factor Xa and β -trypsin, can be directly related to the differences between E97BS and their proposed model. The differences in residue composition between the active site of β -trypsin and Factor Xa are listed in Table 5. One change, Asn⁹⁷ instead of Glu⁹⁷, would affect the binding conformation of the inhibitors. This change would not allow any distortion near the Asn⁹⁷ end of the β -trypsin binding pocket because of the lack of a strong Coulombic/hydrogen bond interaction which is

present between the side chain of Glu⁹⁷ and the acetimidoyl group of DX-9065a. The change in the S1 pocket from Ala¹⁹⁰ in Factor Xa to Ser¹⁹⁰ in β -trypsin would also make the S1 pocket in Factor Xa more lipophilic, allowing for the collapse of the enzyme around the naphthalene ring of DX-9065a. Also, the symmetric lysyl interaction with the propionic group mentioned by Stubbs et al. [9] is absent in Factor Xa. Despite these differences, a crystallographic study in β -trypsin seems to be validated in determining the correct binding conformation of the inhibitor.

The enzyme without the inhibitor after optimization (in magenta) with DX-9065a present is overlaid on the crystal structure [5] of the enzyme (colored by atom type) and is shown in Figure 6. The larger changes are shown in cyan mainly around Glu⁹⁷ and Gln¹⁹². These agree with the key features pointed out in a recent paper by Brandstetter et al. [28]. The changes in the structure of the enzyme optimized with DX-9065a are listed below:

(1) Ala¹⁹⁰–Gly¹⁹³ move towards the naphthalene ring corresponding to the before mentioned collapse of the S1 pocket. The movements of the C α of Cys¹⁹¹ and Gln¹⁹² are 1.374 and 1.703 Å, respectively.

(2) The side chain of Glu⁹⁷ flips and contacts the acetimidoyl group. The final χ_1 , χ_2 and χ_3 values for Glu⁹⁷ are -179.6° , -74.2° and -50.6° , respectively.

(3) There is significant movement (1.04 Å) in the C α of Glu⁹⁷.

Based on these data, the conformations of the inhibitors between the E97B and the E97BS models change, leading to a 'better' model due to the flexibility of the active site of the inhibitor. It is therefore imperative to consider a flexible enzyme model when doing accurate binding mode evaluations.

The propionic group interacts directly with Arg¹⁴³, although in the crystal structure they see the side chain of the Gln¹⁹² forming an 'ionic' bridge between the propionic group and the guanidinium group of Arg¹⁴³. This interaction is on the surface of the protein and solvation effects could play a role in determining the interaction.

The models E97B and E97BS differ slightly only for those molecules that have the acetimidoyl group. Those molecules in the E97BS models have this group raised towards and interacting with the Glu⁹⁷ side chain which would be expected after optimization.

The selectivity against thrombin is also explained. Thrombin and Factor Xa are again overlaid based on the homology alignment of Greer. The insertion loop

Table 4. The detailed PLS statistical results obtained for the four models

Model	Component 1	Component 2	Component 3	Component 4	Component 5
E97BS					
<i>2 Å region</i>					
Standard error of prediction	0.79	0.685	0.63	0.611	0.631
q^2 (cross-validated r^2)	0.324	0.516	0.612	0.654	0.651
F-values (cross-validated run)	1.63	3.622	5.355	6.422	6.346
Prob. ($q^2 = 0$)	0.206	0.021	0.004	0.002	0.002
r^2	0.62	0.828	0.924	0.973	0.983
Standard error of estimate	0.593	0.409	0.278	0.172	0.138
F-values	34.239	47.983	77.085	159.196	201.305
Prob. ($r^2 = 0$)	0	0	0	0	0
Contribution: steric				0.597	
electrostatic				0.403	
<i>1 Å region</i>					
Standard error of prediction	0.802	0.707	0.662	0.664	0.699
q^2 (cross-validated r^2)	0.305	0.485	0.57	0.591	0.571
F-values (cross-validated run)	1.491	3.202	4.514	4.909	4.533
Prob. ($q^2 = 0$)	0.244	0.032	0.008	0.006	0.008
r^2	0.594	0.824	0.918	0.971	0.982
Standard error of estimate	0.612	0.414	0.289	0.176	0.142
F-values	30.745	46.735	71.197	152.785	189.23
Prob. ($r^2 = 0$)	0	0	0	0	0
Contribution: steric				0.569	
electrostatic				0.431	
E97S					
Standard error of prediction	0.843	0.862	0.92	0.876	0.898
q^2 (cross-validated r^2)	0.231	0.234	0.171	0.288	0.293
F-values (cross-validated run)	1.02	1.041	0.703	1.372	1.411
Prob. ($q^2 = 0$)	0.437	0.426	0.629	0.283	0.27
r^2	0.545	0.648	0.907	0.95	0.96
Standard error of estimate	0.649	0.584	0.309	0.233	0.213
F-values	25.131	18.411	61.521	85.026	82.283
Prob. ($r^2 = 0$)	0	0	0	0	0
Contribution: steric					0.589
electrostatic					0.411
E97B					
Standard error of prediction	0.794	0.706	0.71	0.749	0.772
q^2 (cross-validated r^2)	0.317	0.487	0.506	0.479	0.478
F-values (cross-validated run)	1.58	3.222	3.497	3.129	3.11
Prob. ($q^2 = 0$)	0.219	0.032	0.024	0.035	0.036
r^2	0.621	0.812	0.891	0.953	0.97
Standard error of estimate	0.592	0.428	0.333	0.225	0.186
F-values	34.448	43.062	51.951	91.083	109.19
Prob. ($r^2 = 0$)	0	0	0	0	0
Contribution: steric			0.552		
electrostatic			0.448		

Table 4. (Continued)

Model	Component 1	Component 2	Component 3	Component 4	Component 5
E217					
Standard error of prediction	0.797	0.722	0.718	0.734	0.735
q^2 (cross-validated r^2)	0.312	0.463	0.495	0.5	0.527
F-values (cross-validated run)	1.545	2.931	3.337	3.406	3.781
Prob. ($q^2 = 0$)	0.228	0.044	0.028	0.026	0.017
r^2	0.622	0.823	0.913	0.97	0.994
Standard error of estimate	0.591	0.414	0.298	0.181	0.086
F-values	34.531	46.63	66.71	144.26	521.647
Prob. ($r^2 = 0$)	0	0	0	0	0
Contribution: steric					0.579
electrostatic					0.421

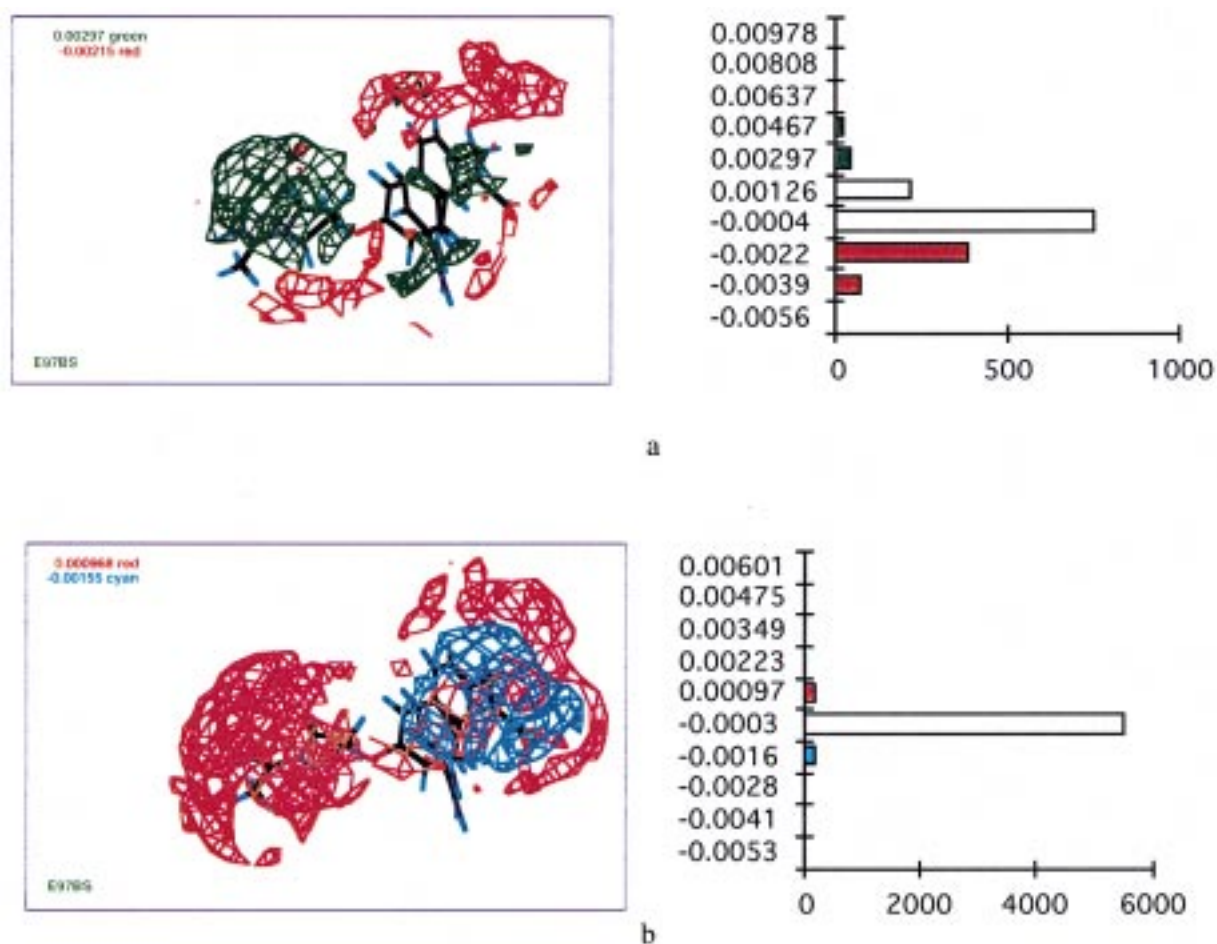


Figure 5. Histograms of the coefficient multiplied by the standard deviation for the point in the CoMFA region together with the corresponding contour plots for the (a) steric and (b) electrostatic fields of the E97BS model. The green contour surface for the steric field implies increased steric substitution in the ligand and red implies decreased or unchanged steric substitution. The cyan contour surface for the electrostatic field implies increased positive charge substitution in the ligand and red implies increased negative charge substitution.

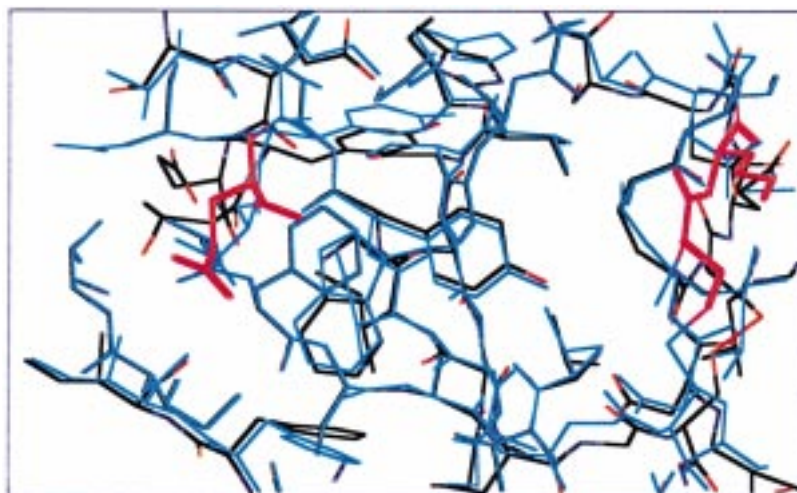


Figure 6. Movements in the enzyme in the E97BS model around the S1 pocket (around Cys¹⁹¹) and around the Glu⁹⁷ area. This is in keeping with the crystal structure.

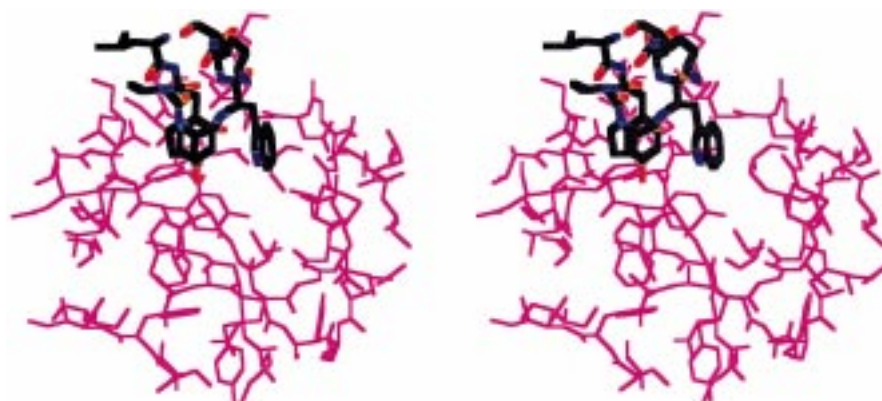


Figure 7. The insertion loop of thrombin shown 'bumping' into DX-9065a when Factor Xa and thrombin are overlaid based on the homology alignment of Greer [24,25].

Table 5. Differences in amino acids in the active site of Factor Xa and β -trypsin

Residues in	
Factor Xa	Bovine β -trypsin
Glu ⁹⁷	Asn ⁹⁷
Tyr ⁹⁹	Leu ⁹⁹
Phe ¹⁷⁴	Gly ¹⁷⁴
Ile ¹⁷⁵	Gln ¹⁷⁵
Ala ¹⁹⁰	Ser ¹⁹⁰
Glu ²¹⁷	Ser ²¹⁷

of thrombin would sterically 'bump' DX-9065a in the E97BS model as can be seen in Figure 7.

Conclusions

In the past, it has been shown in validation studies that a CoMFA model performed on the crystallographic coordinates of inhibitors afforded a statistically significant model [11]. Also, alignments derived from crystallographic information of some inhibitor complexes, have produced statistically significant CoMFAs [12,13]. This paper reports an extension of this approach. The model with the best statistical significance is the same as that shown by crystallographic data. Work is underway with other systems to examine whether the alignment with the best quantitative structure–activity relationship does indeed mirror the binding mode.

Acknowledgements

The authors thank Prof. A. Tulinsky and Dr. K. Padmanabhan (Michigan State University) and Dr. Alan D. Cardin (Hoechst Marion Roussel) for arranging for us to have the Factor Xa coordinates prior to deposition in the Brookhaven Databank.

References

- Stubbs, M.T. and Bode, W., *Curr. Opin. Struct. Biol.*, 4 (1994) 823.
- Claeson, G., *Blood Coagul. Fibrinolysis*, 5 (1994) 411.
- Kelly, A.B., Maraganore, J.M., Bourdon, P., Hanson, S.R. and Harker, L.A., *Proc. Natl. Acad. Sci. USA*, 89 (1992) 6040.
- Furie, B. and Furie, B.C., *Cell*, 53 (1988) 505.
- Padmanabhan, K., Padmanabhan, K.P., Tulinsky, A., Park, C.H., Bode, W., Huber, R., Blankenship, D.T., Cardin, A.C. and Kisiel, W., *J. Mol. Biol.*, 232 (1993) 947.
- Blankenship, D.T., Brankamp, R.G., Manley, G.D. and Cardin, A.D., *Biochem. Biophys. Res. Commun.*, 166 (1990) 1384.
- Katakura, S., Nagahara, T., Hara, T., Kunitada, S. and Iwamoto, M., *Eur. J. Med. Chem.*, 30 (1995) 387.
- Lin, Z. and Johnson, M.E., *FEBS Lett.*, 370 (1995) 1.
- Stubbs, M.T., Huber, R. and Bode, W., *FEBS Lett.*, 375 (1995) 103.
- Nagahara, T., Yokoyama, Y., Inamura, K., Katakura, S., Komoriya, S., Yamaguchi, H., Hara, T. and Iwamoto, M., *J. Med. Chem.*, 37 (1994) 1200.
- Diana, G.D., Kowalczyk, P., Treasurywala, A.M., Oglesby, R.C., Pevear, D.C. and Dutko, F.J., *J. Med. Chem.*, 35 (1992) 1002.
- Waller, C.L., Oprea, T.I., Giolitti, A. and Marshall, G.R., *J. Med. Chem.*, 36 (1993) 4152.
- Cho, S.J., Serrano, M.G., Bier, J. and Tropsha, A., *J. Med. Chem.*, 39 (1996) 5064.
- Cramer III, R.D., Patterson, D.E. and Bunce, J.D., *J. Am. Chem. Soc.*, 110 (1988) 5959.
- Cramer III, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., In Kubinyi, H. (Ed.), *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden, The Netherlands, 1993, pp. 443–455.
- Clark, M. and Cramer, R.D., *Quant. Struct.–Act. Relatsh.*, 12 (1993) 137.
- Böhm, H.J., *J. Comput.-Aided Mol. Design*, 8 (1994) 623.
- Kuntz, I.D., Meng, E.C. and Shoichet, B.K., *Acc. Chem. Res.*, 27 (1994) 117.
- Miller, M.D., Kearsley, S.K., Underwood, D.J. and Sheridan, R.P., *J. Comput.-Aided Mol. Design*, 8 (1994) 153.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., *J. Mol. Biol.*, 112 (1977) 535.
- SYBYL6.22, a molecular modeling system is supplied by Tripos Assoc., St. Louis, MO.
- Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., *J. Am. Chem. Soc.*, 106 (1984) 765.
- Clark, M., Cramer III, R.D. and Van Opdenbosch, N., *J. Comput. Chem.*, 10 (1989) 982.
- Greer, J., *J. Mol. Biol.*, 153 (1981) 1027.
- Greer, J., *J. Mol. Biol.*, 153 (1981) 1043.
- Stewart, J.J.P., MOPAC, Version 5.0, a semi-empirical program supplied by QCPE (No. 455), Indiana University, Bloomington, IN, 1989.
- Katakura, S., Nagahara, T., Hara, T. and Iwamoto, M., *Biochem. Biophys. Res. Commun.*, 197 (1993) 965.
- Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K. and Engh, R.A., *J. Biol. Chem.*, 271 (1996) 29988.
- Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, E.M., Mitchell, G.F., Smith, J.M. and Watson, D.G., *J. Chem. Inf. Comput. Sci.*, 31 (1991) 187.
- Learch, A.R., In Lipkowitz, K.B. and Boyd, D.B. (Eds.), *Reviews in Computational Chemistry*, Vol. II, VCH Publishers, New York, NY, 1991, pp. 1–55.
- Höltje, H.-D. and Folkers, G., *Molecular Modeling, Basic Principles and Applications (Methods and Principles in Medicinal Chemistry, Vol. 5)*, VCH, Weinheim, 1997, pp. 25–28.
- Beusen, D.D., Berkley Shands, E.F., Karasek, S.F., Marshall, G.R. and Dammkoehler, R.A., *J. Mol. Struct. (THEOCHEM)*, 370 (1996) 157.