Accepted Manuscript

Modulating hydrogen-bond basicity within the context of protein-ligand binding: A case study with thrombin inhibitors that reveals a dominating role for desolvation

Nader N. Nasief, Ahmed M. Said, David Hangauer

PII: S0223-5234(16)30768-1

DOI: 10.1016/j.ejmech.2016.09.038

Reference: EJMECH 8904

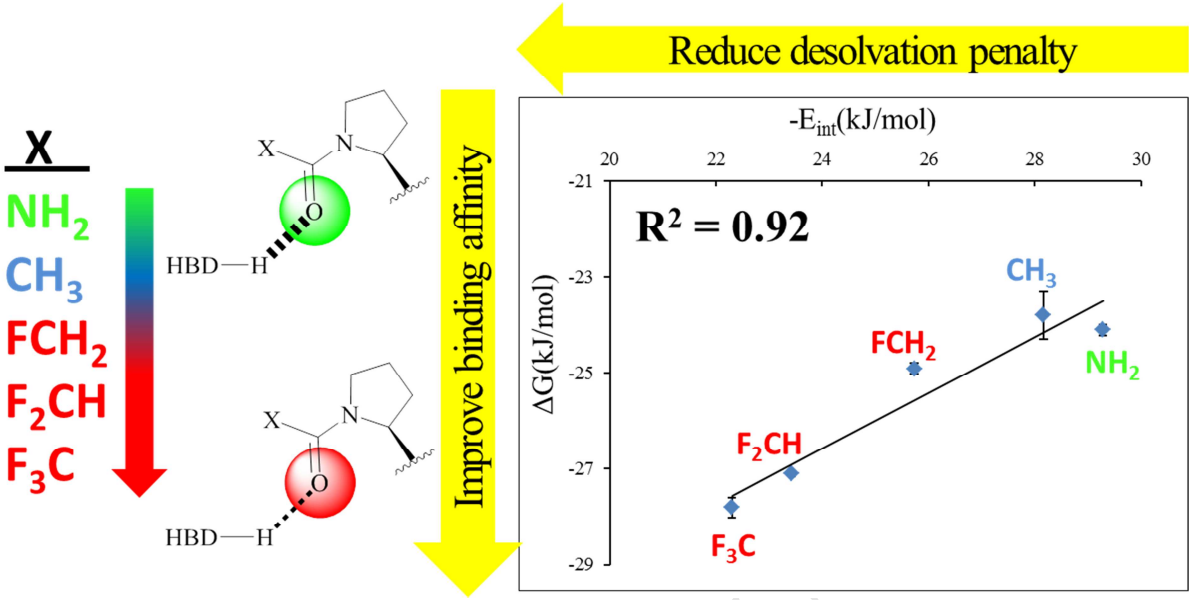
To appear in: *European Journal of Medicinal Chemistry*

Received Date: 14 July 2016

Revised Date: 10 September 2016

Accepted Date: 12 September 2016

Please cite this article as: N.N. Nasief, A.M. Said, D. Hangauer, Modulating hydrogen-bond basicity within the context of protein-ligand binding: A case study with thrombin inhibitors that reveals a dominating role for desolvation, *European Journal of Medicinal Chemistry* (2016), doi: 10.1016/j.ejmech.2016.09.038.



Modulating Hydrogen-Bond Basicity within the Context of Protein-Ligand Binding: A Case Study with Thrombin Inhibitors That Reveals a Dominating Role for Desolvation

Nader N Nasief[a,b*], Ahmed M Said[a,c], and David Hangauer[a*]

[a]Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, NY 14260, [b]The Division of Pharmaceutical Chemistry, National Organization for Drug Control and Research (NODCAR), Giza, Egypt, [c]Pharmaceutical Organic Chemistry Department, Faculty of Pharmacy, Helwan University, Ein Helwan, Helwan, Egypt

KEYWORDS: Thrombin, protein-ligand binding, desolvation, bioisosterism, hydrogen bond, cooperativity, nonadditivity, B3LYP/6-31G*

ABSTRACT

Understanding subtle aspects of hydrogen bonding is a challenging but crucial task to improve our ability to design ligands with high affinity for protein hosts. To gain a deeper understanding of these aspects, we investigated a series of thrombin inhibitors in which the basicity of the ligand's group that accepts an H-bond from Gly216 was modulated via bioisosterism; e.g., a C=O acceptor was made electron deficient or rich via bioisosteric replacements of the adjacent moiety. Although the ligand's binding affinity was anticipated to improve when the H-bond basicity is increased (due to stronger H-bonding with the protein), we herein present data that unexpectedly revealed an opposite trend. This trend was attributed to a dominating role played by desolvation in determining the relative binding affinity. For example, a decrease in the H-bond basicity reduces the desolvation penalty and, as experimentally observed, improves the binding affinity, given that the reduction in the desolvation penalty dominates the change in the contribution of the ligand's H-bond with the protein. The current study, therefore, reveals that desolvation can be a major underlying cause for the apparently counterintuitive structure-activity relationship (SAR) outcomes, and indicates that understanding this factor can improve our ability to predict binding affinity and to design more potent ligands.

* Correspondences should be addressed to N N N at nnnasief@yahoo.com or D H at hangauer@buffalo.edu

1. Introduction

The discovery of novel small molecules with specific biological activity is generally a challenging task that often involves synthesizing a large number of chemical compounds in order to obtain a lead molecule suitable for further development and optimization. One of the major causes for this problem is our insufficient understanding of protein-ligand binding and how such binding is affected by various structural aspects of the designed molecules. This significantly impedes our ability to accurately predict binding affinity, and, in turn, causes extensive experimentation to be the only reliable approach to discriminate between biologically active and inactive compounds. Studies that investigate the intricate molecular aspects of protein-ligand binding are therefore crucial for the field of drug discovery. Recently, various aspects of protein-ligand binding, such as cooperativity [1-5], the role played by water networks and the hydrophobic effect in binding [2][6-7], as well as enthalpy-entropy compensation [7], have been extensively studied using two model biological systems: thermolysin and thrombin.

Thrombin is a serine protease that plays an important role in the coagulation cascade. Thrombin is implicated in pathological coagulation that characterizes several thromboembolic diseases, such as pulmonary embolism, stroke, and heart attacks. These diseases, collectively, are a leading cause of mortality and morbidity, particularly among elderly patients [8]. Thrombin inhibition, therefore, provides a validated approach for anticoagulation therapy; and over the past decades, there have been significant efforts to discover small molecules that act as direct thrombin inhibitors [9-11]. These efforts have resulted in the introduction of dabigatran, a direct thrombin inhibitor, into the market as a prophylaxis to reduce the risk of thromboembolism [12].

Thrombin is an excellent model system to study protein-ligand binding, consequently, it is used herein to investigate subtle aspects of hydrogen bonding that could, and frequently do, result in unpredictable structure-activity relationships (SARs). Hydrogen bonding presented itself as a major non-covalent interaction that is heavily involved in all of the binding aspects we have previously studied [1-7]. Variations in the hydrogen bonding ability of different H-bond forming groups, which is termed hydrogen-bond basicity in the case of H-bond acceptors, can significantly affect protein-ligand binding. In the current study, the hydrogen-bond basicity in a series of closely related thrombin inhibitors was modulated via bioisosteric replacements. The

influence of such modulation on the binding affinity was then investigated and correlated to the desolvation of the H-bond acceptor groups. The study presented herein and similar studies are important for improving our understanding of protein-ligand binding in general, and the binding of thrombin to its inhibitors in particular, which, in turn, could lead to a more efficient design of direct thrombin inhibitors.

1.1. Bioisosterism and hydrogen-bond basicity

Bioisosteric replacement of certain functional groups is an extensively used strategy for optimizing lead compounds in terms of their biological activity, pharmacokinetic properties, and toxicological profile [13-14]. Bioisosteres are substituents or groups that have chemical or physical similarities, and which can produce broadly similar biological properties [15]. There are two types of bioisosterism: classical [16-17] and nonclassical [18]. Classical isosteric replacement includes replacements in which the isosteric atoms/groups have peripheral layers of electrons that can be considered similar (e.g. $–CH_2$-, -NH-, -O-, and –S-) [19-20]. Nonclassical bioisosterism, however, includes broader functional group replacements, which do not satisfy the classical electronic requirements but can produce similar biological activities. Examples of nonclassical bioisosteric replacements are the replacement of C=O with SO_2, COOH with tetrazole, etc.

<Insert Figure 1>

One of the bioisosteric replacements that can be carried out in previously reported thrombin inhibitors [1][4] is the replacement of the C=O group, which accepts an H-bond from the –NH– group of the thrombin Gly216 residue (Figure 1), with an SO_2 moiety, which is anticipated to also accept an H-bond from Gly216 (an amide-sulfonamide bioisosterism). This replacement could be desirable in terms of improving the metabolic stability of a lead compound; consequently, it is a common bioisosteric replacement in drug discovery. It is, however, difficult to accurately predict the relative change in binding affinity caused by such replacement. For example, in the most straightforward cases of protein-ligand binding, the H-bond formed by the SO_2 group needs to be at least as energetically favorable as that formed by the C=O in order to maintain or improve binding affinity. This largely depends on many electronic and geometric parameters that are difficult to evaluate quantitatively (e.g., the relative basicities of the

bioisosteres; the H-bond distance, angles, etc.). In other cases, however, additional parameters, such as desolvation, need to be taken into account. These parameters could significantly affect the outcome of bioisosteric replacements of an H-bonding group, but are difficult to evaluate quantitatively as well.

As noted above, one of the parameters that influence the contribution of an H-bond to the binding free energy is the basicity of the H-bond acceptor. In order to rank various H-bond acceptors according to their relative basicities, several H-bond basicity scales were proposed, such as the pK_{HB} and the $logK_{\beta}$ scales [21-22]. Although an H-bond basicity scale is normally constructed against a common reference H-bond donor that forms 1: 1 H-bonded complexes with the acceptor series, there are basicity scales that were built against multiple donors [23], and scales that considered solutes with polyfunctional acceptor moieties [24]. Recently, a numerical database, the pK_{BHX} database, was built based on the 4-fluorophenol basicity scale in a manner that also took into account the difference in basicity among multiple basic sites in a polyfunctional base [25].

Although the pK_{BHX} database could be useful in determining the relative ability of various basic groups to accept H-bonds, a correlation between H-bond basicity and biological activity is still difficult to establish. This might be because there is a dearth of relevant literature examples, in which the relationships between H-bond basicity and biological activity are systematically explored [25-26]. One of the distinct features of the current study is that it explores the concept of H-bond basicity in the context of protein-ligand binding. For example, the $C{=}O{\rightarrow}SO_2$ replacement described herein modulates the ligand basicity towards Gly216. Additionally, such basicity can be modulated via bioisosteric replacements of groups that are adjacent to the $C{=}O/SO_2$ (e.g., $CH_3CO{\rightarrow}CF_3CO$). The series of inhibitors investigated in the current study, hence, provides a unique opportunity to evaluate the relationship between binding affinity and the H-bond basicity. In the following sections, we describe the rationale behind the bioisosteric replacements we carried out in the current study, as well as the synthesis and the evaluation of various inhibitors in thrombin biochemical assay.

1.2. Thrombin inhibitors with modulated H-bond basicity towards Gly216

<Insert Figure 2>

In order to lay a foundation for this study, the contribution of the C=O group in the thrombin inhibitors shown in Figure 1 to the binding free energy was examined and subsequently used as a reference when the C=O was bioisosterically replaced. This information can be obtained via comparing the binding free energies of ligands **1** and **2** (Figure 2). Although data for these ligands were previously reported [27], the ligands were resynthesized and retested for the sake of consistency of the data presented herein. As shown in Figure 2, ligand **1** does not have the C=O which accepts an H-bond from Gly216; rather it has a CH_2. In contrast, ligand **2** has this C=O. As a consequence of these structural features, the binding free energy contribution of the C=O that forms an H-bond with Gly216 (relative to CH_2) was determined by calculating $\Delta G_{ligand\text{-}2}$ - $\Delta G_{ligand\text{-}1}$ (i.e. the differential binding energy $\Delta\Delta G_{CH2\rightarrow CO}$), and was found to be -3.2 kJ/mol.

<Insert Scheme 1>

Scheme 1 describes the design of the thrombin inhibitors studied herein. Firstly, the C=O group of **2** was bioisosterically replaced with SO_2 to give ligand **3**. While the SO_2 group is likely to form an H-bond with the NH of Gly216, the strength of this H-bond might not be the same as that formed by the C=O, because, based on the pK_{BHX} database, the SO_2 is predicted to be a weaker H-bond acceptor (e.g., sulfonamides are less basic than amides) [25]. Secondly, the H-bond basicities of both the C=O and the SO_2 were further modulated through bioisosteric replacements of adjacent groups. For example, in Scheme 1, the Me group of the terminal acetyl in **2** was replaced by FCH_2, F_2CH, CF_3, and NH_2 to give ligands **4**, **5**, **6**, and **7**, respectively. The substitution of fluorines onto the α-C (e.g., **4**, **5**, and **6**) was anticipated to decrease the basicity of the C=O via an electron withdrawing effect, whereas the replacement of Me with NH_2 in **7** was expected to have an opposite influence on the C=O basicity. Similarly, the Me of the $MeSO_2$ in **3** was replaced by CF_3 and NH_2 to give **8** and **9**, respectively.

In order to further investigate the increase in the basicity of the C=O, the α-CH_2 group in the previously reported thrombin inhibitors, **10** and **11** [27], was replaced by O to give **12** and **13**, and by NH to give **14** and **15**. This set of compounds is particularly important because it provides a systematic comparison among the CH_2, O, and NH substitutions. This comparison is not possible without an R side chain, such as the ones indicated in Scheme 1, because a carbamic acid analogue of **2** (and **7**) is not stable. In addition, compounds **10**, **11**, **14**, and **15**, together with **2** and **7**, offer a basis for investigating the relationship between the R side chain, which binds in

the S3 pocket, and the $CH_2 \rightarrow NH/NH \rightarrow CH_2$ replacement, with regard to additivity/cooperativity of their contribution to the binding affinity/free energy. It should be emphasized herein that the main questions the outlined design probes are whether modulating the H-bond basicity in various ways would affect the binding affinity, and in what manner. To the best of our knowledge, the compounds illustrated in Scheme 1 represent the largest set of systematically modified analogs that investigate the effect of gradually altering the ligand's tendency to form hydrogen bonds on the binding affinity.

2. Results and Discussions

2.1. Chemistry

<Insert Scheme 2>

Ligands **1**, **2**, **3**, **4**, **5**, **6**, **9**, **10**, **11**, **14**, and **15** were synthesized starting with the commercially available L-proline benzyl ester hydrochloride, which was reacted with a number of reagents as follows (Scheme 2):

(1) With ethyl iodide at 80°C in anhydrous DMF, and in the presence of potassium carbonate as a base (N-alkylation). This was followed by hydrogenation of the product in methanol using Pd-C as a catalyst to yield intermediate **1i**, which was then used to synthesize ligands **1**.

(2) With either methanesulfonyl chloride or aminosulfonyl chloride in anhydrous DCM, and in the presence of triethylamine (TEA) as a base. This was followed by hydrogenation of the products in methanol using Pd-C as a catalyst to yield intermediates **3i** and **9i**, which were used to synthesize ligands **3** and **9**, respectively.

(3) With either fluoroacetic acid, 3,3-dimethylbutyric acid, or hydrocinnamic acid in anhydrous DMF using EDCI/HOBt as the coupling reagents and diisopropylethylamine (DIEA) as the base. This was followed by hydrogenation of the products in methanol using Pd-C as a catalyst to yield intermediates **4i**, **10i**, and **11i**, which were used to synthesize ligands **4**, **10**, and **11**, respectively.

(4) With either acetic anhydride, difluoroacetic anhydride or trifluoroacetic anhydride in anhydrous pyridine. This acylation step was followed by hydrogenation of the products in

methanol using Pd-C as a catalyst to yield intermediates **2i**, **5i**, and **6i**, which were used to synthesize ligands **2**, **5** and **6**, respectively.

(5) With either *tert*-butyl isocyanate or benzyl isocyanate in anhydrous DCM, and in the presence of TEA as a base. This was followed by hydrogenation of the products in methanol using Pd-C as a catalyst to yield intermediates **14i** and **15i**, which were used to synthesize ligands **14** and **15**.

<Insert Scheme 3>

Intermediates **1i**, **2i**, **3i**, **4i**, **5i**, **6i**, **9i**, **10i**, **11i**, **14i**, and **15i** were converted to the final products via coupling these intermediates with 4-(aminomethyl)-benzonitrile hydrochloride using EDCI/HOBt as coupling reagents and DIEA as a base, treating the products with hydroxylamine hydrochloride (to give the *N*-hydroxybenzamidines), and stirring the *N*-hydroxyamidines with acetic anhydride in glacial acetic acid followed by hydrogenation using Pd-C as a catalyst (Scheme 2). Similarly, but starting with the commercially available **7i**, **12i**, and **13i**, ligands **7**, **12**, and **13** were synthesized by coupling starting materials with 4-(aminomethyl)-benzonitrile hydrochloride and treating the coupling products with hydroxylamine hydrochloride in presence of DIEA (Scheme 3). The *N*-hydroxybenzamidine were then converted to the final products via either hydrogenation using Pd-C as a catalyst (e.g., **7** and **12**) or reduction using zinc dust in glacial acetic acid (e.g., **13**).

Ligand **8** was synthesized starting with the commercially available L-Proline-*tert*-butyl ester, which was treated with trifluoromethanesulfonyl chloride in presence of TEA. The resulting *tert*-butyl ester was exposed to trifluoroacetic acid at room temperature to yield the corresponding acid, **8i**. The acid was then coupled to 4-(aminomethyl)-benzonitrile hydrochloride, and the product was treated with hydroxylamine hydrochloride to give the corresponding *N*-hydroxybenzamidine. The final product was then obtained via stirring the *N*-hydroxybenzamidine with acetic anhydride in glacial acetic acid and hydrogenation using Pd-C as a catalyst (Scheme 4).

<Insert Scheme 4>

It should be noted that all the final compounds were purified by reverse-phase HPLC to at least 95% purity.

2.2. Thrombin inhibition data

All the ligands reported herein were tested for thrombin inhibition using a kinetic photometric assay [28]. The assay was carried out in a 20 mM Hepes buffer, which contains 0.154 M NaCl, at pH 7.4 using Pefachrome-tPA as a substrate. NaCl was included in the assay buffer as a source of Na^+, which is the most important allosteric thrombin modulator that is required for enzyme activation (i.e., Na^+-bound 'fast' thrombin has higher activity than Na^+-free 'slow' thrombin) [29-30]. The concentration of the Na^+ in the assay buffer is the same as its concentration in the plasma water (154 mEq/L). Initially, the IC_{50} values for the tested ligands were determined using dose-response curves. These values were then utilized to obtain the inhibition constants "K_i's" using the Cheng-Prusoff equation [31], and the inhibition constants were used to calculate the binding free energies "ΔG's" of the tested ligands using the equation $\Delta G = RTlnK_i$. The K_i values and their corresponding binding free energies are listed in Table 1. It is worth noting that even though both the K_i and ΔG values are reported in Table 1, in the following sections, we will analyze relative/differential binding free energies among pairs of compounds; consequently, more emphasis is placed on the free energy values, rather than the K_i's.

<Insert Table 1>

2.3. The Replacement of the -C=O group with a -SO_2- functionality

As illustrated in Scheme 1, the C=O group of each of **2**, **6**, and **7** was replaced with SO_2 to give **3**, **8**, and **9**, respectively. Table 1 reveals that, in each pair of compounds (**2**/**3**, **6**/**8**, and **7**/**9**), there is not much difference in binding affinity/free energy caused by the C=O→SO_2 replacement. For example, **2** and **3** have exactly the same K_i and ΔG values. **6**/**8** and **7**/**9** show only differences in an average range of 2-fold decrease or increase in K_i, which is translated into about 2.0 kJ/mol average free energy change. Notably, the C=O→SO_2 replacement is slightly favorable when an electron withdrawing group, such as the -CF_3, is adjacent to the H-bond acceptor (e.g. ($\Delta G_{ligand\text{-}8}$ - $\Delta G_{ligand\text{-}6}$) or $\Delta\Delta G_{CF3CO\rightarrow CF3SO2}$ = -1.6 kJ/mol: Ligand **8** with CF_3SO_2 is more favorable than ligand **6** with CF_3CO by 1.6 kJ/mol). On the contrary, an electron donating group, such as NH_2,

slightly favors the C=O over the SO_2 (e.g. ($\Delta G_{ligand\text{-}9}$ - $\Delta G_{ligand\text{-}7}$) or $\Delta\Delta G_{H2NCO\rightarrow H2NSO2}$ = +2.5 kJ/mol: Ligand **7** with H_2NCO is the one that is more favorable by 2.5 kJ/mol).

Data, therefore, indicate that the contribution of the SO_2 to the binding affinity is almost the same as the contribution of the C=O. For instance, in the case presented herein, as previously noted, the net favorable contribution of the ligand's hydrogen bonding with the NH of Gly216 is 3.2 kJ/mol no matter whether the acceptor is C=O or SO_2, e.g., $\Delta\Delta G_{CH2\rightarrow SO2}$ ($\Delta G_{ligand\text{-}3}$ - $\Delta G_{ligand\text{-}1}$) = $\Delta\Delta G_{CH2\rightarrow CO}$ = -3.2 kJ/mol. Is this what the H-bond basicity scale for these two groups predicts? Applying the basicity scale, with only the direct ligand interactions with thrombin in mind, one would most likely predict an outcome that is different from what was experimentally observed. For example, according to the pK_{BHX} database, the SO_2 is not as good of an H-bond acceptor as the C=O is [25]. Also, in the PDB database, it has been found that the SO_2 group possesses the dual nature of a weakly polar group that can accept an H-bond, and also a hydrophobic moiety that can exist in close proximity to aliphatic carbon atoms [32]. Considering protein-ligand interactions as the sole, or the major, determinant of binding affinity would therefore lead to the incorrect prediction that the SO_2, for instance in **3**, would cause a reduction in binding affinity. This is, however, not what was observed in the binding affinity data. On the contrary, as will be shown, an in-depth analysis of the other factors that influence binding could lead to a correct prediction of the binding affinity consequences of the C=O→SO_2 replacement.

<Insert Figure 3>

In a C=O/SO_2 comparison, one of the factors that need to be considered is whether the H-bonds formed by the C=O and the SO_2 have similar geometric parameters, because these parameters affect the contribution of an H-bond to the binding affinity. For example, an optimum H-bond would form with a specific distance between the donor and the acceptor atoms, a certain angular preference with regard to the donor-hydrogen-acceptor angle, and a tendency for the H-bond to form along the direction of the donated lone pair of the acceptor [32]. The geometric parameters for the H-bond accepted by the C=O from Gly216 were therefore investigated using PDB ID: 2ZI2, which is a crystallographic complex between thrombin and a ligand similar to **2** except for a terminal butyryl side chain replacing the acetyl moiety. The ligand of this crystal structure was changed to ligand **2**, *in-silico*, through truncating the butyryl to acetyl and minimizing the acetyl group. The H-bond parameters were as follows: NH--O = 3.00 Å (N to O heteroatom distance),

∠N-H-O = 164°, and ∠H-O-C = 160° (Figure 3a). Would a SO_2 moiety retain these parameters? In order to answer this question, a modeling study was performed in which methanesulfonyl replaced the acetyl in ligand **2**'s constructed protein-ligand complex. Several conformations for the methansulfonyl group were then generated and minimized while the rest of the ligand (and the protein) were kept constrained. The minimized conformation that retains geometric parameters closest to those of the acetyl group was selected for further investigations. A protein-ligand complex with the ligand's methansulfonyl adopting this conformation demonstrated the following geometric parameters: NH--O = 2.83 Å (N to O heteroatom distance), ∠N-H-O = 166°, and ∠H-O-S = 139° (Figure 3b). Because these parameters are very close to those of the H-bond formed by the acetyl group, it can be concluded that the SO_2 is able to form an H-bond with Gly216 that is topologically similar to the H-bond formed by the C=O. Given this similarity, would the contributions of the two H-bonds to the protein-ligand interaction component of the binding free energy (or enthalpy) be also similar?

To address this question, two model systems were constructed for the purpose of quantum mechanical (QM) calculation of the energy of the H-bond accepted by the ligand from Gly216 (Note: energy in the context of QM calculations represents enthalpy rather than free energy). One of these models, A_1, consisted of N,N-dimethylacetamide hydrogen bonded to N-methylformamide [33]. The other, B_1, consisted of N.N-dimethyl methanesulfonamide bound to the same N-methylformamide (Models A_1 and B_1 represent ligands **2** and **3**, respectively, accepting H-bonds from Gly216: Table 2). These models were subjected to PM3 [34] and B3LYP/6-31G* [35] calculations in order to determine their equilibrium geometries and energies. The intermolecular interaction energies between the two components of each model (i.e., the N-methylformamide and the N.N-dimethyl methanesulfonamide) were then calculated and were assumed to represent the energies of the H-bonds we are interested in, because, in each model, the investigated H-bond is the major intermolecular interaction in the system. Energy was found to be 31.80 kJ/mol in case of the acetyl moiety and 19.08 kJ/mol in case of the methanesulfonyl (the acetyl forms a stronger H-bond: Table 2). A stronger H-bond formed by the C=O could be attributed to an increase in the availability of the unshared pair of electrons of the C=O due to resonance, which is known to be greater in amides than in sulfonamides. QM calculations are therefore in agreement with the notion that a sulfonamide SO_2 is not as good of an H-bond acceptor as an amide C=O, and had non-covalent protein-ligand interactions been the sole

determinant of the binding affinity, ligand **3** would have been significantly less active than **2**. What then causes the SO_2 to be as good as the C=O in terms of contribution to the binding affinity?

First, it is important to note that the binding affinity/free energy contribution of a ligand functional group is the sum of the contributions of multiple factors of which non-covalent interactions formed between this group and the biological target is one. For example, the amount of binding free energy contributed by a ligand functional group encompasses also a contribution from the desolvation of the part of this group that is buried into the formed complex, as well as contributions from any change induced by this group in the conformation, ionization, and/or hydration status of any other group or amino acid residue in the protein-ligand complex. Desolvation is particularly important in the case presented herein because, as Hunter previously noted, the cost of desolvation nearly cancels out the effect of H-bond formation [36]. This is apparent in the case of the activity improvement shown by the introduction of the carbonyl H-bond acceptor into the ligand ($CH_2 \rightarrow C=O$). This improvement amounts to only 3.2 kJ/mol (i.e., $\Delta\Delta G_{CH2 \rightarrow CO}$), which is significantly smaller than, for example, the H-bond energy calculated in the N,N-dimethyl-acetamide model (i.e., 31.80 kJ/mol). It is likely, therefore, that most of the contribution of the H-bond formed between ligand **2** and Gly216 is canceled out by the desolvation penalty of the C=O. Similar cases in protein-ligand binding have been previously reported. For example, Morgan and co-workers reported that, in thermolysin inhibitors, NH- and CH_2-analogues have almost the same activity because the benefit of the H-bond formed by the NH was canceled out by the desolvation penalty [37].

<Insert Table 2>

As previously described, the C=O is a better H-bond acceptor than the SO_2 in the context of protein-ligand binding, consequently, it can be hypothesized that the C=O should also be a better H-bond acceptor in the context of an unbound hydrated ligand. It follows that the C=O would most likely be harder to desolvate than the SO_2, or, in other words, the C=O would have a larger desolvation penalty. This larger desolvation penalty could cancel out most of the additional benefit contributed to binding by the H-bond formed between the C=O and Gly216. In order to test this hypothesis, and to confirm that the C=O is a better H-bond acceptor towards water, the N,N-dimethylacetamide and the N,N-dimethyl methanesulfonamide test molecules were used

again to construct two QM models (models A_w and B_w), but, this time, instead of being bound to N-methylformamide, each molecule is bound to a water molecule through a single H-bond. The energy of the H-bond in each model was calculated using the B3LYP/6-31G* method, and was found to be -28.16 kJ/mol in case of the C=O and -20.56 kJ/mol in case of the SO_2 (Table 2). Calculations, therefore, indicate that, as expected, the C=O is a better H-bond acceptor towards water. As a consequence, the desolvation penalty of a C=O group is likely larger than that of a SO_2. We, hence, deal with two H-bond acceptor groups: one (i.e., C=O) forms a stronger H-bond with the protein that would contribute more favorably to the binding affinity/free energy were it not for a larger desolvation penalty that cancels out this contribution; and the other (i.e. SO_2) forms a weaker H-bond that does not contribute as favorably to binding but does not cost as much in terms of desolvation energy either. Consequently, the experimental data, which demonstrate that the net contributions of the C=O and the SO_2 to the binding affinity are equivalent, can be rationalized if we look beyond the concept of noncovalent protein-ligand interactions and take into account desolvation.

It should be noted, however, that the analysis presented herein is qualitative, not quantitative, i.e., the calculations performed are not meant for a precise determination of, for example, the desolvation penalty; rather, they were used to reveal qualitative trends. This is simply because, although energy calculations may be correlated with enthalpy, they do not take into account entropic factors. For instance, the -31.80 and -28.16 kJ/mol calculated for the H-bonds of the acetamide models should not be interpreted as the actual amounts that should be added to, or subtracted from, the binding free energy as a consequence of the presence of a hydrogen bonding group in the ligand. The actual contribution to the binding free energies would be much smaller because each of these calculated values is associated with a compensatory opposing entropic contribution (e.g., binding which is favorable in terms of enthalpy is associated with motion restriction which is unfavorable in terms of entropy). Previously, we have reported that up to 80% of the enthalpy associated with hydration water is compensated by entropy [7]. This means that, for example, the -28.16 kJ/mol interaction energy, or enthalpy, calculated in model A_w is likely compensated by about +22.53 kJ/mol entropy, and the amount left over (i.e., -5.63 kJ/mol) is what may be subtracted in the binding free energy master equation [2] as a result of the C=O desolvation. Calculations presented in this paper should therefore be considered within the

context of enthalpy-entropy compensation effects, which often reduce the net contributions to the associated free energy changes.

2.4. The influence of adjacent bioisosteric replacements on the basicity of H-bond acceptors

<Insert Figure 4>

According to the H-bond basicity rule, replacing the Me adjacent to the H-bond acceptor moiety with a more electron withdrawing group (e.g. mono, di, and trifluoromethyl) is anticipated to reduce the availability of the acceptor's lone pair of electrons and, in turn, decrease the tendency to accept H-bonds (Scheme 1). If the free energy contribution of the acceptor were solely governed by the favorable contribution of hydrogen bonding with Gly216, we would experience a gradual decrease in activity going from ligand **2** to ligands **4**, **5**, and **6**, which would be attributable to reducing the basicity and the strength of the H-bond accepted from Gly216. Similarly, ligand **8** would be anticipated to be less active than **3**. Surprisingly, an opposite trend was experimentally observed. For instance, as demonstrated in Table 1 and Figure 4, activity increases going from ligand **2** to **4**, **5**, and **6** (e.g., ΔG goes from -23.8 to -24.9, -27.1, then -27.8 kJ/mol: a negative shift indicates an improvement in activity). Also, ligand **8** was found to be more active than **3**, with a differential binding energy $\Delta\Delta G_{CH3SO2 \rightarrow CF3SO2}$ ($\Delta G_{ligand\text{-}7} - \Delta G_{ligand\text{-}2}$) of -5.6 kJ/mol, an order of magnitude improvement if activity is expressed in terms of inhibition constants.

The same trend of going against the potential direct correlation between the H-bond basicity and the binding affinity was observed when an electron donating group is introduced next to the acceptor moiety. For example, at first glance, it could be anticipated that **7** and **9** would be more active than **2** and **3**, respectively, because of the ability of the NH_2 to donate electrons through resonance to the C=O and the SO_2, and, thereby, to reinforce the hydrogen bonding properties of these acceptors. Experimentally, however, as shown in Figure 4, there is no such improvement in activity, e.g., the -0.3 kJ/mol improvement in the activity of **7** relative to **2** is insignificant. Even **9** was found to be less active than **3** by +2.2 kJ/mol (a positive $\Delta\Delta G$ indicates a decrease in activity; +2.2 kJ/mol is equivalent to about 2-fold on the K_i scale). Because these data are not sufficient to formulate a solid conclusion with regard to the influence of electron donating substitutions on the binding affinity, the triads **10/12/14** and **11/13/15** were also investigated.

Data in Figure 4 reveal that the replacement of the α-CH_2 with O or NH decreases activity. For example, the replacement of CH_2 with O decreases activity by +2.6 kJ/mol in the **10/12** pair, and by +4.2 kJ/mol in the **11/13**. The replacement of CH_2 with NH, which is a more electron donating group, decreases activity even more, e.g., by +5.3 kJ/mol in the **10/14** pair and by +7.5 kJ/mol in the **11/15**. Based on these data, it can be concluded that bioisosteric replacements, which involve the introduction of groups that can donate electrons (e.g., via resonance) to H-bond acceptors, might decrease binding affinity despite the anticipated reinforcement of the H-bonds formed with the protein. On the contrary, groups that are capable of withdrawing electrons from H-bond acceptors might improve binding affinity. The trends revealed by this collection of data are the opposite of what many medicinal chemists might anticipate to observe in SAR studies, and are therefore important to understand in more depth.

<Insert Table 3>

These counterintuitive results motivated us to initially verify our original assumption that adjacent electron withdrawing and donating groups respectively decrease and increase the H-bond basicity of an acceptor moiety. To do so, QM models similar to models A_1 and B_1 (Table 3: QM models L-P1) were constructed (as described in the experimental section) for ligands **4**, **5**, and **6** to account for the "withdrawing" bioisosterism, as well as for **12/13** and **14/15** to account for the "donating" bioisosterism (i.e., NH in **14/15** is a better donor than O in **12/13**). These models were subjected to PM3 and B3LYP/6-31G* calculations in order to determine equilibrium geometries and energies, and the intermolecular interaction energies between the two components of each model were calculated and assumed to represent the energies of the H-bonds we are interested in. Upon evaluating the trend of the interaction/H-bond energies in models A_1, C_1, D_1, and E_1, we noticed that the magnitude of the interaction/H-bond energy inversely correlates with the number of the fluorine atoms on the α-C, and, consequently, with the electron withdrawing effect exerted on the C=O (more fluorine atoms cause a stronger electron withdrawing effect). Considering models F_1 and G_1, we noticed that the presence of an NH, which is a better electron donor than an O, causes an increase in the magnitude of the interaction energy.

These results confirmed that an adjacent electron withdrawing group can indeed decrease the hydrogen bonding basicity of an acceptor moiety resulting in a weaker H-bond with the protein

(e.g. Gly216), whereas an electron donating group has an opposite effect. Because the binding affinity data reveal an opposite trend (i.e., adjacent electron withdrawing and donating groups cause an increase and a decrease in the binding affinity, respectively), it can be concluded that, in the present case, biological activity is inversely correlated with the H-bond basicity of the acceptor and with the strength of H-bond formed between the ligand and Gly216.

Secondly, we investigated if the strength of the H-bond formed by the acceptor moiety (e.g., the C=O) is the only aspect of protein-ligand interaction that changes due to bioisosterism in an adjacent group. This is important because any change in the contribution of protein-ligand interaction to the binding affinity across a series of ligands is determined not by the change in the strength of a single non-covalent interaction, such as the H-bond between the ligand and Gly216, but by the net change in the overall protein-ligand interaction network. To investigate this, we constructed new QM models which incorporate not only the NH of Gly216, as in models L-P1, but also the C=O of this amino acid residue (e.g., models L-P2). This Gly216 C=O is the only additional group that may be strongly influenced by the bioisosteric replacements of the ligand's α-CH_2/CH_3, given that no other group (e.g., besides the NH of Gly216) lies within a 5.00 Å sphere centered on the ligand's α-CH_3 in the crystallographic construct of ligand **2**. Models with formylglycinamide instead of N-methylformamide as the protein representative (i.e., models L-P2) should therefore be a reasonable compromise between accuracy and computational effort when it comes to qualitative investigation of other aspects of protein-ligand interaction besides hydrogen bonding to the NH of Gly216. Table 3 includes these models, which represent ligands **2**, **4**, **5**, and **6** (models A_2, C_2, D_2, and E_2, respectively), and the interaction energies associated with them as calculated using the B3LYP/6-31G* method.

As demonstrated in Table 3, the interaction energies calculated for these models (i.e., L-P2's) are significantly different from those of the corresponding L-P1 models. For example, comparing A_1 to A_2, the magnitude of the interaction energy (i.e., E_{int}) decreases from 31.80 to 22.67 kJ/mol, most likely because of electrostatic repulsion caused by the close proximity of the Gly216 C=O and the ligand's C=O (Gly216-C=O--O=C-ligand **2** = 3.15 Å). Similarly, for ligands **4** and **5** (models C and D), the magnitude of E_{int} decreases going from L-P1 to L-P2, but to a lesser extent, probably because the electron withdrawing effect of the increasing number of fluorine atoms on the α-C decreases the electrostatic charge on the O of the ligand's C=O (e.g., from -

0.57 (CH_3CO) to -0.52 (FCH_2CO) and then to -0.49 (F_2CHCO)) and, consequently, reduces the O-O electrostatic repulsion. The end result is that variations in the magnitude of E_{int} observed in models L-P1 among ligands **2**, **4**, and **5** are almost canceled out in models L-P2 (Table 3). In the case of **6**, the L-P2 model (i.e., E_2) demonstrates dramatic decrease in the magnitude of E_{int} relative to the corresponding L-P1 model (e.g., from -24.75 to -13.32 kJ/mol). This decrease might be attributed to electrostatic repulsion between the additional fluorine on the α-C and the C=O of Gly216 ((Gly216-C=O--F-CF_2-ligand **6** = 2.82 Å). This additional repulsive interaction does not exist in any of ligands **2**, **4**, or **5**, e.g., **4** and **5** likely adopt conformations in which the fluorine atoms are far from the C=O of Gly216. In the case of **6**, despite the significant decrease in the magnitude of E_{int} in model E_2, this ligand is the most active among ligands **2**, **4**, **5**, and **6**; consequently, it could be speculated that this ligand might have a different binding mode which avoids the O--F repulsive interaction. To test this hypothesis, an X-ray crystal structure for this ligand is needed, but this is beyond the scope of the current study. Overall, does the protein-ligand calculated interaction energy correlate with the binding affinity? It appears from the data included in Table 3 that there is no significant correlation. This is more visually depicted in Figure S1 (supplementary data), which represents a plot between E_{int} for five L-P2 models and ΔG for the corresponding ligands. This figure demonstrates an R^2 of 0.54, which indicates a weak correlation at best.

If the binding affinity has no (or little) correlation with protein-ligand noncovalent interactions, what then causes the observed trend in activity? The desolvation factor was investigated next, as was previously done with ligands **2** and **3**. We therefore constructed QM models, each of which consists of a water molecule hydrogen bonded to an acylated N,N-dimethylamine (Table 3: models L-W). The acyl group of the acylated N,N-dimethylamine was varied *in-silico* according to the ligand each model represents (e.g., monofluoroacetyl in model C_w to represent **4**; difluoroacetyl in D_w to represent **5**; etc.). Next, the interaction energy between the acylated N,N-dimethylamine and the water molecule in each model (i.e., the H-bond energy) was calculated using B3LYP/6-31G* and the results are listed in Table 3. Data for models L-W reveal that, among models A_w, C_w, D_w, and E_w, the model that possesses the most favorable H-bond energy is A_w which represents **2** (-28.16 kJ/mol). H-bond energy, then, gradually becomes less favorable as the fluorination of the α-C is increased (e.g. $A_w \rightarrow C_w \rightarrow D_w \rightarrow E_w$: -28.16→-25.73→-23.42→-21.19). These data are in agreement with what was observed in models L-P1, wherein successive

fluorination of the α-C and the resulting increase in the electron withdrawing effect on the C=O cause decease in the basicity and strength of the H-bond accepted by this C=O. Contrary to models L-P1, this effect would impact the binding free energy in an inverse manner, e.g., a more favorable H-bond with water is associated with a weaker binding (i.e., a less negative ΔG) because it signifies a larger desolvation penalty. Conversely, a less favorable H-bond with water signifies a smaller desolvation penalty and would be associated with a better binding. This is exactly what we observe in the **2**, **4**, **5**, and **6** series, where **2**, the weakest binder, demonstrates the most favorable interaction with water (e.g., in model A_w); and **6**, the strongest binder, demonstrates the least favorable interaction with water (e.g., in model E_w). This trend can be clearly perceived when ΔG is plotted in Figure 5 against $-E_{int}$ (representing the desolvation penalty) for A_w, C_w, D_w, E_w, and H_w (an additional model representing the interaction of **7** with water). First, this plot reveals a remarkable linear correlation between ΔG and $-E_{int}$ with an R^2 value of 0.92. Second, the correlation is positive (i.e., a positive slope), which means that when the desolvation penalty increases, ΔG increases, in other words, becomes less negative (i.e., weaker binding); and vice versa.

The data in Table 3 for models F_w and G_w also demonstrate that a more electron donating group, such as an NH relative to an O isostere, causes a stronger H-bond with water (-31.68 vs. -28.46 kJ/mol), which translates into a larger desolvation penalty. As previously described, a larger desolvation penalty results in a weaker binding, and this is what is experimentally observed in **12** vs. **14** and **13** vs. **15** (e.g., **14** and **15** are weaker binders than **12** and **13**, respectively).

<Insert Figure 5>

Overall, there are a number of conclusions that can be drawn based on the data presented in this section. First, a decrease in the basicity of the C=O due to bioisosterism causes not only a decrease in the strength of the H-bond formed with Gly216 in the protein-ligand complex, but also a decrease in the strength of the H-bonds of this group with water and a subsequent reduction in the desolvation penalty (and vice versa). Second, a change in the strength of the H-bond formed with Gly216 in the protein-ligand complex is not the major determinant of the associated change in the binding free energy within the series of the thrombin inhibitor studied herein. Third, the change in desolvation penalty of the C=O dominates the differential/relative binding free energy in this series of thrombin inhibitors and is therefore the factor that produces

the inverse correlation that was observed between the H-bond basicity of the C=O and the binding affinity. For example, the improvement in the binding free energy which accompanies the reduction in the C=O basicity (e.g., **2**→**5**) can be attributed to a reduction in the desolvation penalty, and vice versa (e.g., **12**→**14**). Fourth, the desolvation factor dominates the differential binding because the change in the free energy contribution of the H-bond formed with Gly216 in the protein-ligand complex can be masked by other changes caused by bioisosterism in the array of the noncovalent protein-ligand interactions (e.g., **2** and **5**). This masking effect occurs due to spatial restrictions in the protein-ligand complex, which sometimes prevents groups with unfavorable interactions from moving away from each other (e.g. the two C=O groups of the ligand and Gly216). This is not likely to occur in the unbound state where water molecules can freely move and adopt optimal positions. It should be noted that it is the change in the free energy contribution of the H-bond that is being overcome by the change in the desolvation contribution (e.g., in a series of ligands all of which can form this H-bond such as **2**, **4**, and **5**), not the H-bond's contribution to ΔG itself. When it comes to the absolute contributions of the H-bond with Gly216 and the desolvation of the ligand's H-bond acceptor, there is still enough favorable contribution derived from the H-bond with Gly216 to overcome the desolvation penalty even if this penalty might be large as in **2**. Otherwise, there would be no free energy benefit from having the C=O group in the ligand molecule, which is contrary to what was experimentally observed in **2** vs. **1** (Figure 2). Overall, the data highlight the importance of explicitly considering the desolvation differences among analogs in a SAR study. A further analysis of the data presented herein would require X-ray crystal structures of several analogs, but this is beyond the scope of the current study.

2.5. Functional group cooperativity and bioisosterism

Functional group cooperativity, also termed nonadditivity or synergism/antagonism, is an important concept in drug design and SAR studies. This is mainly because this phenomenon can account for 1-2 or even 3 orders of magnitude unforeseeable increase in binding affinity. We have previously studied functional group cooperativity in both thrombin [1][4] and thermolysin [2-3] and discovered that cooperativity is a common finding in lead optimizations [3], which should be taken into account in scoring functions [1][4], and which can arise from variations in the free energy contributions of different players in the protein-ligand binding process [2]. Other

cases of functional group cooperativity have been reported within the context of SAR studies [38-42]. Functional group cooperativity can be described as the dependency of the functional group contributions to the binding affinity on the structural features of the rest of the ligand molecule. Notably, the ligand C=O group we are evaluating in the current study has previously demonstrated cooperativity. For example, in bicyclic thrombin inhibitors, the contribution of this group to the binding affinity was found to be -3.4 kJ/mol [43], which is the same as the contribution we report herein. In pyrrolidine-based inhibitors with side chains capable of binding in the S3 pocket, however, the contribution of this group was reported to range from -10.0 to -23.0 kJ/mol [4][44-45]. Given that cooperativity is a common phenomenon in SAR studies, albeit typically not recognized as such, it was expected that cases of cooperativity would be discovered in the course of our investigation into the concepts of bioisosterism and H-bond basicity in thrombin inhibitors. In the supplementary materials of this paper, these cases of cooperativity are outlined and evaluated through double mutant (double transformation) cycles. It is important to note that the reason behind the reported cases of cooperativity is currently under investigation and will be reported in due course.

3. Conclusions

In this study, the basicity of the H-bond accepted by thrombin inhibitors from Gly216 was modulated via bioisosteric replacement of either the C=O group, which accepts this H-bond, or an adjacent group that influences the acceptor moiety electronically (via either electron withdrawing or donating effect). A correlation between the basicity of the investigated H-bond acceptor and the ligand's binding affinity was surprisingly found to be absent in case of the C=O replacement with SO_2. For example, ligands featuring an SO_2 moiety, a weaker H-bond acceptor than the C=O, are not significantly different in terms of binding affinity from their C=O analogues. QM calculations revealed that this observed equipotency is largely due to a smaller desolvation penalty in case of the SO_2, and a larger one that cancels out the stronger binding contribution of the H-bond to Gly216 in case of the C=O.

Desolvation was also found to play a dominant role when the basicity of the H-bond acceptor is modulated via bioisosteric replacement of an adjacent group. In this case a counterintuitive inverse correlation between the binding affinity and the H-bond basicity, which could not be

explained by modulating the strength of the ligand's H-bond with the protein, was observed. This inverse correlation could rather be explained only when desolvation is considered. For example, a decrease in the H-bond basicity reduces the strength of the ligand's H-bond(s) not only with the protein but also with hydration waters in the unbound state. This, in turn, reduces the desolvation penalty. If the effect on desolvation is more prominent than the effect on the direct protein-ligand interaction, a decrease in H-bond basicity will be advantageous to the binding affinity (inverse correlation). This explanation was supported by QM calculations wherein ligands with a more favorable binding suffered less of an energy penalty from breaking their H-bonds with a probe solvation water molecule, and vice versa. Furthermore, a plot of the ligand binding free energy vs. the calculated energy for breaking such H-bonds with water yielded a linear relationship with an R^2 value of 0.92.

In summary, this study explores the interconnected roles of bioisosteric replacements, H-bond basicity, and desolvation in the context of protein-ligand binding. The study, therefore, contributes significantly to understanding important aspects of the structure activity relationships of closely related analogues. Although the current study was carried out using thrombin as a model system, the key findings of this study are likely to be applicable to many other protein-ligand drug discovery systems. Three such findings are emphasized herein. First, bioisosterism, an important concept in drug discovery, can yield counterintuitive experimental outcomes; therefore, to predict the outcome of bioisosterism, it is crucial to investigate the changes caused by bioisosteric replacements in the structural and thermodynamic features of both the complexed and uncomplexed ligand systems, e.g., changes in the desolvation, conformational, or ionization status. Second, desolvation is a key factor that, as in the present case, can correlate well with binding affinity but in an unanticipated manner. More attention should therefore be paid to this factor whenever unexpected SAR results are encountered. Furthermore, desolvation should be accurately accounted for in binding affinity prediction algorithms (e.g., scoring functions) that are used for *in-silico* screening of compound libraries. Third, QM models can be successfully employed to study trends in binding affinity (or binding thermodynamics) and determine the factor(s) that may be responsible for these trends. These calculations can also provide important insights into effects that are rarely taken into account in drug design, such as polarizations and charge redistributions.

Notably, the findings presented herein point towards an improved molecular design methodology wherein structural modifications that can reduce the ligand desolvation penalty are explicitly considered when optimizing lead compounds. This methodology could significantly improve the process of lead optimization, and might lead to the discovery of more potent ligands with enhanced biological or therapeutic responses against relevant drug discovery targets.

4. Experimental

4.1. Biological assay

The inhibition constants of the thrombin inhibitors **1-15** were determined photometrically at 405 nm using Pefachrom tPa (CH_3SO_2-D-CHA-Gly-Arg-pNA.AcOH; DSM Nutritional Products Ltd Branch Pentapharm, Switzerland) as a chromogenic substrate. The assay was performed using a SpectraMax Plus 384 UV/VIS microplate spectrophotometer at 25 ± 0.2 ºC. The stock solutions of the enzyme, substrate, and inhibitors were prepared from accurately weighted samples of respective species. All the measurements were carried out in a buffer containing 0.02 M Hepes, 0.154 M NaCl and a final DMSO concentration of 6.5 % at pH 7.4 ± 0.05. The enzyme concentration in all the final assay solutions was approximately 8-12 nM, the substrate concentration in the final assay solutions was 100 μM, and the concentrations of each inhibitor were in the range of 0.5 IC_{50} – 10 IC_{50} (a preliminary IC_{50} for each inhibitor was determined by an initial screening). The assay was always initiated by adding the enzyme to a premixed solution of the substrate and the inhibitor. The absorbance was then monitored kinetically every 20 s. for 30 min, and the reaction rate (OD/s) was used to construct a dose-response curve, which was subsequently used to determine the IC_{50} value. GraphPad Prism was used for data fitting and IC_{50} determination. Inhibition constants (K_i's) were then calculated from the experimentally determined IC_{50}'s using Cheng-Prusoff equation [31] (K_m= 289 ± 6 μM), and were subsequently employed to calculate the binding free energies (ΔG's) using the equation ($\Delta G = RT\ln K_i$, R is the gas constant and T is the temperature in Kelvin). The reported average inhibition constants and binding free energies (K_i's and ΔG's) and their standard deviations were obtained from at least three experimental measurements (e.g. triplicate) for each inhibitor.

4.2. Molecular modeling and QM calculations

4.2.1. Construction of thrombin-2 and thrombin-3 model complexes

The crystal structure 2ZI2 was downloaded and used to construct the thrombin-**2** model complex. Atoms and molecules other than the protein and the ligand were deleted (e.g., water molecules). Using SYBYL-X software, Tripos Inc., hydrogen atoms were added to both the ligand and the protein. The atom and bond types as well as the protonation statuses for the ligand and the active site residues were checked and modified as necessary. Gasteiger-Huckel charges for the ligand and Kollman-All Atom charges for the protein's atoms were then calculated. This was followed by minimizing the added hydrogen atoms with Tripos force field using the default parameters with the exception of using the charges that were previously loaded on the protein-ligand complex atoms, and using a value of 80 for the dielectric constant. It should be noted that during this minimization procedure, all the heavy atoms were kept constrained in order not to lose the crystallographic information (e.g., ligand binding mode, conformations of flexible amino acid residues, etc.). The ligand's terminal butyryl was then replaced with an acetyl moiety through deleting the γ-CH_3 and the β-CH_2, and converting the α-CH_2 to CH_3. Gasteiger-Huckel charges were recalculated for the ligand; then the acetyl moiety, together with all the hydrogen atoms, was minimized with Tripos force field as described above.

To construct the thrombin-**3** model complex, the acetyl group in thrombin-**2** construct was replaced with methanesulfonyl. Gasteiger-Huckel charges were recalculated for the modified ligand; then the methanesulfonyl, together with all the hydrogen atoms, was minimized with Tripos force field as previously described (all other heavy atoms were kept constrained during minimization). This was followed by generating several conformations for the methanesulfonyl group via rotating the CH_2-N-S=O dihedral angle 30° at a time. Conformations that demonstrated steric clashes with the protein were excluded, while other conformations were minimized as described. For each minimized conformation, the distance NH--O (N to O heteroatom distance) and the angles ∠N-H-O and ∠H-O-S were evaluated, and the model with the conformation that demonstrated the closest H-bond distance and angles to those of the thrombin-**2** model was considered the final thrombin-**3** model complex. It is worth mentioning that the chosen model is the only model that satisfied all of the following criteria: NH--O = 3.00 ± 0.20 Å, ∠N-H-O = 164 ± 10°, and ∠H-O-S = 160 ± 25°, noting that 3.00 Å, 164°, and 160° are the NH--O, ∠N-H-O, and ∠H-O-C of the thrombin-**2** model, respectively.

4.2.2. Construction of QM models and interaction energy calculations

Models A_1 and A_2 were constructed using coordinates derived from the thrombin-**2** model complex. Model B_1 was constructed using coordinates derived from the thrombin-**3** complex. Equilibrium geometries were then calculated for the models using the semi-empirical PM3 method [34] (note: heavy atoms whose optimization was likely to significantly change the crystallographic coordinates were constrained). Models L-P2 were constructed via replacing the Me group in the $MeCON(CH_3)_2$ of model A_2 with the corresponding variable moiety (e.g., FCH_2- F_2CH-, etc.). For each model, several conformations were generated for the variable moiety via rotating the αC-CO bond 30° at a time. Duplicate conformations were ignored; e.g. conformations of the CF_3 in E_2 degenerate into only two conformations. Each conformation was then optimized using PM3, energy was calculated using B3LYP/6-31G* method [35], and the conformation that gives the lowest energy (most stable) model was retained and considered final.

Models C_1-E_1 were constructed form the corresponding L-P2 models via truncating the protein representative formylglycinamide to N-methylformamide. Models F_1 and G_1 were constructed via initially replacing the acetyl in thrombin-**2** with either methoxycarbonyl or methylaminocarbonyl and optimizing these groups, then using the coordinates of resulting protein-ligand complexes to build the QM models, which were further subjected to PM3 geometry optimization. Models L-W (e.g. A_w-H_w) were built using the corresponding L-P1 or L-P2 wherein the NH of the protein representative was used to guide placing the water molecule at an H-bond distance from the ligand. Equilibrium geometries were then calculated using the semi-empirical PM3 method. For each model, the interaction energy was calculated in the following manner: first, total energy was calculated using B3LYP/6-31G* method; second, the two components of each model, A and B, were separated (e.g., each component in a file), and the energy of each component was calculated using B3LYP/6-31G*; and, finally, the interaction energy was calculated using the formula $E_{int} = E_{tot} - (E_a + E_b)$, where E_{tot} is the total energy of the model, E_a is the energy of component A, and E_b is the energy of component B. All the QM calculations were performed using Spartan software, Wavefunction, Inc.

4.3. Chemistry

4.3.1. General methods

Reagents were obtained from commercial suppliers and used without further purification. Anhydrous solvents were purchased as sealed bottles from either Fisher-Acros (AcroSeal) or Aldrich (Sure-seal) and were maintained under an argon atmosphere. Dichloromethane (DCM) was distilled from a calcium hydride still and used immediately. Solvent removal was performed on a rotary evaporator equipped with a 20-60 °C water bath and a self-contained aspirator. Thin-layer chromatography (TLC) was performed on Analtech (Newark, DE) 200 micron Silica Gel F coated on polyethylene sheets. Visualization was accomplished with 254 nm UV light or iodine staining. The silica gel used in the flash chromatography was 40-75 μm flash grade purchased from Sorbent Technologies (Atlanta, GA). All amino acids used are L unless otherwise noted. Proton and carbon nuclear magnetic resonance was performed in deuterated solvents purchased from Cambridge Isotope Laboratories, Inc (Andover, MD) on one of the following instruments: Varian Gemini 300 MHz, Varian Inova 400 MHz, or Varian Inova 500 MHz. ^{1}H NMR data are reported in the following format: chemical shift (ppm values in relation to TMS or the appropriate solvent peak), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublet, dt = doublet of triplet, dq = doublet of quartet, m = multiplet, br s = broad singlet), coupling constant(s), and integration. Whenever fractions of chemically equivalent protons appear at widely-spaced chemical shifts, like when the compound exists in multiple conformations, the chemical shifts are reported; followed by the multiplicity(ies) preceded by the number of peaks (e.g. 2 s, 2 d, etc.), the coupling constant(s), and the sum of the integrations of these peaks. Low resolution ESI mass spectrometry was performed on a Thermo Finnigan LCQ Advantage instrument using 60% methanol in water with 1% acetic acid or 60% acetonitrile in water with 0.1% trifluoroacetic acid as the mobile phase. Preparative and semi-preparative HPLC instrumentation included a Milton Roy gm4000 gradient programmer, Milton Roy Constametric I and III pumps, a Rheodyne 7125 injector with a 5.00 mL sample loop, and a Knauer Variable Wavelength Detector set at 254 nm with a preparative flow cell. The HPLC column used was a Phenomenex LUNA C18(2), 5 μm, 100A pore, 21 mm X 250 mm with Security Guard cartridge used with a flow rate of 8 mL/min. All final compounds used to acquire biological data were at least 95% pure as determined by HPLC analysis.

4.3.2. General procedure for sulfonamide and sulfonylurea formation

A solution of the free amine (or its hydrochloride salt: 1.0 equiv.) and TEA (3.0 equiv.) in anhydrous DCM was prepared, cooled to 0 °C, and stirred using a magnetic stir bar. A 0.36 M solution of either methylsulfonyl chloride or aminosulfonyl chloride in anhydrous DCM was gradually added to the amine solution (note: the amount of the added solution was calculated to deliver 1.2 equiv. of the sulfonyl chloride to the reaction medium). The reaction mixture was allowed to warm up to room temperature. It was then stirred for 8 h, diluted with ethyl acetate (50 mL for every 5 mL DCM) and extracted with 1 M HCl (3X), saturated sodium bicarbonate (3X), and brine (2X). The organic layer was then dried with anhydrous sodium sulfate and the solvent was evaporated under vacuum to give the product. The products of this reaction were used in the next step without further purification.

4.3.3. General procedure for amide coupling

A solution or suspension of the carboxylic acid (1.0 equiv.), the free amine (or its hydrochloride salt: 1.0-1.1 equiv.), EDCI.HCl (1.2 equiv.), and HOBt (1.2 equiv.) in anhydrous DMF was prepared, cooled to 0 °C, and stirred using a magnetic stir bar. DIEA (2.2 equiv. in case of coupling to a free amine and 3.3 equiv. in case of coupling to a hydrochloride salt) was gradually added to this solution/suspension. The reaction mixture was then allowed to warm up to room temperature, and was stirred overnight. This was followed by dilution with ethyl acetate (50 mL for every 5 mL DMF) and extraction with 1 M HCl (3X), saturated sodium bicarbonate (3X), and brine (2X). The organic layer was then dried with anhydrous sodium sulfate and the solvent was evaporated under vacuum to give the product. The products of the amide couplings were purified by flash chromatography or semi-preparative reverse phase HPLC whenever needed.

4.3.4. General procedure for acylation using acid anhydride

Either the free amine or its hydrochloride salt (1.0 equiv.) was dissolved in a sufficient amount of anhydrous pyridine. The solution was then cooled to 0 °C using crushed ice. The acid anhydride (3.0 equiv.) was added to the amine solution gradually, and the reaction mixture was allowed to warm up to room temperature. The mixture was then stirred for 5-7 h. This was followed by dilution with ethyl acetate (50 mL for every 3 mL pyridine) and extraction with 1 M HCl (3X), saturated sodium bicarbonate (3X), and brine (2X). The organic layer was then dried with

anhydrous sodium sulfate and the solvent was evaporated under vacuum to give the product. The products of this reaction were used in the next step without further purification.

4.3.5. General procedure for benzyl ester deprotection

The ester was dissolved in methanol in a Parr flask. The flask was charged with 0.05 equiv. of 10% Pd/C, and was then put on a Parr hydrogenation apparatus. The flask was subjected to 3 charge/purge cycles with hydrogen gas, and was then charged with 50-55 psi hydrogen. The reaction mixture was shaken overnight, then filtered using Celite. The flask and the Celite were washed several times with methanol, and the methanol fractions were combined and evaporated under vacuum to give the desired product, which did not normally need any further purification.

4.3.6. General procedure for the conversion of benzonitrile derivatives to *N*-hydroxybenzamidines

The benzonitrile derivative (1.0 equiv.) and hydroxylamine hydrochloride (3.0 equiv.) were dissolved in anhydrous methanol and stirred for 15 min. DIEA (3.0 equiv.) was then added gradually to the reaction mixture, and the mixture was stirred at room temperature overnight. Upon the completion of the reaction, methanol was removed *in vacuo*, and the residue was taken up in *n*-butanol. The *n*-butanol solution was washed with an equal amount of water (3X). Acetonitrile was then added to the *n*-butanol fraction (three times the amount of the *n*-butanol), and the solvents were removed under vacuum to give the desired product which was purified with semi-preparative reverse phase HPLC. Alternatively, upon reaction completion, the mixture was applied directly to semi-preparative reverse phase HPLC to separate the product in a pure form.

4.3.7. General procedure for the conversion of *N*-hydroxybenzamidines to benzamidines using hydrogenation

The *N*-hydroxybenzamidine (1.0 equiv.) was dissolved in glacial acetic acid. Acetic anhydride (3.0 equiv.) was then added, and the reaction mixture was stirred for 30 min. The mixture was then transferred to a barr flask, which was charged with 0.05 equiv. of 10% Pd/C, and was then put on a Parr hydrogenation apparatus. The flask was subjected to 3 charge/purge cycles with hydrogen gas and then charged with 50-55 psi hydrogen. The reaction mixture was shaken

overnight and then filtered using Celite to remove the Pd/C. Acetonitrile was then added to the mixture (3 x the amount of the glacial acetic acid), and the solvents were removed under vacuum to give the desired product, which was purified by reverse phase HPLC (e.g. using a mobile phase that has 1% acetic acid).

4.3.8. Synthesis of compounds 1, 2, 3, 8, 11, 13, and 15

4.3.8.1. (*S*)-1-Ethylpyrrolidine-2-carboxylic acid (1i)

The commercially available L-proline benzyl ester hydrochloride (604 mg, 2.5 mmol) was dissolved in 5.0 ml DMF. Anhydrous potassium carbonate (1.04 g, 7.5 mmol) and ethyl iodide (468 mg, 3.0 mmol) were added, and the mixture was heated to 80˚C and stirred overnight. Upon reaction completion, the mixture was diluted with ethyl acetate (50 ml) and washed with water (2X) and brine (2X). The organic layer was then dried over anhydrous sodium sulfate and concentrated under vacuum. The residue was purified with reverse phase HPLC to give 420 mg of the product (yield: 72%). The product obtained from this reaction, which is the benzyl ester of **1i**, was subjected to hydrogenation to remove the benzyl group according to the general procedure for benzyl ester deprotection. 248 mg of the desired product (**1i**) was obtained and used without further purification (96%). ^{1}H NMR (DMSO-d_6) δ 1.14 (m, 3H), 1.80 (m, 2H), 2.02-2.41 (m, 4H), 2.75 (m, 2H), 4.18 (m, 1H), 10.70 (br, 1H); *m/z* (LCMS, ESI): found 144.1 $[M + H]^+$; $[C_7H_{13}NO_2 + H]^+$ requires 144.1

4.3.8.2. (*S*)-*N*-(4-Carbamimidoylbenzyl)-1-Ethylpyrrolidine-2-carboxamide diacetate (1)

Following the general procedure for amide coupling, with only slight modifications, compound **1i** (243 mg, 1.7 mmol) was reacted with 4-(aminomethyl)benzonitrile hydrochloride (320 mg, 1.9 mmol) in anhydrous DMF (5.0 mL), using EDCI.HCl (384 mg, 2.0 mmol) and HOBt (270 mg, 2.0 mmol) as coupling reagents, and DIEA (720 mg, 5.6 mmol) as a base. Modifications in the general procedure for amide coupling included using *n*-butanol instead of ethyl acetate to dilute the reaction mixture, and performing the extraction with 1 M HCl (2X) and water (2X). 301 mg of the benzonitrile derivative was obtained and used without further purification (yield: 69%). This product was then reacted with hydroxylamine hydrochloride (243 mg, 3.5 mmol) in the presence of DIEA (450 mg, 3.5 mmol) according to the general procedure for the conversion

of benzonitrile derivatives to *N*-hydroxybenzamidines. The *N*-hydroxybenzamidine (232 mg, 0.8 mmol) was then reacted with acetic anhydride (260 mg, 2.5 mmol) and hydrogenated according to the general procedure for the conversion of *N*-hydroxybenzamidines to benzamidines. 258 mg of the final product was obtained after purification with reverse phase HPLC (yield: 82%). ^{1}H NMR (CD_3OD) δ 1.39 (m, 3H) 1.81 (s, 6H), 2.20 (m, 3H), 2.59 (m, 1H), 3.24 (m, 2H) 3.75 (m, 2H) 4.18 (m, 1H) 4.59 (s, 2H) 7.58 (d, J = 8.4 Hz, 2H) 7.82 (d, J = 8.4 Hz, 2H). ^{13}C NMR (CD_3OD) δ 9.96, 22.6, 29.5, 42.5, 50.3, 54.2, 67.1, 127.9, 128.0, 144.8, 167.9, 174.1; *m/z* (LCMS, ESI): found 275.2 $[M + H]^+$; $[C_{15}H_{22}N_4O + H]^+$ requires 275.2

4.3.8.3. (*S*)-1-Acetyl-pyrrolidine-2-carboxylic acid (2i)

Following the general procedure for acylation, the commercially available L-proline benzyl ester hydrochloride (604 mg, 2.5 mmol) was reacted with acid anhydride (770 mg, 7.5 mmol) in anhydrous pyridine (4.0 ml). 315 mg of the product was obtained (yield: 51%). The product obtained from this reaction, which is the benzyl ester of **2i**, was subjected to hydrogenation to remove the benzyl group according to the general procedure for benzyl ester deprotection. 194 mg of the desired product (**2i**) was obtained and used without further purification (97%).^{1}H NMR (CD_3OD) δ 2.06 (m, 6H), 2.27 (m, 1H), 3.50-3.69 (m, 2H), 4.43 (m, 1H); *m/z* (LCMS, ESI): found 158.1 $[M+H]^+$, $[C_7H_{11}NO_3 + H]^+$ requires 158.1

4.3.8.4. (*S*)-1-Acetyl-*N*-(4-Carbamimidoylbenzyl) pyrrolidine-2-carboxamide acetate (2)

Following the general procedure for amide coupling, with only slight modifications, compound **2i** (188 mg, 1.2 mmol) was reacted with 4-(aminomethyl)benzonitrile hydrochloride (222 mg, 1.3 mmol) in anhydrous DMF (5.0 mL), using EDCI.HCl (276 mg, 1.4 mmol) and HOBt (194 mg, 1.4 mmol) as coupling reagents, and DIEA (510 mg, 4.0 mmol) as a base. Modifications in the general procedure for amide coupling included using *n*-butanol instead of ethyl acetate to dilute the reaction mixture, and performing the extraction with 1 M HCl (2X) and water (2X). 260 mg of the benzonitrile derivative was obtained and used without further purification (yield: 80%). The product was then reacted with hydroxylamine hydrochloride (200 mg, 2.9 mmol) in the presence of DIEA (370 mg, 2.9 mmol) according to the general procedure for the conversion of benzonitrile derivatives to *N*-hydroxybenzamidines. The *N*-hydroxybenzamidine (213 mg, 0.7 mmol) was then reacted with acetic anhydride (220 mg, 2.1 mmol) and hydrogenated according

to the general procedure for the conversion of *N*-hydroxybenzamidines to benzamidines. 206 mg of the final product was obtained after purification with reverse phase HPLC (yield: 85%). ^{1}H NMR (DMSO-d_6) δ 1.70 (s, 3H), 1.80 and 2.00 (2s, together 3H), 1.87 (m, 3H), 2.10 (m, 1H), 3.48 (m, 1H), 3.62 (m, 1H), 4.28 (m, 1H), 4.36 (m, 2H), 7.44 (d, J = 7.2 Hz, 2H); 7.74 (d, J = 7.2 Hz, 2H), 8.42 and 8.75 (2t, J = 6.4 Hz, together 1H), 9.50-10.5 (brs, 4H) ; ^{13}C NMR (CD_3OD) δ 19.8, 24.9, 30.8, 42.2, 61.8, 127.3, 128.4, 128.6, 146.0, 167.1, 172.4, 173.0; *m/z* (LCMS, ESI): found 289.2 $[M + H]^+$; $[C_{15}H_{20}N_4O_2 + H]^+$ requires 289.2

4.3.8.5. (*S*)-1-(Methanesulfonyl) pyrrolidine-2-carboxylic acid (3i)

Following the general procedure for sulfonamide formation, the commercially available L-proline benzyl ester hydrochloride (483 mg, 2.0 mmol) was reacted with methanesulfonyl chloride (6.5 mL of the 0.36 M DCM solution, 2.3 mmol) in the presence of TEA (610 mg, 6.0 mmol). 362 mg of the product was obtained after purification with reverse phase HPLC (yield: 64%). The product obtained from this reaction, which is the benzyl ester of **3i**, was subjected to hydrogenation to remove the benzyl group according to the general procedure for benzyl ester deprotection. 238 mg of the desired product (**3i**) was obtained and used without further purification (96%). ^{1}H NMR (DMSO-d_6) 1.88 (m, 3H), 2.00 (m, 1H), 2.96 (s, 3H), 3.34 (t, J = 6.3 Hz, 2H), 4.22 (m, 1H), 12.70 (brs, 1H); *m/z* (LCMS, ESI): found 192.0 $[M - H]^-$; $[C_6H_{11}NO_4S - H]^-$ requires 192.0

4.3.8.6. (*S*)-*N*-(4-Carbamimidoylbenzyl)-1-(methanesulfonyl) pyrrolidine-2-carboxamide acetate (3)

Following the general procedure for amide coupling, with only slight modifications, compound **3i** (193 mg, 1.0 mmol) was reacted with 4-(aminomethyl)benzonitrile hydrochloride (185 mg, 1.1 mmol) in anhydrous DMF (5.0 mL), using EDCI.HCl (230 mg, 1.2 mmol) and HOBt (162 mg, 1.2 mmol) as coupling reagents, and DIEA (430 mg, 3.3 mmol) as a base. Modifications in the general procedure for amide coupling included using *n*-butanol instead of ethyl acetate to dilute the reaction mixture, and performing the extraction with 1 M HCl (2X) and water (2X). 230 mg of the benzonitrile derivative was obtained and used without further purification (yield: 75%). This product was then reacted with hydroxylamine hydrochloride (156 mg, 2.3 mmol) in the presence of DIEA (290 mg, 2.3 mmol) according to the general procedure for the conversion

of benzonitrile derivatives to *N*-hydroxybenzamidines. The *N*-hydroxybenzamidine (204 mg, 0.6 mmol) was then reacted with acetic anhydride (180 mg, 1.8 mmol) and hydrogenated according to the general procedure for the conversion of *N*-hydroxybenzamidines to benzamidines. 198 mg of the final product was obtained after purification with reverse phase HPLC (yield: 86%). ^{1}H NMR (CD_3OD) 1.93 (s, 3H), 2.05 (m, 3H), 2.35 (m, 1H), 2.98 (s, 3H), 3.46 (m, 1H), 3.58 (m, 1H), 4.27 (m, 1H), 4.48 (d, J = 16.0 Hz, 1H), 4.61 (d, J = 16.0 Hz, 1H), 7.57 (d, J = 8.0 Hz, 2H), 7.78 (d, J = 8.0 Hz, 2H), ^{13}C NMR (CD_3OD) 23.23, 24.93, 31.51, 33.94, 42.56, 49.51, 62.60, 127.33, 127.95, 128.02, 145.94, 167.50, 174.15; *m/z* (LCMS, ESI): found 325.2 $[M + H]^+$; $[C_{14}H_{20}N_4O_3S + H]^+$ requires 325.1

4.3.8.7. (*S*)-1-(Trifluoromethanesulfonyl) pyrrolidine-2-carboxylic acid (8i)

The commercially available L-proline *tert*-butyl ester (342 mg, 2.0 mmol) was dissolved in 3.0 ml anhydrous DCM. TEA (610 mg, 6.0 mmol) was added to the solution, which was then cooled to 0°C and stirred. Trifluoromethanesulfonyl chloride (6.5 mL of a 0.36 M DCM solution, 2.3 mmol) was then added gradually to the reaction mixture under inert atmosphere. The mixture was allowed to warm up to room temperature and was then stirred overnight. Upon reaction completion, the solvent was evaporated and the residue was dissolved in acetonitrile and purified by reverse phase HPLC to give 315 mg of the product (yield: 52%). The product, which is the *tert*-butyl ester of **8i**, was then subjected to deprotection of the *tert*-butyl ester using trifluoroacetic acid (TFA). This was carried out by dissolving the ester in 5.0 ml of 50% TFA/DCM solution, and stirring the solution for 5 h. The solution was then evaporated, and the residue was lyophilized and used without further purification (233 mg of **8i** was obtained, 91%); *m/z* (LCMS, ESI): found 246.0 $[M - H]^-$; $[C_6H_8F_3NO_4S - H]^-$ requires 246.0

4.3.8.8. (*S*)-*N*-(4-Carbamimidoylbenzyl)-1-(trifluoromethanesulfonyl) pyrrolidine-2-carboxamide acetate (8)

Following the general procedure for amide coupling, compound **8i** (222 mg, 0.9 mmol) was reacted with 4-(aminomethyl)benzonitrile hydrochloride (168 mg, 1.0 mmol) in anhydrous DMF (5.0 mL), using EDCI.HCl (210 mg, 1.1 mmol) and HOBt (149 mg, 1.1 mmol) as coupling reagents, and DIEA (390 mg, 3.0 mmol) as a base. Extraction was not used to separate the product; rather, the reaction mixture was diluted with acetonitrile and was applied directly to

reverse phase HPLC in order to separate the benzonitrile derivative in a pure form (225 mg, yield: 69%). The product was then reacted with hydroxylamine hydrochloride (132 mg, 1.9 mmol) in the presence of DIEA (250 mg, 1.9 mmol) according to the general procedure for the conversion of benzonitrile derivatives to *N*-hydroxybenzamidines. The *N*-hydroxybenzamidine (158 mg, 0.4 mmol) was then reacted with acetic anhydride (120 mg, 1.2 mmol) and hydrogenated according to the general procedure for the conversion of *N*-hydroxybenzamidines to benzamidines. 144 mg of the final product was obtained after purification with reverse phase HPLC (yield: 82%). ^{1}H NMR (CD_3OD) 1.91 (s, 3H), 2.15 (m, 3H), 2.41 (m, 1H), 3.68 (m, 1H), 3.78 (m, 1H), 4.46 (d, J = 16.0 Hz, 1H), 4.51 (m, 1H), 4.60 (d, J = 16.0 Hz, 1H), 7.57 (d, J = 8.0 Hz, 2H), 7.79 (d, J = 8.0 Hz, 2H); *m/z* (LCMS, ESI): found 379.1 $[M + H]^+$; $[C_{14}H_{17}F_3N_4O_3S + H]^+$ requires 379.1

4.3.8.9. (*S*)-1-(3-Phenylpropanoyl) pyrrolidine-2-carboxylic acid (11i) [1]

Following the general procedure for amide coupling, L-proline benzyl ester hydrochloride (507 mg, 2.1 mmol) was reacted with 3-phenylpropionic acid (300 mg, 2.0 mmol) in anhydrous DMF (10 mL), using EDCI.HCl (460 mg, 2.4 mmol) and HOBt (324 mg, 2.4 mmol) as coupling reagents, and DIEA (850 mg, 6.6 mmol) as a base. 479 mg of the product was obtained after purification with flash chromatography (yield: 71%). The product obtained from this reaction, which is the benzyl ester of **11i**, was subjected to hydrogenation to remove the benzyl group according to the general procedure for benzyl ester deprotection. 337 mg of the desired product (**11i**) was obtained and used without further purification (96%); *m/z* (LCMS, ESI): found 246.2 $[M - H]^-$; $[C_{14}H_{17}NO_3 - H]^-$ requires 246.1

4.3.8.10. (*S*)-*N*-(4-Carbamimidoylbenzyl)-1-(3-phenylpropanoyl) pyrrolidine-2-carboxamide acetate (11) [1]

Following the general procedure for amide coupling, compound **11i** (247 mg, 1.0 mmol) was reacted with 4-(aminomethyl)benzonitrile hydrochloride (185 mg, 1.1 mmol) in anhydrous DMF (5 mL), using EDCI.HCl (230 mg, 1.2 mmol) and HOBt (162 mg, 1.2 mmol) as coupling reagents, and DIEA (430 mg, 3.3 mmol) as a base. 288 mg of the product, which is the benzonitrile derivative, was obtained after purification with reverse phase HPLC (yield: 80%). This product was then reacted with hydroxylamine hydrochloride (166 mg, 2.4 mmol) in the

presence of DIEA (310 mg, 2.4 mmol) according to the general procedure for the conversion of benzonitrile derivatives to *N*-hydroxybenzamidines. The *N*-hydroxybenzamidine (197 mg, 0.5 mmol) was then reacted with acetic anhydride (150 mg, 1.5 mmol) and hydrogenated according to the general procedure for the conversion of *N*-hydroxybenzamidines to benzamidines. 188 mg of the final product was obtained after purification with reverse phase HPLC (yield: 86%); ^{1}H NMR (DMSO-d_6) 1.70 (s, 3H), 1.85 (m, 3H), 2.02 and 2.13 (2m, together 1H), 2.62 (t, J = 7.2 Hz, 2H), 2.81 (t, J = 7.2 Hz, 2H), 3.41 (m, 1H), 3.54 (m, 1H), 4.32 (m, 3H), 7.15 (m, 2H), 7.23 (m, 3H), 7.33 and 7.41 (2d, J = 8.4 Hz, together 2H), 7.65 and 7.72 (2d, J = 8.4 Hz, together 2H), 8.41 and 8.70 (2 t, J = 6.0 Hz, together 1H), 9.60-10.90 (brs, 4H); *m/z* (LCMS, ESI): found 379.2 $[M + H]^+$; $[C_{22}H_{26}N_4O_2 + H]^+$ requires 379.2

4.3.8.11. (*S*)-1-(Benzyloxycarbonyl)-*N*-(4-carbamimidoylbenzyl) pyrrolidine-2-carboxamide acetate (13)

Following the general procedure for amide coupling, 249 mg (1.0 mmol) of Cbz-L-proline (**13i**) was reacted with 4-(aminomethyl)benzonitrile hydrochloride (185 mg, 1.1 mmol) in anhydrous DMF (5.0 mL), using EDCI.HCl (230 mg, 1.2 mmol) and HOBt (162 mg, 1.2 mmol) as coupling reagents, and DIEA (430 mg, 3.3 mmol) as a base. 283 mg of the product, which is the benzonitrile derivative, was obtained after purification with reverse phase HPLC (yield: 78%). This product was then reacted with hydroxylamine hydrochloride (166 mg, 2.4 mmol) in the presence of DIEA (310 mg, 2.4 mmol) according to the general procedure for the conversion of benzonitrile derivatives to *N*-hydroxybenzamidines. The *N*-hydroxybenzamidine (198 mg, 0.5 mmol) was then dissolved in 4.0 ml glacial acetic acid. 163 mg of zinc dust (2.5 mmol) was added to the solution at room temperature, and the mixture was stirred at 60°C overnight. The reaction mixture was then filtered, and the solution was concentrated under vacuum. The concentrate was dissolved in 50% acetonitrile-water containing 1% glacial acetic acid and purified by reverse phase HPLC to give 143 mg of the desired product (yield: 65%); ^{1}H NMR (DMSO-d_6) 1.70 (s, 3H), 1.82 (m, 3H), 2.15 (m, 1H), 3.42 (m, 2H), 4.18-4.32 (m, 3H), 4.95-5.12 (m, 2H), 7.25-7.43 (m, 7H), 7.64 (d, J = 7.2 Hz, 1H), 7.72 (d, J = 7.1 Hz, 1H), 8.58 (m, 1H), 9.40-10.60 (brs, 4H); *m/z* (LCMS, ESI): found 381.2 $[M + H]^+$; $[C_{21}H_{24}N_4O_3 + H]^+$ requires 381.2

4.3.8.12. (*S*)-1-(Benzylcarbamoyl) pyrrolidine-2-carboxylic acid (15i)

To a stirred solution of L-proline benzyl ester hydrochloride (483 mg, 2.0 mmol) and benzyl isocyanate (293 mg, 2.2 mmol) in anhydrous DCM (8.0 mL) was added TEA (610 mg, 6.0 mmol) gradually. The reaction mixture was then stirred at room temperature for 5 h, diluted with ethyl acetate (120 mL) and extracted with 1 M HCl (3X), saturated sodium bicarbonate (3X), and brine (2X). The organic layer was then dried over anhydrous sodium sulfate, and the solvent was evaporated under vacuum to give the product, which was purified by reverse phase HPLC (472 mg, yield: 70%). This product (i.e., the benzyl ester of **15i**) was subjected to hydrogenation to remove the benzyl group according to the general procedure for benzyl ester deprotection. 323 mg of the desired product (**15i**) was obtained and used without further purification (93%); *m/z* (LCMS, ESI): found 247.2 $[M - H]^-$; $[C_{13}H_{16}N_2O_3 - H]^-$ requires 247.1

4.3.8.13. (*S*)-1-(Benzylcarbamoyl)-*N*-(4-carbamimidoylbenzyl) pyrrolidine-2-carboxamide acetate (15)

Following the general procedure for amide coupling, compound **15i** (248 mg, 1.0 mmol) was reacted with 4-(aminomethyl)benzonitrile hydrochloride (185 mg, 1.1 mmol) in anhydrous DMF (5.0 mL), using EDCI.HCl (230 mg, 1.2 mmol) and HOBt (162 mg, 1.2 mmol) as coupling reagents, and DIEA (430 mg, 3.3 mmol) as a base. 286 mg of the product, which is the benzonitrile derivative, was obtained after purification with reverse phase HPLC (yield: 79%). This product was then reacted with hydroxylamine hydrochloride (166 mg, 2.4 mmol) in the presence of DIEA (310 mg, 2.4 mmol) according to the general procedure for the conversion of benzonitrile derivatives to *N*-hydroxybenzamidines. The *N*-hydroxybenzamidine (198 mg, 0.5 mmol) was then reacted with acetic anhydride (150 mg, 1.5 mmol) and hydrogenated according to the general procedure for the conversion of *N*-hydroxybenzamidines to benzamidines. 191 mg of the final product was obtained after purification with reverse phase HPLC (yield: 87%); ^{1}H NMR (DMSO-d_6) 1.69 (s, 3H), 1.85 (m, 3H), 2.03 (m, 1H), 3.27 (m, 1H), 3.46 (m, 1H), 4.16-4.38 (m, 5H), 6.82 (t, J = 6.8 Hz, 1H), 7.18 (m, 1H), 7.26 (m, 4H), 7.40 (d, J = 7.0 Hz, 2H), 7.69 (d, J = 7.0 Hz, 2H), 8.41 (t, J = 6.9 Hz, 1H), 9.20-10.80 (brs, 4H); *m/z* (LCMS, ESI): found 380.2 $[M + H]^+$; $[C_{21}H_{25}N_5O_2 + H]^+$ requires 380.2

Supplementary Data

Supplementary discussions of the correlation between protein-ligand interaction energy and binding affinity, and of functional group cooperativity in relation to bioisosterism are included. Also the synthesis and characterization of compounds **4**, **5**, **6**, **7**, **9**, **10**, **12**, and **14**, and their intermediates, as well as representative NMR spectra for compounds **3**, **6**, **7**, **12**, **14**, and **15** are included.

Corresponding Author

* Correspondences can be addressed to Nader Nasief at nnnasief@yahoo.com or David Hangauer at hangauer@buffalo.edu.

ABBREVIATIONS

Bn: benzyl, *tert*-Bu: tertiary butyl, DCM: dichloromethane; DIEA: diisopropylethylamine, DMF: dimethylformamide, EDCI.HCl: 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride, HOAc: acetic acid, HOBt: hydroxybenzotriazole, HPLC: high pressure liquid chromatography, Me: methyl, MeOH: methanol, QM: quantum mechanical, SAR: structure-activity relationship, TEA: triethylamine, TFA: trifluoroacetic acid.

REFERENCES

[1] L. Muley, B. Baum, M. Smolinski, M. Freindorf, A. Heine, G. Klebe, D. Hangauer, **Enhancement of hydrophobic interactions and hydrogen bond strength by cooperativity: Synthesis, modeling and molecular dynamics simulations of a congeneric series of thrombin inhibitors.** J. Med. Chem. 53 (2010) 2126-2135.

[2] N.N. Nasief, H. Tan, J. Kong, D. Hangauer, **Water mediated ligand functional group cooperativity: The contribution of a methyl group to binding affinity is enhanced by a COO^- group through changes in the structure and thermodynamics of the hydration waters of ligand-thermolysin complexes.** J. Med. Chem. 55 (2012) 8283-8302.

[3] N.N. Nasief, D. Hangauer, **Additivity or cooperativity: Which model can predict the influence of simultaneous incorporation of two or more functionalities in a ligand molecule?** Eur. J. Med. Chem. 90 (2015), 897-915.

[4] A.M. Said, D.G. Hangauer, **Binding cooperativity between a ligand carbonyl group and a hydrophobic side chain can be enhanced by additional H-bonds in a distance dependent manner: A case study with thrombin inhibitors.** Eur. J. Med. Chem. 96 (2015), 405-424.

[5] A.M. Said, D.G. Hangauer, **Ligand binding cooperativity: Bioisosteric replacement of C=O with SO2 among thrombin inhibitors**, Bioorg. Med. Chem. Lett. 26 (2016), 3850-3854.

[6] L. Englert, A. Biela, M. Zayed, A. Heine, D. Hangauer, G. Klebe, **Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: Binding of phosphonamidate to the S_1'-pocket of thermolysin.** Biochem. et Biophys. Acta. 1800 (2010), 1192-1202.

[7] N.N. Nasief, D. Hangauer, **Influence of neighboring groups on the thermodynamics of hydrophobic binding: An added complex facet to the hydrophobic effect.** J. Med. Chem. 57 (2014), 2315-2333.

[8] M.N.D. Di Minno, S. Momi, A. Di Minno, A. Russolillo, **Stroke prevention: From available antiplatelet drugs to novel molecular targets.** Curr. Drug Targ. 14 (2013), 3-12.

[9] A. Straub, S. Roehrig, A. Hillisch, **Oral, direct thrombin and factor Xa inhibitors: The replacement for warfarin, leeches, and pig intestines?** Angew. Chem. Int. Ed. 50 (2011), 4574–4590.

[10] A.Y. Mehta, Y. Jin, U.R. Desai, **An update on recent patents on thrombin inhibitors (2010–2013).** Expert Opin. Ther. Pat. (24) 2014, 47–67.

[11] L-W. He, W-C. Dai, N-G. Li, **Development of orally active thrombin inhibitors for the treatment of thrombotic disorder diseases.** Molecules 20 (2015), 11046-11062.

[12] J. Stangier, K. Rathgen, H. Stähle, D. Gansser, W. Roth, **The pharmacokinetics, pharmacodynamics and tolerability of dabigatran etexilate, a new oral direct thrombin inhibitor, in healthy male subjects.** Br. J. Clin. Pharmacol. 64 (2007), 292–303.

[13] L.M. Lima, E.J. Barreiro, **Bioisosterism: A useful strategy for molecular modification and drug design.** Curr. Med. Chem. 12 (2005), 23-49.

[14] N.A. Meanwell, **Synopsis of some recent tactical application of bioisosteres in drug design.** J. Med. Chem. 54 (2011), 2529–2591.

[15] C.W. Thornber, **Isosterism and molecular modification in drug design.** Chem. Soc. Rev. 8 (1979), 563-580.

[16] A. Burger, **Medicinal chemistry, 3rd Ed.,** Wiley, New York, 1970, pp. 64-80.

[17] A. Korolkovas, **Essentials of molecular pharmacology: Background for drug design,** Wiley, New York, 1970, pp. 54-57.

[18] C.A. Lipinski, **Bioisosterism in drug design.** Annu. Rep. Med. Chem. 21 (1986), 283-291.

[19] H. Erlenmeyer, M. Leo, **On pseudoatoms.** Helv. Chim. Acta 15 (1932), 1171-1186.

[20] G.A. Patani, E.J. LaVoie, **Bioisosterism: A rational approach in drug design.** Chem. Rev. 96 (1996), 3147-3176.

[21] R.W. Taft, D. Gurka, L. Joris, P.v.R. Schleyer, J.W. Rakshys, **Studies of hydrogen-bonded complex formation with p-fluorophenol. V. Linear free energy relationships with OH reference acids.** J. Am. Chem. Soc. 91 (1969), 4801-4808.

[22] M.H. Abraham, P.P. Duce, D.V. Prior, D.G. Barratt, J.J. Morris, P.J. Taylor, **Hydrogen bonding. Part 9. Solute proton-donor and proton-acceptor scales for use in drug design.** J. Chem. Soc. Perkin Trans. 2, 1989, 1355-1375.

[23] M.H. Abraham, P.L. Grellier, D.V. Prior, J.J. Morris, P.J. Taylor, **Hydrogen bonding. Part 10. A scale of solute hydrogen-bond basicity using log K values for complexation in tetrachloromethane.** J. Chem. Soc. Perkin Trans. 2, 1990, 521-529.

[24] M.H. Abraham, J.A. Platts, **Hydrogen bond structural group constants.** J. Org. Chem. (66) 2001, 3484-3491.

[25] C. Laurence, K.A. Brameld, J. Graton, J-Y Le Questel, E. Renault, **The pK_{BHX} database: Toward a better understanding of hydrogen-bond basicity for medicinal chemists.** J. Med. Chem. (52) 2009, 4073-4086.

[26] M.A. Williams, J.E. Ladbury, **Hydrogen bonds in protein-ligand complexes.** Methods Princ. Med. Chem. (19) 2003, 137-161.

[27] L. Muley, **In depth analysis of various aspects of thrombin-ligand binding energy.** Dissertation, State University of New York at Buffalo, 2009, pp. 27-31.

[28] J. Sturzebecher, U. Sturzebecher, H. Vieweg, G. Wagner, J. Hauptmann, F. Markwardt, **Synthetic inhibitors of bovine factor Xa and thrombin comparison of their anticoagulant efficiency.** Thromb. Res. 54 (1989), 245–252.

[29] C.M. Wells, E. Di Cera, **Thrombin is a Na^+-activated enzyme.** Biochem 31 (1992), 11721-11730.

[30] Y. Ayala, E. Di Cera, **Molecular recognition by thrombin. Role of the slow→fast transition, site specific ion binding energetics and thermodynamic mapping of structural components.** J. Mol. Biol. 235 (1994), 733-746.

[31] Y. Cheng, W.H. Prusoff, **Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction.** Biochem. Pharmacol. 22 (1973), 22, 3099–3108.

[32] C. Bissantz, B. Kuhn, M. Stahl, **A medicinal chemist's guide to molecular interactions.** J. Med. Chem. 53 (2010), 5061-5084.

[33] It is important to note that although N-methylformamide can exist in equilibrium with a corresponding formamidic acid tautomer, only the amide tautomer was used in calculations because the imidic acid tautomer is predicted to be significantly less stable than the amide as in: X-C. Wang, J. Nichols, M. Feyereisen, M. Gutowski, J. Boatz, A.D.J. Haymet, J. Simons, **Ab initio quantum chemistry study of formamide-formamidic acid tautomerization.** J. Phys. Chem. 95 (1991), 10419-10424.

[34] (a) J.J.P. Stewart, **Optimization of parameters for semiempirical methods I. Method.** J. Comput. Chem. 10 (1989), 209-220. (b) J.J.P. Stewart, **Optimization of parameters for semiempirical methods II. Applications.** J. Comput. Chem. 10 (1989), 221-264.

[35] (a) A.D. Becke, **Density-functional thermochemistry. III. The role of exact exchange.** J. Chem. Phy. 98 (1993), 5648-5652. (b) C. Lee, W. Yang, R.G. Parr, **Development of the Colle-Salvetti conelation energy formula into a functional of the electron density.** Phys. Rev. B 37 (1988), 785-789. (c) P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, **Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields.** J. Phys.Chem. 98 (1994), 11623-11627. (d) W.J. Hehre, R. Ditchfield, J.A. Pople, **Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules.** J. Chem. Phys. 56 (1972), 2257-2261.

[36] C.A. Hunter, **Quantifying intermolecular interactions: guidelines for the molecular recognition toolbox.** Angew. Chem., Int. Ed. 43 (2004), 5310–5324.

[37] B.P. Morgan, J.M. Scholtz, M.D. Ballinger, I.D. Zipkin, P.A. Bartlett, **Differential binding energy: A detailed evaluation of the influence of hydrogen-bonding and hydrophobic groups on the inhibition of Thermolysin by phosphorus-containing inhibitors.** J. Am. Chem. Soc. 113 (1991), 297-307.

[38] S.H. Kawai, M.D. Bailey, T. Halmos, P. Forgione, S.R. LaPlante, M. Llinas-Brunet, J. Naud, N. Goudreau, **The use of chemical double-mutant cycles in biomolecular recognition studies: Application to HCV NS3 protease inhibitors.** ChemMedChem. 3 (2008), 1654 – 1657.

[39] T. Lübbers, M. Böhringer, L. Gobbi, M. Hennig, D. Hunziker, B. Kuhn, B. Löffler, P. Mattei, R. Narquizian, J.U. Peters, Y. Ruff, H.P. Wessel, P. Wyss, **1,3-Disubstituted 4-Aminopiperidines as Useful Tools in the Optimization of the 2-Aminobenzo[a]quinolizine Dipeptidyl Peptidase IV Inhibitors.** Bioorg. Med. Chem. Lett. 17 (2007), 2966– 2970.

[40] M.S. Malamas, E.S. Manas, R.E. McDevitt, I. Gunawan, Z.B. Xu, M.D. Collini, C.P. Miller, T. Dinh, R.A. Henderson, J.C. Keith, H.A. Harris, **Design and Synthesis of Aryl Diphenolic Azoles as Potent and Selective Estrogen Receptor-B Ligands.** J. Med. Chem. 47 (2004), 5021– 5040.

[41] A. Gangjee, A.P. Vidwans, A. Vasudevan, S.F. Queener, R.L. Kisliuk, V. Cody, R. Li, N. Galitsky, J.R. Luft, W. Pangborn, **Structure-Based Design and Synthesis of Lipophilic 2,4-Diamino-6-Substituted Quinazolines and Their Evaluation as Inhibitors of Dihydrofolate Reductases and Potential Antitumor Agents.** J. Med. Chem. 41 (1998), 3426– 3434.

[42] V. Oza, S. Ashwell, P. Brassil, J. Breed, J. Ezhuthachan, C. Deng, M. Grondine, C. Horn, D. Liu, P. Lyne, N. Newcombe, M. Pass, J. Read, M. Su, D. Toader, D. Yu, Y. Yu, S. Zabludoff, **Synthesis and Evaluation of Triazolones as Checkpoint Kinase 1 Inhibitors.** Bioorg. Med. Chem. Lett. 22 (2012), 2330– 2337.

[43] U. Obst, D.W. Banner, L. Weber, F. Diederich, **Molecular recognition at the thrombin active site: Structure based design and synthesis of potent and selective thrombin inhibitors and the X-ray crystal structures of two thrombin-inhibitors complexes.** Chem. Biol. 4 (1997), 287-295.

[44] N. Balasubramanian, D.R. St. Laurent, M.E. Federici, N.A. Meanwell, J.J. Wright, W.A. Schumacher, S.M. Seiler, **Active site directed synthetic thrombin inhibitors: synthesis, in-vitro and in-vivo activity profile of BMY 44621 and analog[s. An examination of the role of the amino group in the D-Phe-Pro-Arg-H series.** J. Med. Chem. 36 (1993), 300-303.

[45] S.I. Klein, J.M. Dener, B.F. Molino, C.J. Gardner, R.D'Alisa, C.T. Dunwiddie, C. Kasiewski, R.J. Leadley, **O-Benzyl hydroxyproline as a bioisostere for Phe-Pro: Novel dipeptide thrombin inhibitors.** Bioorg. Med. Chem. Lett. 6 (1996), 2225-2230.

Figure, Scheme, and Table Captions

Figure 1: A schematic representation of the general scaffold of previously reported benzamidine thrombin inhibitors [1][4], binding in the thrombin catalytic site. The C=O group which accepts an H-bond from the Gly216 residue is shown in blue and enclosed in a blue rectangle. This C=O group can be replaced by other bioisosteric groups such as SO_2. The hydrophobic pockets of the thrombin active sites are shown as curves and indicated as S1, S2, and S3.

Figure 2: Calculation of the binding free energy contribution of the C=O group that forms an H-bond with Gly216. This contribution was found to be -3.2 kJ/mol.

Figure 3: A) ligand **2** bound to the active site of thrombin. The crystal structure PDB # 2ZI2 was used to construct the thrombin-**2** complex. **B)** Ligand **3** bound to the active site of thrombin. Thrombin-**2** complex, along with molecular modeling, was utilized to construct the thrombin-**3** complex. The geometric parameters of the H-bonds accepted by ligands **2** and **3** from Gly216 are shown (Each of the distances shown is between the oxygen H-bond acceptor and the nitrogen H-bond donor heteroatoms). In thrombin-**3** complex, note that one of the SO_2 oxygens is H-bonded to Gly216, and the other is exposed to the solvent. Also, note that the Me of the $MeSO_2$ does not make significant contacts with any of the protein residues.

Figure 4: Correlating the trend of change in ligand binding affinity for the studied thrombin inhibitors with the electron donating/withdrawing properties of the CH_3/CH_2 bioisosteres when these bioisosteres are adjacent to the C=O or the SO_2 hydrogen bond acceptors; Note: binding affinity is expressed as free energies.

Figure 5: A plot of the free binding energy "ΔG" for ligands **2**, **4**, **5**, **6**, and **7** vs. "$-E_{int}$" of the corresponding L-W models. "$-E_{int}$" represents the desolvation penalty of the C=O and demonstrates positive linear correlation with "ΔG" for the plotted ligands with R^2 value of 0.92

Scheme 1: The strategies used to design the studied ligands: Bioisosteric replacement of the C=O group (e.g., **3**); and modulating the C=O/SO_2 basicity via modifying the adjacent CH_3/CH_2 groups to decrease (**4**, **5**, **6**, and **8**) or increase (**7**, **9**, **12**/**13**, and **14**/**15**) the basicity.

Scheme 2: Synthesis of ligands **1**, **2**, **3**, **4**, **5**, **6**, **9**, **10**, **11**, **14**, and **15**

Scheme 3: Synthesis of ligands **7**, **12**, and **13**

Scheme 4: Synthesis of ligand **8**

Table 1: The values of the inhibition constants (K_i) and the corresponding binding free energies (ΔG) of ligands **1** to **15**

Table 2: QM models representing the H-bonds accepted by ligands **2** and **3** from Gly216 (models A_1 and B_1) and from a water molecule (models A_w and B_w). The calculated energies of these H-bonds are listed in both au and kJ/mol units[a]

Table 3: QM models representing the H-bonds accepted by various ligands from Gly216 (models L-P1 and L-P2) and from a water molecule (models L-W). The calculated energies of these H-bonds (or energies of interaction between model components) are listed.

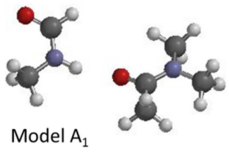
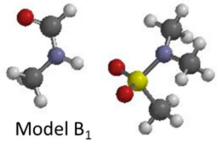
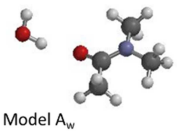
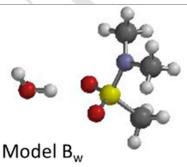
Tables:

Table 1: The values of the inhibition constants (K_i) and the corresponding binding free energies (ΔG) of ligands **1** to **15**

Compound	K_i (µM)	ΔG (kJ/mol)
1	250 ± 60	-20.6 ± 0.6
2	68 ± 15	-23.8 ± 0.5
3	68 ± 9	-23.8 ± 0.4
4	44 ± 2	-24.9 ± 0.12
5	17.9 ± 0.5	-27.1 ± 0.07
6	13.5 ± 1.3	-27.8 ± 0.2
7	59 ± 3	-24.1 ± 0.11
8	7.1 ± 0.9	-29.4 ± 0.3
9	163 ± 17	-21.6 ± 0.3
10	4.2 ± 0.4	-30.7 ± 0.2
11	0.72 ± 0.06	-35.1 ± 0.19
12	11.9 ± 1.6	-28.1 ± 0.3
13	4.0 ± 0.5	-30.9 ± 0.3
14	36 ± 7	-25.4 ± 0.5

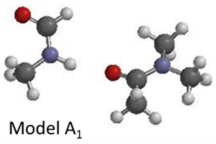
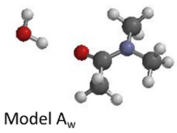
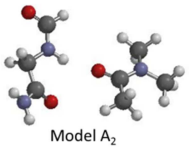
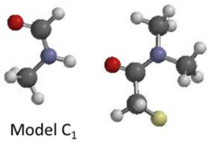
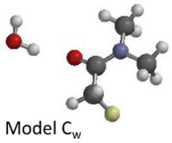
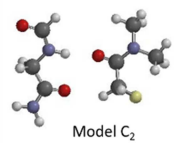
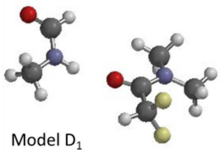
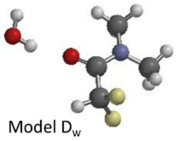
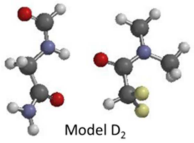
15	14.5 ± 0.5	-27.6 ± 0.09

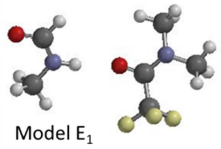
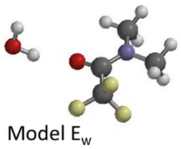
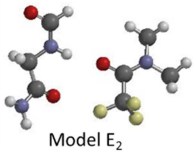
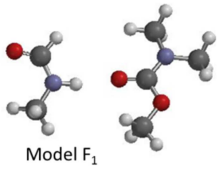
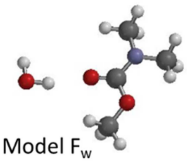
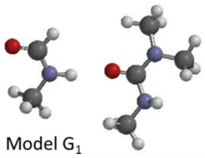
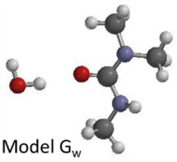
Table 2: QM models representing the H-bonds accepted by ligands **2** and **3** from Gly216 (models A_1 and B_1) and from a water molecule (models A_w and B_w). The calculated energies of these H-bonds are listed in both au and kJ/mol units[a]

QM Model				
$E_{H\text{-}bond} = E_{int}$ (au)	-0.012112	-0.007267	-0.010726	-0.007831
$E_{H\text{-}bond} = E_{int}$ (kJ/mol)	-31.80	-19.08	-28.16	-20.56

[a] E_{int} stands for the calculated energy of interaction between the two molecules in each system

Table 3: QM models representing the H-bonds accepted by various ligands from Gly216 (models L-P1 and L-P2) and from a water molecule (models L-W). The calculated energies of these H-bonds (or energies of interaction between model components) are listed.

QM Model*, L-P1	$E_{H\text{-}bond} = E_{int}$ (au, kJ/mol)	QM Model*, L-W	$E_{H\text{-}bond} = E_{int}$ (au, kJ/mol)	QM Model*, L-P2	E_{int} (au, kJ/mol)
	-0.012112, -31.80		-0.010726, -28.16		-0.008634, -22.67
	-0.010978, -28.82		-0.009798, -25.73		-0.008073, -21.20
	-0.009718, -25.51		-0.008919, -23.42		-0.008624, -22.64

 Model E_1	-0.009427, -24.75	 Model E_w	-0.008072, -21.19	 Model E_2	-0.005075, -13.32
 Model F_1	-0.011663, -30.62	 Model F_w	-0.010849, -28.48	--	--
 Model G_1	-0.013453, -35.32	 Model G_w	-0.012068, -31.68	--	--

*Models A represent ligand **2**, C's represent **4**, D's represent **5**, E's represent **6**; F's represent **12/13**, and G's represent **14/15**

Figures:

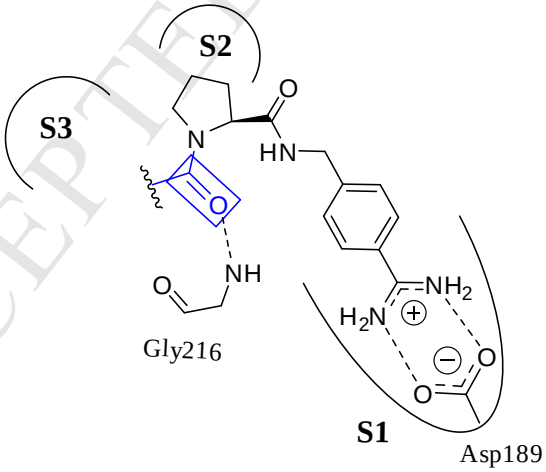


Figure 1: A schematic representation of the general scaffold of previously reported benzamidine thrombin inhibitors [1][4], binding in the thrombin catalytic site. The C=O group which accepts an H-bond from the Gly216 residue is shown in blue and enclosed in a blue rectangle. This C=O group can be replaced by other bioisosteric groups such as SO_2. The hydrophobic pockets of the thrombin active sites are shown as curves and indicated as S1, S2, and S3.

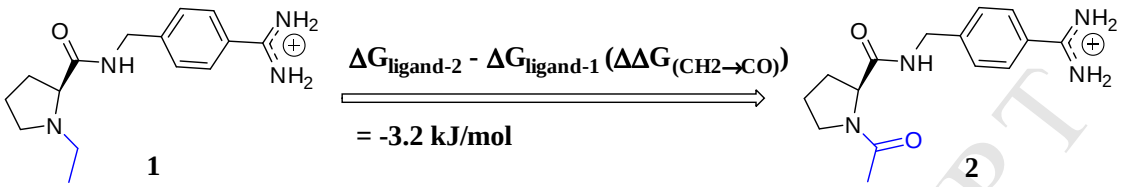


Figure 2: Calculation of the binding free energy contribution of the C=O group that forms an H-bond with Gly216. This contribution was found to be -3.2 kJ/mol.

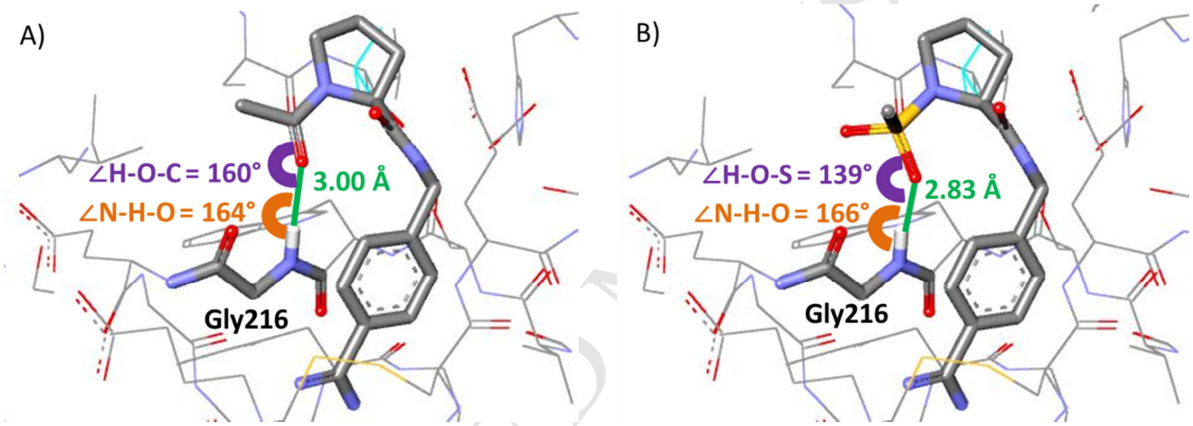


Figure 3: A) ligand **2** bound to the active site of thrombin. The crystal structure PDB # 2ZI2 was used to construct the thrombin-**2** complex. **B)** Ligand **3** bound to the active site of thrombin. Thrombin-**2** complex, along with molecular modeling, was utilized to construct the thrombin-**3** complex. The geometric parameters of the H-bonds accepted by ligands **2** and **3** from Gly216 are shown (Each of the distances shown is between the oxygen H-bond acceptor and the nitrogen H-bond donor heteroatoms). In thrombin-**3** complex, note that one of the SO_2 oxygens is H-bonded to Gly216, and the other is exposed to the solvent. Also, note that the Me of the $MeSO_2$ does not make significant contacts with any of the protein residues.

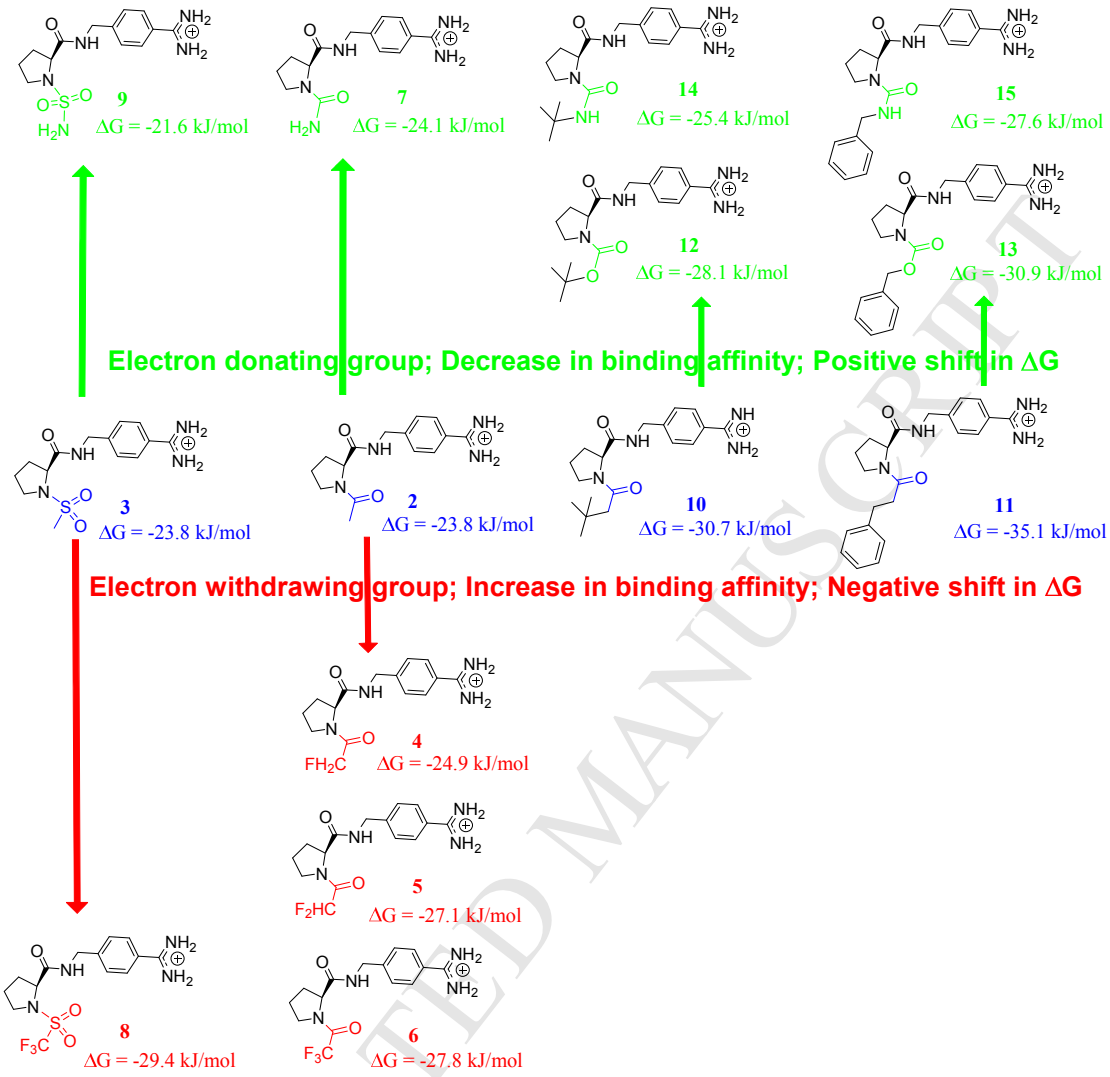


Figure 4: Correlating the trend of change in ligand binding affinity for the studied thrombin inhibitors with the electron donating/withdrawing properties of the CH_3/CH_2 bioisosteres when these bioisosteres are adjacent to the C=O or the SO_2 hydrogen bond acceptors; Note: binding affinity is expressed as free energies.

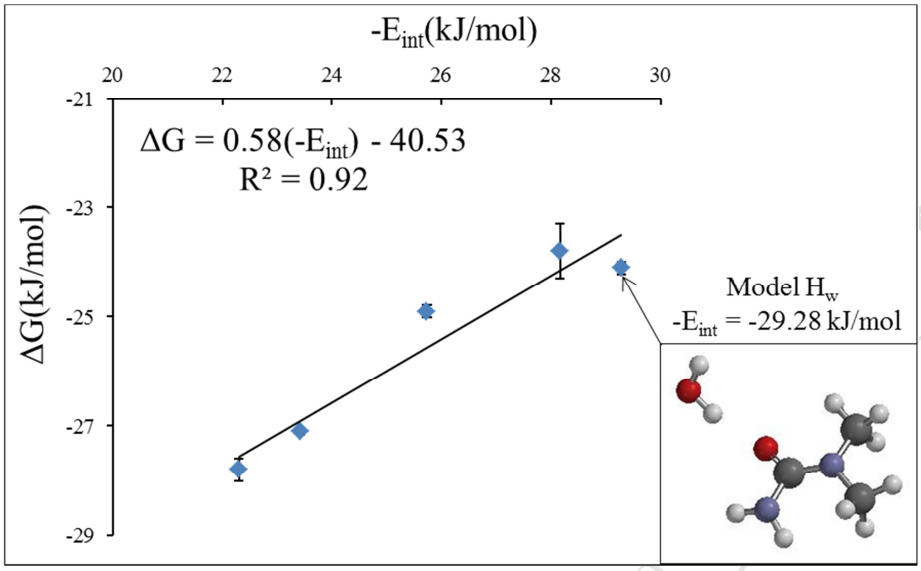
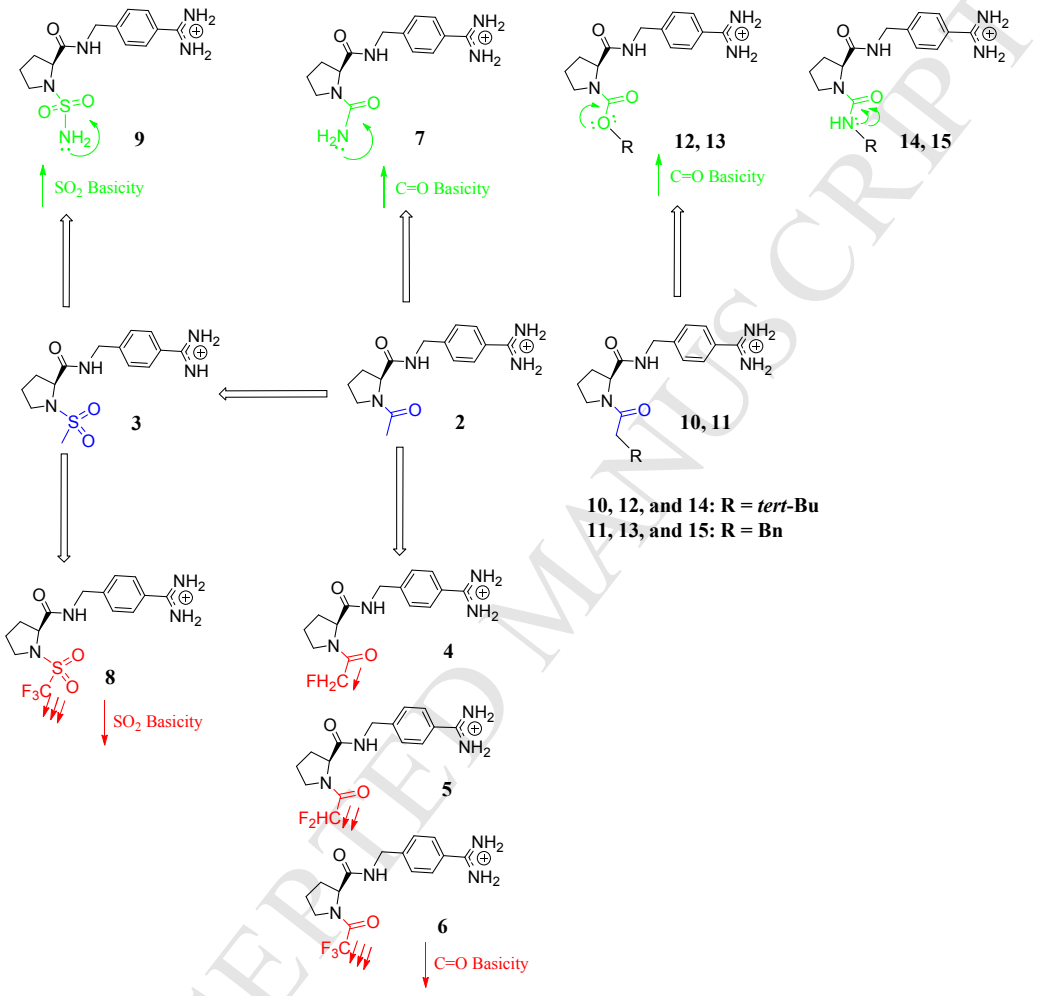


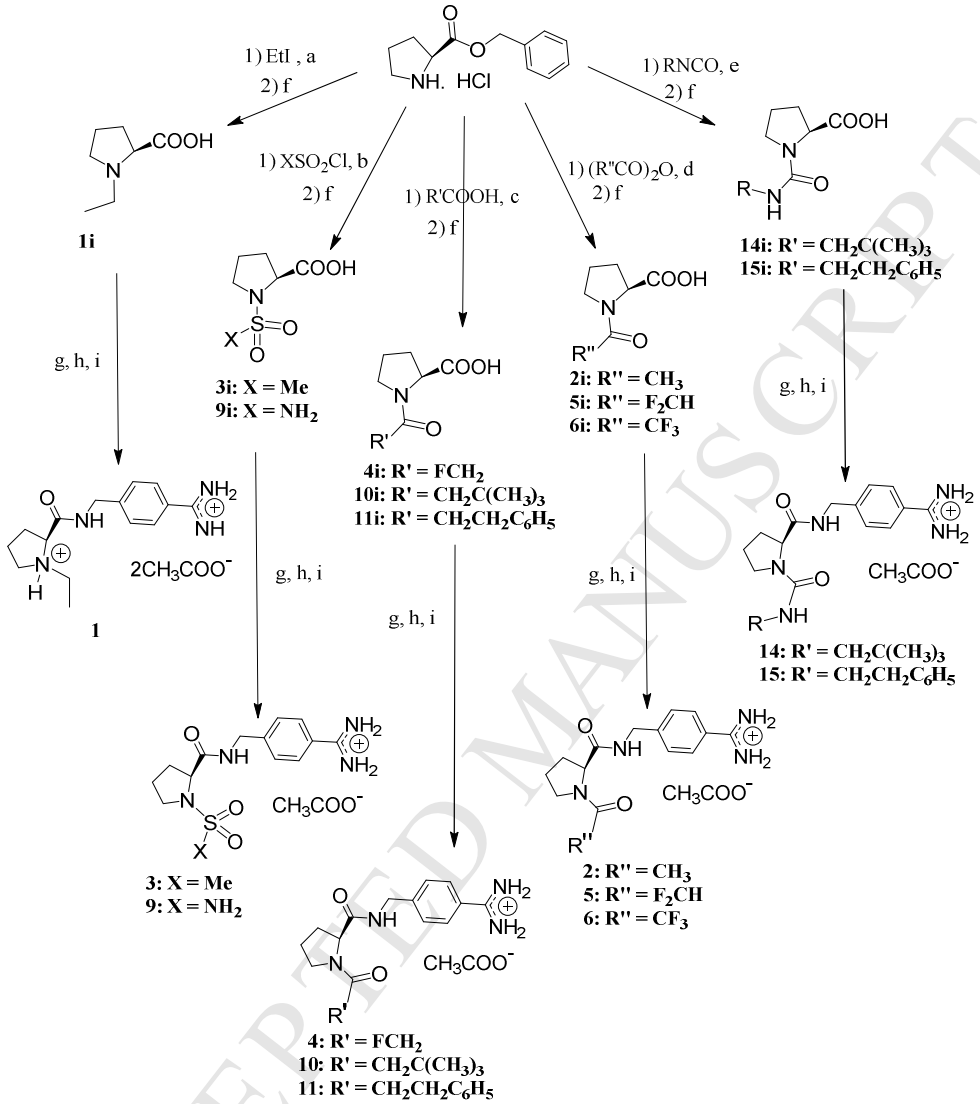
Figure 5: A plot of the free binding energy "ΔG" for ligands **2**, **4**, **5**, **6**, and **7** vs. "-E_{int}" of the corresponding L-W models. "-E_{int}" represents the desolvation penalty of the C=O and demonstrates positive linear correlation with "ΔG" for the plotted ligands with R^2 value of 0.92.

Schemes:

Scheme 1: The strategies used to design the studied ligands: Bioisosteric replacement of the C=O group (e.g., **3**); and modulating the C=O/SO_2 basicity via modifying the adjacent CH_3/CH_2 groups to decrease (**4**, **5**, **6**, and **8**) or increase (**7**, **9**, **12**/**13**, and **14**/**15**) the basicity.

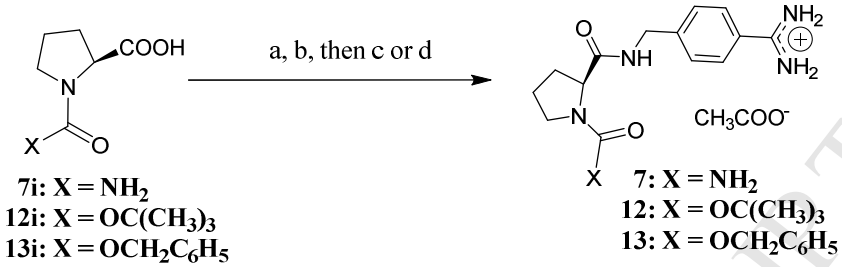


Scheme 2: Synthesis of ligands **1**, **2**, **3**, **4**, **5**, **6**, **9**, **10**, **11**, **14**, and **15**



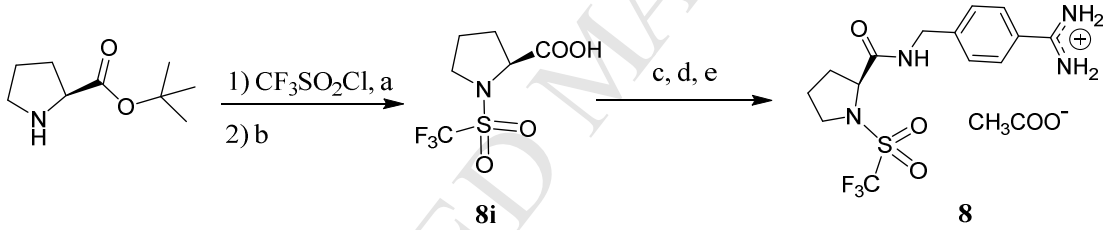
a) 3.0 equiv. K_2CO_3, DMF, 80°C, overnight, 72% **b)** 3.0 equiv. TEA, anhydrous DCM, 0°C→r.t, 8 h, 64-68% **c)** 1.2 equiv. EDCI.HCl, 1.2 equiv. HOBt, 3.3 equiv. DIEA, anhydrous DMF, r.t., overnight, 69-74% **d)** anhydrous pyridine, 0°C→r.t, 5-7 h, 51-59% **e)** 3.0 equiv. TEA, anhydrous DCM, r.t., 5 h, 63-70% **f)** H_2/Pd-C, MeOH, r.t., overnight, 90-97% **g)** 1.1 equiv. 4-(aminomethyl)benzonitrile HCl, 1.2 equiv. EDCI.HCl, 1.2 equiv. HOBt, 3.3 equiv. DIEA, anhydrous DMF, r.t., overnight, 64-82% **h)** 3.0 equiv. $H_2NOH.HCl$, 3.0 equiv. DIEA, anhydrous MeOH, r.t., overnight, 71-88% **i)** 3.0 equiv. $(CH_3CO)_2O$, glacial HOAc, r.t., 30 min, followed by H_2/Pd-C, glacial acetic, r.t., overnight, 82-91%.

Scheme 3: Synthesis of ligands **7**, **12**, and **13**



a) 1.1 equiv. 4-(aminomethyl)benzonitrile HCl, 1.2 equiv. EDCI.HCl, 1.2 equiv. HOBt, 3.3 equiv. DIEA, anhydrous DMF, r.t., overnight, 65-78% **b)** 3.0 equiv. $H_2NOH.HCl$, 3.0 equiv. DIEA, anhydrous MeOH, r.t., overnight, 66-73% **c)** 3.0 equiv. $(CH_3CO)_2O$, glacial HOAc, r.t., 30 min, followed by H_2/Pd-C, glacial acetic, r.t., overnight, 78-88% **d)** for **13**: 5.0 equiv. Zn dust, glacial HOAc, r.t.→60°C, overnight, 65%.

Scheme 4: Synthesis of ligand **8**



a) 3.0 equiv. TEA, anhydrous DCM, r.t., overnight, 52% **b)** TFA, DCM, r.t., 5 h, 91% **c)** 1.1 equiv. 4-(aminomethyl)benzonitrile HCl, 1.2 equiv. EDCI.HCl, 1.2 equiv. HOBt, 3.3 equiv. DIEA, anhydrous DMF, r.t., overnight, 69% **d)** 3.0 equiv. $H_2NOH.HCl$, 3.0 equiv. DIEA, anhydrous MeOH, r.t., overnight, 76% **e)** 3.0 equiv. $(CH_3CO)_2O$, glacial HOAc, r.t., 30 min, followed by H_2/Pd-C, glacial acetic, r.t., overnight, 82%.

Highlights

- The role of H-bond basicity in SAR studies was explored in thrombin inhibitors.
- H-bond basicity/strength was systematically modulated via bioisosterism.
- Improved strength of H-bonds to the protein did not improve binding affinity.
- Decreased strength of H-bonds to water of desolvation lead to better inhibitors.
- Bioisosteric replacements were found to produce synergism in SAR modifications.