DOI: 10.1002/cmdc.200600029

Enzyme Isoselective Inhibitors: A Tool for Binding-Trend Analysis

Rachel Ozeri, Netaly Khazanov, Nurit Perlman, Michael Shokhen,* and Amnon Albeck*[a]

A transition-state analogue inhibitor that covalently reversibly binds to an enzyme formally consists of two parts: the chemical site, CS and the recognition site, RS. We have experimentally and theoretically demonstrated that the trend of binding affinity in a series of isoselective inhibitors (with identical RS and different CS fragments) depends mainly on their CS fragments. Isoselective inhibitors have the same affinity trend toward different enzymes of the same family with a common catalytic mechanism. Thus, very good correlation between experimentally determined and theoretically calculated K_i values was demonstrated. A practical outcome is the application of the described method as a tool for an expert analysis in virtual screening of inhibitor libraries and in the design of new enzyme inhibitors.

Introduction

Many drugs that are currently in use or in development are enzyme inhibitors.[1–5] Thus, understanding enzyme mechanism at the molecular level is of high priority both as a fundamental science and for a variety of practical applications, especially in the field of medicinal chemistry. Enzyme inhibitors can be divided into two groups: noncovalent and covalent. In this report we concentrate on the latter group. Covalent inhibitors can bind their target enzyme either reversibly or irreversibly. In a previous theoretical study of transition-state (TS) analogue inhibitors of serine proteases, we observed that the covalent bond formed in a thermodynamically stable enzyme–inhibitor tetrahedral complex (TC) is about 30 kcal mol^{-1} stronger than the analogous bond formed by a native substrate in the catalytic reaction intermediate.[6,7] Moreover, the trend in experimental binding constants in series of such inhibitors was well correlated with the energy of the enzyme–inhibitor covalent bond.[6] This conclusion is also supported by the kinetic data for chymotrypsin interacting with a series of peptide amide substrates and trifluoromethyl ketone (TFK) TS analogue inhibitors, reported by Brady and Abeles.[8] For each pair of TFK and amide with the same peptidyl fragment, the value of $\Delta G_{(NH_2-CF_3)}$ was estimated as $\Delta G_{(NH_2-CF_3)} = G^{\neq}_{(NH_2)} - G_{TC(CF_3)}$, for which $G^{\neq}_{(NH_2)}$ is the free energy of activation of amide substrate hydrolysis and $G_{TC(CF_3)}$ is the corresponding free energy of formation of the covalent TC, both measured at 25 °C. In the series of five different peptidyl fragments examined, the standard deviation from the average value of $\Delta G_{(NH_2-CF_3)}$ (30.0 kcal mol^{-1}) is only 0.5 kcal mol^{-1}. Analysis of the contribution of various interactions in the enzyme active site to the stability of the TC demonstrated that the dominating factor is the enzyme–inhibitor covalent bond.[7]

In continuing our mechanistic studies of enzyme–inhibitor interactions, we consider herein a very specific class of covalent inhibitors—*isoselective* inhibitors. They are used here as a tool to analyze fundamental principles that control the trend in the efficiency of enzyme covalent TS analogue inhibitors.

Results and Discussion

The concept of isoselective inhibitors

In this work we consider the reversible binding of covalent TS analogue inhibitors to the active site of an enzyme, quantitatively described by the equilibrium constant $K_i = ([E][I])/[EI]$, which characterizes the thermodynamic equilibrium between the reactants (R) and the products (P): R⇌P. The reactants correspond to the free enzyme and inhibitor (E and I, respectively) and the product is an enzyme inhibitor complex, EI. Thus, according to the classical definition of thermodynamics, the free energy of product P formation from reactants R (inhibitor binding energy) is calculated as:

$$RT\ln K_i = \Delta G_P = G_P - G_R \quad (1)$$

A ligand for an enzyme (either a substrate or an inhibitor) formally consists of two parts. One part, the CS region, contains the reactive chemical site and the covalent surroundings that control its chemical reactivity. The CS is a localized compact structural fragment. The rest of the ligand, the recognition site (RS), participates only in various noncovalent interac-

[a] Dr. R. Ozeri, N. Khazanov, Dr. N. Perlman, Dr. M. Shokhen, Prof. A. Albeck
The Julius Spokojny Bioorganic Chemistry Laboratory
Department of Chemistry
Bar Ilan University
Ramat Gan 52900 (Israel)
Fax: (972) 3-5351250
E-mail: shokhen@mail.biu.ac.il
albecka@mail.biu.ac.il

tions with the enzyme. Its role is to properly align the ligand in the enzyme active site.[6] The binding energy ΔG_P of each covalent inhibitor defined in Equation (1) can be approximated as:

$$\Delta G_P = \Delta G_P(CS) + \Delta G_P(RS) \quad (2)$$

We consider a special class of ligands with identical RS and different CS fragments. The RS fragment bears the main contribution to inhibitor selectivity (recognition), so we classify such inhibitors as *isoselective*. Thus, at a good level of approximation, the relative free energy of enzyme–inhibitor binding, $\Delta\Delta G_{P,k}$:

$$\Delta\Delta G_{P,k} = \Delta G_{P,k} - \Delta G_{P,o} \quad (3)$$

measured for any inhibitor k relative to a reference inhibitor o in the isoselective inhibitor series (all having the same RS fragment), depends mainly on the varied CS fragment:

$$\Delta\Delta G_{P,k} \approx \Delta\Delta G_{P,k}(CS) \quad (4)$$

In other words, the trend in binding affinity in a series of isoselective ligands depends mainly on their CS fragments.

Table 1. Apparent inhibition constants (K_i^{app} [μM]) for TS analogue inhibitors of serine proteases.[a]

Inhibitor	Chymotrypsin	Subtilisin	CPY[b]
Cbz-Phe-OH	340	6010	3600
Cbz-Phe-CH_3	208	630	70
Cbz-Phe-H	19	26	41
Cbz-Phe-$COCH_3$	0.75	19	1.07
Cbz-Phe-CF_3	0.22	0.64	0.04

[a] Measured at 25 °C. [b] carboxypeptidase Y.

Table 2. Apparent inhibition constants (K_i^{app} [μM]) for TS analogue inhibitors of serine proteases.[a]

Inhibitor[b]	Chymotrypsin	Subtilisin	CPY[c]	Elastase
RS-OH	64	404	126	1380
RS-CH_3	48	80	50	1000
RS-H	1.96	2.69	4.0	67
RS-$COCH_3$	0.011	0.053	0.59	31

[a] Measured at 25 °C. [b] RS = Cbz-Gly-Leu-Phe for chymotrypsin, subtilisin, and CPY; RS = Cbz-Ala-Ala for elastase. [c] carboxypeptidase Y.

Experimental validation of the concept

We synthesized three series of isoselective TS analogue inhibitors of serine proteases and measured their inhibition constants. Two sets of the general formula RS–X (RS = Cbz-Phe; X = OH, H, CH_3, $COCH_3$, CF_3 and RS = Cbz-Gly-Leu-Phe; X = OH, H, CH_3, $COCH_3$) were examined with three serine proteases: chymotrypsin, subtilisin, and carboxypeptidase Y. The third RS–X series (RS = Cbz-Ala-Ala; X = OH, H, CH_3, $COCH_3$) was studied on the serine protease elastase. The syntheses of the aldehyde, the methyl ketone, and the methyl diketone derivatives (X = H, CH_3, and $COCH_3$, respectively) followed the approach of Angelastro et al.[9] The trifluoromethyl ketone TS analogue (X = CF_3) was prepared according to Walter et al.[10] The corresponding inhibition constants are presented in Tables 1 and 2. The same trend is observed for all three different recognition sites Cbz-Phe, Cbz-Gly-Leu-Phe, and Cbz-Ala-Ala in all examined enzymes; the stability of the TC as a function of varied CS fragment decreases in the following order:

$$CF_3 > COCH_3 > H > CH_3 > OH$$

Thus, the CS fragment controls not only the extremely large difference in stability of the TC formed by a good inhibitor versus a substrate (CF_3 versus NH_2),[7,8] but also the difference in the binding trend observed between inhibitors with similar electron-withdrawing ability such as CF_3 and $COCH_3$. The above experimental data can be generalized in a conceptual conclusion: *the trend in binding affinity of any two isoselective ligands to an enzyme is independent of variation in the RS fragment*.

Data in Tables 1 and 2 demonstrate that the same trend is observed for all enzymes examined. This result experimentally demonstrates the validity of the second basic concept: *a series of isoselective inhibitors has the same affinity trend toward different enzymes belonging to the same family with a common catalytic mechanism.*

Theoretical analysis

Different members of the same enzymatic family with a common catalytic mechanism differ by the recognition features of their active sites. Therefore, noncovalent enzyme–RS interactions are different for different enzymes from the same family. Analogously, the enzyme–RS interactions between an enzyme and two inhibitors with different RS fragments would also be different. On the other hand, the experimental data demonstrate that reversible covalent inhibitors with varied CS fragments have the same binding trend for various representatives of the same enzyme family and in different series of isoselective inhibitors. A reasonable explanation is that the dominant trend-determining factor is not the recognition contribution of the RS fragment, but rather the energy released from the enzyme–inhibitor chemical reactivity of the CS. This idea can be examined by molecular modeling that ignores the RS fragment of inhibitors as well as the noncovalent interactions of the CS fragment in the enzyme active site. The latter approximation is justified, as the energy of the enzyme–inhibitor covalent bond is one or two orders of magnitude higher than the energy of noncovalent recognition interactions in the enzyme active site for such a small fragment as CS.[7,11] In fact, our earlier modeling studies (mentioned above) in which the RS fragments were avoided indeed reproduced the experimental trend.[6,7] Nevertheless, more precise modeling, especially with systems that are energetically similar, would require the proper accounting of all environmental effects. We studied a series of substituted carbonyl groups (CH_3–CO–X; X = H, OH, CH_3, CF_3, $COCH_3$) to generate the CS fragments of the exam-

ined isoselective inhibitors. These functional groups of the water-solvated inhibitors are subjected to various chemical interactions with water. Thus, the peptidyl bound carboxyl group (X = OH) has $pK_a \sim 5$ and therefore is ionized at neutral pH. The carbonyl group with H, $COCH_3$, and CF_3 substituents could be hydrated.[8,12] The products of such competitive reactions, the carboxylate anion and the hydrate, could not interact with the Ser nucleophile in the active site of serine proteases. Thus, the competitive reactions consume free inhibitor and considerably decrease its concentration. Therefore, the experimentally measured values in Tables 1 and 2 are in fact the apparent binding constants, K_i^{app}.[8] These competing reactions that lead to the formation of a nonreactive product C are expressed in the thermodynamic equilibrium:

$$C \rightleftharpoons R \rightleftharpoons P \quad (5)$$

In this case, the formation of product P is determined by K_i^{app}, the expression of which can be derived from statistical thermodynamics as:

$$\ln K_i^{app} = \Delta G_P/RT + \ln[1+e^{(-\Delta G_C/RT)}] \quad (6)$$

$\ln K_i = \Delta G_P/RT$, $\ln K_C = -\Delta G_C/RT$; $\Delta G_P = G_P - G_R$ and $\Delta G_C = G_C - G_R$, for which G_R, G_P, and G_C are the free energies of states R, P, and C in the equilibrium, respectively. If the formation of the competitive product C is an exothermic process with a large equilibrium constant, $K_C > 1$, the value of K_i^{app} can be approximated by a simpler formula such as that used by Brady and Abeles[8] for the hydration of TFK inhibitors, for which $K_h = 4500$:

$$K_i^{app} = K_C K_i \quad (7)$$

Another complication is the fact that as electrophiles, the COOH and $COCH_3$ groups are too weak, and thus the overwhelming fraction of the enzyme–inhibitor complex is represented by a noncovalent Michaelis complex (MC), and only a negligible fraction is in the form of the covalent TC.[13] Therefore, the average relative free energy of binding $\langle \Delta\Delta G_{P,k} \rangle$ of inhibitor k defined in Equation (3) can be expressed as:

$$\langle \Delta\Delta G_{P,k} \rangle = x_{MC,k}\,\Delta\Delta G_{MC,k} + x_{TC,k}\,\Delta\Delta G_{TC,k} \quad (8)$$

The relative energies of formation are $\Delta\Delta G_{MC,k} = \Delta G_{MC,k} - \Delta G_{MC,o}$ and $\Delta\Delta G_{TC,k} = \Delta G_{TC,k} - \Delta G_{TC,o}$ [Eq. (3)], for which the molar fractions $x_{MC,k}$ and $x_{TC,k}$ can be estimated from standard formulas of Boltzmann statistics: $x_i = [e^{(-\Delta\Delta G_i/RT)}]/Z$; Z is a partition function.

In the Michaelis complex formed by any inhibitor k, the CS fragment of the inhibitor is not involved in any covalent chemical interaction with the enzyme. Therefore, if we ignore the noncovalent interactions of the CS fragment with the enzyme active site (see discussion above), we can conclude that $\Delta\Delta G_{MC,k} = 0$. Indeed, according to Equation (2), in the absence of covalent binding the only contribution to the free energy of inhibitor binding is the ΔG_{MC}(RS) noncovalent recognition interaction: $\Delta G_{MC,k} = \Delta G_{MC,k}$(RS). As all inhibitors in a series of isoselective inhibitors have identical RS fragments, their binding energy values $\Delta G_{MC,k}$(RS) are equivalent. Consequently, within a series of isoselective inhibitors, the relative binding energies in the Michaelis complex are zero: $\Delta\Delta G_{MC,k} = 0$. In other words, the binding trend in a series of isoselective inhibitors, $\Delta\Delta G_{P,k}$, is in fact a function of $\Delta\Delta G_{TC,k}$ values only, even when considering a very small molar fraction of TC, as in the case of CH_3 and OH substituents at the carbonyl group, for which $x_{MC,k} \gg x_{TC,k}$ (see ref. [13] and discussion above).

We have examined the above formulated principles of isoselective inhibition through molecular modeling of TS analogue inhibitors[8,14,15] of serine proteases. We used a series of substituted carbonyl groups (CH_3–CO–X; X = H, OH, CH_3, CF_3, $COCH_3$) to generate the CS fragments of the isoselective inhibitors examined. The variation of X covers a wide range of covalent TC stability, from very unstable in the case of CH_3 and OH, up to highly stable TCs formed by TFK inhibitors.[8] The serine nucleophile was simulated by the CH_3O^- anion. Water is the bulk solvent for the enzymatic reaction considered. To accommodate both processes, the formation of the anionic TC product and the competitive hydration of the carbonyl group or acid dissociation in the case of X = OH, we considered the following three reactions:

$$CH_3O^- + CH_3{-}CO{-}X \rightarrow CH_3{-}C(O^-)(OCH_3){-}X \quad (9)$$

$$H_2O + CH_3{-}CO{-}X \rightarrow CH_3{-}C(OH)_2{-}X \quad (10)$$

$$H_2O + CH_3{-}COOH \rightarrow CH_3{-}COO^- + H_3O^+ \quad (11)$$

According to the formulated concept, the relative binding affinity of any two isoselective ligands is independent of the structural constants of the reaction series, such as common reactants (serine nucleophile) and common structural fragments (the RS fragment of inhibitors). Specifically, the nucleophiles CH_3O^- and H_2O [Eq. (9) and Eq. (10)], the basic H_2O, and the product H_3O^+ [Eq. (11)] are structural constants in the modeling reaction series. Therefore, they do not contribute to the corresponding relative free energies of formation ($\Delta\Delta G_{TC,k}$ and $\Delta\Delta G_{C,k}$) of TC and C, the competitive product from the inhibitor I_k:

$$\Delta\Delta G_{TC,k} = [G_{TC,k} - G_{I,k}] - [G_{TC,o} - G_{I,o}] \quad (12)$$

$$\Delta\Delta G_{C,k} = [G_{C,k} - G_{I,k}] - [G_{C,o} - G_{I,o}] \quad (13)$$

for which $G_{TC,k}$, $G_{C,k}$, and $G_{I,k}$ are the quantum mechanically calculated free energies of TC, C, and the inhibitor I, respectively. The subscript "$_k$" marks the varied substituent in CS and "$_o$" corresponds to the referenced substituent (the aldehyde in this case: X = H) used as a zero point on the scale of relative $\Delta\Delta G_{TC,k}$ and $\Delta\Delta G_{C,k}$ energies. The competitive reaction of hydration in Equation (13) is considered for all but X = OH substituents. In the latter case, the side reaction is the acid dissociation reaction. With pK_a 4.76 for acetic acid and an equilibrium constant $K = [A^-]/[AH]$ obtained by the Henderson–Hasselbalch equation:[16]

$$pH = pK_a + \log([A^-]/[AH]) \quad (14)$$

the derived absolute free energy, $\Delta G_{A(-),OH}$ of CH_3COO^- competitive product formation is $-3.04\ kcal\,mol^{-1}$ at 298 K and pH 7. In the specific case of X = OH, the relative free energy of the competitive product formation, $\Delta\Delta G_{C,k}$ [Eq. (13)] is $\Delta\Delta G_{A(-),OH}$ and can be calculated by:

$$\Delta\Delta G_{A(-),OH} = \Delta G_{A(-),OH} - \Delta G_{hyd,CF_3} + \Delta\Delta G_{hyd,CF_3} \quad (15)$$

In Equation (15) $\Delta G_{hyd,CF_3}$ is the absolute free energy of hydrate formation for $X = CF_3$, and $\Delta\Delta G_{hyd,CF_3}$ is the corresponding relative value defined in Equation (13). The value of $\Delta G_{hyd,CF_3} = -4.96\ kcal\,mol^{-1}$ was derived from the experimental value of the hydration constant of TFK inhibitors.[8] The relative free energy $\Delta\Delta G_{hyd,CF_3}$ of the corresponding hydrate formation was determined by Equation (13).

Based on our experimental results and the theoretical analysis discussed above, we have suggested that the trend in TC stability in a series of isoselective inhibitors is mainly determined by the covalent interaction of the inhibitor CS fragment with Nuc, the nucleophilic functionality of the enzyme (Nuc = Ser residue in serine proteases, considered herein). The modeling approach used reflects this feature and considers only the reaction core explicitly: the CS–Nuc covalent interaction. Moreover, as mentioned above, such a simplification was accepted in order to examine by modeling whether the CS–Nuc covalent interaction is the dominant factor in determining the trend in a series of isoselective inhibitors. Nevertheless, despite the local character of such a CS–Nuc interaction, it should be strongly influenced by environmental effects: polar and charged groups in the enzyme active site and solvation by bulk water. We used the QM/SCRF(VS) approach aimed at quantum mechanical simulation of chemical reactions in enzyme active sites, treating a water-solvated enzyme as a two-layer system.[17,18] The inner layer, the molecular cluster model of the reaction core, was calculated in gas-phase approximation by high level DFT.[19–21] The outer layer includes the rest of the protein and the bulk water, so it is very heterogeneous. The QM/SCRF(VS) approach avoids the complications of heterogeneity by considering the protein/water outer layer as a "virtual solvent", a uniform continuum dielectric medium characterized by an empirical parameter, the effective dielectric constant ε_{eff}; $1 \le \varepsilon_{eff} \le 80$. The interval borders correspond to the extreme cases of either a vacuum- or water-embedded molecular cluster. The ε_{eff} parameter considered is a local characteristic of the polarization effect of the outer layer at the small area of the reaction core (molecular cluster). The polarization level at the reaction core caused by the outer layer is specific for every enzyme, so ε_{eff} can vary widely between different enzymes. In its physical sense, ε_{eff} is very close to ε_p, the local protein dielectric constant introduced by Warshel and coworkers; ε_p is not an absolute constant, but rather simply a parameter that depends on the model used.[22–24] Therefore, it may even vary for the same enzyme, depending on the explicit model or specific inhibitor used, as exemplified in the present study by the effect of different modeling on the ε_{eff} value for a given enzyme (Figure 1 versus Figure 3).

To fit the empirical parameter ε_{eff}, the effective dielectric constant of the virtual solvent, the solvation energy was calculated for different ε values to best fit between sets of experimental and theoretically calculated values of $\ln K_{i,k}^{app}$ for a series $\{I_k\}$ of isoselective inhibitors. The total free energy G of any

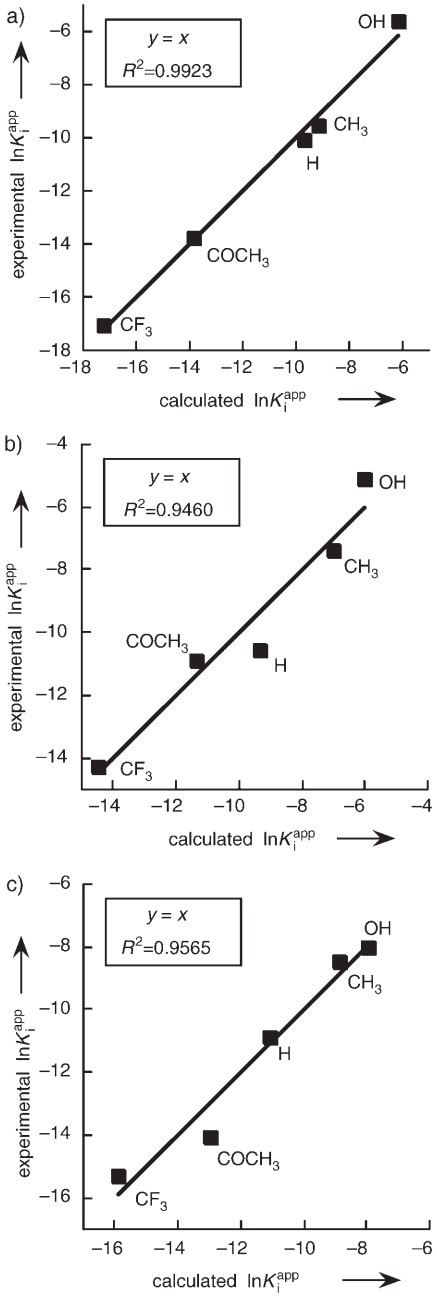

Figure 1. Linear correlation according to Equation (18) between experimental and theoretically calculated values of $\ln K_{i,k}^{app}$ for the Cbz-Phe-X series of isoselective inhibitors forming anionic TCs with chymotrypsin, subtilisin, and carboxypeptidase Y (CPY). The values of $\ln K_{i,k}^{app\,(rel)}$ were calculated by Equation (17). The best fits correspond to the following coefficients and parameters: a) CPY: $a = 0.2484$, $b = -9.8879$, SE = 0.445, $\varepsilon_{eff} = 1$; b) subtilisin: $a = 0.2405$, $b = -9.5204$, SE = 0.940, $\varepsilon_{eff} = 4$; c) chymotrypsin: $a = 0.2263$, $b = -11.2468$, SE = 0.792, $\varepsilon_{eff} = 4$.

considered state of the modeling systems was quantum mechanically calculated as:

$$G = G^g + G^s(\varepsilon) \quad (16)$$

where G^g and $G^s(\varepsilon)$ are the absolute gas-phase and solvation free energies, respectively. The calculated numerical values of relative free energies $\Delta\Delta G_{TC,k}$ and $\Delta\Delta G_{C,k}$ are collected in Table 3. $\Delta\Delta G_{C,k}$ values are calculated at $\varepsilon = 80$, reflecting the fact that the competitive reaction of inhibitor hydration or acid dissociation occurs when the free inhibitor is fully water-solvated. $\Delta\Delta G_{TC,k}$ values are calculated at different values of ε. Combining Equation (6) with Equations (12) and (13) gives the desired expression of the relative $K^{app}_{i,k\ (rel)}$ for any inhibitor k in the isoselective series:

Table 3. Relative free energies[a] of formation of hydrates ($\Delta\Delta G_{C,k}$)[b] and TCs ($\Delta\Delta G_{TC,k}$)[c].

X	Hydrate $\varepsilon=80$	TC $\varepsilon=1$	TC $\varepsilon=2$	TC $\varepsilon=3$	TC $\varepsilon=4$
CH_3	5.01	1.69	3.63	4.28	6.20
OH[d]	−0.04	8.49	8.38	8.27	8.19
H	0.00	0.00	0.00	0.00	0.00
CF_3	−1.96	−19.42	−15.95	−14.77	−14.16
$COCH_3$	2.16	−9.48	−6.12	−4.98	−4.45
X	TC $\varepsilon=6$	TC $\varepsilon=8$	TC $\varepsilon=10$	TC $\varepsilon=20$	TC $\varepsilon=40$
CH_3	6.95	7.47	7.75	8.36	8.44
OH[d]	8.11	8.11	8.08	8.04	8.01
H	0.00	0.00	0.00	0.00	0.00
CF_3	−13.56	−13.17	−12.98	−12.58	−12.63
$COCH_3$	−3.85	−3.56	−3.39	−3.19	−3.10

[a] In kcal mol^{-1}. [b] Calculated at $\varepsilon = 80$. [c] Calculated at different values of ε. [d] In the case of X = OH, the numerical value corresponds to the relative free energy of acetic acid dissociation, calculated by Equation (15).

$$\ln K^{app}_{i,k\ (rel)} = \Delta\Delta G_{TC,k}/RT + \ln[1 + e^{(-\Delta\Delta G_{C,k}/RT)}] \quad (17)$$

The advantage of theoretically calculating $K^{app}_{i,k\ (rel)}$ by Equation (17) instead of absolute values $K^{app}_{i,k}$ [Eq. (6)] is that the recognition contribution to the binding energy can be ignored, as $\Delta\Delta G_{P,k}(RS) = 0$ and $\Delta\Delta G_{P,k} = \Delta\Delta G_{P,k}(CS)$. Thus the molecular cluster used for modeling is decreased to the reaction core CS–Nuc only. Moreover, $K^{app}_{i,k\ (rel)}$ gives comprehensive information if one is interested in the binding trend only. Comparison of Equations (6) and (17) shows that $\ln K^{app}_{i,k}$ can be well approximated as a linear function of $\ln K^{app}_{i,k\ (rel)}$:

$$\ln K^{app}_{i,k} = a \ln K^{app}_{i,k\ (rel)} + b \quad (18)$$

Importantly, the linearity of Equation (18) originates in the approximation that in a series of isoselective inhibitors, the binding contributions of CS and RS fragments can be separated as independent variables. Thus, the intercept b in Equation (18) has a physical sense of the constant contribution of the RS fragment to the total binding energy. The slope a can be interpreted as a quantitative measure of the correspondence of the used RS recognition fragment to the active site of the specific enzyme. In other words, the numerical values of the linear coefficients a and b are strictly unique for every enzyme. Improved recognition by RS corresponds to higher sensitivity of the binding energy of any pair of inhibitors to the difference between their CS fragments. The reaction core CS–Nuc contributes maximally to the binding energy if the alignment of CS and Nuc in the enzyme active site optimally reflects their intrinsic electronic properties. A poor alignment of CS and Nuc requires geometrical deformation in the active site to provide the chemical interaction, leading to an energy "penalty" to the overall inhibitor binding energy.

The absolute values of $K^{app}_{i,k}$ can be calculated if the linear coefficients a and b are known. We used standard linear regression analysis to optimize the a and b coefficients by RMS fitting of the theoretically calculated value of $\ln K^{app}_{i,k}$ to the experimentally determined values of $\ln K^{app}_{i,k}$ derived from the $K^{app}_{i,k}$ values presented in Tables 1 and 2. The quality of estimation can be considerably improved by variation of the calculated value of $\ln K^{app}_{i,k\ (rel)}(\varepsilon)$ as a function of the dielectric constant ε of the virtual solvent. Thus, we have provided regression analysis with different sets of values of $\ln K^{app}_{i,k\ (rel)}(\varepsilon)$ calculated with the corresponding $\Delta\Delta G_{TC,k}(\varepsilon)$ values presented in Table 3. The best-fit results of coefficients a and b in Equation (18) selected by minimum SE and maximum R^2 for different ε values are presented in Figures 1 and 2.

Equation (18) is based on the analytically derived expression for $\ln K^{app}_{i,k\ (rel)}$ [Eq. (17)]. Formulation of an analytical dependence of $\ln K^{app}_{i,k\ (rel)}$ that takes the role of various factors into account can be too complicated. Therefore, for practical inhibitor design, it may be better to use an alternative simplified numerical approach. This is illustrated in Figure 3, which shows the linear regression fitting of experimental versus theoretically calculated absolute values of $K^{app}_{i,k}$ calculated on two independent variables $\Delta\Delta G_{TC,k}$ and $\Delta\Delta G_{C,k}$ (the factors accounted for herein). Thus, our modeling based on inhibitor CS fragments alone, reproduces quite well the experimentally observed binding trend for all enzymes considered and three isoselective inhibitor series. This result confirms the suggested dominant role of the covalent enzyme–CS interactions in the binding trend of the isoselective inhibitors.

Conclusions

We have introduced isoselective inhibitors as an expert tool for inhibition trend analysis. The binding trend in a series of isoselective inhibitors is determined by the varied CS fragments. A practical outcome of this analysis is that universal libraries of CS fragments and enzyme families with tabulated binding increments can be prepared. These libraries will be based on experimental K_i values for "training" series of isoselective inhibitors, measured for commercially available representative enzymes for each enzyme family. Such libraries will serve as data sources for given CS fragments.

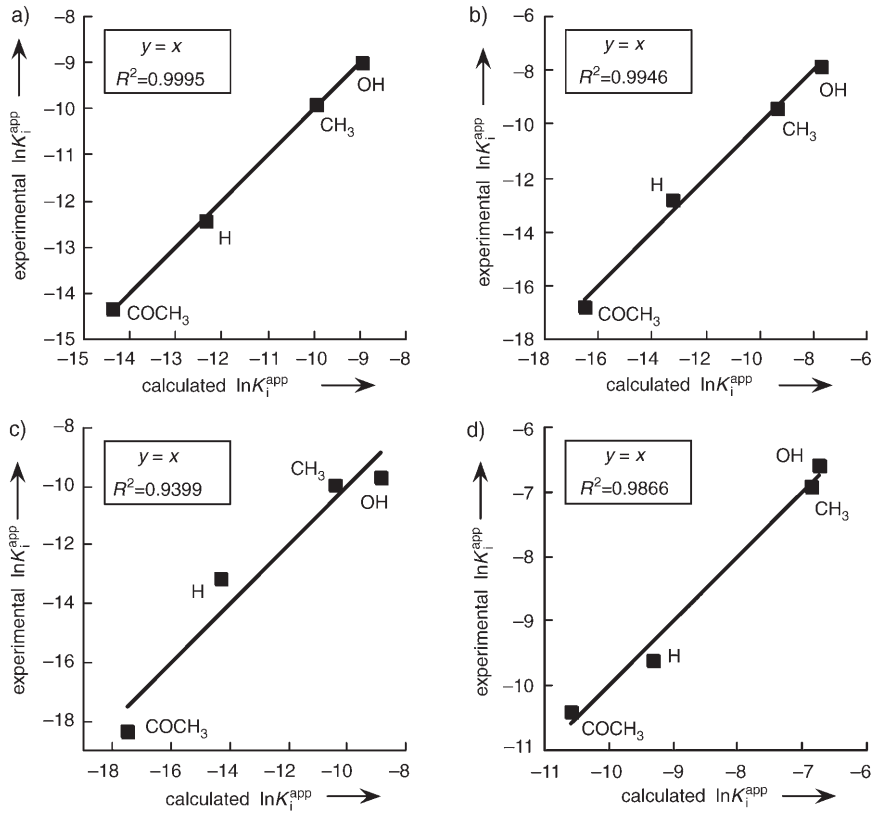

Figure 2. Linear correlation according to Equation (18) between experimental and theoretically calculated values of $\ln K_{i,k}^{app}$ for the Cbz-Gly-Leu-Phe-X and Cbz-Ala-Ala-X series of isoselective inhibitors forming anionic TCs with chymotrypsin, subtilisin, carboxypeptidase Y (CPY), and elastase. The values of $\ln K_{i,k}^{app}{}_{(rel)}$ were calculated by Equation (17). The best fits correspond to the following coefficients and parameters: a) CPY: $a=0.2450$, $b=-12.5329$, SE = 0.070, $\varepsilon_{eff}=4$; b) subtilisin: $a=0.3969$, $b=-13.5250$, SE = 0.359, $\varepsilon_{eff}=4$; c) chymotrypsin: $a=0.3926$, $b=-14.5664$, SE = 1.209, $\varepsilon_{eff}=4$; d) elastase: $a=0.1985$, $b=-9.4571$, SE = 0.272, $\varepsilon_{eff}=10$.

Experimental Section

General: Enzymes, their substrates, amino acids and peptides were purchased from Sigma–Aldrich. NMR spectra were recorded at 600, 300 or 200 MHz (^{1}H), 150, 75 or 50 MHz (^{13}C), and 188 MHz (^{19}F) in $CDCl_3$. Chemical shifts are reported in δ (ppm) with TMS as internal standard. All ^{1}H NMR assignments were supported by COSY experiments, whereas ^{13}C NMR assignments were supported by DEPT or hetero-COSY experiments. MS data were recorded in CI mode with either isobutane or ammonia as the reagent gas. TLC was performed on Merck 0.2 mm precoated silica gel F-254 plates and viewed by either UV light or Cl_2/KI-tolidine.[26] Flash column chromatography[27] was carried out on silica gel 60 (230–400 mesh ASTM, Merck).

Synthesis: The N-protected tripeptide Cbz-Gly-Leu-Phe-OH was prepared and characterized as previously described (Cbz = benzyloxycarbonyl).[28] The aldehyde, methyl ketone, and methyl diketone were synthesized according to the general procedure of Angelastro et al.[9,29] Cbz-Phe-H,[30] Cbz-Phe-CH_3,[31] and Cbz-Phe-$COCH_3$[29] were characterized according to published data.

Our inhibitor trend analysis, based on the isoselectivity principles formulated above, uses modeling on small molecular systems that simulate the interactions of the reaction centers of the inhibitor and the enzyme. Such simulations can examine, in reasonable time, the affinity of dozens of different CS fragments in high-level ab initio quantum mechanics. Another advantage of this approach, which focuses only on the reactive functional groups, is that it does not require any 3D structural information about the enzyme. This latter feature is very important when such information is not available, as is the case for most of membrane-bound enzymes. On the other hand, our approach does not take into account, at the explicit molecular level, any possible local environmental effects such as hydrogen bonds and van der Waals interactions between the inhibitor CS fragment and the enzyme active site. Nevertheless, it gives an excellent correlation between theoretically and experimentally determined $\ln K_i^{app}$ values, as the enzyme–inhibitor covalent bond formed is by far the most trend-dominating factor.[7] This is further demonstrated in the successful application of the isoselective inhibition trend analysis methodology to some medicinally important enzymes, published separately.[25]

Cbz-Phe-CF_3 was prepared according to Walter et al.[10] and characterized according to Schofield et al.[32]

Cbz-Gly-Leu-Phe-N(CH_3)OCH_3 (41% yield): ^{1}H NMR: $\delta=0.77$ (d, $J=4.2$ Hz, 3H), 0.78 (d, $J=4.2$ Hz, 3H), 1.37 (ddd, $J=13.5$, 7.0, 2.8 Hz, 1H), 1.46 (obscured, 1H), 1.47 (nonet, $J=5.4$ Hz, 1H), 2.82 (dd, $J=13.7$, 7.5 Hz, 1H), 2.97 (dd, $J=13.7$, 6.3 Hz, 1H), 3.05 (s, 3H), 3.55 (s, 3H), 3.77 (bd, 2H), 4.56 (dt, $J=8.3$, 4.7 Hz, 1H), 5.04 (s, 2H), 5.16 (q, $J=7.2$ Hz, 1H), 5.98 (t, $J=6.9$ Hz, 1H), 6.89 (d, $J=7.8$ Hz, 1H), 7.05–7.27 (m, 10H), 7.29 ppm (d, $J=7.0$ Hz, 1H); ^{13}C NMR: $\delta=22.75$, 24.53, 25.34, 31.93, 38.07, 41.30, 50.27, 51.51, 60.25, 61.40, 66.86, 126.67, 127.91, 127.97, 128.15, 128.35, 129.34, 136.24, 156.64, 168.91, 171.5, 171.72 ppm; HRMS (*m/z*) calcd for $C_{27}H_{37}N_4O_6$ [MH^+]: 513.1774, found: 513.1766.

Cbz-Gly-Leu-Phe-H (82% yield): ^{1}H NMR: $\delta=0.82$ (d, $J=7.6$ Hz, 3H), 0.84 (d, $J=7.6$ Hz, 3H), 1.33–1.50 (m, 2H), 1.50–1.63 (m, 1H), 2.97 (dd, $J=14.0$, 7.9 Hz, 1H), 3.10 (dd, $J=14.0$, 5.28 Hz, 1H), 3.80 (d, $J=4.5$ Hz, 2H), 4.52 (q, $J=6.8$ Hz, 1H), 4.60 (q, $J=6.6$ Hz, 1H), 5.06 (s, 2H), 5.95 (bd, 1H), 7.11–7.32 (m, 10H), 9.51 ppm (d, $J=6.6$ Hz, 1H); ^{13}C NMR: $\delta=22.06$, 22.75, 24.72, 41.01, 41.18, 44.37, 51.71, 59.81, 67.14, 127.02, 128.05, 128.25, 128.58, 128.66, 129.29, 136.04, 136.24, 156.81, 169.56, 172.63, 198.99 ppm.

Cbz-Gly-Leu-Phe-CH_3 (77% yield): ^{1}H NMR: $\delta=0.87$ (d, $J=6.1$ Hz, 3H), 0.89 (d, $J=6.0$ Hz, 3H), 1.43 (ddd, $J=14.1$, 8.1, 6.0 Hz, 1H), 1.49 (obscured, 1H), 1.56 (nonet, $J=6.08$ Hz, 1H), 2.10 (s, 3H), 2.97

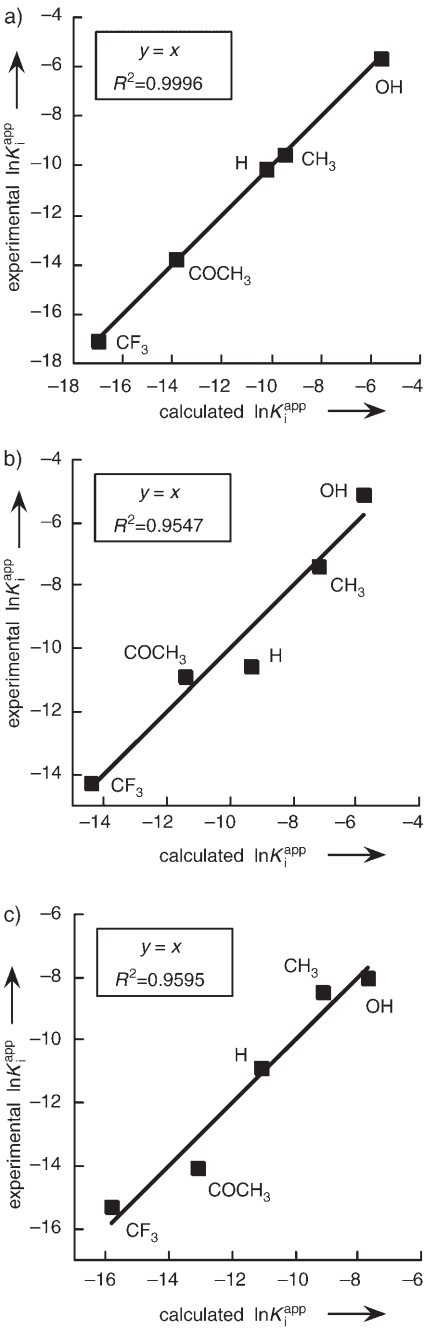

Figure 3. Linear correlation between experimental and theoretically calculated values of $\ln K_{i,k}^{app}$ by the linear approximation $\ln K_{i,k}^{app} = a\,\Delta\Delta G_{TC,k} + b\,\Delta\Delta G_{C,k} + c$ for the Cbz-Phe-X series of isoselective inhibitors forming anionic TCs with chymotrypsin, subtilisin, and carboxypeptidase Y (CPY). The best fits of *a*, *b*, and *c*, the linear regression coefficients, and ε values are: a) CPY: $\varepsilon_{eff} = 4$, SE = 0.128, $a = 0.5572$, $b = -0.5504$, $c = -10.1859$; b) subtilisin: $\varepsilon_{eff} = 40$, SE = 1.054, $a = 0.4483$, $b = -0.31931$, $c = -9.3682$; c) chymotrypsin: $\varepsilon_{eff} = 20$, SE = 0.9362, $a = 0.4221$, $b = -0.3057$, $c = -11.0906$.

(dd, J = 6.9, 14.1 Hz, 1H), 3.09 (dd, J = 6.3, 14.1 Hz, 1H), 3.81 (d, J = 5.4 Hz, 2H), 4.49 (dt, J = 5.4, 8.4 Hz, 1H), 4.77 (q, J = 6.9 Hz, 1H), 5.10 (s, 2H), 5.79 (t, J = 5.1 Hz, 1H), 6.89 (d, J = 8.1 Hz, 1H), 7.10–7.34 ppm (m, 10H); ^{13}C NMR: δ = 22.3, 23.04, 25.01, 28.26, 37.12, 41.30, 44.69, 51.99, 59.85, 67.45, 127.28, 128.34, 128.52, 128.85, 129.48, 136.4, 136.47, 156.97, 169.97, 172.08, 206.59 ppm; HRMS (*m/z*) calcd for $C_{26}H_{34}N_3O_5$ [MH$^+$]: 468.2498, found: 468.2462.

Cbz-Gly-Leu-Phe-COCH$_3$ (60% yield): ^{1}H NMR: δ = 0.86 (d, J = 6.3 Hz, 3H), 0.90 (d, J = 6.3 Hz, 3H), 1.43 (t, J = 6.37 Hz, 2H), 1.55 (nonet, J = 6.37 Hz, 1H), 2.27 (s, 3H), 2.94 (dd, J = 14.1, 8.7 Hz, 1H), 3.18 (dd, J = 14.1, 5.1 Hz, 1H), 3.83 (bd, 2H), 4.42 (q, J = 7.2 Hz, 1H), 5.10 (s, 2H), 5.4 (dt, J = 8.1, 6.7 Hz, 1H), 5.63 (bt, 1H), 6.74 (d, J = 8.1 Hz, 1H), 7.09–7.35 ppm (m, 10H); ^{13}C NMR: δ = 22.09, 22.96, 23.98, 24.78, 36.64, 40.49, 44.60, 51.39, 54.91, 67.44, 127.26, 128.22, 128.44, 128.71, 128.82, 129.40, 136.01, 136.14, 156.83, 169.5, 172.12, 195.75, 196.78 ppm; HRMS (*m/z*) calcd for $C_{27}H_{34}N_3O_6$ [MH$^+$]: 496.2447, found: 496.2434.

Cbz-Ala-Ala-N(CH$_3$)OCH$_3$ (76% yield): ^{1}H NMR: δ = 1.31 (d, J = 7.2 Hz, 3H), 1.34 (d, J = 6.9 Hz, 3H), 3.18 (s, 3H), 3.73 (s, 3H), 4.37 (quintet, J = 6.9 Hz, 1H), 4.97 (quintet, J = 7.1 Hz, 1H), 5.05 (d, J = 12.3 Hz, 1H), 5.11 (d, J = 12.3 Hz, 1H), 5.95 (d, J = 7.5 Hz, 1H), 6.05 (d, J = 9.0 Hz, 1H), 7.38 ppm (m, 5H); ^{13}C NMR: δ = 18.07, 19.07, 32.10, 45.47, 50.37, 61.46, 66.62, 127.91, 128.17, 128.34, 136.33, 155.73, 172.09, 172.79 ppm.

Cbz-Ala-Ala-H (69% yield): ^{1}H NMR: δ = 1.21 (d, J = 6.9 Hz, 3H), 1.31 (d, J = 6.9 Hz, 3H), 4.31 (quintet, J = 6.9 Hz, 1H), 4.51 (sextet, J = 7.2 Hz, 1H), 5.01 (s, 2H), 5.72 (d, J = 6.6 Hz, 1H), 5.74 (d, J = 9.3 Hz, 1H), 7.19–7.25 (m, 5H), 9.39 ppm (d, J = 7.5 Hz, 1H); ^{13}C NMR: δ = 18.6, 18.77, 50.46, 54.46, 67.08, 128.27, 128.41, 128.58, 136.15, 156.13, 172.85, 199.2 ppm; HRMS (*m/z*) calcd for $C_{14}H_{19}N_2O_4$ [MH$^+$]: 279.1344, found: 279.1332.

Cbz-Ala-Ala-CH$_3$ (86% yield): ^{1}H NMR: δ = 1.25 (d, J = 7.2 Hz, 3H), 1.30 (d, J = 6.9 Hz, 3H), 2.11 (s, 3H), 4.24 (quintet, J = 6.9 Hz, 1H), 4.46 (quintet, J = 6.9 Hz, 1H), 5.02 (s, 2H), 5.64 (d, J = 7.2 Hz, 1H), 6.97 (bd, 1H), 7.21–7.25 ppm (m, 5H); ^{13}C NMR: δ = 17.10, 18.70, 26.49, 50.53, 54.64, 67.00, 128.06, 128.21, 128.56, 136.26, 155.97, 172.16, 206.73 ppm; HRMS (*m/z*) calcd for $C_{15}H_{21}N_2O_4$ [MH$^+$]: 293.1501, found: 293.1504.

Cbz-Ala-Ala-COCH$_3$ (95% yield): ^{1}H NMR: δ = 1.26 (d, J = 7.2 Hz, 3H), 1.34 (d, J = 6.9 Hz, 3H), 2.33 (s, 3H), 5.02 (obscured, 1H), 5.08 (s, 2H), 5.94 (bquintet, 1H), 6.73 (d, J = 6.9 Hz, 1H), 7.28–7.32 (m, 5H), 7.41 ppm (bd, 1H); ^{13}C NMR: δ = 16.28, 18.41, 26.77, 49.05, 50.47, 66.95, 127.93, 128.15, 128.48, 136.11, 155.97, 172.33, 204.98, 206.86 ppm; HRMS (*m/z*) calcd for $C_{16}H_{21}N_2O_5$ [MH$^+$]: 321.1450, found: 321.1452.

Enzyme inhibition assays: All enzymatic assays were carried out at 25 °C by following substrate hydrolysis spectrophotometrically. For competitive inhibition assays, solutions of substrate (at the indicated concentration in 20 μL DMSO) and inhibitor (varying concentrations in 20 μL DMSO) were dissolved in buffer (940 μL). The catalytic reaction was initiated by the addition of enzyme solution (20 μL). Enzyme concentration was set such that under conditions of substrate saturation (V_{max}) and absence of inhibitor, the initial velocity of hydrolysis was $\approx 10^{-3}$ ODU sec^{-1}. For slow tight binding inhibition assays, the discontinuous method was applied.[33] Typically, the enzyme studied was incubated at 25 °C with the particular inhibitor dissolved in DMSO (volume of organic solvent not exceeding 5% of the total volume). Aliquots were removed periodically, diluted into assay solution containing the substrate, and the residual enzymatic activity was measured. A control pre-incubation solution containing all of the ingredients except for the inhibitor itself was assayed in parallel.

Chymotrypsin (bovine pancreas, EC 3.4.21.1) was assayed in potassium phosphate buffer (100 mM, pH 7.0) by following the hydrolysis of *N*-benzoyl-L-tyrosine ethyl ester (BTEE, 20 mM in DMSO) at λ = 256 nm.[34]

Subtilisin (protease type XXVII, EC 3.4.21.62) was assayed in potassium phosphate buffer (100 mM, pH 7.0) by following the hydrolysis of succinyl-Ala-Ala-Pro-Phe-*p*-nitroanalide (Succ-AAPF-*p*NA, 5 mM in DMSO) at $\lambda = 404$ nm.[35]

Carboxypeptidase Y (*S. cerevisiae*, EC 3.4.16.1) was assayed in Tris·HCl buffer (100 mM, pH 8.0) by following the hydrolysis of BTEE (20 mM in DMSO) at $\lambda = 256$ nm ($K_M = 0.35$ mM).

Elastase (porcine pancreas, EC 3.4.21.36) was assayed in Tris·HCl buffer (100 mM, pH 8.0) by following the hydrolysis of Succ-AAPF-*p*NA (4.4 mM in DMSO) at $\lambda = 412$ nm.

Computational details: The molecular structures of the reactants and products were fully optimized in the gas-phase by the DFT method, applying B3LYP functional and cc-pvdz basis set implemented in the Jaguar 4.1 package.[36] The absolute gas-phase free energy values G^g of reactants and products were calculated in harmonic approximation. The continuum reaction field solvation model implemented in Jaguar, SCRF,[37,38] was used in B3LYP/SCRF//cc-pvdz level of the DFT method to calculate $G^s(\varepsilon)$, free energies of solvation.

Acknowledgements

This research was supported by the "Marcus Center for Pharmaceutical and Medicinal Chemistry" at Bar Ilan University. The Kamea scholarship granted to M.S. by the Israeli Ministry of Absorption is gratefully acknowledged.

Keywords: binding trends · drug design · quantum chemistry · serine proteases · transition-state analogues

[1] J. G. Robertson, *Biochemistry* **2005**, *44*, 5561 – 5571.
[2] J. Zhong, W. C. Groutas, *Curr. Top. Med. Chem.* **2004**, *4*, 1203 – 1216.
[3] R. I. Christopherson, S. D. Lyons, P. K. Wilson, *Acc. Chem. Res.* **2002**, *35*, 961 – 971.
[4] E. De Clerq, *Nat. Rev. Drug Discovery* **2002**, *1*, 13 – 25.
[5] C. T. Supuran, A. Scozzafava, A. Mastrolorenzo, *Expert Opin. Ther. Pat.* **2001**, *11*, 221 – 259.
[6] M. Shokhen, D. Arad, *J. Mol. Model.* **1996**, *2*, 390 – 398.
[7] M. Shokhen, A. Albeck, *Proteins Struct. Funct. Genet.* **2000**, *40*, 154 – 167.
[8] K. Brady, R. H. Abeles, *Biochemistry* **1990**, *29*, 7608 – 7617.
[9] M. R. Angelastro, S. Mehdi, J. P. Burkhart, N. P. Peet, P. Bey, *J. Med. Chem.* **1990**, *33*, 11 – 13.
[10] M. W. Walter, R. M. Adlington, J. E. Baldwin, C. J. Schofield, *J. Org. Chem.* **1998**, *63*, 5179 – 5192.
[11] H.-J. Böhm, G. Klebe, *Angew. Chem.* **1996**, *108*, 2750 – 2778; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2588 – 2614.
[12] P. Greenzaid, Z. Luz, D. Samuel, *J. Am. Chem. Soc.* **1967**, *89*, 749 – 756.
[13] S. R. LaPlante, D. R. Cameron, N. Aubry, P. R. Bonneau, R. Deziel, C. Grand-Maitre, W. W. Ogilvie, S. H. Kawai, *Angew. Chem.* **1998**, *110*, 2858 – 2860; *Angew. Chem. Int. Ed.* **1998**, *37*, 2729 – 2732.
[14] L. Hedstrom, *Chem. Rev.* **2002**, *102*, 4501 – 4524.
[15] R. E. Babine, S. L. Bender, *Chem. Rev.* **1997**, *97*, 1359 – 1472.
[16] D. Voet, J. G. Voet, *Biochemistry*, 3rd ed, Wiley, New York, **2004**, p. 46.
[17] M. Shokhen, A. Albeck, *Proteins Struct. Funct. Genet.* **2004**, *54*, 468 – 477.
[18] M. Shokhen, A. Albeck, *Proteins Struct. Funct. Genet.* **2004**, *55*, 245 – 250.
[19] "An introduction to Density Functional Theory": L. J. Bartolotti, K. Flurchick in *Reviews in Computational Chemistry, Vol. 7* (Eds.: K. B. Lipkowitz, D. B. Boyd), VCH Publishers, New York, **1996**, chap. 4, pp. 187.
[20] R. A. Friezner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz, Y. Cao, *J. Phys. Chem. A* **1999**, *103*, 1913.
[21] R. B. Murphy, Y. Cao, M. D. Beachy, M. N. Ringalda, R. A. Friesner, *J. Chem. Phys.* **2000**, *112*, 10131.
[22] Y. Y. Sham, Z. T. Chu, A. Warshel, *J. Phys. Chem. B* **1997**, *101*, 4458.
[23] A. Warshel, S. T. Russell, *Q. Rev. Biophys.* **1984**, *17*, 283.
[24] C. N. Schutz, A. Warshel, *Proteins Struct. Funct. Genet.* **2001**, *44*, 400.
[25] M. Shokhen, N. Khazanov, A. Albeck, *ChemMedChem* **2006**, *1*, 639.
[26] K. G. Krebs, D. Heusser, H. Wimmer in *Thin Layer Chromatography*, 2nd ed (Ed.: E. Stahl), Springer, New York, **1969**, pp. 862 – 863.
[27] W. C. Still, M. Kahn, A. Mitra, *J. Org. Chem.* **1978**, *43*, 2923 – 2925.
[28] A. Albeck, R. Persky, *Tetrahedron* **1994**, *50*, 6333 – 6346.
[29] M. R. Angelastro, N. P. Peet, P. Bey, *J. Org. Chem.* **1989**, *54*, 3913 – 3916.
[30] L. De Luca, G. Giacomelli, A. Porcheddu, *J. Org. Chem.* **2001**, *66*, 7907 – 7909.
[31] F. Matsuda, M. Kawatsura, F. Dekura, H. Shirahama, *J. Chem. Soc. Perkin Trans. 1* **1999**, *16*, 2371 – 2375.
[32] C. J. Schofield, M. W. Walter, R. M. Adlington, J. E. Baldwin, J.-M. Frere, A. Felici, PCT Int. Appl. WO 9719681A1 19970605, **1997**, (CODEN: PIXXD2) [*Chem. Abstr.* **1997**, 127, 95610 AN 1997:467739, 1997].
[33] R. Kitz, I. B. Wilson, *J. Biol. Chem.* **1962**, *237*, 3245 – 3249.
[34] B. C. W. Hummel, *Can. J. Biochem. Physiol.* **1959**, *37*, 1393 – 1399.
[35] E. G. DelMar, C. Largman, J. W. Broderick, M. C. Geokas, *Anal. Biochem.* **1979**, *99*, 316 – 320.
[36] *JAGUAR 4.1*, Schrödinger Inc., Portland, OR (USA) **2000**.
[37] D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitcoff, A. Nicholls, B. Honig, M. Ringnalda, W. A. Goddard III, *J. Am. Chem. Soc.* **1994**, *116*, 11875 – 11882.
[38] B. Marten, K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnalda, D. Sitcoff, B. Honig, *J. Phys. Chem.* **1996**, *100*, 11775 – 11788.

Received: February 7, 2006
Published online on April 4, 2006

