DOI: 10.1002/cmdc.200600030

Enzyme Isoselective Inhibitors: Application to Drug Design

Michael Shokhen,* Netaly Khazanov, and Amnon Albeck*[a]

*Common methodologies of computer-assisted drug design focus on noncovalent enzyme–ligand interactions. We introduced enzyme isoselective inhibition trend analysis as a tool for the expert analysis of covalent reversible inhibitors. The methodology is applied to predict the binding affinities of a series of transition-state analogue inhibitors of medicinally important serine and cysteine hydrolases. These inhibitors are isoselective: they have identical noncovalent recognition fragments (*RS*) and different reactive chemical fragments (*CS*). Furthermore, it is possible to predict the binding affinities of a series of isoselective inhibitors toward a prototype enzyme and to extrapolate the data to a target medicinally important enzyme of the same family. Rational design of* CS *fragments followed by conventional* RS *optimization could be used as a novel approach to drug design.*

Introduction

Most computer-assisted drug design (CADD) methodologies focus on enzyme–ligand recognition as determined by noncovalent interactions.[1] This approach can be very useful in the design of drugs for metabolic diseases. On the other hand, well-optimized inhibitors of enzymes that are implicated in infectious diseases and cancer rapidly loose their in vivo activity through the development of mutational drug resistance.[2] A study of FDA-approved anti-AIDS drugs that function as noncovalent inhibitors of HIV protease demonstrated that about 75% of the enzyme's amino acid residues are subjected to mutations that facilitate drug resistance.[3] Structural analysis of the enzyme–inhibitor interactions led to the conclusion that mutations that give rise to drug resistance never occur at well-conserved residues that constitute the catalytic machinery of the enzyme. This conclusion led us to suggest that the problem of mutational drug resistance could be avoided if the main contribution to inhibitor binding affinity came from interactions with residues that constitute the enzyme's catalytic machinery. This would be achieved by covalent inhibitors, which can bind their target enzyme either reversibly or irreversibly. Analysis of the U.S. FDA Orange Book supports the pharmaceutical importance of covalent inhibitors, demonstrating that 65% of the 317 marketed drugs which are enzyme inhibitors either undergo reactive chemistry in the active site of the target enzyme or contain a structural motif related to the substrate.[4] The candidates best suited to maximally use interactions with the catalytic residues of the enzyme for binding energy are probably transition state (TS) analogue inhibitors.[5]

A covalent inhibitor is formally composed of two parts: the chemical site (CS) and the recognition site (RS). The former includes the atoms that covalently interact with the enzyme active site and the immediate neighboring atoms that influence their electronic state. The RS part includes the rest of the inhibitor and is generally considered to be responsible for the selectivity of the inhibitor toward the target enzyme.

In a previous theoretical study of TS analogue inhibitors of serine proteases, we demonstrated that the covalent bond formed between an inhibitor CS fragment and the enzyme Ser nucleophile 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] Wolfenden stressed that TS analogue inhibitors are very specific for the enzyme whose activated complex they resemble, just as the transition state is unique to that reaction.[5c] Therefore, even a relatively small TS analogue inhibitor with a small CS and a small RS fragment that occupies only the S_1 and S_1' sites of the target enzyme can be not only very potent but also sufficiently selective. At the same time, such an inhibitor will suffer much less from mutational drug resistance, as its recognition part is greatly decreased.

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.[7] The relative binding affinities of any two isoselective ligands are independent of variations of the RS fragment. Isoselective inhibitors have the same affinity trend toward different enzymes with a common catalytic mechanism. In the present study we demonstrate that these principles of isoselective inhibition can be used to introduce a novel method of theoretical analysis of the binding trend of inhibitor CS fragments.

[a] *Dr. M. Shokhen, N. Khazanov, 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

Results and Discussion

We examined the method of isoselective inhibition trend analysis on TS analogue inhibitors of serine and cysteine hydrolases in which the inhibitor forms a reversible covalent TC with the enzyme. In all series considered, the electrophilic center of the inhibitor is a carbonyl group modified by different covalent substituents. Hydration of the carbonyl group of free water-solvated inhibitor was considered as a competing reaction, and is accounted for in the apparent binding constant, K_i^{app}. We correlated the theoretically calculated K_i^{app} values with the corresponding experimental K_i^{app} values for inhibitors of five medicinally important enzymes: human neutrophil elastase (HNE),[8] fatty acid amide hydrolase (FAAH),[9] human α-thrombin,[10] hepatitis C viral NS3 protease,[11] and human rhinovirus 3C protease.[12] Each inhibitor series was based on a different RS fragment: the Cbz-Val-Pro-Val peptide in inhibitors of HNE (Figure 1a); the unsaturated hydrocarbon tail of the α-keto derivatives of oleamide, the natural substrate of FAAH (Figure 2a); Me-D-Phe-Pro-Arg in inhibitors of thrombin (Figure 3a); a peptidomimetic in the NS3 inhibitors (Figure 4a); and a substituted isatin heterocycle in the 3C inhibitors (Figure 5a). The varied substituent at the carbonyl group of the CS fragment is marked by X. The trend of binding affinity of CS fragments in a series of isoselective inhibitors is controlled by the energy of the covalent bond formation in the TC.[6] Therefore, experimental values of K_i^{app} are well reproduced by a simplified small-molecular modeling scheme for TC formation in the active site of a serine hydrolase [Eq. (1)] and the hydration side reaction [Eq. (2)].[7] The cysteine hydrolase reaction can similarly be modeled by Equation (3). In the latter model, the inhibitor CS carbonyl oxygen atom is protonated in the TC, as has been suggested by a few published studies.[13] This model also gave much better correlation between the theoretically calculated and experimentally reported K_i values than the corresponding ionized (oxyanion) model (data not shown).

$$CH_3O^- + CH_3{-}CO{-}X \rightarrow CH_3{-}C(O^-)(OCH_3){-}X \quad (1)$$

$$H_2O + CH_3{-}CO{-}X \rightarrow CH_3{-}C(OH)_2{-}X \quad (2)$$

$$CH_3S^- + CH_3{-}CO{-}X + H^+ \rightarrow CH_3{-}C(OH)(SCH_3){-}X \quad (3)$$

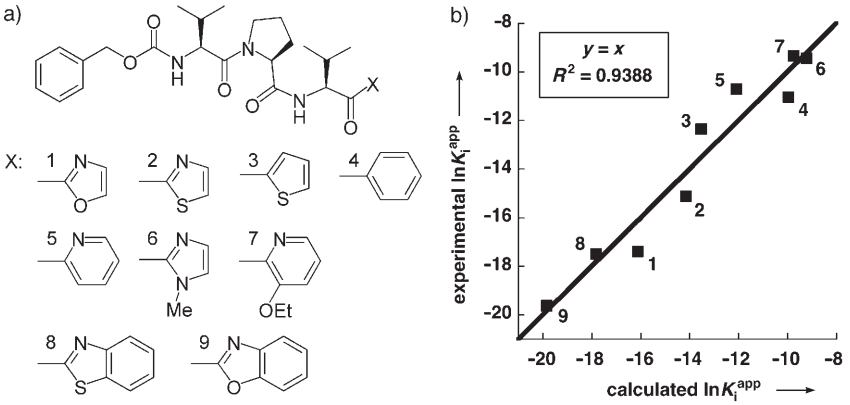

Figure 1. a) Cbz-Val-Pro-Val peptide inhibitors of human neutrophil elastase (HNE); Cbz = benzyloxycarbonyl. b) Linear correlation according to Equation (4) between experimental and theoretically calculated values of lnK_i^{app} for the inhibition of HNE. The best fit corresponds to the following coefficients and parameters: $\varepsilon_{eff} = 6$, $a = 1.2888$, $b = -0.2661$, $c = -9.9641$, SE = 1.107.

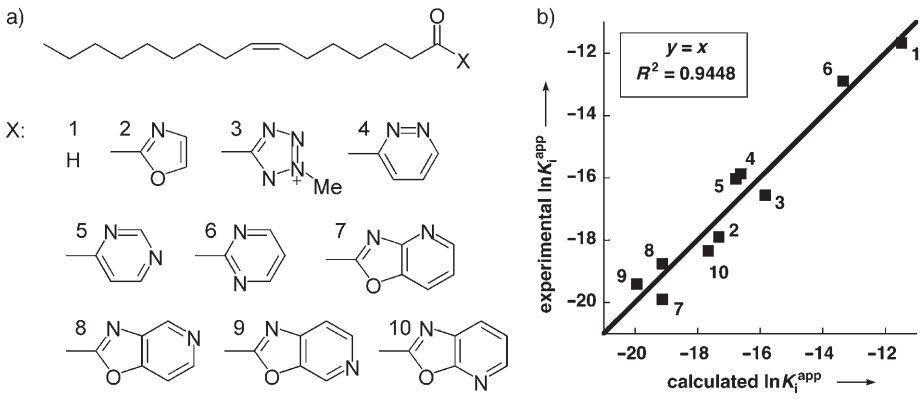

Figure 2. a) α-Keto derivatives of oleamide, the natural substrate of fatty acid amide hydrolase (FAAH). b) Linear correlation according to Equation (4) between experimental and theoretically calculated values of lnK_i^{app} for the inhibition of FAAH. The best fit corresponds to the following coefficients and parameters: $\varepsilon_{eff} = 20$, $a = 0.7983$, $b = -0.7286$, $c = -11.4807$, SE = 0.725.

Linear-regression fitting of experimental values of $K_{i,k}^{app}$ was calculated by two independent variables: the relative free energy of TC formation ($\Delta\Delta G_{TC,k}$) and the relative free energy of competitive product formation ($\Delta\Delta G_{C,k}$) for a given inhibitor k (see ref. [7] for details):

$$\ln K_{i,k}^{app} = a\,\Delta\Delta G_{TC,k} + b\,\Delta\Delta G_{C,k} + c \quad (4)$$

The optimized linear regression coefficients *a*, *b*, and *c* were used to calculate theoretical values of lnK_i^{app}. The corresponding plots of experimental lnK_i^{app} versus calculated lnK_i^{app} values for enzyme inhibition are presented in Figures 1b–5b.

The heterogeneous combined protein/water environmental effect on inhibitor binding affinity was taken into account by our QM/SCRF(VS) approach[14] with empirically fitted dielectric constant of the virtual solvent, ε_{eff}, at the B3LYP/cc-pvdz level of nonempirical quantum mechanical DFT calculations. Such small-molecule simulations can examine, in reasonable time, the affinity of dozens of different CS fragments in high-level ab initio quantum mechanics. Another

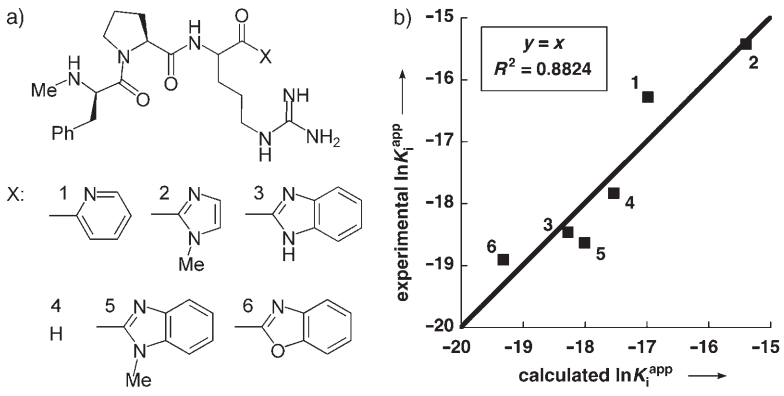

Figure 3. a) Me-D-Phe-Pro-Arg inhibitors of human α-thrombin. b) Linear correlation according to Equation (4) between experimental and theoretically calculated values of $\ln K_i^{app}$ for the inhibition of thrombin. The best fit corresponds to the following coefficients and parameters: $\varepsilon_{eff} = 1$, $a = 0.2925$, $b = 0.1896$, $c = -17.5364$, SE = 0.629.

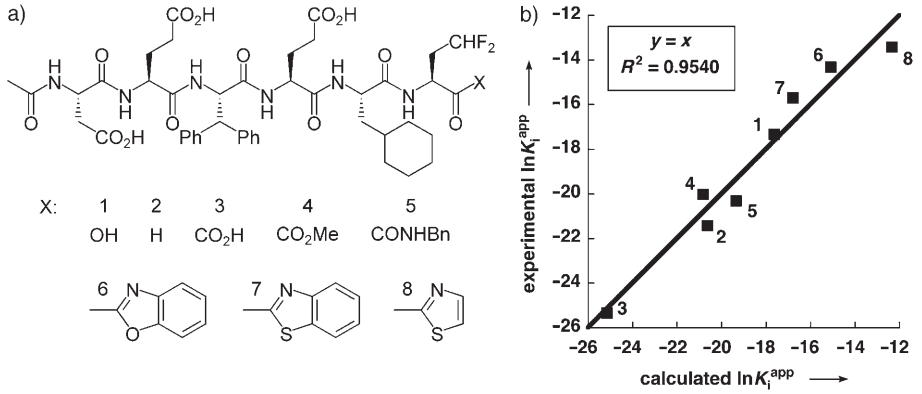

Figure 4. a) Peptidomimetic inhibitors of hepatitis C viral NS3 protease; Bn = benzyl. b) Linear correlation according to Equation (4) between experimental and theoretically calculated values of $\ln K_i^{app}$ for the inhibition of NS3 protease. The best fit corresponds to the following coefficients and parameters: $\varepsilon_{eff} = 8$, $a = 0.3745$, $b = 0.9129$, $c = -20.6420$, SE = 1.018.

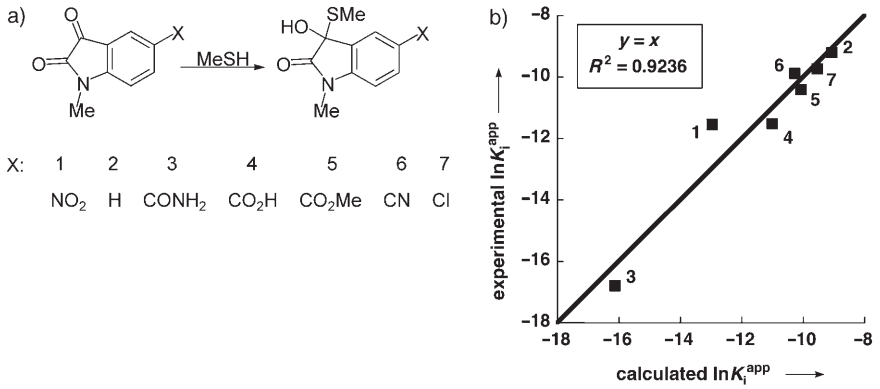

Figure 5. a) Substituted isatin heterocycle inhibitors of human rhinovirus 3C protease. b) Linear correlation according to Equation (4) between experimental and theoretically calculated values of $\ln K_i^{app}$ for the inhibition of 3C cysteine protease. The TC is protonated, as presented in the reaction scheme of its formation and in Equation (3). The best fit corresponds to the following coefficients and parameters: $\varepsilon_{eff} = 1$, $a = -0.2932$, $b = 0.3588$, $c = -9.0766$, SE = 0.872.

advantage of this approach, which focuses only on the reactivity functional groups, is that it does not require any 3D structural information about the enzyme. This latter feature is very important in cases for which such information is not available, as is the case for most membrane-bound enzymes. On the other hand, this 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 excellent correlation between theoretical and experimental $\ln K_i^{app}$ values, as the enzyme–inhibitor covalent bond formed is by far the most trend-dominating factor.[6] The small deviations from linearity may indeed reflect various local noncovalent interactions. Taking these interactions into account with additional indexes in the regression analysis [Eq. (4)] in a statistically significant way would require a much larger data set.

Gleeson et al. applied a QM/MM computational approach, the current state of the art in the computational biochemistry of enzyme active sites,[15] for modeling the binding of the series of peptidyl α-ketoheterocyclic inhibitors in the active site of HNE.[16] Semiempirical quantum mechanics with PM3 Hamiltonian was used for the active-site fragment, molecular mechanics with AMBER force field for the rest of the protein, and TIP3P solvent model (with explicit water molecules in the periodic box) for the bulk water solvation of the free inhibitor and the enzyme–inhibitor complex. Nevertheless, these calculations produced poor correlation between the calculated and experimental binding energies for the set of the inhibitors. The failure was attributed to insufficient accounting of the solvation effects.[16] Our opinion is that all factors are important. The poor prediction could be the result of intrinsic deficiencies of the semiempirical PM3 Hamiltonian, of ignoring the competitive reactions, and clearly of the solvation model used. The graph of experimental versus theoretical $\ln K_i^{app}$ values, calculated by our method for the same set of inhibitors, is presented in Figure 1 b. Comparison of the quality of correlation of the QM/MM approach with our method demonstrates the crucial role of the CS fragment in the binding trend of isoselective inhibi-

tors. Thus, inadequate calculation of the factors related to CS reactivity cannot be compensated by extensive molecular mechanics accounting of the protein body and the noncovalent enzyme–inhibitor interactions because the latter play a minor role in the binding trend of covalent isoselective inhibitors.

We have presented very good correlations between experimentally measured and theoretically calculated binding affinities (Figures 1–5), but the real challenge is in the prediction of binding affinities of new inhibitors with different CS fragments. To demonstrate this, we considered a subset of four HNE inhibitors that span a wide range of binding affinities as a "training set" with known experimental data and calculated their theoretical K_i^{app} values [Eq. (4)]. The optimized linear regression coefficients *a*, *b*, and *c* obtained for this "training set" were then used for prediction of the K_i^{app} values of the other five inhibitors. The resulting predicted binding affinities correlate very well with the experimental data (Figure 6a). A more demanding task is the prediction of binding affinities that are far outside the range of affinities of the training set (extrapolation). Figure 6b exhibits such a successful application for two inhibitors of FAAH.

The given medicinally important target enzyme may be unavailable for various reasons. The above formulated principles of isoselective inhibitors provide a solution for this problem. The inhibitor binding trend for the target enzyme can be predicted for an alternative, readily available enzyme (a "prototype") if both enzymes belong to the same mechanistic enzymatic family. Thus, the first step of the method is to prepare a "training set" of inhibitors—a representative set of varied CS fragments (that cover a wide range of binding constants) of isoselective inhibitors for the prototype enzyme. This step is followed by the experimental measurement of binding constants for the training set and the prototype enzyme. An optimized QSAR model that correlates the experimental binding constants with the reactivity descriptors such as $\Delta\Delta G_{TC,k}$ and $\Delta\Delta G_{C,k}$, or any other variables, will be generated. Based on this QSAR equation, it is now possible to predict binding constants for the prototype enzyme of any other CS fragment that does not belong to the training set.

To illustrate and validate this methodology, we used chymotrypsin as a prototype enzyme and Cbz-Phe-X (X = H, OH, CH_3, CF_3, and $COCH_3$) as the training set of isoselective inhibitors (see ref. [7] for details; Figure 7a). Equation (4) was used as a QSAR model. HNE and FAAH are medicinally important en-

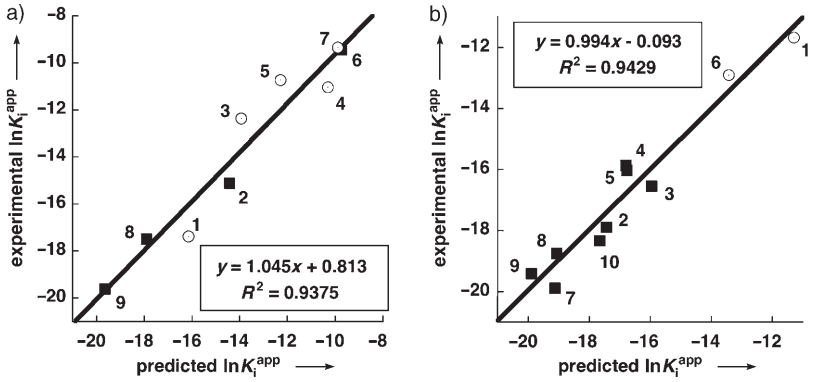

Figure 6. a) Prediction (interpolation) of binding affinities of "new" CS fragments in the inhibition of HNE by Cbz-Val-Pro-Val peptide inhibitors. The set of isoselective inhibitors is the same as that shown in Figure 1a. Filled squares indicate the "training set" isoselective inhibitors; open circles indicate the predicted isoselective inhibitors. The best fit of the "training set" corresponds to the following coefficients and parameters: $\varepsilon_{eff} = 6$, $a = 1.238$, $b = -0.3180$, $c = -10.3073$, SE = 0.8647. b) Prediction (extrapolation) of binding affinities of "new" CS fragments in the inhibition of FAAH by α-keto derivatives. The set of isoselective inhibitors is the same as that shown in Figure 2a. Filled squares indicate the "training set" isoselective inhibitors; open circles indicate the predicted isoselective inhibitors. The best fit of the "training set" corresponds to the following coefficients and parameters: $\varepsilon_{eff} = 20$, $a = 0.7796$, $b = -0.7819$, $c = -11.2889$, SE = 0.8211.

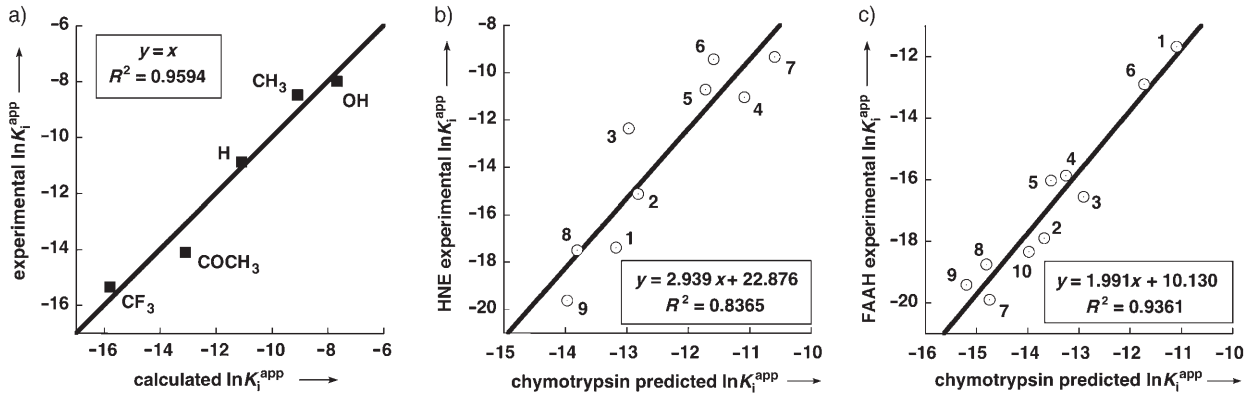

Figure 7. a) Inhibition of chymotrypsin by the Cbz-Phe-X (X = H, OH, CH_3, CF_3, and $COCH_3$) training set of inhibitors; $\varepsilon_{eff} = 20$, $a = 0.4221$, $b = -0.3057$, $c = -11.0906$, SE = 0.9362. b) Prototype-based prediction for HNE inhibition. The set of isoselective inhibitors is the same as that shown in Figure 1a. The calculations are based on the QSAR parameters for chymotrypsin and its training set (Figure 6a). c) Prototype-based prediction for FAAH inhibition. The set of isoselective inhibitors is the same as that shown in Figure 2a. The calculations are based on the QSAR parameters for chymotrypsin and its training set (Figure 6a).

zymes that belong to the mechanistic family of chymotrypsin. The corresponding predictions of inhibition constants of isoselective inhibitors of these two enzymes are presented in Figures 7b and c; they are based on the QSAR model obtained for chymotrypsin and its own training set of isoselective inhibitors. Comparison of Figures 1 and 2 with Figures 7b and c, respectively, shows that very good correlations between the theoretically calculated and the experimentally measured K_i values were obtained not only for direct modeling of the target enzyme, but also for correlation of the experimental data with calculations carried out on the "prototype" enzyme. They both provide practically the same results, justifying the idea of extrapolating trend data from one enzyme to another target enzyme of the same mechanistic family.

Conclusions

The methodology presented herein is a tool for the expert analysis of any set of new CS fragments that can be arranged according to their binding affinity toward the target enzyme. It allows prediction of the binding affinity of new inhibitors based on the analysis of the binding trend of a "training set" of inhibitors to the same enzyme. Furthermore, it is even possible to predict their binding to a prototype enzyme and to extrapolate it to the target medicinally important enzyme. The rational design of CS fragments followed by conventional RS optimization[1] could be used as a novel approach toward new classes of enzyme inhibitors.

Experimental Section

Computational methods: The most stable rotational conformers of reactants and product TCs were identified by the Monte Carlo Multiple Minimum (MCMM) method[17,18] implemented in the Macromodel package.[19] The energies were calculated with the MM2 force field. These conformers were used for the following quantum mechanical calculations.

The molecular structures of the reactants and the 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.[20] The absolute gas-phase free energies of reactants and products were calculated in harmonic approximation. The continuum reaction field solvation model implemented in Jaguar, SCRF,[21,22] was used in B3LYP/SCRF//cc-pvdz level of the DFT method to calculate 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: enzyme inhibition · hydrolases · proteases · QSAR · quantum mechanical calculations

[1] a) P. M. Dean, D. G. Lloyd, N. P. Todorov, *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 347–353; b) A. N. Jain, *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 396–403; c) A. M. Davis, S. J. Teague, G. J. Kleywegt, *Angew. Chem.* **2003**, *115*, 2822–2841; *Angew. Chem. Int. Ed.* **2003**, *42*, 2718–2736; d) J. Krumrine, F. Raubacher, N. Brooijmans, I. Kuntz, *Methods Biochem. Anal.* **2003**, *44*, 443–476; e) A. C. Anderson, *Chem. Biol.* **2003**, *10*, 789–797; f) Y. C. Martin, *J. Comb. Chem.* **2001**, *3*, 231–250.

[2] a) R. Kantor, D. Katzenstein, *J. Clin. Virol.* **2004**, *29*, 152–159; b) D. B. Longley, P. G. Johnston, *J. Pathol.* **2005**, *205*, 275–292; c) C. Therrien, R. C. Levesque, *FEMS Microbiol. Rev.* **2000**, *24*, 251–262.

[3] W. Wang, P. A. Kollman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14937–14942.

[4] J. G. Robertson, *Biochemistry* **2005**, *44*, 5561–5571.

[5] a) R. Wolfenden, *Nature* **1969**, *223*, 704–705; b) G. E. Lienhard, *Science* **1973**, *180*, 149–154; c) R. Wolfenden, *Biophys. Chem.* **2003**, *105*, 559–572.

[6] a) M. Shokhen, D. Arad, *J. Mol. Model.* **1996**, *2*, 390–398; b) M. Shokhen, A. Albeck, *Proteins Struct. Funct. Genet.* **2000**, *40*, 154–167.

[7] R. Ozeri, N. Khazanov, N. Perlman, M. Shokhen, A. Albeck, *ChemMedChem* **2006**, *6*, 631.

[8] P. D. Edwards, D. J. Wolanin, D. W. Andisik, M. W. Davis, *J. Med. Chem.* **1995**, *38*, 76–85.

[9] D. L. Boger, H. Sato, A. E. Lerner, M. P. Hedrick, R. A. Fecik, H. Miyauchi, G. D. Wilkie, B. J. Austin, M. P. Patricelli, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5044–5049.

[10] M. J. Costanzo, H. R. Almond, Jr., L. R. Hecker, M. R. Schott, S. C. Yabut, H.-C. Zhang, P. Andrade-Gordon, T. W. Corcoran, E. C. Giardino, J. A. Kauffman, J. M. Lewis, L. de Garavilla, B. J. Haertlein, B. E. Maryanoff, *J. Med. Chem.* **2005**, *48*, 1984–2008.

[11] F. Narjes, K. F. Koehler, U. Koch, B. Gerlach, S. Colarusso, C. Steinkucler, M. Brunetti, S. Altamura, R. De Francesco, V. G. Matassa, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 701–704.

[12] S. E. Webber, J. Tikhe, S. T. Worland, S. A. Fuhrman, T. F. Hendrickson, D. A. Matthews, R. A. Love, A. K. Patick, J. W. Meador, R. A. Ferre, E. L. Brown, D. M. DeLisle, C. E. Ford, S. L. Binford, *J. Med. Chem.* **1996**, *39*, 5072–5082.

[13] a) A. E. Howard, P. A. Kollman, *J. Am. Chem. Soc.* **1988**, *110*, 7195–7200; b) D. Arad, R. Langride, P. A. Kollman, *J. Am. Chem. Soc.* **1990**, *112*, 491–502; c) A. Albeck, S. Fluss, R. Persky, *J. Am. Chem. Soc.* **1996**, *118*, 3591–3596.

[14] a) M. Shokhen, A. Albeck, *Proteins Struct. Funct. Genet.* **2004**, *54*, 468–477; b) M. Shokhen, A. Albeck, *Proteins Struct. Funct. Genet.* **2004**, *55*, 245–250.

[15] A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

[16] M. P. Gleeson, I. H. Hiller, N. A. Burton, *Org. Biomol. Chem.* **2004**, *2*, 2275–2280.

[17] G. Chang, W. C. Guida, W. C. Still, *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.

[18] M. Saunders, K. N. Houk, Y. D. Wu, W. C. Still, M. Lipton, G. Chang, W. C. Guida, *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.

[19] *Macromodel 7.1*, Schrödinger Inc., Portland, OR (USA) **2000**.

[20] *JAGUAR 4.1*, Schrödinger Inc., Portland, OR (USA) **2000**.

[21] 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.

[22] 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 March 24, 2006

