# Query Generation to Search for Inhibitors of Enzymatic Reactions†

Martin Reitz, Alexander von Homeyer, and Johann Gasteiger*

University of Erlangen-Nuernberg, Computer-Chemie-Centrum and Institute for Organic Chemistry, Naegelsbachstr. 25, 91052 Erlangen, Germany

Received November 17, 2005

A method for the generation of intermediates of enzyme-catalyzed reactions is presented. These intermediates can be used as three-dimensional structural queries for searching for inhibitors of enzymatic reactions. The intermediates can be considered as being structurally quite close to transition-state analogues. For this application, a database containing detailed chemical information on metabolic reactions is used. The likely three-dimensional structure of the intermediates of enzyme-catalyzed reactions can be generated from the information in the database. For three reactions catalyzed by the enzymes AMP deaminase (EC code 3.5.4.6), triose phosphate isomerase (EC code 5.3.1.1), and arginase II (EC code 3.5.3.1), we show how a 3D model of these intermediates can be superimposed onto known inhibitors of these enzymes by a program that uses a genetic algorithm. For this, we test different methods for the superimposition using information on the enzymatic binding site, using information on physicochemical properties calculated from the molecular structure, or without having any information in the superimposition process. We show that these inhibitors are most similar to the corresponding intermediates regarding the 3D structure.

## INTRODUCTION

In recent years, much research has been put into elucidating how genes control biochemical reactions, the endogenous metabolism. Genes express proteins which are often enzymes that catalyze these biochemical reactions. This catalysis is highly efficient, leading to rate enhancements of up to $10^{20}$ compared to those of the uncatalyzed reactions.[1] These rate enhancements are caused by a variety of factors encompassing geometric, electronic, and bonding effects. Current studies point out that there might also be covalent bonding involved between the transition state and the enzyme to explain such proficient rate enhancements.[2]

Clearly, an enzyme must bind the substrate of a reaction, but beyond that, it must even more favorably bind the transition state of a reaction, leading to a substantial lowering of the activation energy. This was pointed out quite some time ago by Linus Pauling,[3,4] who stated that enzymes stabilize the transition states of biochemical reactions by binding them very tightly and, thus, lowering the energy barrier of the reaction. He further postulated that analogues to these transition states should act as potent inhibitors of enzymatic reactions. The inhibitor of an enzyme should be quite similar to the transition state of the reaction catalyzed by this enzyme in terms of the geometric arrangement and the physicochemical effects. However, in contrast to the transition state, an inhibitor cannot undergo the bond-breaking and -making process observed in the enzymatic reaction of the natural substrate. Thus, the transition-state analogue occupies the catalytic site of the enzyme and blocks it from processing the natural substrate, leading to inhibition.

Biochemical reactions catalyzed by enzymes form a network that is regulated by signaling pathways and, most importantly, by the expression of enzymes. Many dysfunctions of metabolic pathways in human and other species result from an unbalancing of these reaction networks, and it is, therefore, of high interest to interfere in the regulation of pathways. The inhibition of enzymes is, thus, an important tool in drug and agrochemical research.[5]

An understanding of the structure of the transition state and the intermediates of an enzyme-catalyzed reaction asks for an atomic resolution in the analysis of substrates and its reactions. Clearly, the structures of transition states and reaction intermediates can be calculated by quantum mechanical methods of various degrees of sophistication. This often asks for substantial amounts of computational efforts. We, however, were interested in developing a fast method that can be applied to large data sets of molecules. That is where chemoinformatics has to come in, to model the 3D structure of substrates and to analyze physicochemical effects that bind small molecules in proteins and that make bonds break and new ones form.

To support this endeavor, we have developed BioPath, a database of biochemical reactions, that stores molecules and reactions at atomic resolution.[6] Specifically, molecules are stored as connection tables, as lists of all atoms and all bonds.

Such a standard representation of chemical structures by connection tables allows the interfacing of automatic 3D-structure generators for obtaining 3D molecular models. The bond-breaking and -making events in the biochemical reactions are indicated by marking the reaction center and by mapping the atoms of the reactants onto those of the products.

The marking of the reaction center plays a crucial role in the studies reported here as it allows the generation of

* Corresponding author tel.: +49-9131-8526570; fax: +49-9131-8526566; e-mail: gasteiger@chemie.uni-erlangen.de; Web site: http://www2.chemie.uni-erlangen.de.

† Part of the Professor Johann Gasteiger special issue.

10.1021/ci050503u CCC: $33.50 
Published on Web 06/02/2006

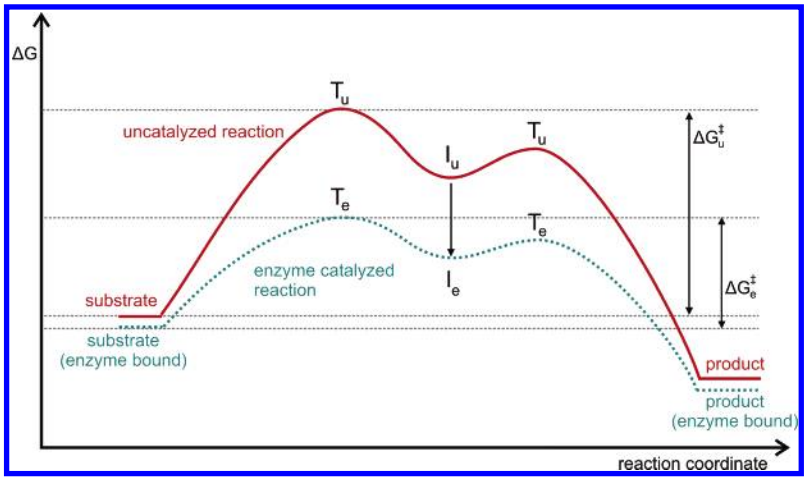


**Figure 1.** Energy diagram of an uncatalyzed reaction compared to an enzyme-catalyzed reaction ($\Delta G_u^{\ddagger}$ vs $\Delta G_e^{\ddagger}$) with the corresponding transition states $T_u$ and $T_e$ and reaction intermediates $I_u$ and $I_e$.

intermediates of enzymatic reactions. This, in conjunction with the 3D modeling of all molecules, puts us in a position to explore how inhibitors of enzymes match in 3D space with the starting materials, intermediates, and products of enzyme-catalyzed reactions.

On the basis of this, the generation of intermediates from the information contained in the BioPath database provides a 3D structural query for searching for inhibitors of enzyme-catalyzed reactions. We are testing this methodology with several enzymatic reactions for which inhibitors are known. This should provide a proof of concept for then using only information on the structure of a reaction intermediate to search for inhibitors in 3D structure databases.

## MATERIALS AND METHODS

*General Outline.* The structure of a transition state can be determined best by quantum mechanical calculations. The exact determination where the transition state lies on the reaction coordinate, that is, which geometry, bond lengths, and energy the transition state has, requires quite sophisticated quantum mechanical calculations. To avoid determining the exact geometry and energy of a transition state by time-consuming calculations, we simplify the problem by first investigating those reactions that proceed through a reaction intermediate. Such reactions are predominantly observed when the reaction occurs through an attack at a $C_{sp^2}$ atom involving first an addition and then an elimination step. When the energy of such a reaction intermediate is appreciably above the substrate, the structure of the transition state should be quite close to that of the reaction intermediate according to the Hammond postulate.[7]

Figure 1 shows the energy diagram of an uncatalyzed and an enzyme-catalyzed reaction proceeding through an intermediate. In this diagram, it is assumed that the binding of the substrate leads to an energy decrease, but the energy decrease for the binding of the reaction intermediate, $I_e$, is much more pronounced, in accordance with the Pauling hypothesis.[3,4] Such intermediates of a reaction can automatically be generated if an appropriate data source is available. A suitable database for this task, the BioPath database, is presented in the next section. The general outline of the approach for generating reaction intermediates as transition-state models and then searching for transition-state analogues

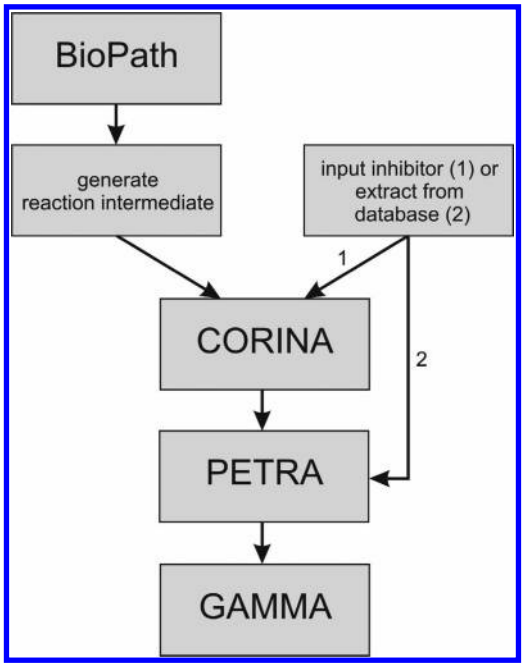


**Figure 2.** General outline of the process of comparing reaction intermediates with enzyme inhibitors indicating the different steps and the software programs used.

is presented in Figure 2. The various steps and the software involved are presented now in detail.

*BioPath Database.* The BioPath biochemical pathways database is a database of molecules involved in endogenous metabolism and the reactions interconverting them. The database was produced from the information which is contained on the famous wall chart distributed by Boehringer Mannheim, now Roche.[8] To make the wealth of data contained on the poster and the corresponding atlas[9] accessible by computational methods, an effort was made to input all of the information into a database. For this purpose, all structures were entered as connection tables, lists of all atoms and their bonds. Reactions were represented by their starting materials and the products and cofactors involved, giving the full stoichiometry of the reaction including even protons. Furthermore, all atoms of the starting materials were mapped onto those of the products, indicating their correspondence by the numbers of their atoms, and all reaction sites where bonds are broken, made, or altered were marked. This latter

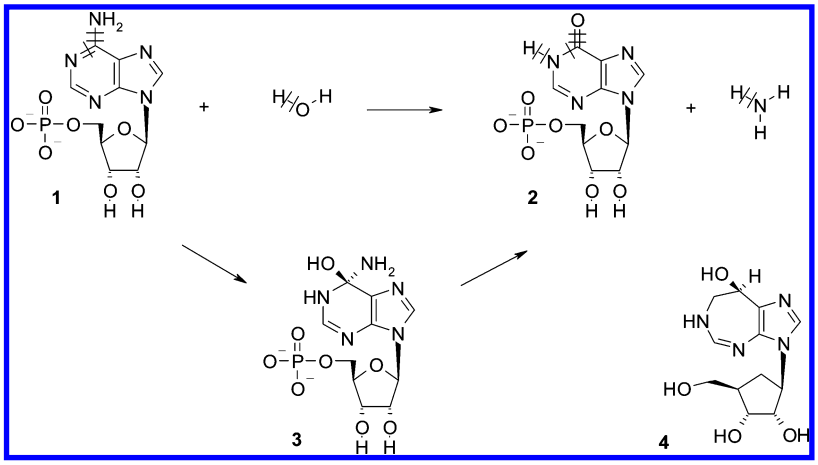

**Figure 3.** Hydrolysis of AMP, **1**, into IMP, **2**, and ammonia by AMP deaminase as stored in the BioPath database. The bonds broken and made are marked by lines crossing the bonds. The reaction intermediate, **3**, as generated from this reaction center information, and carbocyclic coformycin, **4**, an inhibitor of AMP deaminase, are also shown.

feature makes our database unique among all other databases of metabolic pathways such as, for example, KEGG[10] or those on the BioCyc[11] Web page. Additionally, each reaction was enriched by supplementary information such as enzyme name, EC number, the pathway the reaction is part of, and the organism in which it occurs.

The BioPath database presently consists of about 2200 reactions and more than 1500 structures. BioPath has been made accessible through the C@ROL[12] retrieval system on the Web at http://www2.chemie.uni-erlangen.de/services/biopath/index.html and http://www.mol-net.com/databases/biopath.html.

Of eminent importance for the application reported here is that all reactions in BioPath have their reaction centers marked; that is, the bonds broken and made in a reaction are indicated, and the atoms of those bonds are mapped from the starting materials onto those in the products. This allows the automatic construction of reaction intermediates. Figure 3 illustrates this for the reaction catalyzed by AMP deaminase (AMPDA; EC 3.5.4.6), converting adenosine monophosphate (AMP), **1**, into inosine monophosphate (IMP), **2**. From the information on which bonds are broken in this reaction, the reaction intermediate, **3**, can be generated.

To generate the reaction intermediate, the BioPath database is loaded into the CACTVS system.[13] This program offers an extensive scripting interface which allows the manipulation of data. For this application, a program was implemented which allows the generation of intermediates for several reaction types.

This is done by a simple algorithm which uses the information on the bonds broken and made in the reaction center for a specific reaction type. It allows the generation of intermediates for all reactions matching a specific reaction type. First, the reaction center for a specific reaction is defined, and then, the BioPath database is scanned for all reactions matching this defined reaction center. The retrieved reactions are stored in a hit list. The reactions from the hit list are then split into a substrate handle and a product handle. The handle which is closer to the intermediate (reaction center and transformation to build the intermediate) is then modified by making and breaking the bonds that are part of the reaction center according to the intermediate. The generated intermediates are saved in a file.

*CORINA.* Compared to the vast amount of known compounds, experimental 3D structure information from X-ray data is available only for a small fraction of compounds. To obtain 3D structure information also for those compounds where no experimental data are on-hand, computational methods are necessary. The 3D structure generator CORINA (COoRdINAtes) converts the constitution of a molecule as laid down in a connection table into a 3D structure.[14,15] This 3D molecular model is a single, low-energy conformation of a molecule. It should be emphasized that this conformation does not necessarily correspond to the biologically active conformation. This problem will be addressed later in the superimposition process by the program GAMMA.

*PETRA.* The program package PETRA (Parameter Estimation for the Treatment of Reactivity Applications)[16] allows the calculation of a variety of physicochemical effects in organic molecules by using various empirical methods. PETRA can calculate properties for atoms, bonds, or the whole molecule. In our studies, we used a variety of atom properties such as total charges, lone pair electronegativities, effective polarizabilities, and the lipophilicity represented by the octanol/water coefficient. These values can be used in the superimposition process by the program GAMMA.

*GAMMA.* The method developed for the superimposition of three-dimensional structures is based on atom−atom matching of the non-hydrogen atoms, including conformational flexibility of the compounds. The key algorithms have been described elsewhere in detail.[17]

In this approach, two functions are additionally used to automatically superimpose molecules. First, atoms can optionally be characterized by physicochemical properties. The atoms to be overlaid must then conform to a given interval of the physicochemical property. For example, if the matching criterion is chosen to be total atomic charges, $q_{tot}$, and the interval is selected to be $q_{tot} = \pm 0.05\ e$, then, for an atom of the first molecule with $q_{tot} = -0.2\ e$, only atoms in the interval of $q_{tot} = [-0.25, -0.15]$ are allowed to build match tuples with this first atom. Combinations of several physicochemical properties have to be valid at the same time. The physicochemical properties are calculated by the program package PETRA.[14]

Second, GAMMA allows the selection of sets of atom tuples that can be enforced to match. Therefore, indices have

to be given for all those atoms of the molecules that must build match tuples with each other. The remainder of the atoms have to fit the resulting spatial or, if given, physicochemical demands.

The quality of a superposition is scored by the root-mean-square (RMS) error and the size of the achieved substructure.

## RESULTS AND DISCUSSION

In the following, the method for generating intermediates of enzyme-catalyzed reactions and their superimposition with inhibitors of these reactions was tested for three enzymes. The examples were chosen so as to cover quite different substrates and reaction types.

The study with AMP deaminase (EC code 3.5.4.6) investigates a hydrolysis reaction in an aromatic heterocyclic system.

With triose phosphate isomerase (TIM; EC code 5.3.1.1), an isomerization reaction in an aliphatic system showing a fair amount of conformational flexibility was investigated. Furthermore, the study of two different inhibitors provided deeper insights into the structure of the intermediate of this reaction and important features of the binding pocket.

The last example, arginase II (EC code 3.5.3.1), investigates a hydrolysis reaction of an aliphatic system having substantial conformational flexibility. Furthermore, the hydrolysis of a guanidine group can serve as a model reaction for a large group of hydrolysis reactions involving ester and amide groups. Three different inhibitors were studied in order to gain deeper insights into the validity of the transition-state hypothesis and the performance of our approach.

The studies presented here also investigate the use of different types of 3D structures. In the first three investigations, all 3D structures, of the starting materials, of the products, of the intermediates, and of the inhibitors, were generated by CORINA and then submitted to 3D structure superimposition by GAMMA (see the Materials and Methods section). This is the approach which has to be taken when searching in databases for new potential enzyme inhibitors where no information about existing drugs or the enzyme binding site is available. In the last study with arginase II, the 3D structures of the inhibitors were taken from the 3D experimental observations as stored in the Protein Data Bank (PDB).[18] Structures of the intermediates, on the other hand, had to be generated by CORINA.

In this last study, how the quality of the superimposition is affected if different levels of knowledge are given in the superimposition process was also explored. For this, three different overlay procedures were performed. First, the atoms which are known to participate in the hydrogen bondings of the inhibitor into the catalytic pocket of the enzyme were forced to be matched to the corresponding atoms of the intermediate. In the second approach, constraints were provided that allowed only atoms of both molecules with similar physicochemical properties to be matched to each other. In the third superimposition process, no further constraints were provided in order to see how the program can find a solution if no information on binding is available.

For all shown superimpositions, the intermediate structure generated from BioPath was handled as flexible, while the superimposition partner (the substrate, the product, or the inhibitor) served as a rigid template.

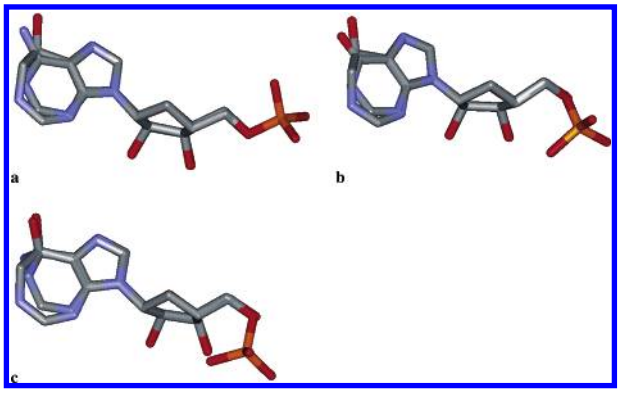


**Figure 4.** Superimposition of the inhibitor carbocyclic coformycin, **4**, (a) with the reaction substrate AMP, **1**, (b) with the reaction product IMP, **2**, and (c) with the AMP deaminase reaction intermediate, **3**.

*AMP Deaminase.* As a first example, we selected the enzyme AMP deaminase (EC code 3.5.4.6) from the database, an enzyme which provides a relevant function in purine nucleotide metabolism. AMPDA catalyzes the conversion of adenosine-5′-monophosphate (AMP), **1**, to inosine-5′-monophosphate (IMP), **2**, by hydrolytic deamination and plays a key role in maintaining the relative concentrations of adenylate nucleotides. During this conversion, a tetrahedral intermediate, **3**, is formed, where the leaving ammonia group and the attacking water molecule are simultaneously bound (addition mechanism, see Figure 3).

This enzyme has been well-examined in recent years, and there are efficient transition-state analogues known. A deficiency of AMPDA leads to a disruption of muscle energy production. Symptoms are rapid fatigue, pain, and cramps. In plants, an inhibition of AMPDA results in a strong herbicidal effect.[19] One of these known inhibitor molecules is carbocyclic coformycin, **4** (Figure 3), a fermentation product of *Saccharothrix* spp. with herbicidal activity.[20]

In our investigations, the inhibitor, carbocyclic coformycin, **4**, was superimposed by the program GAMMA with the starting material, AMP, **1**; with the product, IMP, **2**; and with the tetrahedral intermediate, **3**. As explained in the Materials and Methods section, we always selected the reaction intermediate for our investigations as, in contrast to a transition state, the bonds in a reaction intermediate are clearly defined. In this reaction, the structure of the transition state should be structurally close to the intermediate according to the Hammond postulate.[7] Figure 4 shows the resulting superimpositions obtained from GAMMA.

Clearly, in all three cases, the ribose part and the five-membered imidazol part of the bicyclic ring system match nearly exactly the corresponding parts of the carbocyclic coformycin. It is also clear that the seven-membered ring of the inhibitor cannot completely match the six-membered ring of AMP, IMP, or the reaction intermediate. Differences in the three superimpositions show up at the reaction site where the $NH_2$ group is exchanged against an OH group (into its tautomeric form, to be exact). In the superimposition of AMP and the inhibitor, the $NH_2$ group of AMP and the OH group of the inhibitor point to substantially different directions in space (Figure 4a). The RMS value for this superimposition is 0.19 Å with a substructure size of 16 matching atoms. Also, for the superimposition of the inhibitor with the reaction product IMP, there are discernible structure differ-

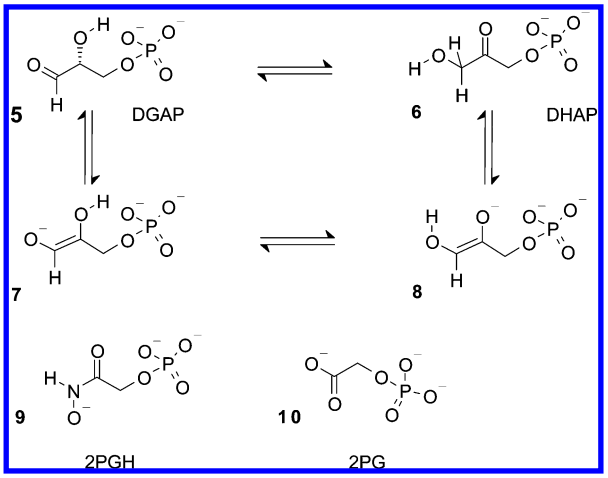


**Figure 5.** Isomerization of D-glyceraldehyde-3-phosphate (DGAP), **5**, into dihydroxyacetonephosphate (DHAP), **6**, catalyzed by triose phosphate isomerase (TIM). In this reaction, two intermediates, **7** and **8**, are formed, differing in the position of one proton. At the bottom, two inhibitors of TIM are shown: 2-phosphoglycolohydroxamate (2PGH), **9**, and 2-phosphoglycolate (2PG), **10**.

ences: the OH group of the inhibitor and the carbonyl group of IMP clearly point to different positions in space (Figure 4b). Here, an RMS value of 0.20 Å was obtained with a substructure size of 16 matching atoms. In contrast, the superimposition of the tetrahedral intermediate with the inhibitor shows a good match of both OH groups (Figure 4c). For this superimposition, an RMS value of 0.13 Å was reached with a substructure size of 16 atoms. This emphasizes the close geometric correspondence of the two structures and points out that the OH group—and its direction in space—apparently is very important for the binding of the reaction intermediate and inhibitor. Thus, in fact, the geometry of the intermediate of this reaction and that of the inhibitor are in close correspondence.

*Triose Phosphate Isomerase.* TIM (EC code 5.3.1.1) is an enzyme consisting of two identical subunits, each comprising about 250 amino acids which form the active enzyme.[21] TIM plays an important role in glycolysis, converting D-glyceraldehyde-3-phosphate (DGAP), **5**, into dihydroxyacetone phosphate (DHAP), **6** (sometimes also called glycerone phosphate), and vice versa (Figure 5). The conversion passes through two enediolate intermediates, **7** and **8**, which differ in the localization of one proton, taking into account that this is only one of several proposed reaction mechanisms as the reaction mechanism of TIM is not fully elucidated.[22]

The catalytic mechanism of the TIM reaction has been studied in detail.[23,24] TIM deficiency leads to a severe multisystemic disease with hemolytic anemia and neurological disorders.[25] Several inhibitors of this enzyme are known; two of them are 2-phosphoglycolohydroxamate (2PGH), **9**,[24] and 2-phosphoglycolate (2PG), **10**,[24] shown in Figure 2. Both have been proposed to be analogues of the enediolate intermediate state.[22,23]

First, the situation with the inhibitor 2PGH, **9**, will be investigated. For this analysis, three superimpositions with this inhibitor were performed: first, with the substrate DGAP, **5**; second with the product DHAP, **6**; and finally, with the reaction intermediate, **7**. For the superimposition, only one intermediate was modeled, as transition states **7** and **8** have the same 3D structure and differ only in the position of one proton. The first superimposition of the substrate DGAP with the inhibitor shows that both molecules are quite similar in shape, as they have the same number of non-hydrogen atoms (Figure 6a). The phosphate groups of both molecules match very well, but the other end of the molecules do not match exactly, as the configuration at the C-2 atom of DGAP and that of the C-1 atom of 2PGH are not the same: there is $sp^3$ hybridization in the case of DGAP and $sp^2$ hybridization in the case of 2PGH. This results in a different geometry for the two molecules. The RMS value for this superimposition is 0.79 Å at a substructure size of 10 atoms, which is the maximum substructure size that can be reached in this case as DGAP and 2PGH both have only 10 non-hydrogen atoms.

The superimposition of the product DHAP and the inhibitor 2PGH (Figure 6b), on the other hand, shows a quite close matching between both molecules, especially regarding the terminal O atoms. But, to obtain the close matching of the terminal atoms, the matching of the medial atoms is not perfect, which results in a RMS value of 1.64 at a substructure size of 10 atoms.

The superimposition of the inhibitor 2PGH with the intermediate, in contrast, shows a perfect match (Figure 6c) as the hydroxamic acid group is an excellent bioisostere of the enediolate transition state,[26] resulting from the partial double-bond character of the amide bond. For this superimposition, an RMS value of 0.04 Å for a substructure size of 10 atoms was obtained.

The superimposition of the second inhibitor, 2PG, **10**, with the starting material D-glyceraldehyde-3-phosphate, **5** (Figure 6d), clearly shows the structural differences between both molecules.

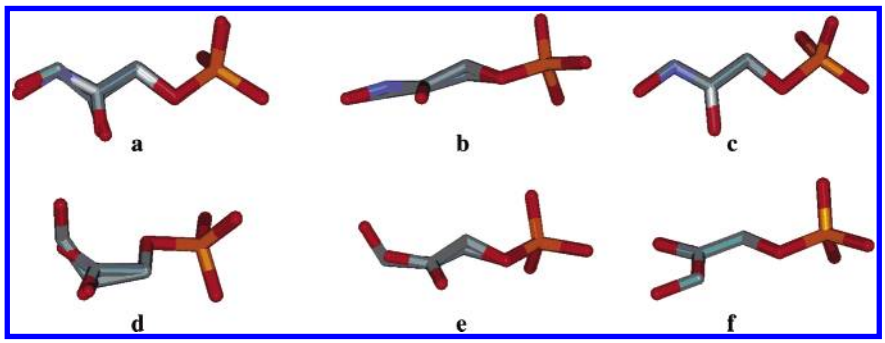


**Figure 6.** At the top, the superimposition of the inhibitor 2PGH, **9**, with the substrate DGAP, **5** (a), with the product DHAP, **6** (b), and with the reaction intermediate, **7** (c), of the triose phosphate isomerase catalyzed reaction is shown. At the bottom, the superimposition of the inhibitor 2PG, **10**, with the substrate DGAP, **5** (d), with the product DHAP, **6** (e), and with the reaction intermediate, **7** (f), of the triose phosphate isomerase catalyzed reaction is shown.

**Table 1.** RMS Values Obtained for the Superimpositions of Triose Phosphate Isomerase[a]

| | molecule species from the reaction | | |
|---|---|---|---|
| inhibitor | DGAP | DHAP | intermediate |
| 2PGH | 0.79 | 1.64 | 0.04 |
| 2PG | 0.76 | 1.55 | 0.03 |

[a] All values are given in Å. The substructure size for 2PGH is 10 atoms, and for 2PG, it is nine atoms.

The first, and a quite obvious, difference is in the number of atoms between DGAP, **5**, having 10 non-hydrogen atoms, and the inhibitor 2PG, **10**, having nine such atoms. This does not allow a full superimposition of both molecules. A further difference lies in the hybridization state of carbon atom C-2 of D-glyceraldehyde-3-phosphate and that of carbon atom C-1 of 2PG: while the C-2 atom of D-glyceraldehyde-3-phosphate is tetrahedral, the matching carboxygroup of the 2PG is planar; thus, both cannot be completely matched. The RMS value for this superimposition is 0.76 Å at a substructure size of nine atoms.

The superimposition of the reaction product, DHAP, **6**, with the inhibitor 2PG, **10** (Figure 6e), shows the same difference in molecular size as before (10 vs 9 atoms), but in this case, the configurations at the C atoms are the same in both molecules, and thus, the OH group of 2PG matches the OH groups of DHAP quite well. Additionally, as in the superimposition of DHAP with the first inhibitor 2PGH, the matching of the medial atoms is not perfect. The RMS value for this superimposition is 1.55 Å for a substructure size of nine atoms.

The superimposition of the intermediate with the inhibitor 2PG (Figure 6f) shows a perfect match. The carboxylate function in 2PG mimics the planar enediolate in the intermediate state.[27] The RMS value for this superimposition is 0.03 Å for a substructure size of nine atoms.

A summary of the RMS values obtained for the superimpositions of both inhibitors is given in Table 1.

The study of the superimpositions of both inhibitors of TIM, 2PGH and 2PG, leads to the following conclusions: (1) The reaction involves two intermediate states that differ in the arrangement of a proton, and accordingly, the reaction progresses through three transition states. (2) The geometry at the C-1 atom of dihydroxyacetone phosphate is not important for the transition state of this reaction. (3) The C-1 atom can even be deleted and still an inhibitor can be obtained. Thus, it seems that the geometry on carbon atom C-1 of DGAP, which is no longer present in 2PG, is not that important for the binding.

*Arginase II.* The previously presented studies work with the assumption that no 3D information is available from experimental sources, as this is the case when searching for new inhibitors in a database. Therefore, CORINA-generated models have been used for the intermediate as well as for the inhibitor compounds. In contrast, in the following example, we studied the situation where experimentally derived 3D structural information is available and brought into the superimposition process. This may give us an indication of whether we can indeed work only with computed 3D information for the alignments.

Additionally, we were interested, in this experiment, whether the quality of the alignment can noticeably be increased if we incorporate known information into the superimposition.

For the first case, atoms that are known to interact with the binding pocket of the enzyme through hydrogen bonds were forced to match together. In the second case, physicochemical properties, calculated by PETRA (see the Materials and Methods section), which were assumed to be relevant for the receptor−ligand interaction, were introduced. In the last and simplest case, only 3D structural information of the inhibitor was used.

Arginase is a binuclear manganese metalloenzyme that catalyzes the hydrolytic cleavage of L-arginine, **11**, into L-ornithine, **12**, and urea, **13**, through a metal-activated hydroxide mechanism.[28] In mammals, two isoenzymes are identified: arginase I is found predominantly in hepatocytes, and arginase II occurs extrahepatic. The arginase isoenzymes differ from each other in terms of their catalytic, molecular, and immunological properties. Human penile arginase is a potential target for the treatment of sexual dysfunction in males.[29] The reaction and the invoked intermediate, **14**, are given in Figure 7.

In this experiment, we have used the 3D structures of three inhibitors—(*S*)-2-amino-6-boronohexanoic acid (ABH), **15** (PDB: 1D3V);[30,31] *S*-(2-boronoethyl)-L-cysteine (BEC, sometimes also S2C), **17** (PDB: 1HQ5);[30,31] and *S*-(2-sulfonamidoethyl)-L-cysteine (SDC), **19** (PDB: 1R1O)[28]—from *Rattus norvegicus*, all shown in Figure 8.

ABH and BEC are slow-binding competitive inhibitors belonging to the class of boronic acid inhibitors, while SDC contains a sulfonamide group. Bound into the active site of the enzyme, ABH and BEC form tetrahedral boronate anions, **16** and **18**, respectively. These mimic the tetrahedral

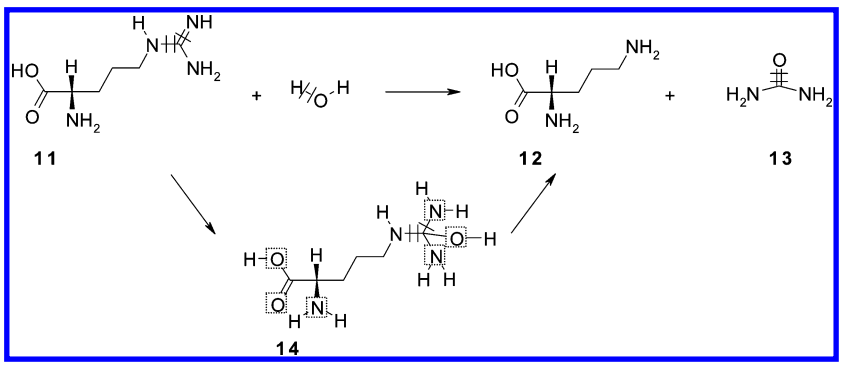


**Figure 7.** Hydrolysis of L-arginine, **11**, to L-ornithine, **12**, and urea, **13**, catalyzed by arginase II with the corresponding reaction intermediate, **14**, shown.

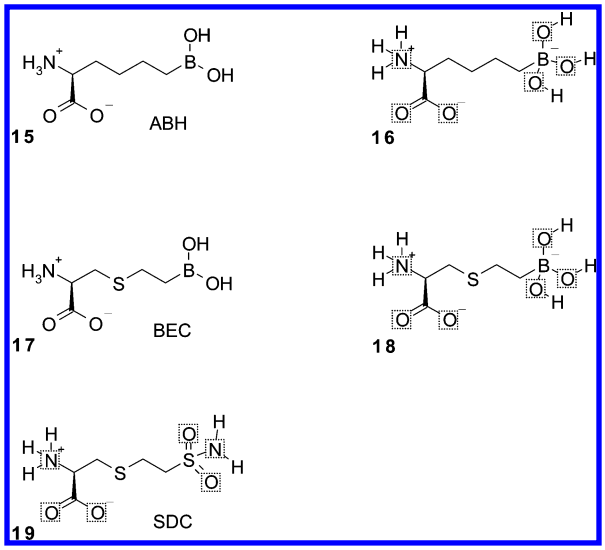


**Figure 8.** Inhibitors of arginase II: ABH, **15**, and in its active form as hydrated ABH, **16**; BEC, **17**, and in its active form as hydrated BEC, **18**; SDC, **19**.

**Table 2.** Ranges, $\Delta p$, of Physicochemical Properties Assigned to the Superimposition Process

| | inhibitor | | |
|---|---|---|---|
| physicochemical property | ABH | BEC | SDC |
| lone pair electronegativity (eV) | 2.10 | 2.10 | not used |
| $\sigma$-electronegativity (eV) | 2.10 | 2.10 | not used |
| effective atom polarizability (Å$^3$) | 0.60 | 1.00 | not used |
| total charge (eu) | 0.25 | 0.25 | 0.35 |
| octanol/water partition coefficient | 0.50 | 0.50 | 0.60 |

intermediate of the arginase hydrolysis reaction. The same function is fulfilled by the sulfonamide group of SDC. For all three inhibitors, the experimentally derived 3D structure as bound into arginase is available from the PDB.[18]

After extraction of the 3D structures of the inhibitors from the PDB, the data were processed for the superimposition conserving the original 3D structural data. Then, five physicochemical properties used for the superimposition were calculated by PETRA: These properties comprise lone-pair electronegativity, $\sigma$-electronegativity, effective atom polarizability, total charge, and the octanol/water partition coefficient. The same steps were performed for the intermediate of the reaction which was generated from the BioPath database. For each inhibitor, three superimpositions were performed.

In the first superimposition, the atoms of the inhibitor and of the intermediate that should match were assigned as constraints for the superimposition process. This information was derived from refs 28, 29, and 31. The atoms assigned to match between the intermediate and each inhibitor are indicated in structures **14**, **16**, **18**, and **19** by dashed boxes.

For the second kind of superimposition, similarity ranges regarding physicochemical properties which describe the electronic effects for the binding into the binding pocket of the enzyme were taken as matching criteria. This allows only those atoms to match which are similar in regard to these properties and should, therefore, bind into the same region of the binding pocket. The physicochemical values used in the superimposition and the defined ranges are given in Table 2. The ranges, $\Delta p$, in the physicochemical properties were used such that only those atoms were allowed to be superimposed whose properties, $p$, had values that deviated by less than $\Delta p$.

The ranges were set by initial inspection of the properties of the atoms given as match tuples in the first superimposition experiment.

For the third superimposition, no constraints were specified for the superimposition process, providing a match totally adjusted to the geometry of the molecules. This is the case which is typically true for scanning a database of compounds for new potential inhibitors.

First, the superimposition with the inhibitor ABH, **16**, was analyzed. All three experiments showed a good overlap between the inhibitor and the reaction intermediate. A look at the RMS values shows how close the superimpositions lie together: With given match tuples, the RMS value is 0.30 Å (Figure 9a); with given constraints on physicochemical properties, it is 0.24 Å (Figure 9b), and without any constraints on the superimposition, it is 0.29 Å (Figure 9c). As can be seen, the superimposition without constraints

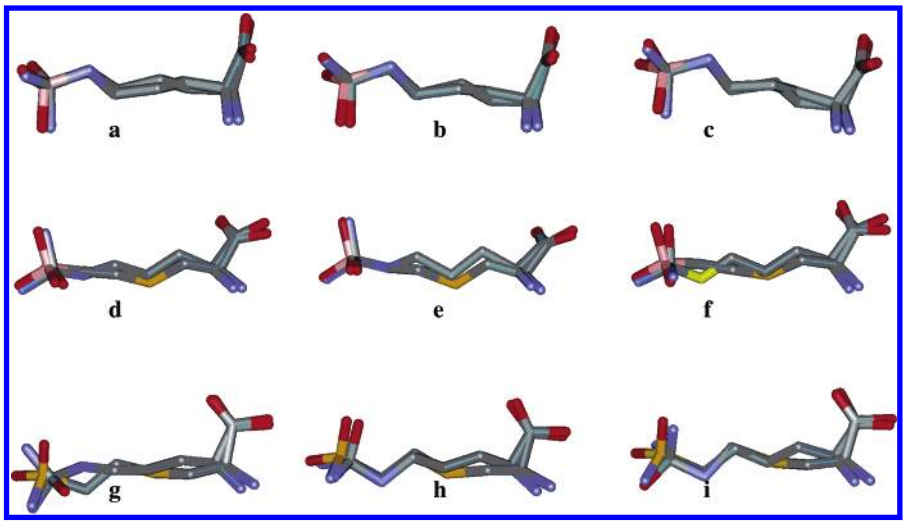


**Figure 9.** Superimposition of the arginase reaction intermediate, **14**, with ABH, **16**, with given match tuples (a), based on physicochemical properties (b), and without any constraints (c). Superimposition of the arginase reaction intermediate, **14**, with BEC, **18**, with given match tuples (d), based on physicochemical properties (e), and without any constraints (f). Superimposition of the arginase reaction intermediate, **14**, with SDC, **19**, with given match tuples (g), based on physicochemical properties (h), and without any constraints (i).

**Table 3.** RMS Values Obtained in the Superimposition Experiments with Arginase II[a]

| | information given into the superimposition process | | |
|---|---|---|---|
| inhibitor | match tuples given | ranges of physicochemical properties given | no constraints given |
| ABH | 0.30 (13) | 0.24 (13) | 0.29 (13) |
| BEC | 0.31 (13) | 0.23 (10) | 0.43 (13) |
| SDC | 1.00 (13) | 0.41 (13) | 0.50 (13) |

[a] The RMS values are given in Å. The substructure size for all superimpositions is given in parentheses.

performs as well as with given match tuples. For all three superimpositions, the maximum substructure size is 13 atoms.

For the inhibitor BEC, **18**, the RMS value of the superimposition onto the intermediate is 0.31 Å when matching tuples are given (Figure 9d), 0.23 Å when physicochemical properties are given (Figure 9e), and 0.43 Å without any given constraints (Figure 9f). For all superimpositions, a substructure size of 13 atoms was obtained except in the case with superimposition based on physicochemical properties: here, a maximum substructure size of 10 atoms was obtained. In this case, the sulfur atom and both flanking C atoms of BEC were not recognized as match partners to the corresponding atoms of the intermediate as they exceeded the given property ranges. Here, the purely geometric superimposition performs slightly poorer than the others.

For the last inhibitor, SDC, **19**, in the superimposition with matching atoms given, the RMS value is 1.00 Å (Figure 9g); with given physicochemical properties, the RMS is 0.41 Å (Figure 9h), and without any constraints, the RMS is 0.50 Å (Figure 9i). Here, again, the geometric superimposition is better than that with given match tuples but also slightly poorer than that with physicochemical properties.

An overview of the RMS values for all superimpositions is given in Table 3. For all three inhibitors, the differences between the three methods can hardly be recognized by visual inspection.

## CONCLUSIONS

In this paper, it was shown that 3D molecular models of intermediates of enzyme-catalyzed reactions can automatically be generated from a database of biochemical reactions and can serve as templates for matching inhibitors of the enzymes that catalyze the corresponding reaction. It was shown by superimposing these generated intermediates onto known transition-state analogue inhibitors that the similarity between both is sufficient to use the intermediate as a template to search for new transition-state analogue inhibitors. This can be performed by a superimposition method which uses a genetic algorithm enriched with a numerical optimization method. If there is no experimental 3D information on the inhibitors available, it is also possible to use computed 3D molecular information, which still delivers good results. As the superimposition process also allows conformational changes, detailed information on the steric requirements of enzyme-catalyzed reactions can be gained. The consideration of physicochemical effects in the superimpositions allows one to draw conclusions on the electronic effects operating in the enzyme pocket. The comparison of several inhibitors of specific enzymes also allows for accumulating knowledge on the crucial features of an inhibitor for the enzyme. This approach provides a three-dimensional structure query that can be used for searching in databases of chemical structures for new potential enzyme inhibitors without using elaborate and time-consuming ab initio methods. This opens the prospects for finding new drugs and agrochemicals.

## ACKNOWLEDGMENT

We gratefully acknowledge funding for this project by the "Bundesministerium fuer Bildung und Forschung" (BMBF), projects 031U112D and 031U212D, and "Bioinformatics for the Functional Analysis of Mammalian genomes" (BFAM), which is part of the "Nationales Genomforschungsnetz Deutschland" (NGFN). The work on this project was also supported by KONWIHR (ENZYMECH), funded by the state of Bavaria through the "High-Tech-Offensive Bayern".

## REFERENCES AND NOTES

(1) Borman, S. Much Ado About Enzyme Mechanisms. *Chem. Eng. News* **2004**, *82*, 35−39.
(2) Zhang, X.; Houk, K. N. Why Enzymes Are Proficient Catalysts: Beyond the Pauling Paradigm. *Acc. Chem. Res.* **2005**, *38*, 379−385.
(3) Pauling, L. The Nature of Forces between Large Molecules of Biological Interest. *Nature* **1948**, *161*, 707−709.
(4) Pauling, L. Molecular Architecture and Biological Reactions. *Chem. Eng. News* **1946**, *24*, 1375−1377.
(5) Robertson, J. G. Mechanistic Basis of Enzyme-Targeted Drugs. *Biochemistry* **2005**, *44*, 5561−5571.
(6) Reitz, M.; Sacher, O.; Tarkhov, A.; Trümbach, D.; Gasteiger, J. Enabling the Exploration of Biochemical Pathways. *Org. Biomol. Chem.* **2004**, *2*, 3226−3237.
(7) Hammond, G. S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.
(8) Michal, G. *Biochemical Pathways Wall Chart*; Boehringer Mannheim (now Roche): Mannheim, Germany, 1993. http://www.expasy.org/tools/pathways (accessed Apr 2006).
(9) Michal, G. *Biochemical Pathways Biochemistry Atlas*; Spektrum Akademischer Verlag: Heidelberg, Germany, 1999.
(10) Goto, S.; Okuno, Y.; Hattori, M.; Nishioka, T.; Kanehisa, H. LIGAND: Database of Chemical Compounds and Reactions in Biological Pathways. *Nucleic Acids Res.* **2002**, *30*, 402−404.
(11) BioCyc Database Collection. http://www.biocyc.org (accessed Apr 2006).
(12) *C@ROL*. Molecular Networks GmbH: Erlangen, Germany. info@mol-net.com, http://www.mol-net.com (accessed Apr 2006).
(13) Ihlenfeldt, W. D.; Takahashi, Y.; Abe, H.; Sasaki, S. Computation and Management of Chemical Properties in CACTVS: An Extensible Networked Approach toward Modularity and Compatibility. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 109−116.
(14) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(15) *CORINA*; Molecular Networks GmbH: Erlangen, Germany. info@mol-net.com, http://www.mol-net.com. CORINA can be tested on the Internet at http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html (accessed Apr 2006).
(16) Gasteiger, J. In *Physical Property Prediction in Organic Chemistry*; Jochum, C., Hicks, M. G., Sunkel, J., Eds.; Springer-Verlag: Heidelberg, Germany, 1988; pp 119−138.
(17) Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220−232.
(18) Bernstein, F. C.; Koetzle, T. F.; Williams, G. F.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(19) Bojack, G.; Earnshaw, C. G.; Klein, R.; Lindell, S. D.; Lowinski, C.; Preuss, R. Design and Synthesis of Inhibitors of Adenosine and AMP Deaminase. *Org. Lett.* **2001**, *3*, 839−842.

(20) Dancer, J. E.; Hughes, R. G.; Lindell, S. D. Adenosine-5′-Phosphate Deaminase − A Novel Herbicide Target. *Plant Physiol.* **1997**, *114*, 119−129.
(21) Witmans, C. J. An Approach to the Rational Design of New Inhibitors for *Trypanosoma brucei* Triosephosphate Isomerase. PhD dissertation, University Groningen, Netherlands, 1995. http://irs.ub.rug.nl/ppn/139946616. (accessed Apr 2006).
(22) Cui, Q.; Karplus, M. Catalysis and Specifity in Enzymes: A Study of Triosephosphate Isomerase and Comparison with Methyl Glyoxal Synthase. In *Advances in Protein Chemistry*; Daggett, V., Ed.; Elsevier Academic Press: San Diego, CA, 2003; pp 315−320.
(23) Aqvist, J.; Fothergill, M. Computer Simulation of the Triosephosphate Isomerase Catalyzed Reaction. *J. Biol. Chem.* **1996**, *271*, 10010−10016.
(24) Collins, K. D. An Activated Intermediate Analogue. The Use of Phosphoglycolohydroxamate as a Stable Analogue of a Transiently Occurring Dihydroxyacetone Phosphate-Derived Enolate in Enzymatic Catalysis. *J. Biol. Chem.* **1974**, *249*, 136−142.
(25) Livet, M. O. Triose-Phosphate Isomerase Deficiency. *Orphanet Encyclopedia* 2003. http://www.orpha.net/data/patho/GB/uk-TPI.pdf (accessed Apr 2006).
(26) Lewis, D. J.; Lowe, G. Phosphoglycollohydroxamic Acid: An Inhibitor of Class I and II Aldolases and Triosephosphate Isomerase. A Potential Antibacterial and Antifungal Agent. *J. Chem. Soc., Chem. Commun.* **1973**, *19*, 713−715.
(27) Wolfenden, R. Transition State Analogues for Enzyme Catalysis. *Nature* **1969**, *233*, 704−705.
(28) Cama, E.; Hyunshun, S.; Christianson, D. W. Design of Amino Acid Sulfonamides as Transition-State Analogue Inhibitors of Arginase. *J. Am. Chem. Soc.* **2003**, *125*, 13052−13057.
(29) Kim, N. N.; Cox, J. D.; Baggio, R. F.; Emig, F. A.; Mistry, S. K.; Harper, S. L.; Speicher, D. W.; Morris, S. M., Jr.; Ash, D. E.; Traish, A.; Christianson, D. W. Probing Erectile Function: *S*-(2-Boronoethyl)-L-cysteine Binds to Arginase as a Transition State Analogue and Enhances a Smooth Muscle Relaxation in Human Penile Corpus Cavernosum. *Biochemistry* **2001**, *40*, 2678−2688.
(30) Cox, J. D.; Kim, N. N.; Traish, A. M.; Christianson, D. W. Arginase−Boronic Acid Complex Highlights a Physiological Role in Erectile Function. *Nat. Struct. Biol.* **1999**, *6*, 1043−1047.
(31) Cama, E.; Colleluori, D. M.; Emig, F. A.; Hyunshun, S.; Kim, S. W.; Kim, N. N.; Traish, A. M.; Ash, D. E.; Christianson, D. W. Human Arginase II: Crystal Structure and Physiological Role in Male and Female Sexual Arousal. *Biochemistry* **2003**, *42*, 8445−8451.

CI050503U