Journal of Computer-Aided Molecular Design (2005) 19: 69–82
DOI 10.1007/s10822-005-1485-7

A computational tool to optimize ligand selectivity between two similar biomacromolecular targets

Deliang L. Chen & Glen E. Kellogg*
Department of Medicinal Chemistry & Institute for Structural Biology and Drug Discovery, School of Pharmacy, Virginia Commonwealth University, Richmond, Virginia, 23298-0540, USA

Received 14 September 2004; accepted in revised form 27 January 2005

Key words: binding specificity, hydropathic analysis, selectivity

Summary

Algorithms for a new computer program designed to increase ligand–receptor selectivity between two proteins are described. In this program ligand–receptor selectivity is increased by functional modifications to the ligand so as to increase the calculated binding affinity of it to one protein and/or decrease the calculated binding affinity of it to the other protein. The structure of the ligand is modified by selective replacement of atoms and/or functional groups *in silico* based on a specific set of steric and/or hydropathic complementarity rules involving atoms and functional groups. Relative binding scores are calculated with simple grid-based steric penalty, hydrogen bond complementarity, and with the HINT score model. Two examples are shown. First, modifying the structure of the ligand CB3717 is illustrated in a number of ways such that the binding selectivity to wild type *L. casei* thymidylate synthase or its E60Q mutant may be improved. Second, starting with a non-selective lead compound that had been co-crystallized with both plant and mammalian 4-hydroxyphenylpyruvate dioxygenases, new compounds (similar to selective ligands discovered by screening) to improve the selectivity of (herbicidal) inhibitors for the plant enzyme were designed by the program.

Introduction

The design of new potent and *selective* ligands, drugs and enzyme inhibitors is one of the most important applications in medicinal chemistry research. In the course of drug discovery much attention is paid currently to optimizing potency or binding efficacy, while much less effort is expended (at least in the early stages) on understanding and optimizing binding selectivity. Ideally a ligand acts through binding at a single site of a single protein. However, in reality a ligand may bind at different sites of the same biomacromolecule, bind with different biomacromolecules, or even with biomacromolecules from different species if the target is a bacterial or viral infection, or even a herbicidal target. Any of these cases may cause undesirable side effects, some of which may be too significant for the ligand to be useful or safe as a drug or other product. Side effects can be reduced by improving the selectivity of a ligand for one site over another, i.e., by increasing the binding affinity of a ligand to one (desirable) site/target and/or decreasing the binding affinity to another (undesirable) site/target.

One immediate and obvious application for understanding drug selectivity is as part of investigations of ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity). ADMET has historically not been considered as part of the drug design process, but instead during later stage

*To whom correspondence should be addressed. Tel.: +1-804-828-6452; Fax: +1-804-827-3664; E-mail: glen.kellogg@vcu.edu

pre-clinical and clinical investigations, where the costs of failure are very high [1]. Clearly, understanding the distribution, with respect to the energetics of binding, of a chemical entity within a human has significant implications. Recently, however, early stage ADMET investigations have come in vogue with both new biochemical screening assays [2] and computational research technologies being reported [3]. On a larger scale, understanding the distribution of a chemical entity within an ecosystem, e.g., an agrochemical or chemical manufacturing byproduct, is also an issue where understanding and exploiting 'ligand' selectivity can be significant.

We became interested in ligand selectivity as two recent projects unfolded in our laboratories. First, we have been investigating the sequence specificity of anthracycline antibiotic intercalators of double helix DNA [4]. Second, we have an ongoing project to target β-ketoacyl synthase III (FabH) from a variety of bacterial sources, e.g., *E. Coli*, *M. tuberculosis*, etc., as a new mode for antibiotic development [5]. In both of these cases we felt the need for a computational tool that would, in effect, integrate structural information from two or more competing biomacromolecular sites with structure-activity information of known ligands or inhibitors in order to *de novo* improve binding selectivity.

This report describes the algorithms and technology of a new computer program for enhancing binding selectivity. It is based, in part, on the HINT model for biological interactions [6] and was written using the HINT toolkit [7]. By way of example we have examined two cases. First, the selectivity of compounds related to CB3717 bound to *L. casei* thymidylate synthase and its E60Q mutant is described. Second, we have examined the herbicidal inhibitor selectivity for compounds bound to plant and mammalian 4-hydroxyphenylpyruvate dioxgenases. Further communications will illustrate additional applications of this technology as development of the program continues.

Algorithms

In this paper, we assume that a ligand **L** can bind to two proteins ($\mathbf{P}_1$ and $\mathbf{P}_2$) with initially similar binding affinities. We wish to increase the selectivity of **L** for $\mathbf{P}_1$ over $\mathbf{P}_2$ and in order to do that we may modify **L** to create a new ligand **L′** such that **L′** has a higher binding affinity to $\mathbf{P}_1$ than **L** and/or has a lower binding affinity to $\mathbf{P}_2$ than **L**. Three methods of scoring are used to simulate binding affinity/selectivity in our program: a simple steric interaction score, and two HINT-based functional group interaction scores. These three approaches are presented as separate algorithms in this paper.

The first, steric, algorithm simply represents that if the target biomacromolecule does not have enough space to hold the new ligand, then there would be unfavorable steric interactions between the new ligand and the protein and the binding affinity would be greatly decreased. The second and third algorithms, based on the HINT functional group interaction score algorithm [6, 7], characterize the binding affinity/selectivity in terms of polar (favorable and unfavorable) and hydrophobic matches and mismatches as derived from solvent partitioning data. The second algorithm emphasizes favorable polar–polar interactions while the third algorithm emphasizes favorable hydrophobic–hydrophobic interactions. Consideration of the energy for removing the solvent from ligand and protein, especially for potent polar–polar interactions such as between a carboxyl group and an ammonium group and forming particularly strong hydrogen bonds, is a major emphasis of the second algorithm. The HINT binding interaction score has been shown in a number of recent publications [6b, c] to correlate well with the free energy of binding.

Algorithm 1: modify ligand based on steric complementarity

To modify the ligand based on steric complementarity, the algorithm identifies which atom or group should be replaced and then identifies which functional group from an available pool should replace it.

First, structural data of the two competing proteins are read from PDB files. At least one of the structure data sets should include coordinate data for the initial ligand. In some rare cases both data sets will include the same bound ligand, otherwise the ligand will have to manually or semi-manually docked in the second protein. From the ligand data we identify all atoms or groups that can be replaced by other groups and that will not cause significant

change in the ligand framework and electronic structure. In our program, we assume that all C–H, O–H, N–H, C–(OH), C–(NH_2), and C–(CH_3) bonds are 'breakable' and that the corresponding 'leaving' atoms or groups (i.e., H, H, H, OH, NH_2, CH_3) can be replaced by other groups. However, because not all of these replacements will increase the selectivity, the program next identifies the steric selectivity of the attachment bonds for these leaving atoms or groups. We define the steric selectivity of a bond here to be the available free space *difference* around the bond in the two proteins. If there is large free space available around the bond in one protein while there is small free space in the other protein, the steric selectivity of the bond would then be high.

To estimate the steric selectivity of each bond and to compare the selectivity of different bonds, origin-centered grid boxes where the bond of interest is aligned with the z-axis, are used (see Figure 1a). Three grid boxes are created, all with spacing of 0.3 Å. Each box extends from −8 to +8 Å in the x-, y-directions and from −2 to +16 Å in the z-direction. The first box is for protein one, the second box is for protein two and the third box is used to compare box one and box two. Grid points within the van der Waals radius of protein atoms are referred to as occupied grid points (OGP) (see Figure 2a, b). Other points, not within the van der Waals radius of any atom, are referred to as unoccupied grid points (UGP). If there are more UGPs for either protein, then there is higher available free space around the bond. The difference between the UGPs in box 1 and box 2 determines the steric selectivity of the bond. The following items are calculated to estimate the steric selectivity:

$$\text{free space ratio } r_f^i = (\text{Total number UGP})/(\text{Total number OGP}) \text{ for protein } i; \quad (1)$$

$$\text{steric selectivity for protein 1: } S_1 = r_f^1/r_f^2; \quad (2)$$

$$\text{steric selectivity for protein 2: } S_2 = r_f^2/r_f^1; \quad (3)$$

In our program, if $S_1 > 2$, it may be possible to modify the bond to increase selectivity for protein 1. If $S_2 > 2$, it may be possible to modify the bond to increase selectivity for protein 2. If neither S_1

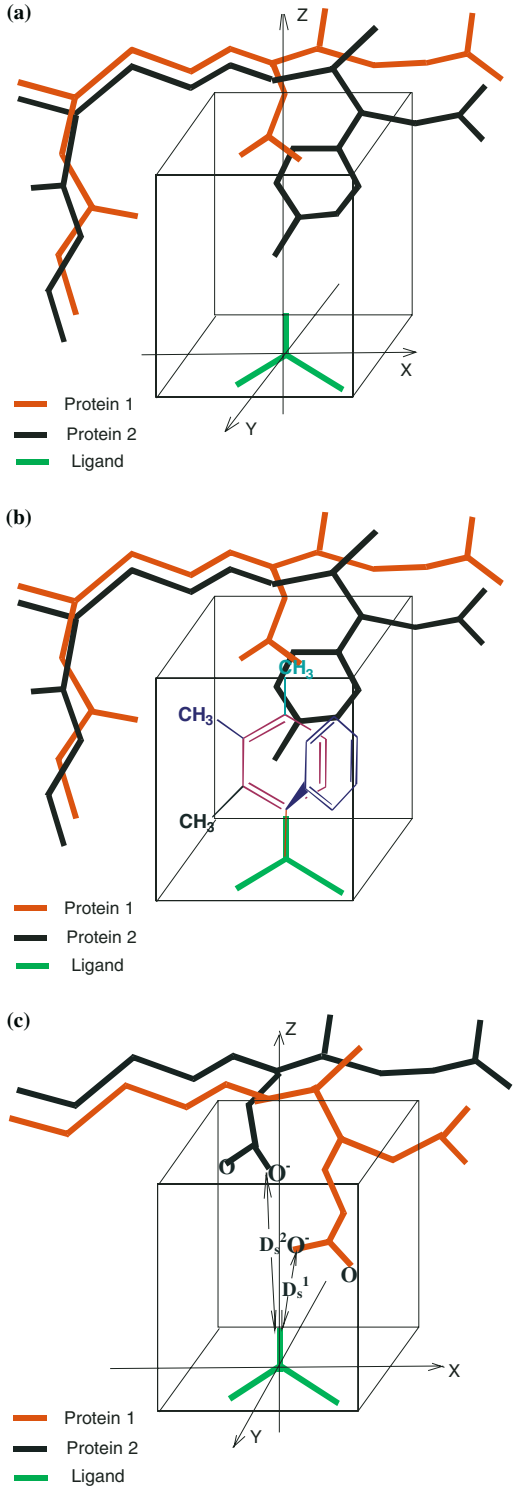

Figure 1. (a) The grid box is created with z-axis in the direction of the 'breakable' bond. (b) Four ways to connect toluene to the 'breakable' bond of the ligand; after connection the toluene is placed in the grid box and connected as described in the text. (c) Grid box for algorithm 2 and selective distances (D_s).

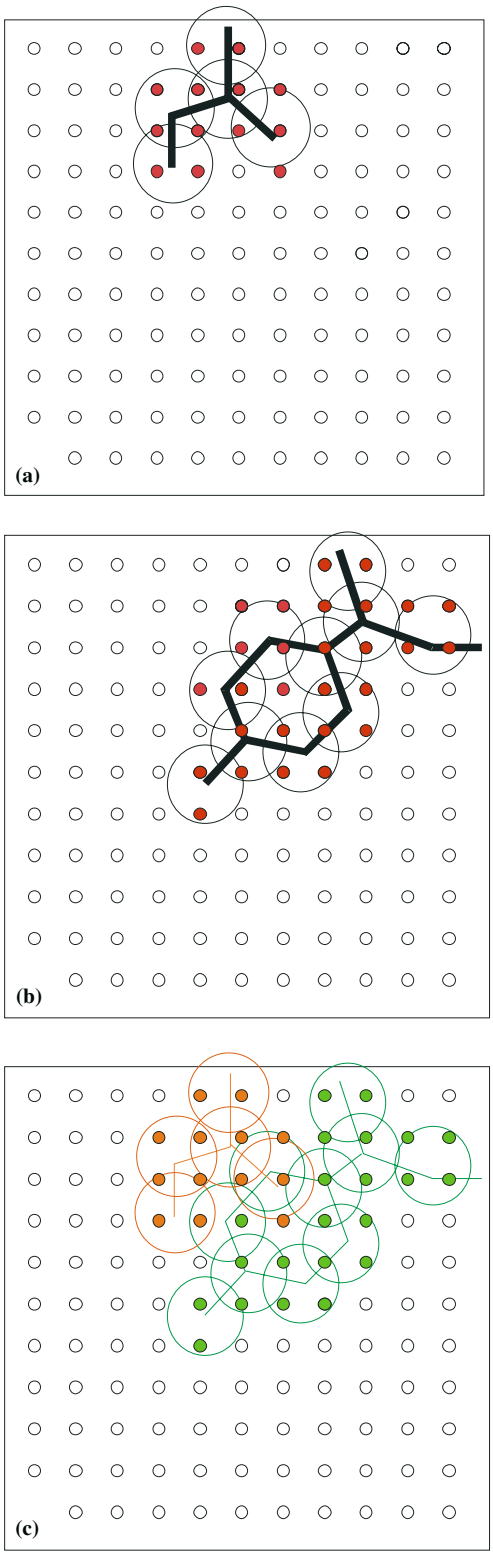

Figure 2. Illustration of the grid properties OGP, UGP, PSP, USP and NSP when the goal is to increase ligand binding at protein 1 and/or decrease ligand binding at protein 2. (a) Part of grid box 1; red grid points are OGP; white grid points are UGP. (b) Part of grid box 2; red grid points are OGP; white grid points are UGP. (c) Part of grid box 3; white grid points are USP, green grid points are PSP, orange grid points are NSP.

nor S_2 is greater than 2, no modifications are made to the bond. After the selectivity of each bond is estimated with a value for S_1 or S_2, all S_1 and S_2 are sorted in descending order. Bonds are modified according to this order: the bond with highest S_1 is assumed to be the best one for modification to increase the selectivity of a ligand for protein 1 over protein 2.

The most important function of grid box 3 is to identify the grid points at which the atoms of the ligand have steric interaction with one protein and do not have steric interaction with the other protein. Here we illustrate the algorithm with a bond whose $S_1 > 2$. Because $S_1 > 2$ this bond may be modified to increase the selectivity of a ligand for protein 1 over protein 2. The grid point property of a point in grid box 3 can be expressed as one of three types: potentially selective grid point (PSP), unselective grid point (USP), or negative selective grid point (NSP). Figure 2 illustrates these grid properties for a case where the goal is to increase binding at protein 1 relative to protein 2. If the property of a point in grid box 1 (Figure 2a) is UGP and the property of the identical point in grid box 2 (Figure 2b) is OGP, then the property of the identical point in grid box 3 (Figure 2c) is PSP. If an atom of the ligand is at or near a grid point of PSP, the atom may have steric interaction with protein 2 and not have a steric interaction with protein 1. In this case we may be able to increase the ligand's selectivity for protein 1 over protein 2. If the property of a point in grid box 1 (Figure 2a) is UGP and the property of the identical point in grid box 2 (Figure 2b) is also UGP, then the property of the identical point in grid box 3 (Figure 2c) is USP. If an atom of the ligand is at or near a grid point of USP, that atom does not increase the ligand's selectivity for protein 1 over protein 2. If the property of a point in grid box 1 is OGP, regardless of the property of the identical point in grid box 2, the property of that

point in grid box 3 is NSP. If an atom of the ligand is at a grid point of NSP, it will decrease the binding affinity of the ligand to protein 1. For a new group, its atoms can be at grid points of PSP and USP, but cannot be at grid points of NSP. Furthermore, only the atoms at the grid points of PSP can increase the selectivity. The more atoms at PSP, the larger the selectivity may be potentially increased.

However, because the bond attaching the new added group is rotatable and the initial orientation of that group may be arbitrary, we must perform a systematic search through the angles of the rotatable bond. For generality of the algorithm, only added groups that are sterically allowed for all rotations about the attachment point are allowed. Grid points that meet these criteria are designated real selective grid points (RSP). We find RSP by rotating protein 2 around the z-axis with increments of $360°/n$ (e.g., for $n = 12$ the rotation is 30°). For each rotation increment whether a grid point is PSP or not is recorded. After a complete rotation, we have n sets of data. If a grid point is PSP in every case, that grid point is RSP; otherwise the grid point is USP. Atoms of the ligand at the positions of more interior RSPs should have stronger steric interaction with protein 2 and such ligands should improve the selectivity.

The new added groups are small molecules that are read from a database. Some examples of small molecules in the database are shown in Figure 3. For the largest increase in computational selectivity for protein 1, small molecules that do not have steric interaction with protein 1 but have strong steric interactions with protein 2 are desirable. The algorithm identifies the most appropriate small molecule to use as a new added group and how to attach that small molecule to the ligand. This is illustrated for the small molecule toluene (8) in

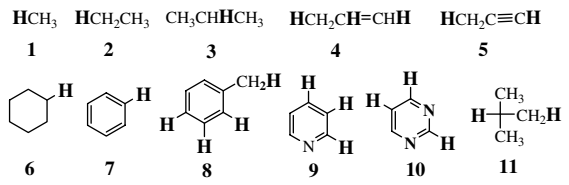

Figure 3. Some examples of small molecules used as new added groups for algorithm 1. The bold hydrogens indicate atoms at which the small molecule can attach to the ligand.

Figure 1b where there are four possible ways to connect it to the ligand. Bonds that connect the highlighted hydrogens (Figure 3) can connect the selected 'breakable' bond of the ligand. After a bond of the small molecule is overlapped with the selected 'breakable' bond of the ligand, the small molecule is placed in the grid box 3. For the group to increase binding selectivity of the new ligand none of its atoms can be at grid points of NSP and at least one atom is at a grid point of RSP. If the group can increase selectivity, the strength of the steric interaction between the new added group and protein 2 is estimated based on how many atoms are at points of RSP and their locations. Comparison of these results for different groups reveals which group should be added for the greatest increase in selectivity. Finally, a new molecule is created by adding the group to the ligand, removing the atoms to be substituted and then forming new bond between the original ligand and added group.

Algorithm 2: modify ligand based on polar–polar functional group interactions

In this algorithm, the new added group is designed to have strong favorable polar–polar functional group interactions with residues of protein 1 and to have unfavorable functional group interactions with residues of protein 2. The algorithm first estimates the selectivity of each 'breakable' bond and then adds a suitable functional group to the bonds with highest selectivity.

Similar to algorithm one, the program first reads the structural data of the two proteins and available ligand(s) from PDB files and identifies the 'leaving' atoms or groups and 'breakable' bonds. Which 'leaving' atoms or groups should be replaced depends on selectivity of the polar–polar functional group interaction of the 'breakable' bonds, which in turn depend on the differences of the hydropathic properties of the protein atoms around the bonds. The selectivity of a 'breakable' bond is estimated based on a ratio of the polar–polar functional group interactions between the new added group and residues of the two 'competing' proteins. An estimation of the strength of a favorable functional group interaction is based on the distance between the closest active atom and the 'breakable' bond. We refer to the atoms of the protein that can form especially strong

polar–polar functional group interactions as 'active' atoms (*vide infra*). To calculate the selectivity of each bond and to compare the selectivity of different bonds, grid boxes are used (See Figure 1c). Two grid boxes are created, all with spacing of 0.3 Å. As before, each box extends from −8 to 8 Å in the x, y-directions and from −2 to 16 Å in the z-direction around the origin-centered bond. The first box is for protein 1; the second box is for protein 2.

For an atom of the protein to form a strong hydrogen bond or ion–ion interaction with an atom of the ligand, the atom must be solvent accessible using the definition of Richards [8]. The solvent accessible atoms that can form especially strong polar–polar functional group interactions with the ligand are called 'active' atoms in this algorithm. These atoms include the H^+ atom in the Arg and Lys residues and the O^- in Glu and Asp residues. The strength of hydrogen bonds is very dependent on the distance between the two hydrogen bond forming atoms. In some cases, there is more than one active atom in the grid box, and the shortest distance between an active atom and the 'leaving' atom is referred to as the selective distance (D_s) (see Figure 1c). If there are no active atoms in the box, D_s is assumed to be a high value, i.e., 20.

The D_s for protein 1 is expressed as D_s^1 and for protein 2 as D_s^2. For the added group of the ligand to form a strong functional group interaction with one protein and not to form a functional group interaction with the other protein, the difference of the selective distances for the two proteins must be greater than a certain value that we are assuming in this work to be 3 Å. If $D_s^2 - D_s^1 > 3$ Å, it should be possible to add a functional group to this 'breakable' bond to increase the selectivity of the ligand for protein 1 over protein 2. If the difference of the selective distances is too small, the added functional group may form strong interaction with both proteins and the selectivity would likely not be increased. We use the following equation to estimate the selectivity of a 'breakable' bond:

$$S_1 = (D_s^2 - D_s^1)/D_s^1 \tag{4}$$

S_1 is the selectivity for protein 1 over protein 2. As S_1 becomes larger, the likelihood of modifying the ligand at the 'breakable' bond to increase the selectivity of protein 1 over protein 2 increases. If $D_s^1 - D_s^2 > 3$ Å, a similar equation is used to estimate the bond selectivity of protein 2 over protein 1:

$$S_2 = (D_s^1 - D_s^2)/D_s^2 \tag{5}$$

After the selectivity of all bonds is estimated with values of S_1 and S_2, all S_1 and S_2 are sorted in descending order and bonds are modified according to this order. The functional group modifications are determined by the closest active atoms for each modified bond and are most complementary to that atom, e.g., if the active atom is from a Arg or Lys residue, the added functional group should contain the carboxyl group and if the atom is from a Glu or Asp residue, the added functional group should contain the ammonium group. We are defining carboxyl and ammonium groups as the 'fundamental' functional group units for ligand design (Later, more subtle ligand design steps would incorporate a larger set of functional group substitutions). In order to calculate the coordinates of the atoms in the added group, we first place the key atoms of the fundamental functional groups using algorithms in the HINT (Hydropathic INTeractions) program [6]. HINT utilizes experimental solvent partitioning data as a basis for an empirical molecular interaction model that calculates free energy scores previously shown to fairly accurately reproduce experimental measurements of binding [6c, d and e]. The HINT calculation [6a] is a summation of hydropathic interactions between all atom pairs:

$$B = \sum \sum b_{ij}; \tag{6}$$

$$b_{ij} = S_i\ a_i\ S_j\ a_j\ R_{ij}\ T_{ij} + r_{ij}; \tag{7}$$

where b_{ij} is the interaction score between atoms i and j, which can be summed for all i and j, S is the solvent accessible surface area, a is the hydrophobic atom constant, and R_{ij} and r_{ij} are the functions of distance between i and j. T_{ij} is a logic function with value of 1 or −1, depending on the character of interacting polar atoms.

To calculate the best coordinates of the fundamental functional groups, we first treat CH_3COO^- or $CH_3NH_3^+$ as 'small ligands' placed between the 'breakable' bond and the closest active atom of the protein, with atom 1 (C of COO^- or N of NH_3^+)

directed at the active atom of the protein and atom 2 (C of methyl) directed at the 'breakable' bond. The distance between atom 1 and the active atom is initially set at about 2 Å. The HINT score for the interaction between the 'small ligand' and protein is then optimized by translations and rotations as described in an algorithm reported previously [9]. To form the new ligand, it is necessary to attach the fundamental functional group to the original ligand. A chain of 0, 1, or 2 carbons is usually used for this connection. The coordinates of the connecting carbons are dependent on the coordinates of the fundamental functional groups but it is also important that each new formed bond angle should be close to the normal angle for that bond. For example, the bond angle for an sp^3 carbon should be close to 109.5°. However, if more than 2 carbons are needed for the connection, the chain may be too flexible, which may in turn reduce the desired selectivity. This is an issue we are currently investigating.

A new molecule is created by adding the carbon chain atoms and the fundamental functional group(s) to the original ligand, removing the 'leaving' atoms or group and then forming new bonds among the original ligand, the connecting carbons and fundamental functional group. It is important to note that in this algorithm we are only interested in 'special' polar–polar functional group interactions such as the interactions between the carboxyl group and ammonium. Some hydrogen bond forming atoms in proteins such as in the −OH group of Serine, Threonine or Tyrosine can also form hydrogen bonds, but the hydrogen bonds formed by these atoms may not increase the binding affinity significantly because the strength of the hydrogen bonds between the −OH and ligand is not substantially different than that between solvent (water) and the ligand. In other words, the number of hydrogen bonds is conserved as shown in the following equation:

$$\begin{aligned} &\mathrm{D-H\cdots OH_2 + HOH\cdots B} \\ &= \mathrm{D-H\cdots B + HOH\cdots OH_2} \end{aligned} \quad (8)$$

The hydrogen bond formed by the hydrogen bond forming atoms in the −OH group of Serine, Threonine or Tyrosine is almost as strong as the hydrogen bond formed by water and the binding process in these cases is more or less isoenthalpic [10]. Thus, the relatively small contribution to the binding arises mainly from entropy. In contrast, a hydrogen bond formed by an ammonium group and a carboxyl group is ionically reinforced and much stronger than a hydrogen bond formed with water.

Algorithm 3: modify ligand based on hydrophobic–hydrophobic functional group interactions

Adding a carboxyl group or ammonium group to a ligand would result in increment of H-bond donors and/or H-bond acceptors and the new ligand may become too polar or, for example, violate the Lipinski rules [11]. We have also explored ligand modifications that involve hydrophobic–hydrophobic functional group interactions. In this algorithm, the new added groups are designed to have strong favorable interactions (mainly hydrophobic–hydrophobic) with residues of protein 1 and have unfavorable interactions (polar–hydrophobic or steric) with residues of protein 2. The estimate of interaction is based on HINT score calculations.

The hydrophobic-enabled algorithm is very similar to algorithm 1 with some differences in the following aspects: the calculation of the hydrophobic selectivity of a 'breakable' bond is based on the difference of hydrophobic properties of the atoms of the two 'competing' proteins around the bond. We first replace the 'leaving' atom or group with a methyl group and then calculate the HINT scores of the interaction between the new ligand and proteins. The hydrophobic selectivity of the 'breakable' bond is estimated by comparing the scores of the new ligand with the scores of the original ligand. The following equation is used to estimate the hydrophobic selectivity of a 'breakable' bond:

$$S_1 = (B_{11} - B_{10}) * B_{20}/B_{10} - (B_{21} - B_{20}) \quad (9)$$

S_1 is the hydrophobic selectivity of the bond for protein one. B_{10} is the HINT score of the interaction between protein 1 and the original ligand. B_{11} is the HINT score of the interaction between protein 1 and the modified ligand. B_{20} is the HINT score of the interaction between protein 2 and the original ligand. B_{21} is the HINT score of the interaction between protein 2 and the modified ligand. For a 'breakable' bond to be considered as a possible bond that can be modified to increase the selectivity of protein 1 over protein 2, its B_{11}

must be greater than B_{10} and its B_{21} must be less than B_{20}; otherwise nothing would be accomplished by modifying that 'breakable' bond.

The database of the small molecules used to replace the 'leaving' atoms or groups are different from the database used in algorithm 1. Figure 4 illustrates some examples of the small molecules from the algorithm 3 database.

The identification of the best small molecule is based on HINT score calculations such that the interaction score between the best new ligand (i.e., best small molecule) and protein 2 is less than that between original ligand and protein 2 and the interaction score between the best new ligand and protein 1 is higher than that between any other new ligand (from any other small molecule in any attachment position) and protein 1.

Applications

We plan to describe additional examples of using these algorithms in upcoming communications, but two simple applications that we have used for testing are: (1) we have automated the design modification of 'CB3717', a ligand that can bind with two very similar proteins: *L. casei* Thymidylate Synthase (TS) (wild type) and its E60Q mutant; and (2) we have applied our program to a previously published account of designing herbicidal inhibitor selectivity for compounds bound to plant and mammalian 4-hydroxyphenylpyruvate dioxgenases (HPPD). In both cases our program has discovered plausible compounds that have *calculated* selectivities much higher than the non-selective seed compounds.

Selective inhibition of L. casei thymidylate synthase

Thymidylate synthase catalyzes a step in the *de novo* synthesis of 2′-deoxythymidine 5′-monophosphate (dTMP) using the substrate 2′-deoxyuridine 5′-monophosphate (dUMP) and a cofactor, 5,10-methylene-5,6,7,8-tetrathydrofolate [12]. Crystal structures for both the wild type (pdb code: 1LCA) [12] and mutant (pdb code: 1VZE) [12] complexes of *L. casei* thymidylate synthase (TS) are available. Although wild type TS and E60Q TS are very similar, their crystal structures are noticeably different at the active (binding) site for CB3717 (**A** in Figure 5). Interestingly, the E60Q site mutation is not directly at the active site, but induces several residues to shift positions, thus changing the shape and hydropathic profile of the site. However, the Glu60 sidechain of the wild type TS exerts a catalytic effect through the mediation of a well-ordered water molecule [12c]. Our selectivity program modifies CB1717, by our choice, either to increase its computational selectivity of wild type TS over the E60Q mutant or to increase its computational selectivity of the E60Q mutant over the wild type TS. Some examples of the new ligands designed by this program are shown in Figure 5. The modifications to the original CB1717 ligand are indicated in the boxed areas.

The first ligand, **B** (Figure 5), is designed based on steric interaction where the highlighted portion is the new added group. This group has strong

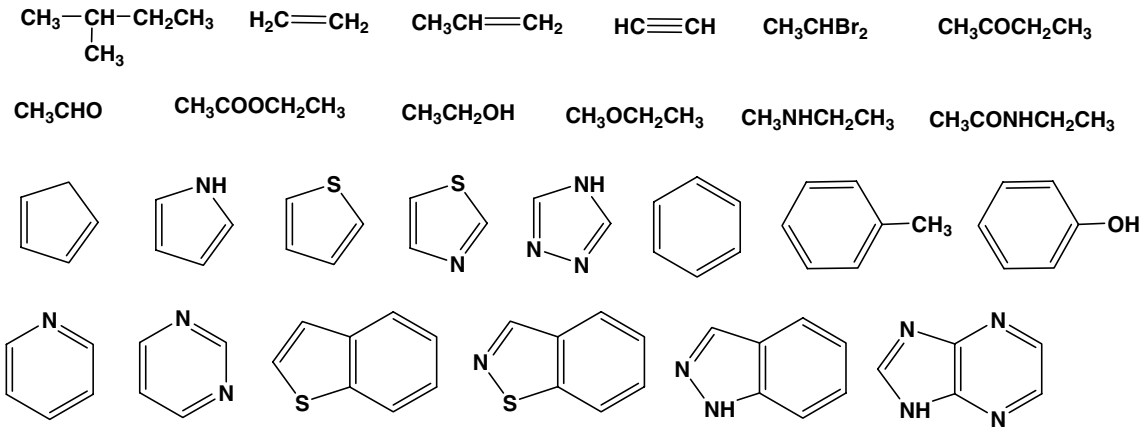

Figure 4. Some examples of small molecules as potential new added groups for algorithm 3.

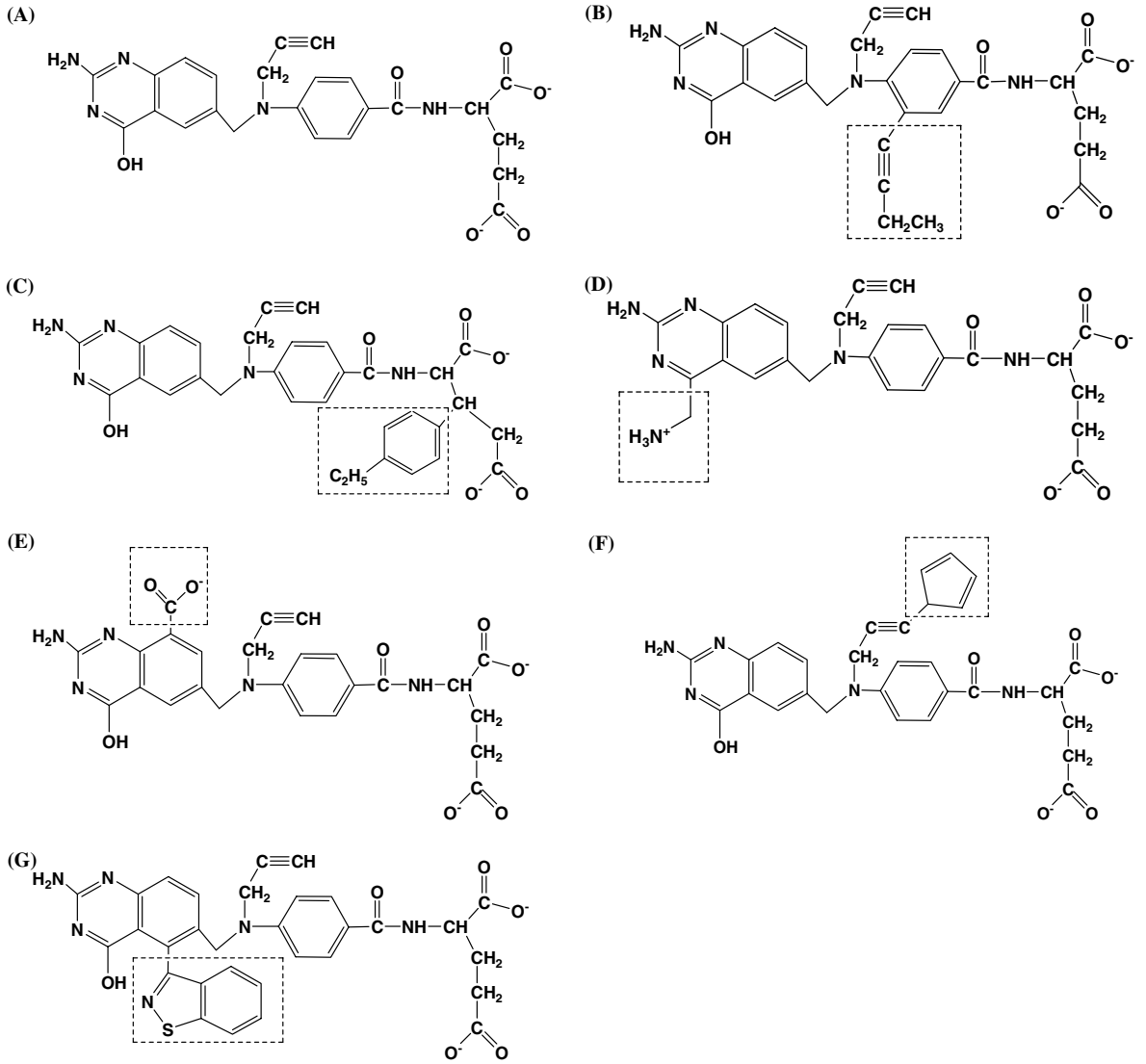

Figure 5. Structures of some new ligands (b–g) created by modifying the structure of CB3717 (a). The modifications are indicated in the boxed areas.

steric interaction with some residues of TS, in particular Trp82 and Ile81 as shown in Figure 6a, but does not have similar interactions with the E60Q mutant as shown in Figure 6b. Thus, ligand **B** is expected to be very selective for the mutant thymidylate synthase protein over the wild type protein. The second new ligand (**C**) (Figure 5) is also designed based on steric interaction differences. The new added group has strong (unfavorable) steric interactions with some residues of wild type TS such as Pro227 and Phe228 as illustrated in Figure 6c, but has no unfavorable interactions with the E60Q mutant as illustrated in Figure 6d.

Ligand **D** (Figure 5) is designed based on favorable polar–polar functional group interactions. The hydroxyl group at C5 of CB3717 is replaced by an ammonium group. This new added group has a strong favorable interaction (hydrogen bond) with the residue Asp221 of the wild type TS with a HINT score of about 520 ($\sim$1 kcal mol^{-1}) [6a, b and c]. There is no analogous hydrogen bond with residues in the E60Q mutant TS. Figure 7a indicates the interactions between the new group and Asp221 of TS (wild type).

Ligand **E** (Figure 5) is also designed based on favorable polar–polar functional group

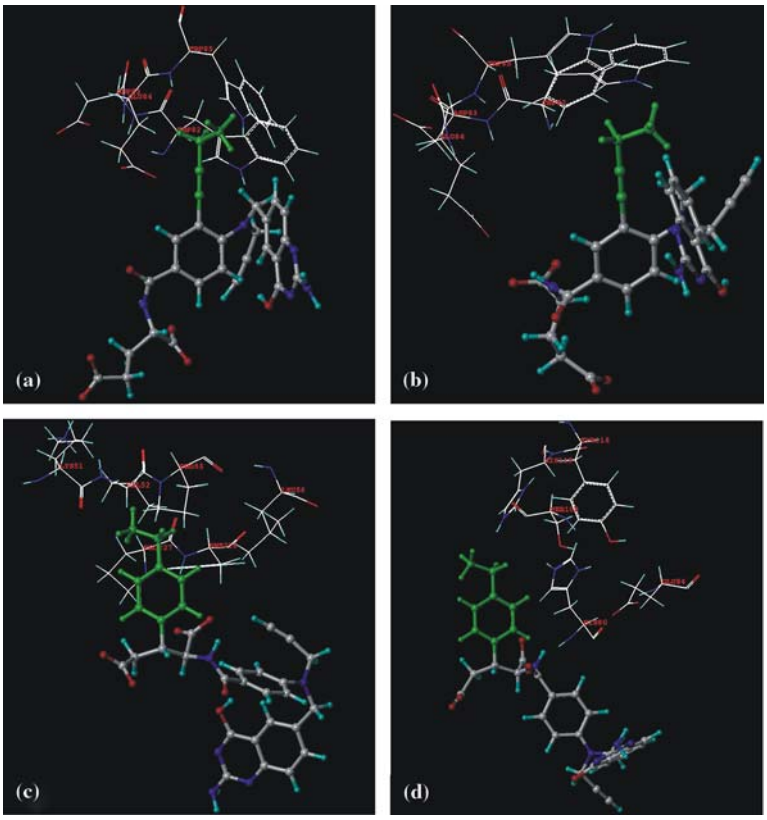

Figure 6. (a) New ligand **B** and the residues of *L. casei* thymidylate synthase (TS) in the region of the new added group (green atoms). (b) Ligand **B** and the residues of the mutant E60Q *L. casei* TS in the region of the new added group. (c) New ligand **C** and the residues of *L. casei* thymidylate synthase (TS) in the region of the new added group (green atoms). (d) Ligand **C** and the residues of the mutant E60Q *L. casei* TS in the region of the new added group.

interactions: a carboxylate group is added to C11 of CB3717. This new added group has a strong favorable interaction (hydrogen bond) with the residue Arg23 of the E60Q mutant TS and the HINT score of this interaction is about 1770 ($\sim$3.5 kcal mol^{-1}). The added group does not have a strong interaction with residues in the wild type TS. Figure 7b shows the interaction between

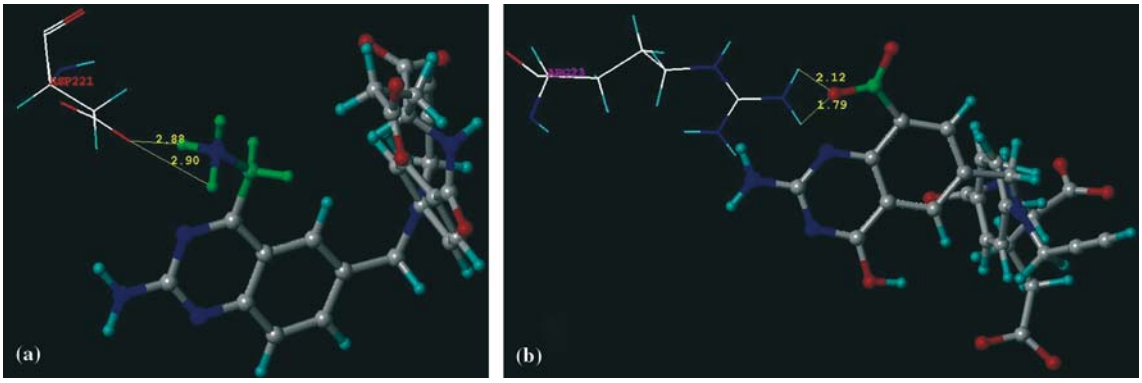

Figure 7. (a) Bound structure and hydrogen bonds formed between the new ligand **D** and residue Asp221 of *L. casei* TS. (b) Hydrogen bond formed between the new ligand **E** and residue Arg23 of E60Q *L. casei* TS.

the new added group and Arg23 of the E60Q mutant TS.

Ligands **F** and **G** (Figure 5) are designed based on favorable hydrophobic–hydrophobic functional group interactions. **F** has higher binding affinity with wild type TS and lower binding affinity with the E60Q mutant than the original ligand. On the other hand, **G** has higher binding affinity to the E60Q mutant and lower binding affinity to the wild type TS than the original ligand. Table 1 lists the HINT score comparison of **A–G** with the two proteins. Since about 515 HINT score units roughly correspond to 1 kcal mol^{-1} binding energy [6a,b and c], these changes should be significant enough to noticeably enhance binding selectivity. Note that since the HINT score correlates with free energy of binding, a ~500 unit increase in HINT score roughly corresponds to factor of 10 decrease in K_i.

Selective inhibition of (*plant*) *Arabidopsis 4-hydroxyphenylpyruvate dioxgenase*

The 4-hydroxyphenylpyruvate dioxygenase (HPPD) enzyme catalyzes the oxygenation of 4-hydroxyphenylpyruvate to homogentisate and releases CO_2. In mammals HPPD is important for catalytically metabolizing tyrosine, while in plants HPPD is critical in the synthesis of photoprotectant carotenoids [13]. Yang et al. [13] recently reported crystal structures for mammalian (rat: 1sqi) and plant (*Arabidopsis thaliana*: 1tfz) HPPD bound with the same non-selective inhibitor, as well as a crystal structure of a new 1600-fold selective inhibitor bound to the *Arabidopsis* HPPD enzyme (1tg5). This ligand was discovered through an extensive screening campaign. As a second example of applications for our selectivity enhancement program, we have investigated this problem. We were interested to see whether our program would generate similar or identical chemical entities as selective enzyme inhibitors to the plant HPPD.

Table 1. HINT interaction scores for TS protein–ligand complexes.

	A (CB1717)	B	C	D	E	F	G
Wild Type	409		<0[a]	<0[a]	929	654	899	<0[a]
Mutant	217		304	241	331	1987	<0[a]	1034

[a]Because of significant steric clashes, these scores are essentially meaningless.

The inhibitor **H** (DAS869) (Figure 8) binds to plant and rat HPPDs with I_{50}s of 7 and < 20 nM, respectively. Inhibitor **I** (DAS645) is highly selective for *Arabidopsis* HPPD with I^{50} of 12 nM and no detectable inhibition of rat HPPD at concentrations up to 20 μM [13]. HINT scores (see Table 2) for **H** are consistent: 1099 for *Arabidopsis* and 1029 for rat. However, because of steric clashes, **I** can not be scored within the active site induced by **H** in either plant (1tfz) or mammal (1sqi) HPPD [13], and must, instead, be scored within its own ligand-HPPD (plant) crystal structure (1tg5). In this environment, **I** has a HINT score of 931. The differences in active-site architecture between the **H**-induced and **I**-induced structures are an important point that will be discussed below.

Inhibitors **J**, **K** and **L** (Figure 8) were designed from the differences in active site structure between the two structures where **H** is bound. All have higher scores than **H** for plant HPPD and lower scores than **H** for mammal HPPD (see Table 2). The added carboxyl group of **J** appears to be in position to form a very strong hydrogen bond with Lys400 of *Arabidopsis* HPPD. However, none of these complexes, or a number of others that were generated, indicate the key hydrophobic substitution at the pyrazole heterocyclic ring of **H** as in the highly selective inhibitor **I**. This is because there are a number of flexible chains and residues at the active site that are induced to reposition by the inhibitor **I**. In particular, the sidechain of Phe403 undergoes a significant rotation [13].

To explore this effect and sample this region of the active site, we performed a second selectivity optimization run on the difference in active site structure between the mammal HPPD and the **I**-induced plant HPPD structure. Inhibitors **M** and **N** were thus designed by our program (Figure 8 and Table 2) using a seed compound based on inhibitor **I** with the pyrazole hydrophobic substituent deleted. These two compounds (**M** and **N**) have very similar HINT scores to inhibitor **I**. Most significant, this change in starting structure allowed the program to make hydrophobic chemical substitutions at the pyrazole ring. Our database does not at present include *m*-dichlorobenzene, so we were not able to actually design compound **I**. Lastly, we artificially adapted the

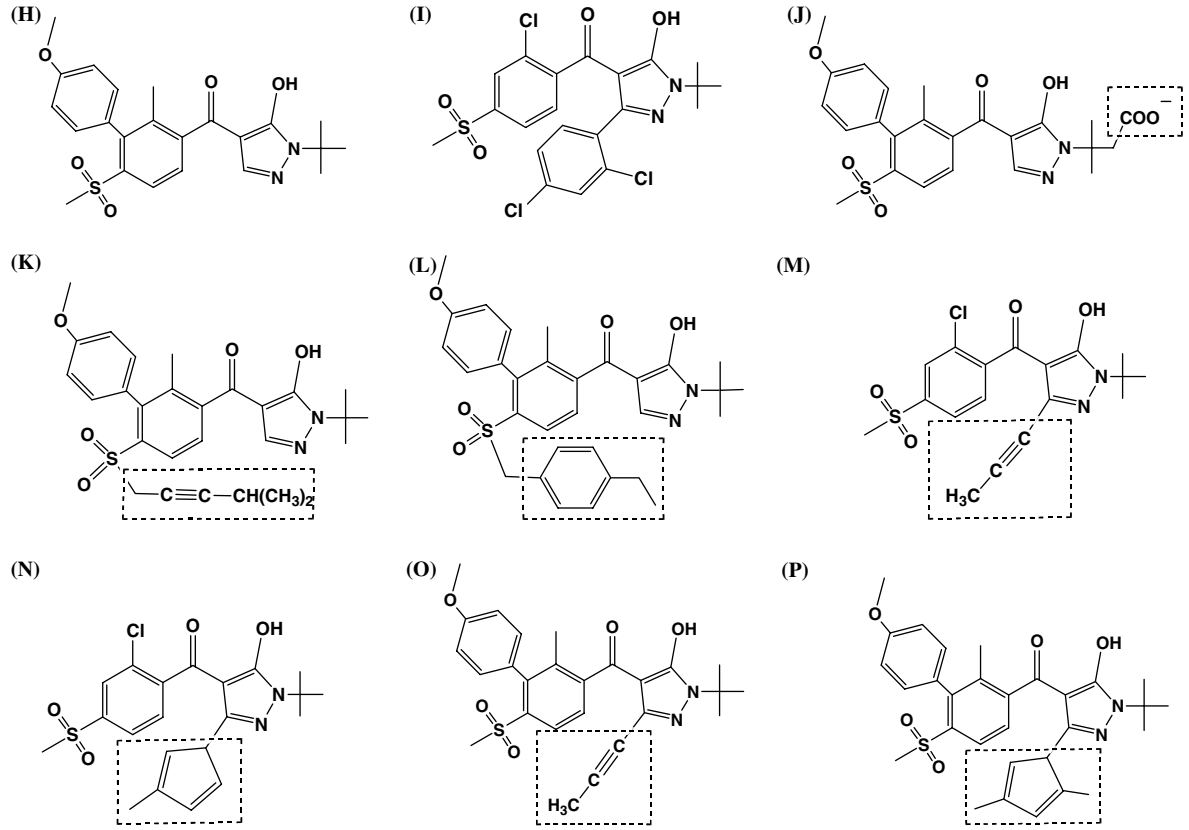

Figure 8. Structures of some new ligands (j–p) created by modifying the structure of DAS869 (h). The ligand DAS645 (**i**) was identified by screening.

active sites of the **H**-induced plant and mammal enzymes by rotating Phe403 (1tfz) and the corresponding Phe364 (1sqi) to maximize the volume of the active sites. Phe403 was rotated 150° about the Cα-Cβ bond in 1tfz and Phe364 was rotated 255° in 1sqi. Inhibitors **O** and **P** (Figure 8 and Table 2) were generated from these two starting protein structures using **H** as the seed. These compounds are similarly substituted on the pyrazole ring as **M** and **N** and would also appear to be quite selective for the *Arabidopsis* HDDP enzyme.

Table 2. HINT interaction scores for HPPD protein-ligand complexes.

Enzyme/Host Structure	PDB code	H	I	J	K	L	M	N	O	P
Plant HPPD (bound with **H**)	1tfz	1099	<0[a]	1948	1264	1308	–	–	–	–
Mammal HPPD (bound with **H**)	1sqi	1029	<0[a]	279	−622	−1757	<0[a]	<0[a]	–	–
Plant HPPD (bound with **I**)	1tg5	575	931	1438	612	840	1014	1068	–	–
Plant HPPD (**H**/Phe403 rotated)	1tfz[c]	785	908	–	–	–	–	–	1190	1272
Mammal HPPD (**H**/Phe364 rotated)	1sqi[d]	865	<0[a]	–	–	–	–	–	−1484	−4536

[a]Because of significant steric clashes, these scores are essentially meaningless.
[b]Although these ligands were designed for the **H**-induced plant vs. mammal structures, scores have been calculated for the **I**-induced plant active site.
[c]The Phe403 sidechain was rotated to maximize the active site volume.
[d]The Phe364 sidechain was rotated to maximize the active site volume.

Discussion and summary

Some inspiration for our technology of "automated" computational design of selective inhibitors is the family of programs that build *in situ* at the binding site new and possibly novel compounds possessing idealized contacts with the site atoms. This is termed "*de novo* ligand design" and a number of such programs designed for this purpose have been described, e.g., LUDI [14], LEAPFROG [15] and AlleGrow (Bohacek, unpublished), but to date none have been more than modestly successful in designing new, useful and accessible drug leads. The core technologies of these programs are largely based on a somewhat random process of iteratively building and scoring. There is, however, an inherent limitation in that molecules thus created may end up being neither thermodynamically reasonable nor synthetically accessible. In an attempt to circumvent this problem, Paul Bartlett and coworkers [16] have designed 3D libraries of tricyclic hydrocarbons (TRIAD) and acyclic molecules (ILIAD) that are chemically-reasonable starting points and also developed a complementary software package: CAVEAT. Our program incorporates this latter database feature, albeit in a much more limited form, but also has a much more modest goal of only modifying the chemical "decorations" of a proven template, rather than complete *de novo* design.

In the present work a *de novo* computational program used to increase ligand–receptor selectivity was designed and implemented. This program is designed to increase a ligand's binding selectivity by creating modified ligands such that the binding affinity of the ligand to one protein is increased while the binding affinity to the other protein is decreased. In this paper we described the algorithms of our program and illustrated the types of ligand structural modifications that the program creates with an example based on binding to thymidylate synthase. Being able to computationally design specificity is a significant milestone in drug design because of the importance of reducing side effects and improving ADMET properties before undertaking expensive clinical studies of drug candidates. Our program is a small step towards this goal, but much work remains, including investigations of systems where the computational predictions can be experimentally validated. Such studies are now underway.

Acknowledgements

We gratefully acknowledge partial support of this research by Virginia Commonwealth University and NIH NIAID grant 5R01AI052330 to Dr. Kevin A. Reynolds. We also thank Drs. Reynolds, Derek Cashman and Micaela Fornabaio for helpful discussions.

References

1. (a) Li, A.P., Curr. Top. Med. Chem., 4 (2004) 701; (b) Lin, J., Sahakian, D.C., de Morais, S.M., Xu, J.J., Polzer, R.J. and Winter, S.M., Curr. Top. Med. Chem., 3 (2003) 1125.
2. Riley, R.J. and Kenna, J.G., Curr. Opin. Drug. Discov. Devel., 7 (2004) 86.
3. (a) Cheng, A. and Merz, K.M., Jr., J. Med. Chem., 46 (2003) 3572; (b) Gombar, V.K., Silver, I.S. and Zhao, Z., Curr. Top. Med. Chem., 3 (2003) 1205; (c) Vermeulen, N.P., Curr. Top. Med. Chem., 3 (2003) 1227; (d) Krejsa, C.M., Horvath, D., Rogalski, S.L., Penzotti, J.E., Mao, B., Barbosa, F. and Migeon, J.C., Curr. Opin. Drug Discov. Devel., 6 (2003) 470; (e) Stouch, T.R., Kenyon, J.R., Johnson, S.R., Chen, X.Q., Doweyko, A. and Li, Y., J. Comput. Aided Mol. Des., 17 (2003) 83.
4. (a) Cashman, D.J. and Kellogg, G.E., J. Med. Chem., 47 (2004) 1360; (b) Cashman, D.J., Scarsdale, J.N. and Kellogg, G.E., Nucleic Acids Res., 31 (2003) 4410.
5. (a) He, X., Reeve, A.M., Desai, U., Kellogg, G.E. and Reynolds, K.A., Antimicrob. Agents Chemother., 48 (2004) 3093; (b) He, X. and Reynolds, K.A., Antimicrob. Agents Chemother. 46 (2002) 1310; (c) Scarsdale, J.N., Kazanina, G., He, X., Reynolds, KA. and Wright, H.T., J. Biol. Chem., 276 (2001) 20516.
6. (a) Kellogg, G.E. and Abraham, D.J., Eur. J. Med. Chem., 35 (2000) 651; (b) Burnett, J.C., Kellogg, G.E. and Abraham, D.J., Biochemistry, 39 (2000) 1622; (c) Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D.J., Kellogg, G. E. and Mozzarelli A., J. Med. Chem., 45 (2002) 2469; (d) Fornabaio, M., Cozzini, P., Mozzarelli, A., Abraham, D.J. and Kellogg, G. E., J. Med. Chem., 46 (2003) 4487; (e) Fornabaio, M., Spyrakis, F, Mozzarelli, A., Cozzini, P., Abraham, D.J. and Kellogg, G. E., J. Med. Chem., 47 (2004) 4507.
7. (a) Kellogg, G.E., Fornabaio, M., Chen, D.L. and Abraham, D.J., J. Chem. Info. Comput. Sci. (submitted). (b) Kellogg, G.E; eduSoft LC Programmers' Toolkit Manual http://www.edusoft-lc.com/toolkits/manuals.
8. Richards, F.M., Ann. Rev. Biophs. Bioeng., 6 (1977) 151.
9. Kellogg, G.E. and Chen, D.L., Chem. & Biodivers., 1 (2004) 98.
10. (a) Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P, Waye, M M.Y. and Wiinter, G., Nature, 314 (1985) 235; (b)

Jencks, W.P., Catalysis in Chemistry and Enzymology, McGraw-Hill, New York, 1969.
11. (a) Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., Adv. Drug Deliv. Rev., 46 (2001) 3; (b) Oprea, T. I., J. Comput. Aided Mol. Des., 14 (2000) 251–264.
12. (a) van Laar, J.A., Rustum, Y.M., Ackland, S.P., van Groeningen, C.J. and Peters, G.J., Eur. J. Cancer, 34 (1998) 296–306; (b) Finer-Moore, J., Fauman, E.B., Foster, P.G., Perry, K.M., Santi, D.V. and Stroud, R.M., J. Mol. Biol., 232 (1993) 1101–1116; (c) Birdsall, D.L., Finer-Moore, J. and Stroud, R.M., J. Mol. Biol., 255 (1996) 522–535.
13. Yang, C., Pflugrath, J.W., Camper, D.L., Foster, M.L., Pernich, D.J. and Walsh, T.A., Biochemistry, 43 (2004) 10414–10423.
14. Bohm, H.J., Persp. Drug. Discov. Design, 3 (1995) 21–33.
15. Tripos, Inc., St. Louis, MO, USA.
16. (a) Lauri, G. and Bartlett, P.A., J. Comput.-Aided Mol. Design, 8 (1994) 51–66; (b) Bartlett, P.A., In Chatgilialoglu, C. and Snieckus, V. (Eds.), Organic Synthesis, From Gnosis to Prognosis (NATO Advanced Study Institute), Kluwer, Dordrecht, 1986, pp. 137–173.