Journal of Computer-Aided Molecular Design, **15:** 911–933, 2001.
KLUWER/ESCOM

A genetic algorithm for structure-based de novo design

Scott C.-H. Pegg[a], Jose J. Haresco[b] & Irwin D. Kuntz[a,*]
[a]*Department of Pharmaceutical Chemistry and* [b]*Program in Biological and Medical Informatics, University of California, San Francisco, San Francisco, CA 94143, USA*

Received 2 February 2001; accepted 18 July 2001

Key words: ADAPT, combinatorial chemistry, de novo design, DOCK, genetic algorithm

Summary

Genetic algorithms have properties which make them attractive in de novo drug design. Like other de novo design programs, genetic algorithms require a method to reduce the enormous search space of possible compounds. Most often this is done using information from known ligands. We have developed the ADAPT program, a genetic algorithm which uses molecular interactions evaluated with docking calculations as a fitness function to reduce the search space. ADAPT does not require information about known ligands. The program takes an initial set of compounds and iteratively builds new compounds based on the fitness scores of the previous set of compounds. We describe the particulars of the ADAPT algorithm and its application to three well-studied target systems. We also show that the strategies of enhanced local sampling and re-introducing diversity to the compound population during the design cycle provide better results than conventional genetic algorithm protocols.

Introduction

The ultimate goal of structure-based de novo drug design is the ability to develop new ligands using information from the three dimensional (3D) structure of a protein target without the prior knowledge of other ligands. The search space of all possible chemical compounds is far too large to be enumerated [1], while the locations of bioactive drugs within this space tend to be sparse, non-contiguous, and very difficult to predict a priori. Current computational methods use the 3D structural information of the target to narrow this search space. One strategy is to join compound fragments which fit favorably into the active site [2, 3] either by constructing linking atoms or by searching a database of possible linking structures. [4–9] While this strategy leads to many novel and potentially useful structures, the resulting compounds are very dependent on the placement and selection of the fragments [10]. Another strategy is to construct a molecule within a binding site by 'growing' it from a seed atom or fragment, using the site's steric and/or energetic properties to guide the growth of the compound [11–15]. This strategy can also result in useful structures, but the molecules built are sensitive to the position and choice of the seed atom used and often lead to molecules that are difficult to synthesize [10, 16].

Genetic algorithms have recently been applied to the problem of de novo drug design. Based on the concepts of natural selection, genetic algorithms operate by creating a set of proposed solutions ('population') to a problem of interest, evaluating them ('fitness pressure'), and then using the best solutions to develop a new set of proposed solutions ('breeding'). Genetic algorithms have several properties which make them attractive in de novo ligand design, the most important being the ability to perform well when the search space is large and incompletely understood, the fitness functions are not exact, and a global minimum is not required. Unlike incremental growing schemes, genetic algorithms can use properties of whole compounds to guide the generation of new compounds. Properties such as absorption, distribution, metabolism, and excretion (ADME) cannot easily be calculated from molecular fragments or partially constructed compounds, and they are of increasing

*To whom correspondence should be addressed. E-mail: kuntz@cgl.ucsf.edu

importance in today's drug development climate [17]. Another potentially attractive feature of genetic algorithms is the production of an ensemble of solutions. A diverse set of proposed compounds can help to avoid patent issues and provide 'back up' compounds of different chemical classes which can be used if the initial compounds fail during drug development.

As with other computational de novo design strategies, genetic algorithms require a method of reducing the allowable search space. Most of the currently published genetic algorithm strategies use fitness functions that rely upon knowledge of a template ligand or pharmacophore to restrict the space of allowable compounds. Such fitness functions include similarity to a specific known ligand [18], fitness to a pharmacophore template derived from putative or known ligands [19–21], and the presence of properties previously determined by structure-activity relationships [22]. These strategies use implicit information about the target, and are not solely structure-based strategies. Exceptions include the method of Glen and Payne in which a set of pharmacophoric constraints is built directly from a protein active site [20] and a genetic algorithm implementation which used experimentally measured binding affinities as a fitness function [23]. It has been suggested that current molecular docking methods, in which the steric and energetic fit of a ligand to a target site is estimated, could be used as a fitness function [24]. Such a function would restrict the search space using information from the 3D structure of the target without using knowledge of existing ligands. Recently, Sheridan et. al. published a further development of their genetic algorithm for designing combinatorial chemistry compound libraries which incorporates docking calculations into the fitness function [25, 26]. While similar to our approach, their algorithm has been demonstrated on compounds built only via linear combinations of fragments, and uses a docking calculation in which only a small number of rigid conformations of each compound are scored.

We present here the ADAPT program, a genetic algorithm for structure-based de novo drug design which builds compounds as acyclic graphs of fragments and incorporates the use of flexible docking calculations into the fitness function. We first give the details of the genetic algorithm implementation, and then show its application to ligand design problems in which we demonstrate the searching ability of the genetic algorithm, the effects of forcing local sampling, and the performance enhancement of adding extra diversity to the compound population while the genetic algorithm is running. We also demonstrate application of the program to three protein target systems in which we can recover fundamental properties of ligands known to bind to these targets.

Methods

Genetic algorithms

Genetic algorithms are a type of stochastic search algorithm originally developed by Holland [27, 28] based on the concepts of Darwinian evolution [29]. All genetic algorithms start with a population of proposed solutions to a given problem. In the de novo drug design case, the population is a set of small molecules and the problem at hand is that of building bioactive compounds with desirable properties. The population is subjected to a fitness pressure in which the individuals are evaluated in terms of how well they solve the given problem. The individuals are then bred to create a new population, with the more 'fit' individuals being chosen to breed more often. This breeding is usually done using an adaptation of the crossover and mutation events observed in DNA replication. The genetic algorithm cycle continues though generations of evaluation and breeding until sufficiently good solutions have been found. Figure 1 shows a genetic algorithm cycle and its application to the de novo design problem. The following are the specifics of our particular adaptation of genetic algorithms to structure-based de novo ligand design.

Compound representation

In the ADAPT program, compounds are represented by an *acyclic* graph of at most 16 fragments, with the user defining an upper and lower bound on the number of fragments per molecule. The fragments themselves (including chiral centers) are defined by the user via an adapted subset of the SMILES line notation [30], and each fragment can have, at most, 8 user-defined points to which other fragments may be attached. The fragments reside in a linear order in a file such that chemically similar fragments are located next to each other in the fragment file (this order can also be hand-edited). This order is determined using a combination of binning and 'gray-scaling' based on Daylight fingerprints [31]. Fragments are first binned by the number of 1's in their fingerprint and the bins are sorted according to the number of 1's per fingerprint. The fragments are then sorted using a walk,

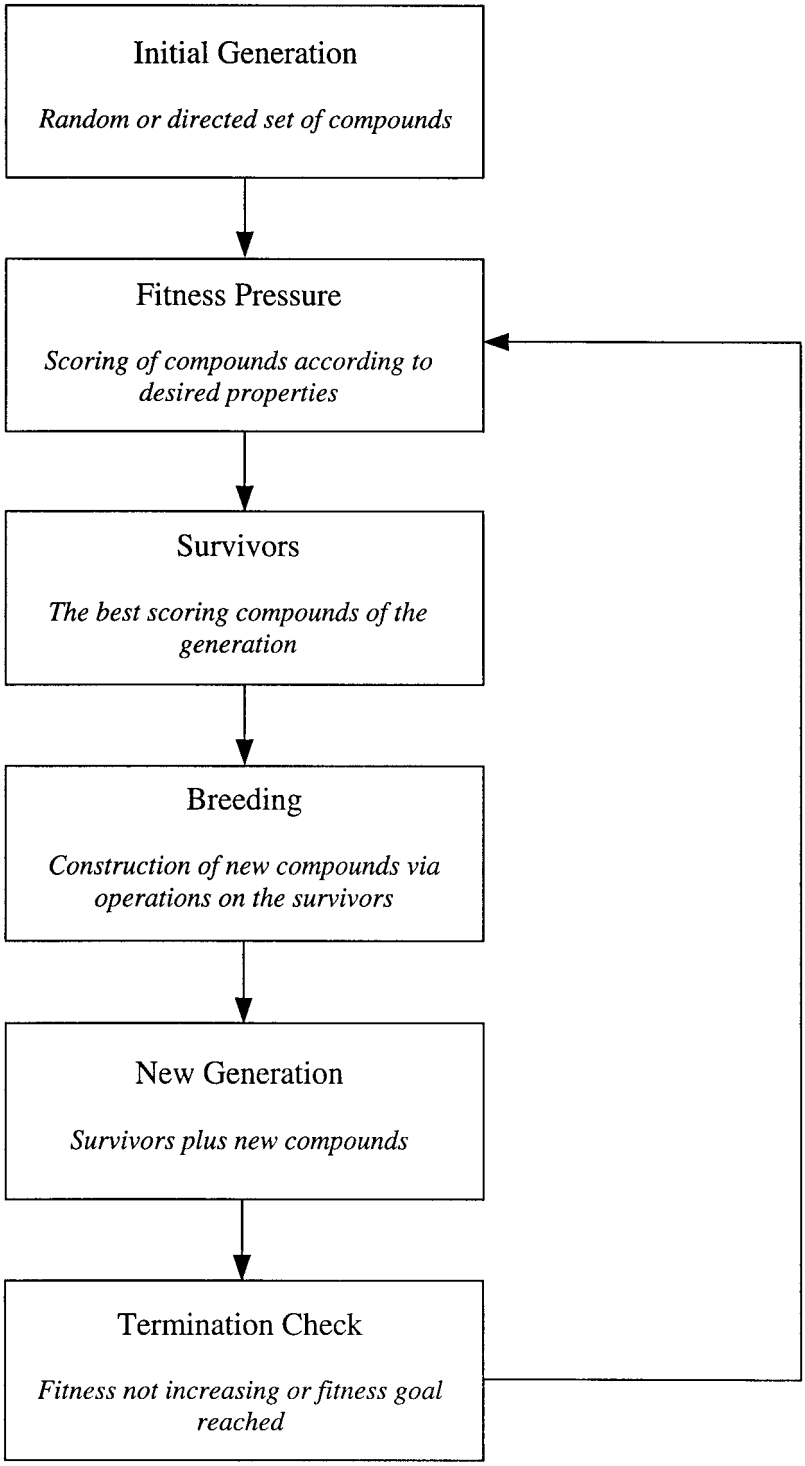

Figure 1. The genetic algorithm cycle and its application to drug design. An initial set of compounds is scored according to a set of desired properties, such as estimated binding affinity. The best scoring compounds of the generation are considered 'survivors' and are used in the construction of new compounds. The next generation, consisting of the survivors and these new compounds, is then subjected to the fitness pressure. The cycle typically continues for either a set number of generations, or until the fitness scores no longer improve.

which starts with a randomly chosen fragment from the bin holding fragments with the fewest number of 1's. The next fragment in the walk is chosen from the same bin as the fragment with the largest number of 1's in common with current fragment. When the current bin is empty, the next fragment is chosen from the next bin. The user may designate one fragment to be a 'scaffold', in which case each compound must have one (and only one) occurrence of that particular fragment.

Starting generation

The starting generation can either be a diverse set of compounds, or it can be based on a user-defined compound. A diverse population of compounds is generated by choosing fragments at random and assembling them at random attachment points one at a time. Iteratively adding a node to the graph via a single new edge creates a connected graph with N nodes and N−1 edges, which must be acyclic. When the starting generation is seeded with a user-defined compound, the initial generation is built by randomly adding, subtracting, or swapping at most two fragments from the original structure. The user must define the original compound as an acyclic graph in terms of fragments (from the fragment file) and their connectivities. A copy of the original compound is included unchanged in the first generation.

Fitness evaluation

The major fitness evaluation of the ADAPT program is performed by the DOCK4.0 program [32] which gives an estimate of the steric and energetic fit of a ligand to a receptor site, taking into account ligand flexibility. We chose the DOCK package based on its immediate availability and our familiarity with its operation. Compounds are first converted to SMILES strings by our genetic algorithm, and then to a single three dimensional conformer using the CONCORD (version 4.0.2) [33] module of the SYBYL package (version 6.5) with default settings [34]. Under some circumstances, such as a higher than allowed number of rotatable bonds, CONCORD will fail to produce a structure, in which case the compound receives a DOCK score of zero. The DOCK score is computed as a pairwise potential between atoms of the ligand and atoms of the active site. The scoring function contains two terms, a standard 6–12 Van derWaals term and a 1/r electrostatic term [32]. The user may specify the number of times the docking calculations are performed, using different random number seeds for each run and taking the best score across all runs as the DOCK score, S_{dock}.

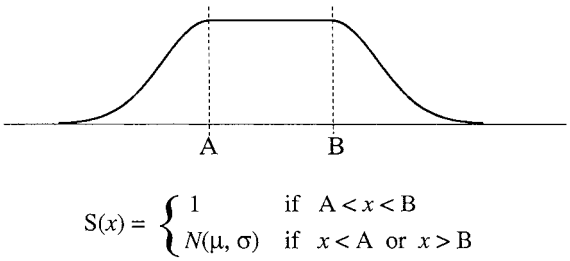

Figure 2. The plateau fitness function. Values of properties within the user defined limits A and B return a score of 1. For values outside this range, the scoring function returns the value of the normal distribution (with mean and standard deviation defined by the user) which has been split in half and moved to the edges of the plateau.

Also implemented are functions to measure simple physical properties of the assembled molecules such as *C* log *P* (using Daylight's 'clogp' program), molecular weight, number of rotatable bonds, and number of hydrogen bond donors/acceptors. These functions are scored according to a plateaued normal distribution (see Figure 2) in which values between the user-defined plateau range return 1.0 and otherwise return values between 0 and 1 according to a user-defined normal distribution that makes up the edges of the plateau. Given the value x for a compound property n, the score returned is

$$S_n(x) = \begin{cases} 1 & \text{if} \quad A_n \le x \le B_n, \\ N(\mu_n, \sigma_n) & \text{if} \quad A_n > x \text{ or } B_n < x, \end{cases}$$

where A_n and B_n are user-defined boundaries and $N(\mu_n, \sigma_n)$ is a normal distribution with mean μ_n and standard deviation σ_n as defined by the user. In the case where a property measurement function returns an error (not uncommon for *C* log *P* calculations), the score returned is zero. Each compound property thus has its own scoring function defined by the user. Several of the properties which can be used in the ADAPT fitness measurement are identical to those used to describe the properties of 'drug-like' compounds; users may set the boundaries A_n and B_n to the values recommended by Lipinksi et al. [17] or to values of their own choosing.

The overall fitness of a compound is simply the linear sum of each fitness measure multiplied by its user-defined coefficient,

$$\text{fitness} = C_{\text{dock}}S_{\text{dock}} + C_{mw}S_{mw} + C_{\text{rot}}S_{\text{rot}} + \\ + C_{h\text{bond}}S_{h\text{bond}} + C_{c\log p}S_{c\log p},$$

where C_n is the coefficient for score S_n. The total fitness score of each compound is saved, yet if an identical compound is created via breeding of the population the score for the newly created compound is recalculated (the redundant computing of a few scores is typically faster than performing exhaustive comparisons of each new compound to each of those previously seen). Once each compound has an overall fitness score, the top scoring fraction (as defined by the user) of compounds is used as the parents for the next generation. The user may also choose whether the parents are kept unchanged in the next generation, or whether only their children continue. In the experiments described below, we chose to keep the parents to avoid losing a good compound due to the high rates of mutation.

Breeding

Breeding of a set of compounds to produce the next generation is done via adaptations of crossover and mutation to the graph representation of the compounds. Crossover occurs by swapping a subset of nodes between two graphs. The number of nodes is chosen at random according to a decaying exponential distribution (biasing the exchange towards a smaller number of nodes), and the nodes themselves are chosen by choosing a random node and then performing a random walk through the graph to build a connected set of nodes. The sets of nodes swapped are either of equal size, where the crossover process is isomorphic to a recoloring of nodes according to type, or are of unequal sizes, in which case the crossover process is equivalent to a recoloring of nodes followed by the addition or removal of nodes with one edge per node being added or removed. Because the sets of nodes swapped are themselves acyclic graphs, the cases of equal and unequal set sizes both lead to acyclic graphs. The current implementation of crossover in ADAPT is analgous to a 'single-point' crossover of DNA. Currently being investigated is an implementation of 'multiple-point' crossover, which requires the swapping of multiple disjoint sets of fragments (nodes) between compounds (graphs). The parents for crossover are chosen via an 'elitist' scheme, in which the probability of being chosen is proportional to a compound's relative fitness within a generation, the more fit being chosen more often.

Mutation can occur to either the identity of the nodes or the connectivity of the graph. The connectivity is mutated by choosing an edge at random and reconnecting it to a node with an open attachment point at random as long as it doesn't induce a cycle in the graph. Because there are N−1 edges for N nodes in the graph, checking for a cycle is equivalent to looking for a disconnected node, an algorithm which has complexity O(N). Node identity is mutated by moving up or down in the fragment definition file. The probability of a mutation is determined by the user, while the distance moved is determined by a decaying exponential also defined by the user, such that short distance moves are more likely. When the fragments are ordered in terms of chemical similarity, this scheme causes the genetic algorithm to preferentially sample local chemical space.

Stopping

The ADAPT cycle will stop when the fitness of the generations ceases to improve, the program has found a generation with fitness better than a defined goal, or the program has gone a set number of generations. The user can set a threshold for the required improvement of the fitness, as well as a window over which these values can be averaged. When using a fitness goal, the algorithm may get stuck in a local minimum where the fitness of the generations is no longer improving, but the fitness value is not near the goal. In this case, the ADAPT program adds 'diversity' to the current generation by randomly adding, subtracting, or swapping at most two fragments from each compound. This allows the program to 'jump' to another random (but not too distant) point in chemical space from which it may have a better chance of reaching the fitness goal. The user defines the number of generations the ADAPT program may progress before this diversity is added.

Results

We chose three well studied protein targets to test the ADAPT program on: cathepsin D, dihydrofolate reductase, and HIV-1 reverse transcriptase. The cathepsin D target was chosen because we have experimental binding data from a previous combinatorial chemistry project [35] to which we could compare the results of ADAPT applied to the same combinatorial scheme. Dihydrofolate reductase was chosen because it would provide a larger search space than a simple combinatorial chemistry scheme, but one in which there exists

a well-studied ligand (methotrexate) that would allow us to test the effects of seeding the initial population with a known ligand. HIV-1 reverse transcriptase was chosen because the non-nucleoside binding site accepts a wide variety of known ligands, representing a large search space, but with a specific structural theme that we could attempt to rediscover with the ADAPT program.

In all three cases we wished to investigate how well ADAPT performed when using molecular docking calculations as the fitness function. In our cathepsin D experiments, DOCK was used as the only fitness function. In the dihydrofolate reductase and HIV-1 reverse transcriptase experiments, where molecular size was not restricted by a scaffold-based design, we used both the DOCK scoring function and a loose molecular weight restriction in order to adjust for the DOCK scoring function's bias towards larger compounds. All docking calculations were performed using flexible ligands.

Table 1. ADAPT parameter settings used in the cathepsin D experiments.

Parameter	Value
population_size	12
fragments_per_compound_max	3
fragments_per_compound_min	3
use_scaffold	TRUE
dock_fitness	TRUE
num_dock_runs	3
dock_fitness_coefficient	3
clogp_fitness	FALSE
weight_fitness	FALSE
rotatable_bond_fitness	FALSE
h_bond_fitness	FALSE
breeding_fraction	0.5
keep_parents	TRUE
breeding_types	crossover, mutation
mutation_types	identity
mutation_probability	0.5
mutation_lambda	0.5

Cathepsin D

In a structure-based design collaboration, a peptidomimetic scaffold with three substituent positions was combined with a $10 \times 10 \times 10$ set of side chain groups in an effort to develop non-peptide inhibitors to Cathepsin D [35]. All 1000 resulting compounds were synthesized and tested experimentally for inhibition. Figure 3 shows the set of fragments and the scaffold. Our interest here was to see if we could determine, using the ADAPT program, which of the side chain groups, and at which positions, would produce the best inhibitors.

With current computational molecular docking methods, 1000 compounds is a small number to evaluate, and represents a small chemical search space. In order to create a larger, yet still practical search space, we let any of the 25 fragments be at any of the 3 scaffold attachment sites, giving us 15 625 unique compounds in our search space. We performed ten runs of the ADAPT program, each going through 50 generations, using DOCK as the only measure of compound fitness, and with the fragments grayscaled according to Daylight fingerprints. The values of other ADAPT parameters used in our experiments with cathepsin D are given in Table 1. By the 50th generation, the average fitness of the compounds within a generation was no longer improving rapidly (see Figure 4). Also, by the 50th generation, the number of unique compounds was much lower, indicating that we were converging on specific structures in our chemical search space (see Figure 5).

Figure 6 shows a histogram of the fragments occurring at each position in the final generations of the ten runs of ADAPT. At positions R1, R2, and R3, there were 8, 7, and 7 fragments, respectively, which occurred more than once in the final generations. A combinatorial library constructed by using the fragments and positions identified by ADAPT would result in a $8 \times 7 \times 7 = 392$ compound library. This library would contain four of the seven inhibitors identified experimentally at 100 nM in the synthesized 1000 compound library, including the single compound with the lowest inhibition constant. There could, in fact, be more inhibitors in the ADAPT based library, but experimental data exists for only 24 of the 392 compounds. Evaluated on the basis of just DOCK scores, the ADAPT based library has a better score distribution than the synthesized library (see Figure 7).

None of the 23 inhibitors identified in the synthesized library at 330 nM were found as individuals in the final generations of the ADAPT runs, however. This is not surprising, since the fitness space defined by the DOCK score does not mirror precisely the binding landscape of the cathepsin D active site, and the chemical space searched by ADAPT was roughly 15 times larger than the synthesized library. In fact, the known inhibitors, while still mostly in the top 15%

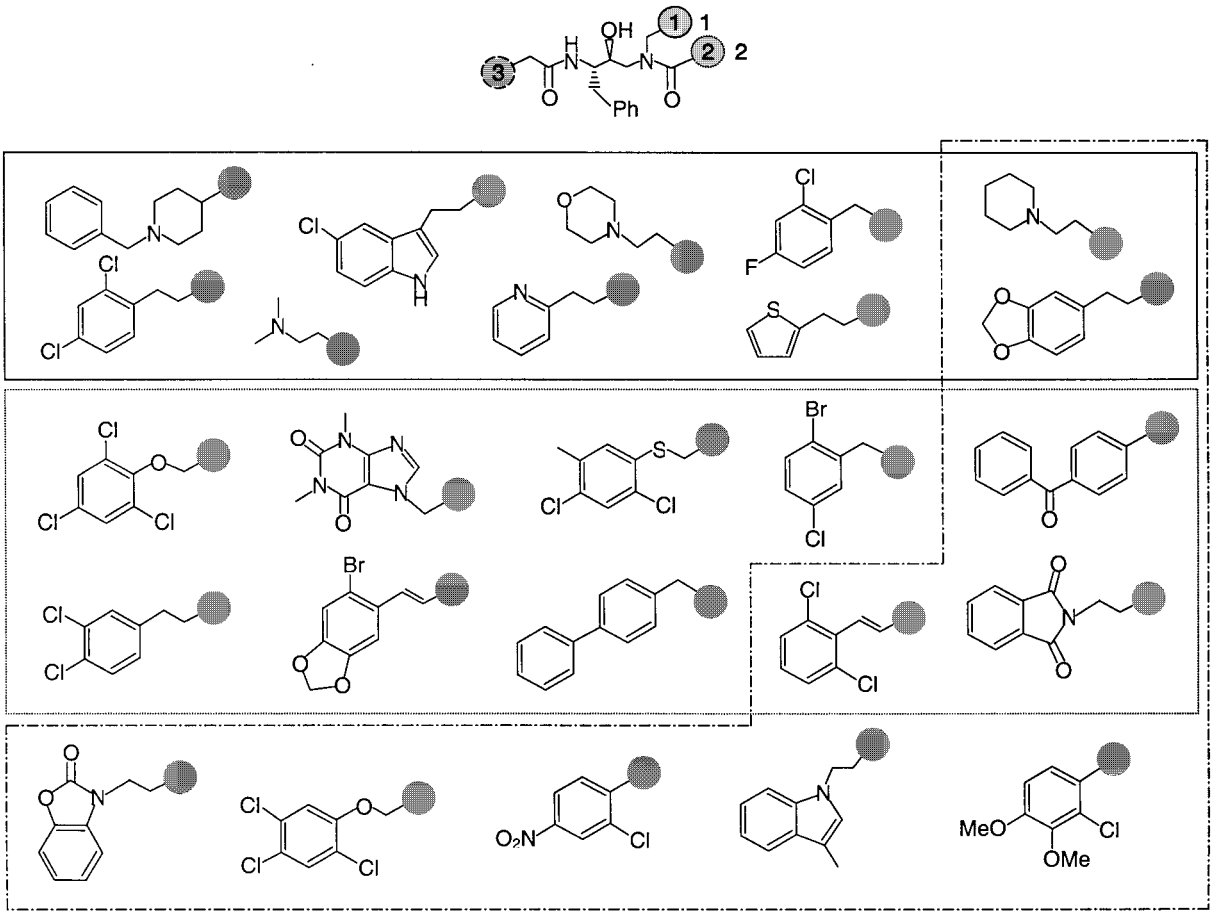

Figure 3. The fragment set used in the cathepsin D experiments. At the top is the scaffold with the three R groups labeled. The fragments bounded by the solid lines were tested experimentally at R1, those bounded by the dashed lines were tested experimentally at R2, and those bounded by the dashed and dotted line were tested experimentally at R3. Fragments in multiple regions were tested at two (or three) positions. In the ADAPT experiments, all 25 fragments were allowed to be at any of the 3 scaffold positions.

of the landscape have an overall worse DOCK score distribution than the ADAPT produced compounds, which are all within the top 1% (see Figure 8). Nevertheless, the DOCK score provided enough information to allow ADAPT to identify the fragments present in the majority of the best inhibitors. To investigate how well our genetic algorithm is sampling the space of possible compounds, we scored the fitness of all 15 625 possible compounds using DOCK. This allowed us to see the overall ranking of any compound generated by the genetic algorithm. Figure 9 compares the genetic algorithm with a random search in terms of the fractional ranking expected at random divided by the fractional ranking of the worst scoring survivor of a given population. This gives us a measure of 'enrichment' over a random search. We also ran our genetic algorithm ten times under the same conditions, but with a fragment set that was randomly ordered instead of grayscaled. While both the grayscaled and randomized fragment sets result in a positive enrichment over a random search, the grayscaling of the fragment set to force local sampling produced much better enrichment.

Overall, the ADAPT program was able to select the fragments (and their positions) that were consistently present in the best inhibitors determined experimentally. The program's inability to directly produce the known inhibitors in the final generations was likely due to their ranking in the binding landscape being slightly different than their ranking in the landscape defined by DOCK scores. Our experiments with the cathepsin D system also show that forcing local sampling by allowing fragments to mutate only to other

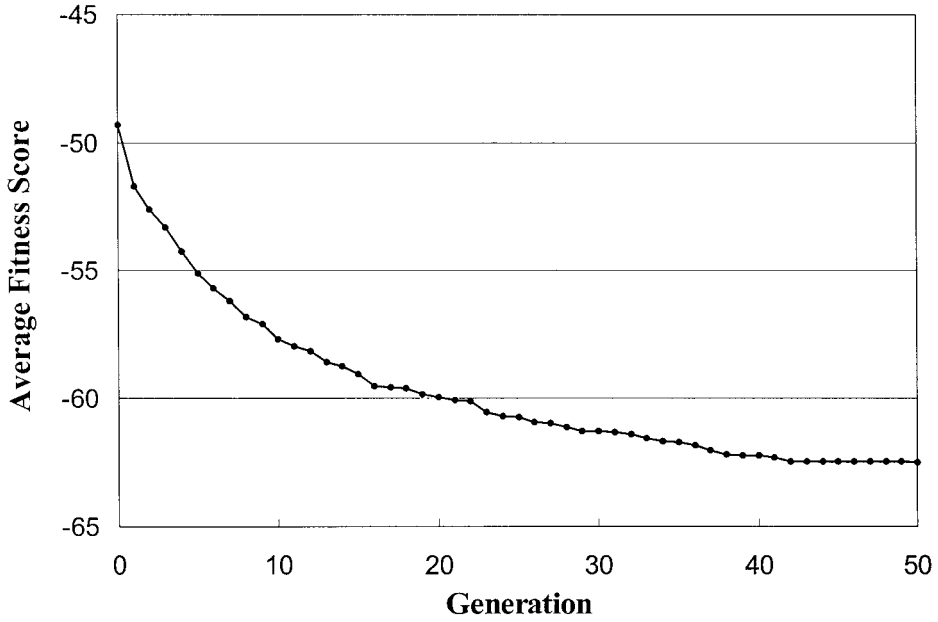

Figure 4. Improvement of the fitness score in the cathepsin D experiments. The values are the fitness scores of the surviving compounds, averaged over all ten runs of the ADAPT program.

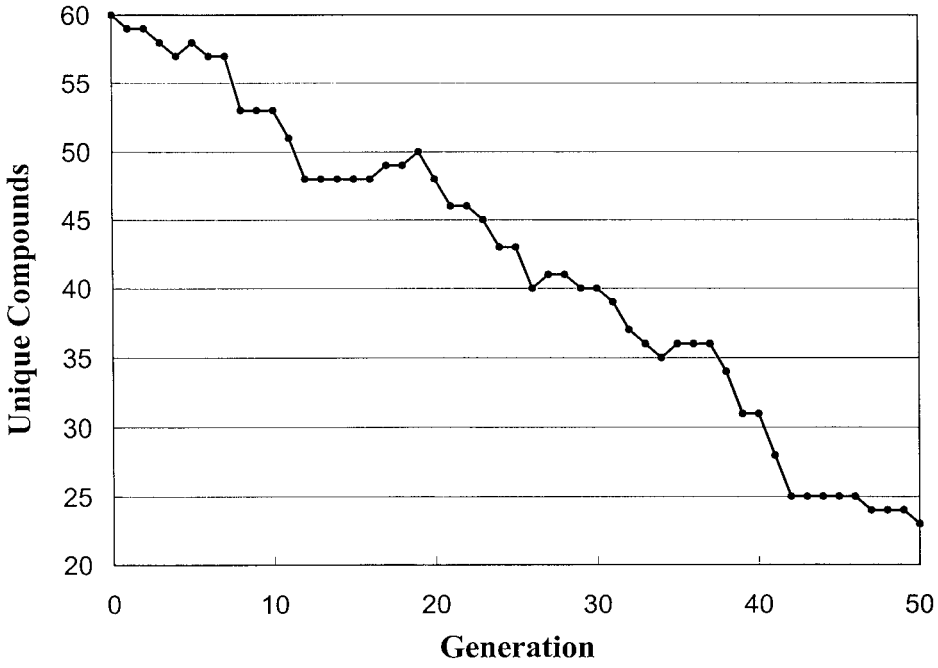

Figure 5. Number of unique compounds per generation in the cathepsin D experiments. The values represent the number of unique survivors across all ten runs of the ADAPT program.

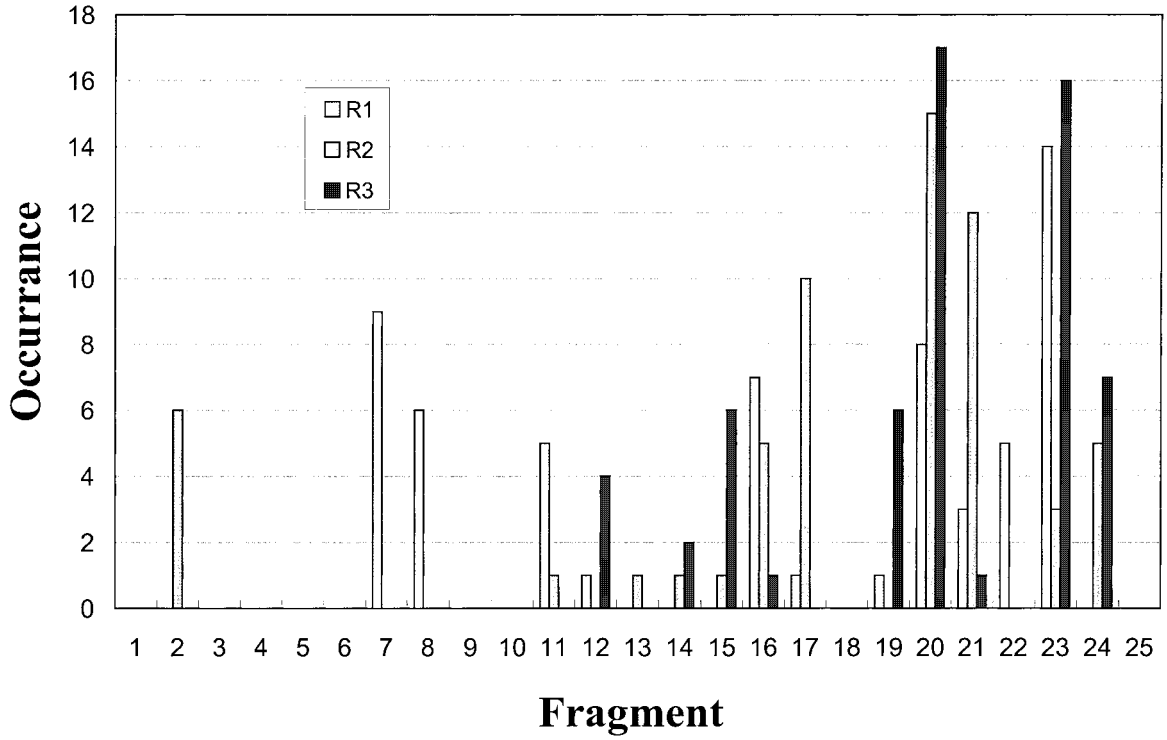

Figure 6. Fragments observed in the final generations of the ten ADAPT runs in the cathepsin D experiments.

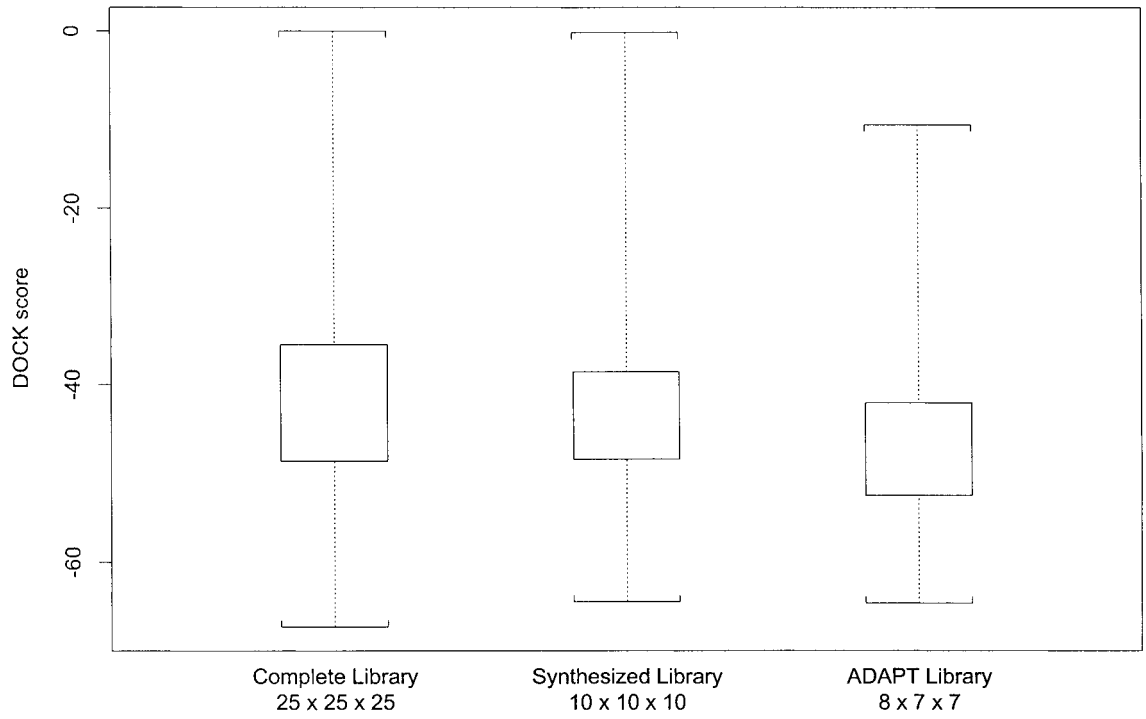

Figure 7. The distributions of rank (defined by DOCK score) of the complete set of possible compounds ($25 \times 25 \times 25$), the synthesized compound library ($10 \times 10 \times 10$), and the library based on fragment identified by ADAPT ($8 \times 7 \times 7$). The distance between the top and bottom of the shaded box represents the inter-quartile distance, the dark line represents the median, and the brackets represent the extreme values of the distribution.

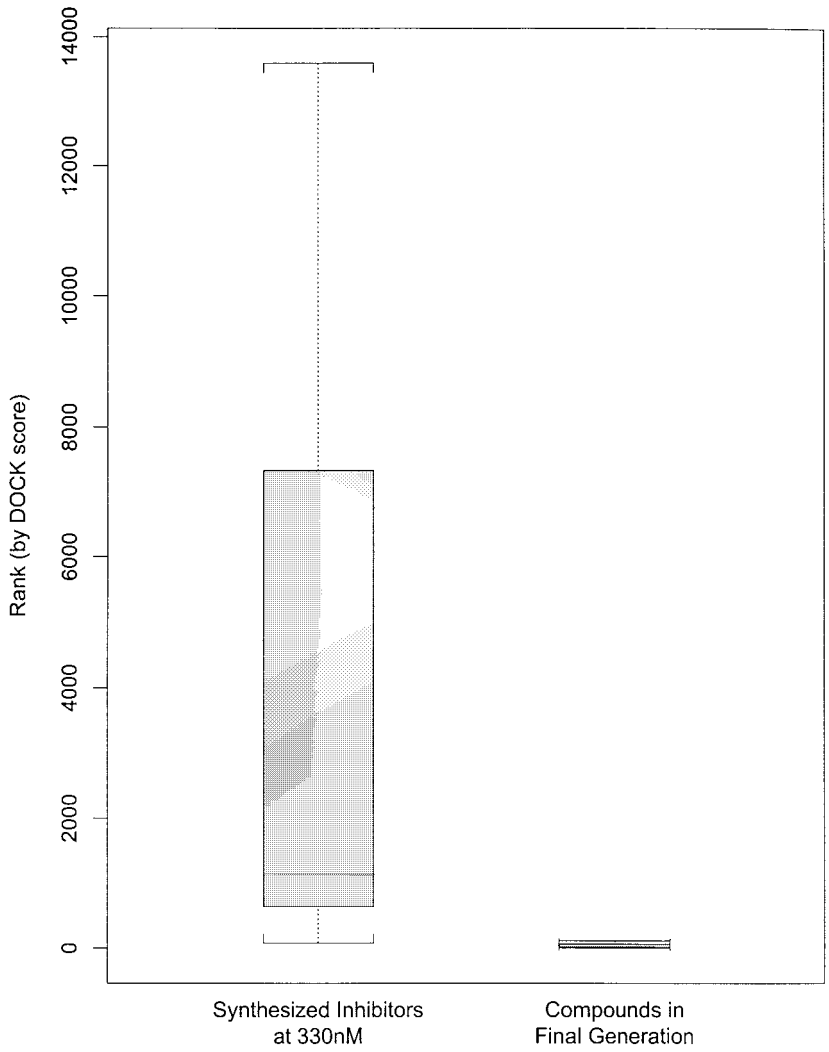

Figure 8. The distributions of rank (defined by DOCK score) of the 23 experimentally determined inhibitors at 330 nM, and the 23 compounds from the final generations of the ten ADAPT runs.

fragments which are chemically similar produces a more efficient search for high scoring compounds.

Dihydrofolate reductase

The enzyme dihydrofolate reductase (DHFR) with the small molecule inhibitor methotrexate represents a well-studied system in structure based drug design [36–38]. We applied the ADAPT program to this system to investigate three topics: how well DOCK would guide the construction of compounds in a larger search space, the effects of seeding the initial population with a known ligand, and the effects of adding diversity to the population during longer runs of the program. The fragment set consisted of 13 fragments which make up methotrexate, logical modifications to those fragments (e.g. replacement of a nitrogen in the pteridine ring with a carbon atom), and some basic 'building blocks', fragments such as single carbon and nitrogen atoms. Figure 10 shows methotrexate and some of the fragments from the fragment set. Included in the set are building blocks general enough to allow virtually any acyclic hydrocarbon to be constructed. With 32 fragments to choose from and compounds allowed to have between 3 and 13 fragments, a lower bound for the size of the chemical space is approximately 3.5×10^8 unique compounds.

We performed three sets of ten runs each with ADAPT, using DOCK as the primary fitness function and a mild weight restriction to compensate for

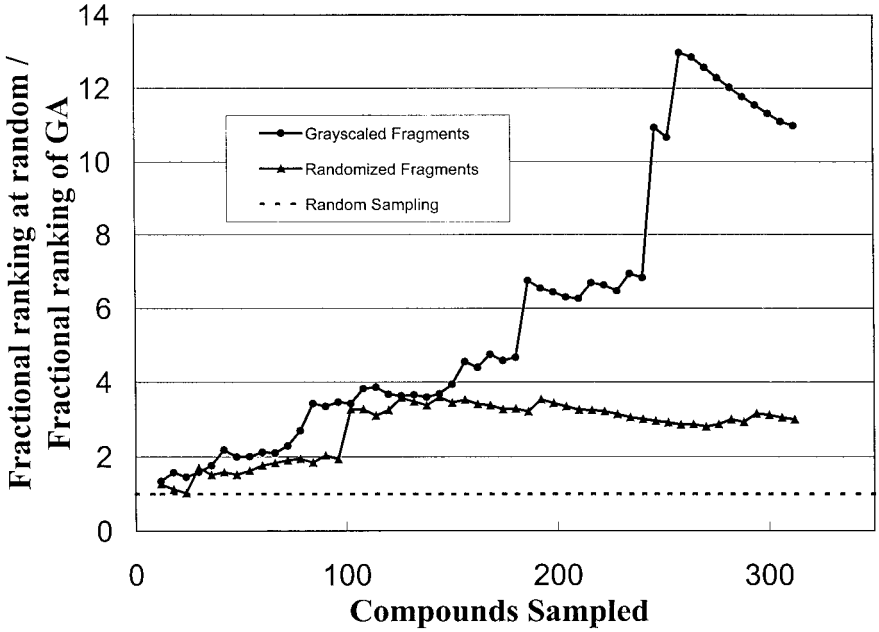

Figure 9. Enrichment of sampling using the ADAPT program over sampling at random. The enrichment is calculated by dividing the fractional ranking expected via random sampling (in this case sampling 6 at a time without replacement and taking the rank of the worst scoring individual) by the fractional ranking of the worst scoring survivor of each generation, averaged over ten runs.

the DOCK score function's bias towards large compounds. Table 2 lists the ADAPT parameter values which remained identical across all runs. For the first set, the initial generation was based on methotrexate and the ADAPT program was forced to run a minimum of 30 generations before stopping based on a lack of improvement in the fitness score. For the second and third sets, the initial generation was built randomly but the second set was forced to run a minimum of 30 generations while the third set was forced to run exactly 100 generations. Each run used a population size of 20, of which half were kept as survivors in each generation, and used both crossover and mutation in breeding.

Figure 11 shows the distributions of the DOCK scores for the compounds in the final generations of each of the three sets of runs, plus a distribution of the DOCK scores of (unoptimized) randomly assembled fragments. All three ADAPT runs create compounds that score much better than random, with the runs seeded with methotrexate having both the narrowest and best scoring distribution. No compounds in the final generations of the unseeded runs to 30 generations had DOCK scores better than methotrexate while 94% of the compounds in the final generations of the seeded runs had scores better than methotrexate. However, 28% of the compounds in the final generations of the unseeded runs which were run to 100 generations had DOCK scores better than methotrexate.

Despite generating a wide variety of structures, our genetic algorithm identified a general motif present in known DHFR inhibitors. There were 98 unique structures among the survivors in the final generations of the seeded runs, a selection of which is shown in Figure 12a, along with their DOCK scores. All of the structures had an essentially chain-like structure with a fused ring fragment at one end. Of the 98 compounds, all but two retained the pteridine ring fragment at one end, while the chain extending from it varied widely. In the second set of runs, in which the initial generations were not based on methotrexate, there was still a large bias towards a fused ring structure at one end and a flexible chain coming off of it, but the pteridine ring fragment was not seen in a high fraction of the compounds in the final generations (21/100). Figure 12b shows a selection of good scoring structures generated by the second set of runs. The DOCK scores of these compounds were all poorer than that of methotrexate and the compounds generated by the first run. These results are in accordance with work by Verkhivker et al. [39] who demonstrated that the DHFR/methotrexate binding landscape would accept mutations to the end of the chain away from the pteridine fragment, and would accept virtually no changes

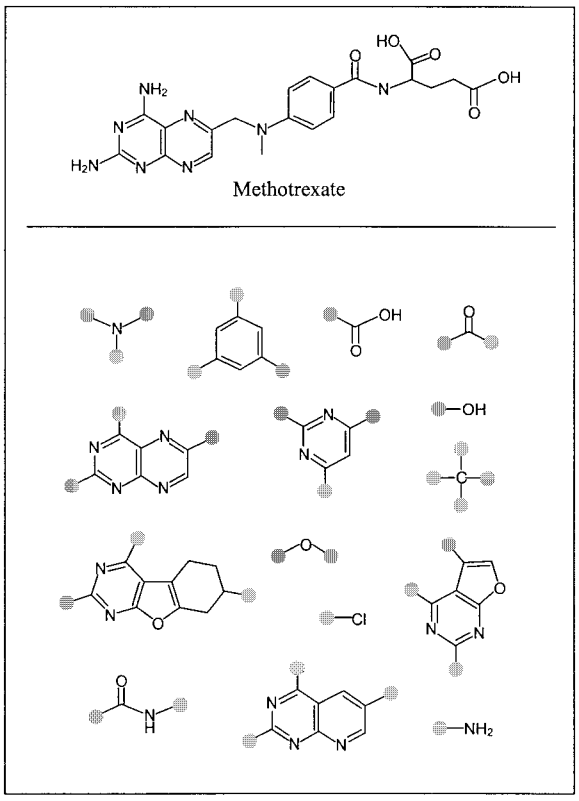

Figure 10. The known DHFR inhibitor methotrexate and fragments from the ADAPT runs. Fragments shown represent enough fragments to reconstruct methotrexate, as well as diverse selection of general fragments that were added to the set. The gray circles represent attachment points.

in the pteridine structure. In the set of longer unseeded runs, compounds were generated with DOCK scores better than methotrexate, along with a higher occurrence of the pteridine fragment (56/100). Figure 12c shows a selection of compounds generated from the third set of runs. The basic 'fused-ring plus flexible chain' motif common to known DHFR inhibitors was present in the runs seeded with the methotrexate structure, and ADAPT stuck with this motif in the compounds it subsequently generated. This motif was not present in the starting generations of the unseeded runs, yet was also observed, to a lesser extent, in the final populations. Many more generations of the unseeded runs were required for the motif to appear, however.

To investigate the effects of adding diversity to the population during a long run of the program, we ran the ADAPT program for 1000 generations, adding diversity back to the population every 200 generations. We compared this to doing 5 separate runs of 200 generations each. Figure 13 shows the fitness score distribution of the 5 final generations of the individual runs and the fitness score distribution of the 5 generations just before diversity was added to the population. The distribution of compounds at the generations just before diversity was added has a better average fitness score and best score.

Table 2. ADAPT parameter settings used in the dihydrofolate reductase experiments.

Parameter	Value
population_size	20
fragments_per_compound_max	13
fragments_per_compound_min	3
use_scaffold	FALSE
dock_fitness	TRUE
num_dock_runs	3
dock_fitness_coefficient	1.0
clogp_fitness	FALSE
weight_fitness	TRUE
weight_range	410 to 450
weight_std_dev	60
weight_fitness_coefficient	−10.0
rotatable_bond_fitness	FALSE
h_bond_fitness	FALSE
breeding_fraction	0.5
keep_parents	TRUE
breeding_types	crossover, mutation
mutation_types	connectivity, identity
mutation_probability	0.5
mutation_lambda	1.0

The results of our experiments using the DHFR/methotrexate system show that the ADAPT program can evolve compounds with the basic motif of known ligands. However, in a search for compounds with better fitness than a known ligand, using the ligand to seed the initial population results in the faster production of better scoring, yet still diverse compounds. Compounds with fitness scores better than the known ligand can still be produced by ADAPT, but it requires runs with significantly more generations. Our results also show that when performing long runs in search of good scoring compounds, it can be more efficient to add diversity to a long run instead of combining multiple runs.

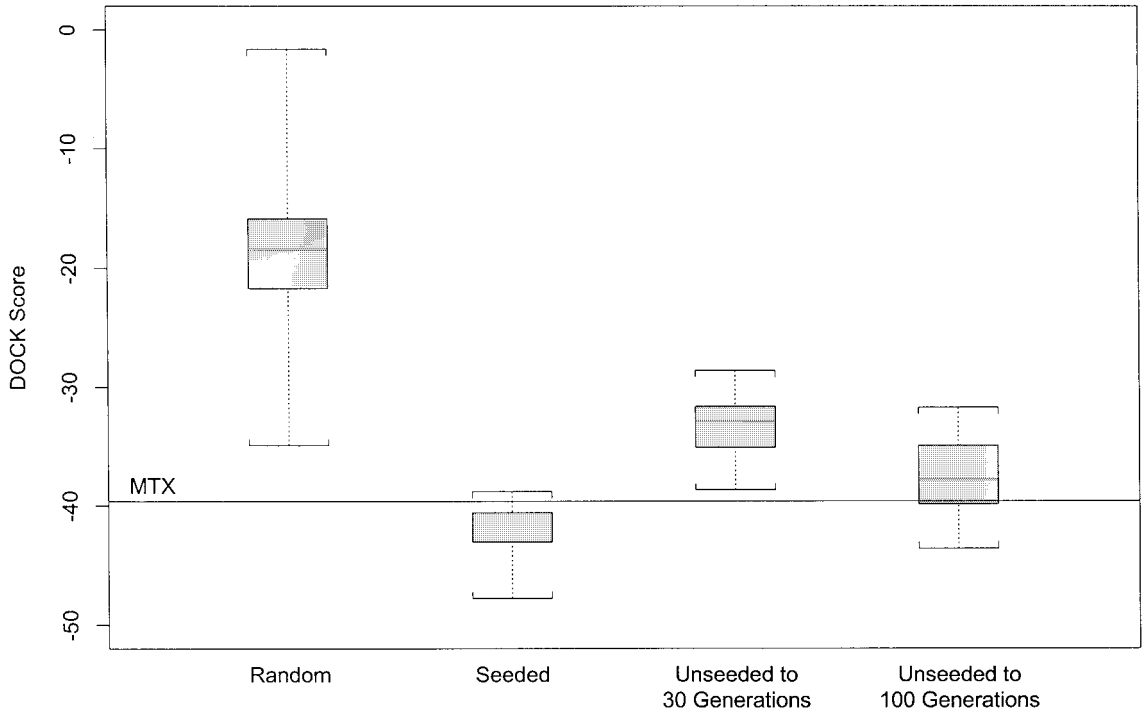

Figure 11. The distributions of DOCK scores for the compounds in the final generations of the ADAPT runs and of randomly assembled compounds. The distance between the top and bottom of the shaded box represents the inter-quartile distance, the dark line represents the median, and the brackets represent the extreme values of the distribution. The random distribution represents 1000 randomly assembled compounds, while the other three represent approximately 100 compounds (10 from each of 10 runs). The horizontal line shows the DOCK score of methotrexate (−39.61).

HIV-1 reverse transcriptase

The non-nucleoside inhibitors (NNI) of HIV Reverse Transcriptase (HIV-RT) are a chemically diverse class of compounds that non-competitively inhibit enzymatic activity by binding to a specific site on the enzyme. Previous work indicates that a vast majority of these inhibitors adopt a 'butterfly-like' geometry when bound in the NNI pocket [40, 41]. The axis of the butterfly's body is aligned almost parallel to the β9-β10 hairpin. Wing I is defined as the portion of the NNI binding site that lies closest to the polymerase active site, while the section that lies distant from the active site is wing II [41]. We applied the ADAPT program to the HIV-RT system in order to evaluate its ability to reproduce scaffolds with 'butterfly-like' shapes in their docked geometries, as well as its ability to rediscover known scaffolds.

We chose a model of HIV-1 RT complexed with 1-[(2-hyrdoxyethoxy)methyl]-6-(phyenylthio)thymine (HEPT) (pdb identifier 1RTI) as the target for our application. In this model HEPT adopts butterfly geometry in its bound state. The NNI binding site was characterized using *sphgen* and GRID as previously described [32]. Twenty-six chemically diverse non-nucleoside inhibitors were fragmented into rigid segments in order to form a library of 65 chemical fragments which ADAPT used to build compounds. With 65 fragments and 4 to 12 fragments per compound, a lower bound for the number of compounds that can be constructed using this library is approximately 5×10^{12}. Table 3 lists all the NNI's used to create the fragment library. Figure 14 shows how Thiocarboxyanilide UC781 was fragmented into its rigid fragments. We performed ten runs of ADAPT, using an unseeded population size of 30. The DOCK score and a loose molecular weight constraint based on the range of weights of known NNIs (214 to 415) were used as the fitness function. Table 4 lists the other ADAPT parameter values used. After ten runs a total of 5250 compounds had been generated.

In order to characterize the 'butterfly' geometry criterion, five other HIV-RT models complexed with different non-nucleoside inhibitors that adopt the butterfly shape were superimposed onto the 1RTI NNI binding site containing HEPT. The MidasPlus

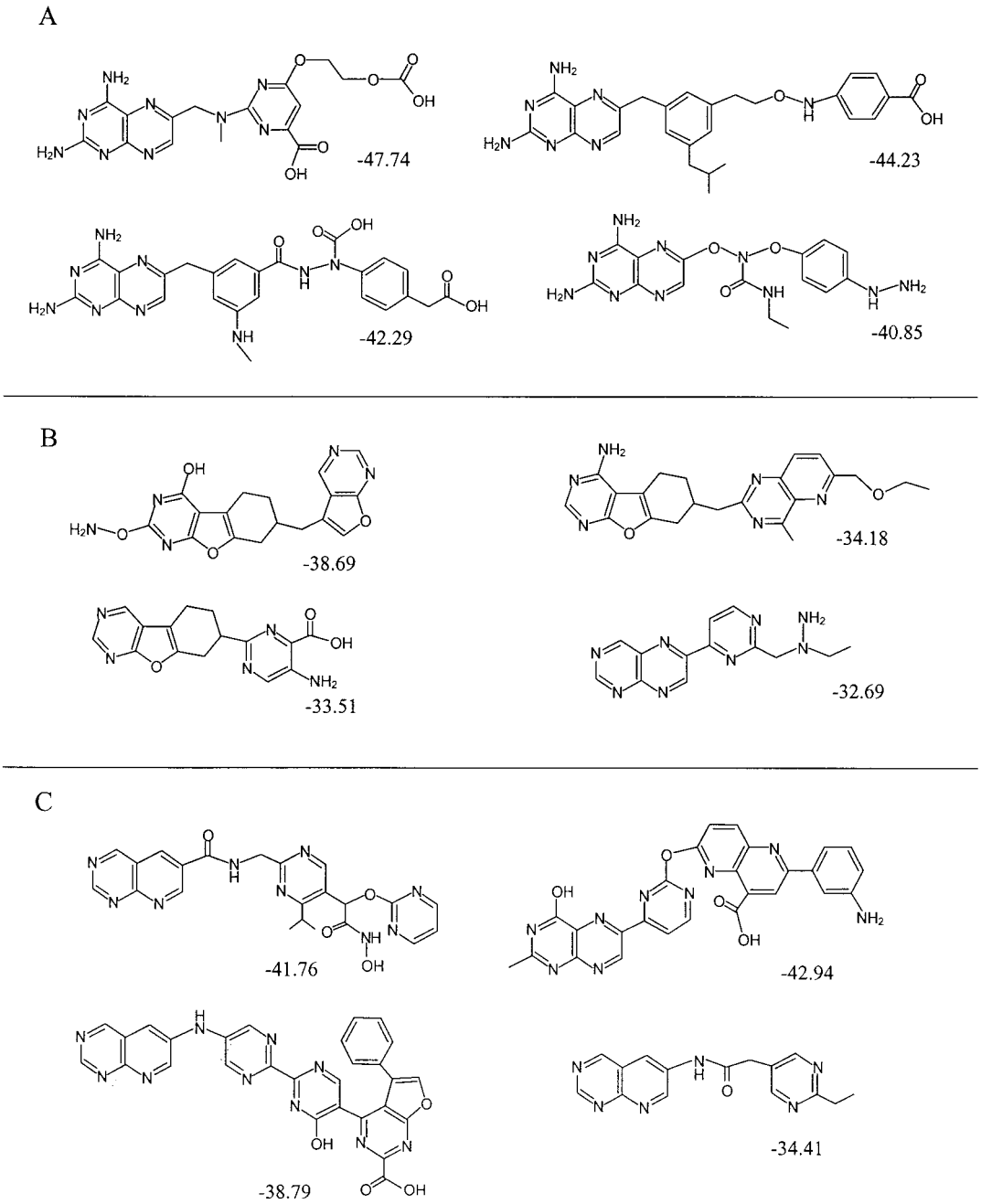

Figure 12. Representative structures and their DOCK scores from the final generations of the three sets of ADAPT runs in the DHFR experiments. (A) Compounds from the runs seeded with methotrexate. (B) Compounds from the unseeded runs to 30 generations. (C) Compounds from the unseeded runs to 100 generations.

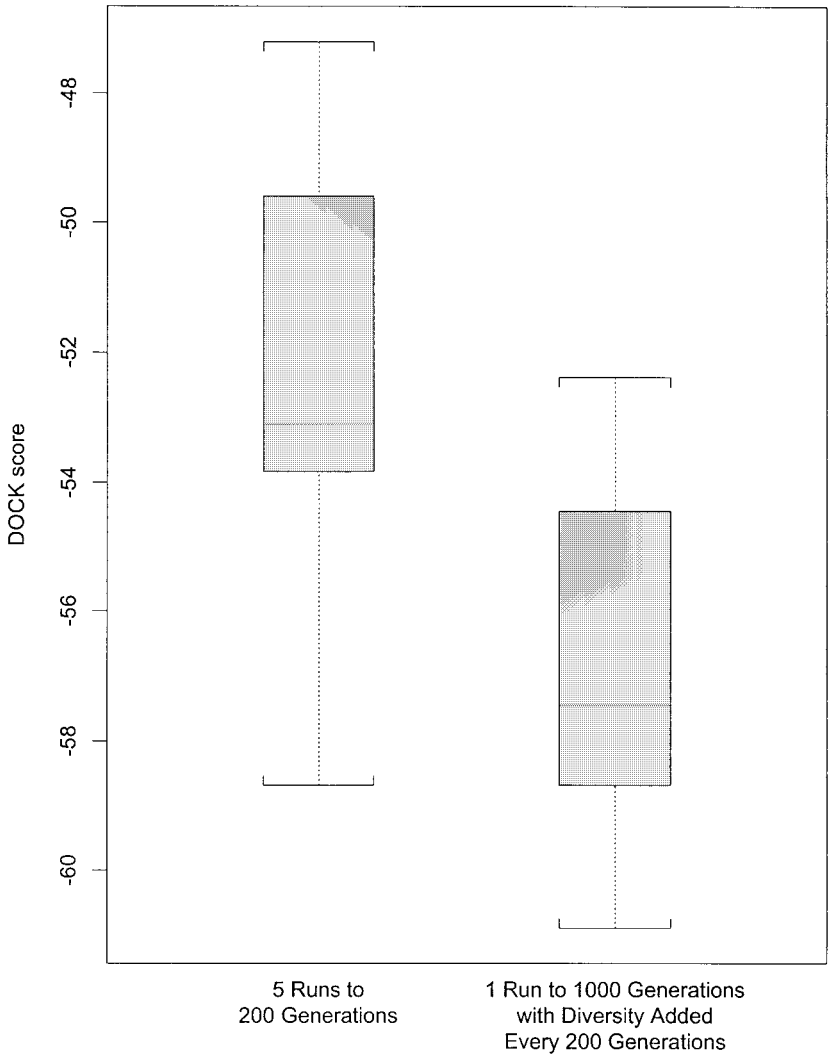

Figure 13. The distributions of DOCK scores for the compounds from the final generations of 5 runs to 200 generations and the compounds from the generations just before diversity was re-introduced in a run to 1000 generations, adding diversity every 200 generations. Both distributions represent 100 compounds. The distance between the top and bottom of the shaded box represents the inter-quartile distance, the dark line represents the median, and the brackets represent the extreme values of the distribution.

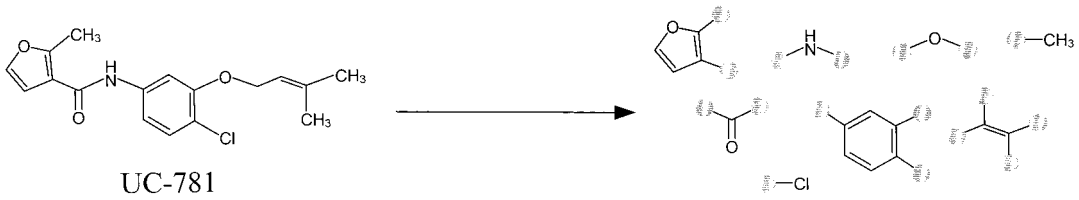

Figure 14. Decomposition of UC-781 into rigid fragments. A gray circle indicates an attachment point.

Table 3. The 26 non-nucleoside inhibitors used to create the HIV-RT fragment library.

TNK-651	Thiadozyl dialkylcarbamate RD 4-2024
α APA	Arylpyridothiodiazepine MEN 10979
8-Chloro TIBO (Trivirapine)	DABO 12e
BHAP U90152 (Delavirdine)	HEPT Pyridinone Hybrid 8a
Pyridinone L-697, 661	Benzyloxymethylpyridinone 18a
Thiocarboxyanilide UC781	Alkyl(arylthio)uracil
Indole carboxamide L-737,126	Phenylthiazolylthiourea (PETT)
Benzathiodiazine-1-oxide NSC 287474	Diarylsulfone NPPS
Quinaxolinone 13a	Pyrrl Aryl Sulfone
Benzoxazinone DMP 266 (Sustiva)	Highly Substituted Pyrrole
Pyrrolobenzoxazepinone 4	I-EBU (MKC-442)
Imidazopyridazine 33	Imidazodipyridodiazepine UK-129,485

Table 4. ADAPT parameter settings used in the HIV-1 reverse transcriptase experiments.

Parameter	Value
population_size	30
fragments_per_compound_max	12
fragments_per_compound_min	3
use_scaffold	FALSE
dock_fitness	TRUE
num_dock_runs	1
dock_fitness_coefficient	1.0
clogp_fitness	FALSE
weight_fitness	TRUE
weight_range	214 to 415
weight_std_dev	100
weight_fitness_coefficient	−10.0
rotatable_bond_fitness	FALSE
h_bond_fitness	FALSE
breeding_fraction	0.5
keep_parents	TRUE
breeding_types	crossover, mutation
mutation_types	connectivity, identity
mutation_probability	0.5
mutation_lambda	1.0

[42] program was used to superimpose 1BQM (HIV-1 RT complexed with Quinoxaline), 1DTT (HIV-1 RT complexed with a phenylethyl thioureathiazole (PETT) derivative), 1HNI (HIV-1 RT complexed with an α-APA derivative), 1REV (HIV-1 RT complexed with 8-Choloro TIBO), and 1VRT (HIV-1 RT complexed with Nevirapine) into 1RTI's reference frame. Thirty-one residues (G93a-P97a, L100a-L101a, L103a, S105a-T107a, V179a-Y183a, Y188a-S191a, P225a-W229a, G131a-D237a, Y318b) immediately surrounding the NNI binding site served as the primary reference for orienting all the models into the correct reference frame. The root-mean-squared-deviations for the 31 residues in 1BQM, 1DTT, 1HNI, 1REV, and 1VRT to 1RTI are 2.86, 1.85, 2.33, 1.56, and 1.78 Å, respectively. Once superimposed, each atom in all six inhibitors was classified as belonging to either wing I or wing II of the butterfly shape. We performed the classification by dividing the butterfly along a plane containing N11 of Nevirapine, S′ of HEPT, and the Cδ of L100a, and partitioning the heavy atoms of the ligands into each wing. Figure 15 is a 2D projection of the atoms onto a plane perpendicular to that used to partition the heavy atoms. The origin represents an end-on view of a vector connecting the N11 of Nevirapine and the S′ atom of HEPT. The angle between best fit lines through all the wings in each atom is 138°. In order to calculate the angle between the wings of each ligand, the positions of all atoms in each wing are averaged to yield a 'wing-point' description of the wing's size position. The angle between the wings is represented by the angle between the line connecting the wing-point in wing I to the origin, and the line connecting the origin to the wing-point in wing II. The angles for the six known non-nucleoside inhibitors range from 105°–167° (HEPT: 141°, Quinoxaline: 106°, PETT: 149°, α-APA: 157°, 8-Chloro TIBO: 167°, Nevirapine: 133°). In addition to facilitating angle calculations, the distance of each wing-point from the origin can be considered a measure of the relative size of each wing.

We analyzed the compounds generated by the ADAPT program in order to determine how well the compounds fulfill the butterfly geometry. To count as

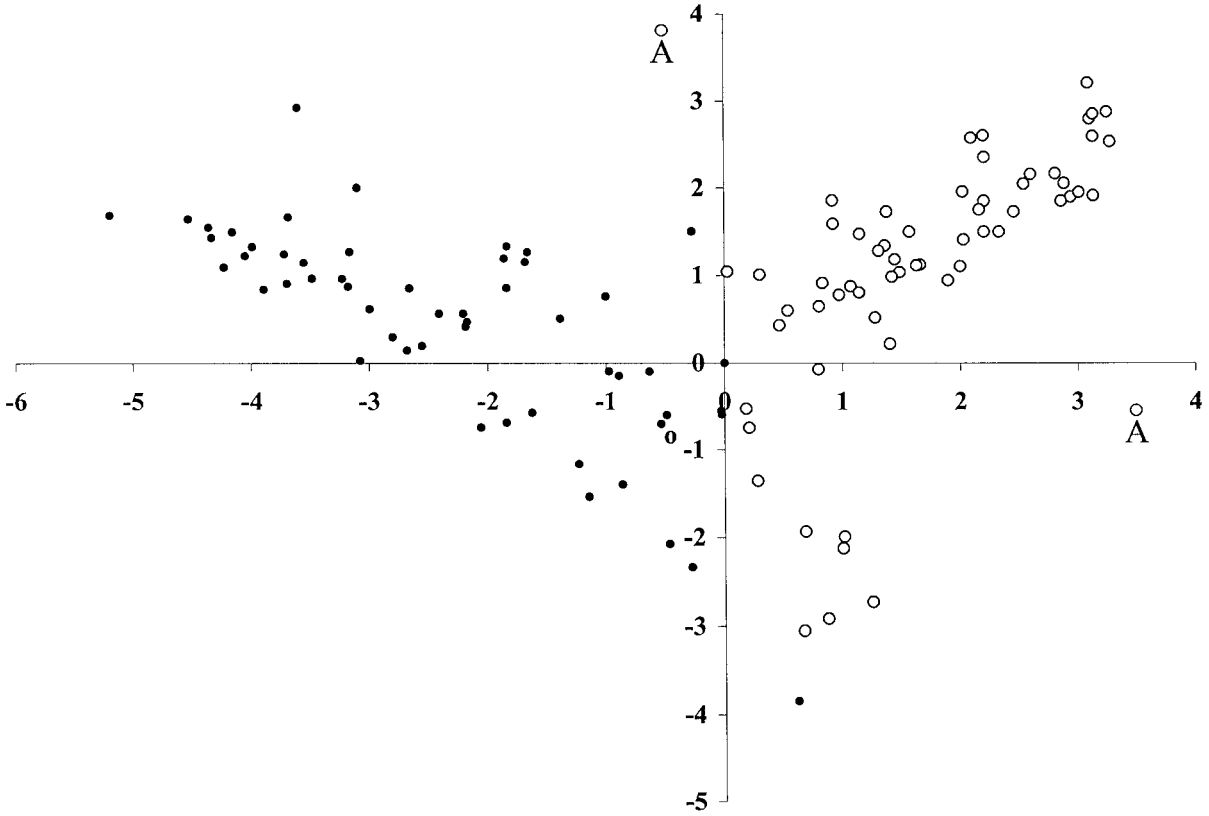

Figure 15. 2D projection of heavy atoms from HEPT, Quinoxaline, PETT, a-APA, Nevirapine, and 8-Chloro TIBO bound into the NNI binding site. Filled in circles (●) are atoms that occupy the wing I section, and empty circles (◯) occupy wing II. The origin represents a vector between N11 of Nevirapine and S1 of 8-Cholor TIBO. The angle between linear best fit lines through each all the atoms in each wing is 138°.

a butterfly compound, 50% of a ligand's atoms must fill both wings; this was done to eliminate those compounds that fill only one of the wings of the NNI binding site. 3363 of the 5250 (64.1%) compounds fulfilled this criteria. The atoms in the six known NNI's examined are all found within 3.5 Å of the plane used to divide the NNI binding site. Further, these same atoms can all be found within 2 Å of a best fit plane through the atoms in each wing. We used these distance criteria to assign the atoms in each of the 3363 compounds as belonging to either wing I or wing II of the NNI binding site. These atoms were used to calculate the wing-point representing the atoms of each wing in each compound. These points were projected onto the 2D plane perpendicular to the partition between the wings. The angle between the lines connecting each wing-point to the origin represents the ensemble average angle between the atoms in each wing of the compound.

Figure 16 is a 2D projection of all the wing-points; as before, the origin represents an end-on view of the vector connecting the N11 of Nevirapine and the S′ of HEPT. Each point represents the average coordinate of a ligand's wing. Lines drawn through each wing intersect at an angle of 120°. This value agrees with the angle measured previously by Ding et al. [41] Figure 17 is a distribution of the angles in Figure 16. The distribution has a mean of 142° and a standard deviation of 21°. Figure 18 is a plot of the angles between the wings as viewed against the distances of the wing points from the origin. The distance of a wing-point from the origin is a rough measure of a wing's size. The wings of a non-nucleoside inhibitor can adopt a range of angles, however the data shows that the angle between the wings decreases as the average wing size increases. This is to be expected since the geometric constraints of the binding site on the non-nucleoside inhibitors are more pronounced when the compounds occupy more space within the site.

The ADAPT program was able to reproduce four known NNI scaffolds. Effavirenz (Sustiva^{TM}), Pyrrolobenzodiazepinone, PETT, Dyarryl Sulfone-

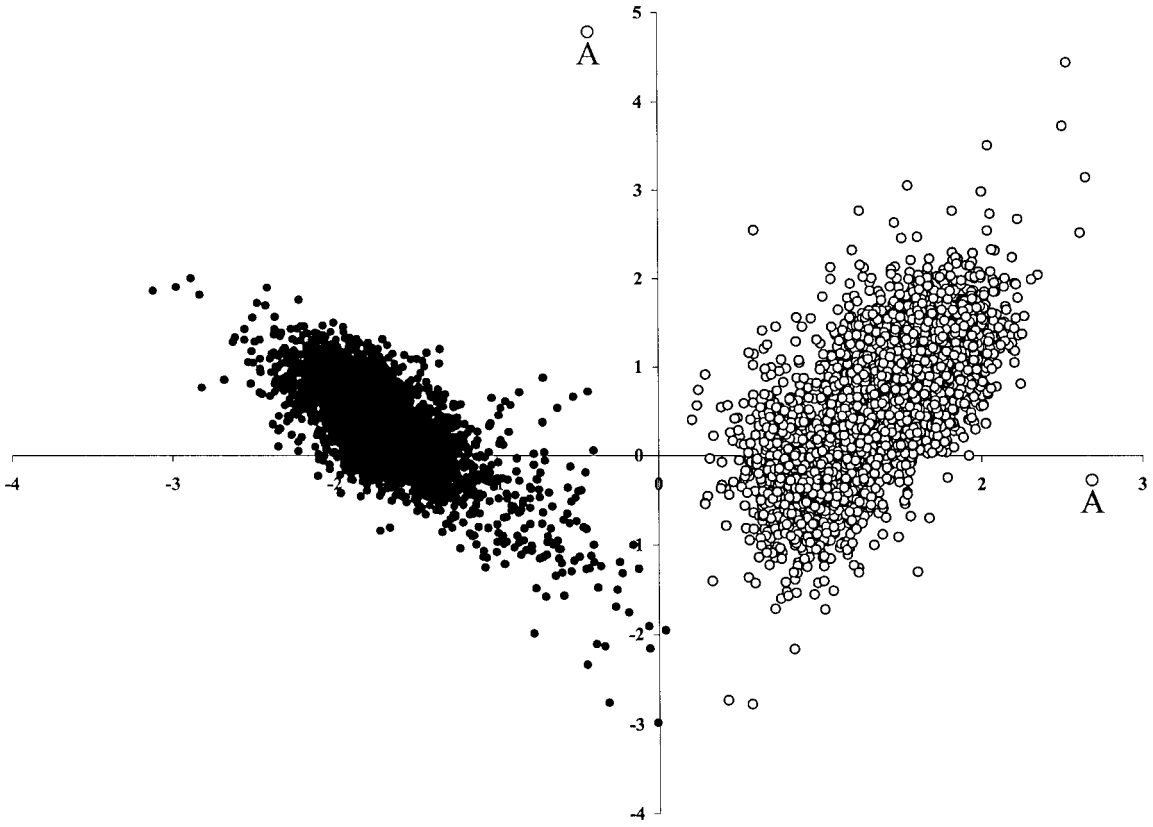

Figure 16. 2D projection of the wings of all ligands that fulfill have a butterfly geometry. Filled circles (●) are wing-points that occupy the wing I section, and empty circles (◯) occupy wing II The angle between the best fit lines through each wing is 120°.

like scaffolds were found among the butterfly like compounds (Figure 19). Of particular note is the PETT-like compound generated by the algorithm. Its real-world counterpart, MSC-127, was the result of an extensive synthetic chemistry program that explored over 750 variations of the lead PETT scaffold [43]. Despite the lack of a structural motif in the initial, unseeded populations, the ADAPT program was able to reproduce a geometric constraint, the 'butterfly' motif of known NNI's from the use of a molecular docking fitness function. We also had some modest success in reproducing known NNI scaffolds.

Discussion

We have developed the ADAPT program, a genetic algorithm for structure-based de novo design which uses molecular docking as a fitness function. We discuss here the utility of design algorithms which use a binding estimate as a fitness function compared to the prior practice of a fitness function based on analysis of known ligands. We then consider the benefits and hazards of forced local sampling with the ADAPT program, and the difficulty in measuring the success of de novo design algorithms.

The ADAPT program shares the advantages of other genetic algorithms for de novo design in that it can optimize under many constraints simultaneously and provide a rich set of diverse solutions, however the use of molecular docking as a fitness function introduces both particular advantages and disadvantages. The major advantage is that information about known ligands to a particular active site is not required. This is an increasingly important advantage as genomic sequencing projects and post-genomic analysis produce a growing number of new protein targets for which no binding information is known. Another feature of avoiding the use of known ligands is an increased ability to discover more diverse compounds, a clear advantage in situations where there may be competing patent issues, or where it is desirable to explore backup candidates of different chemical classes. Information

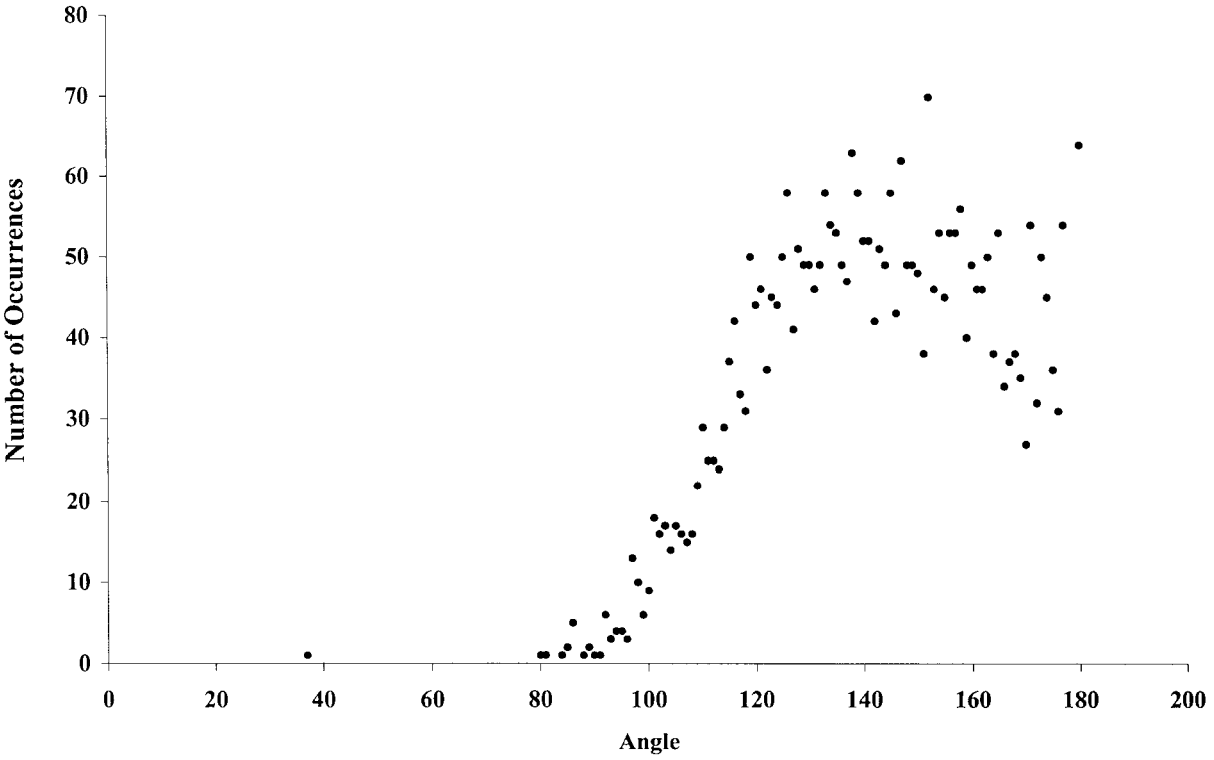

Figure 17. Distribution of angles between the wings of 3363 compounds generated by the genetic algorithm that fulfill the butterfly geometry. The distribution has a mean of 142° and a standard deviation of 21°.

about known ligands can, of course, be useful. The flexibility of ADAPT allows us to explore the value of this additional information, as in our experiments with DHFR, where compounds based on methotrexate were readily generated.

The major disadvantages to using molecular docking fitness functions are that the calculations can be fairly slow, especially when they include ligand flexibility, and that any errors in the scoring function are propagated into the compound generation process. The computational overhead forces the population sizes of the ADAPT program to be relatively small and thus may cause us to loose some sampling ability. The genetic algorithm of Sheridan et. al. opts for a larger population and fewer generations. The optimal settings of parameters such as population size remains an area of open research in genetic algorithm theory, with some arguing for high population sizes [44] and others for small populations, especially when the cost of fitness evaluation is high [45, 46]. As with most genetic algorithms, ADAPT spends the vast majority of its time evaluating the fitness function. Even when docking with relatively low orientational sampling, it took approximately 5 hours to complete 100 generations with a population of size 20 on a Silicon Graphics R10000 processor (the housekeeping of the genetic algorithm, including crossover and mutation, took less than 10 s per generation). Scoring errors from the docking calculations may neglect certain important molecular interactions (e.g. hydrogen bonding, desolvation [47, 48]) and produce compounds that 'score' well, but do not necessarily have good experimental binding affinities. Genetic algorithms based on pharmacophore and template models suffer from essentially the same problem, however. Fortunately, the ease of programming genetic algorithms means that as new scoring functions become available, they can be readily implemented and tested.

Fitness functions using molecular docking calculations can be designed to take advantage of information not readily available in fitness functions based on pharmacophore or molecular templates. One can, for example, build a fitness function which advocates specificity by rewarding compounds which dock well into the active site of one target while penalizing compounds which dock well into the active site of another target. One could also search for compounds which

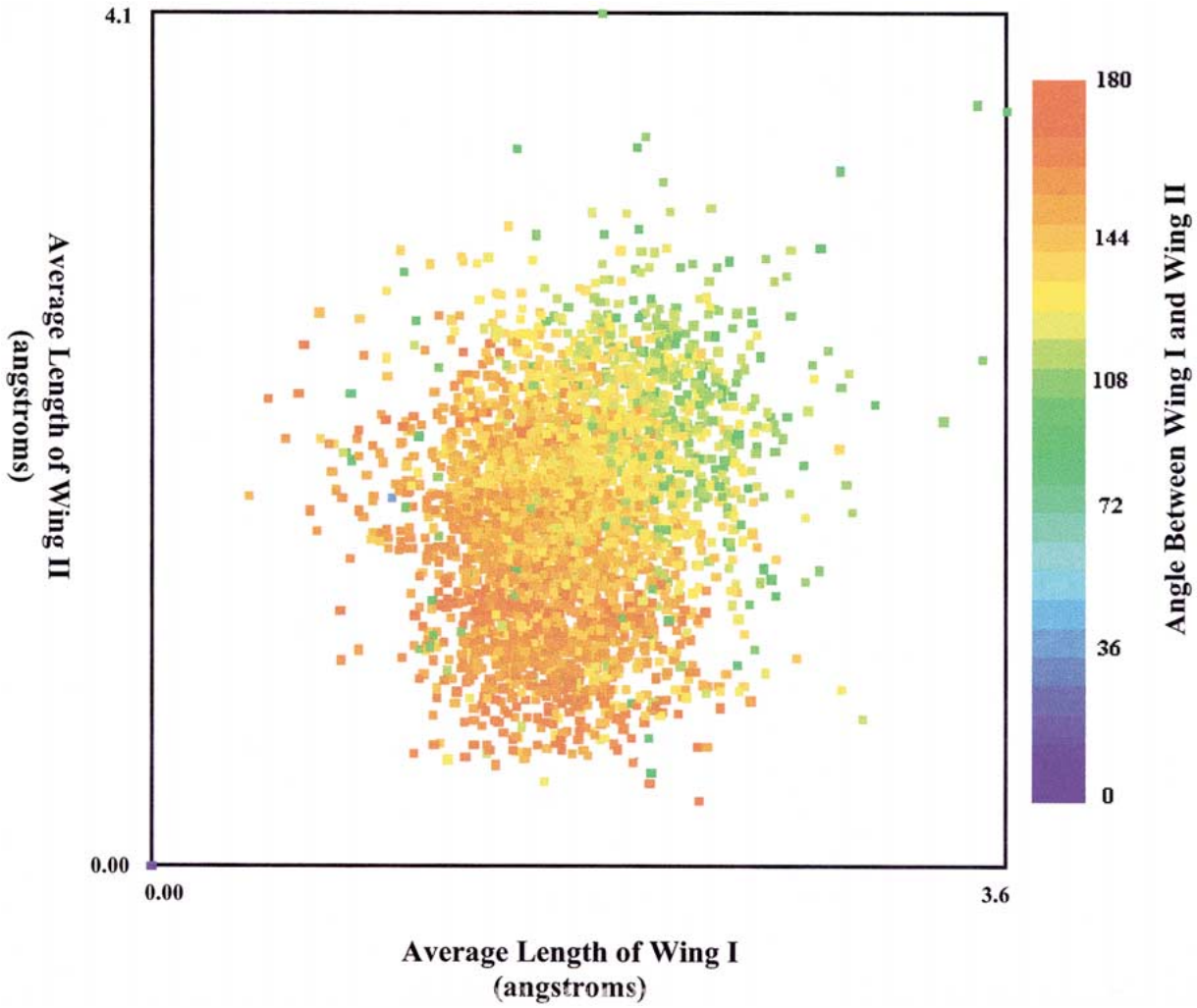

Figure 18. Angles between wing I and wing II as viewed against the distance of each wing-point from the origin. Distances from the origin provide a rough measure of a wing's size. As size of the wings increase, they are forced to adopt smaller angles.

dock well to entire families of targets by favoring only compounds which bind to multiple active sites.

Another advantage of the ADAPT program is the ability to let the user specify not only the specific fragment set used to build the compounds, but also the fragment attachment points. This feature allows the user a fair amount of control over the possible chemistry and can help avoid the generation of synthetically difficult compounds. Combinatorial chemistry schemes, such as that used in our experiments with cathepsin D, are a clear example of this utility.

Our experiments with the cathepsin D system also show the advantage of forcing a genetic algorithm to sample locally in chemical space. This makes sense, because it leads to focused sampling near optimal solutions. We enforced local sampling by only allowing mutations of fragment types to fragments similar in chemistry. Our implementation of this is somewhat crude, however, and consideration of alternate strategies such as binning fragments such that fragments can only be mutated or swapped with fragments within the same bin is part of future work. Sheridan and Kearsley demonstrated a performance improvement in their genetic algorithm through the use of a type of binning scheme [25]. It may also prove fruitful to combine the genetic algorithm search strategy with a more localized branch-and-bound style search.

Forcing mutations to be local in nature not unexpectedly increases the chance of the optimization process being trapped in local minima. We have found

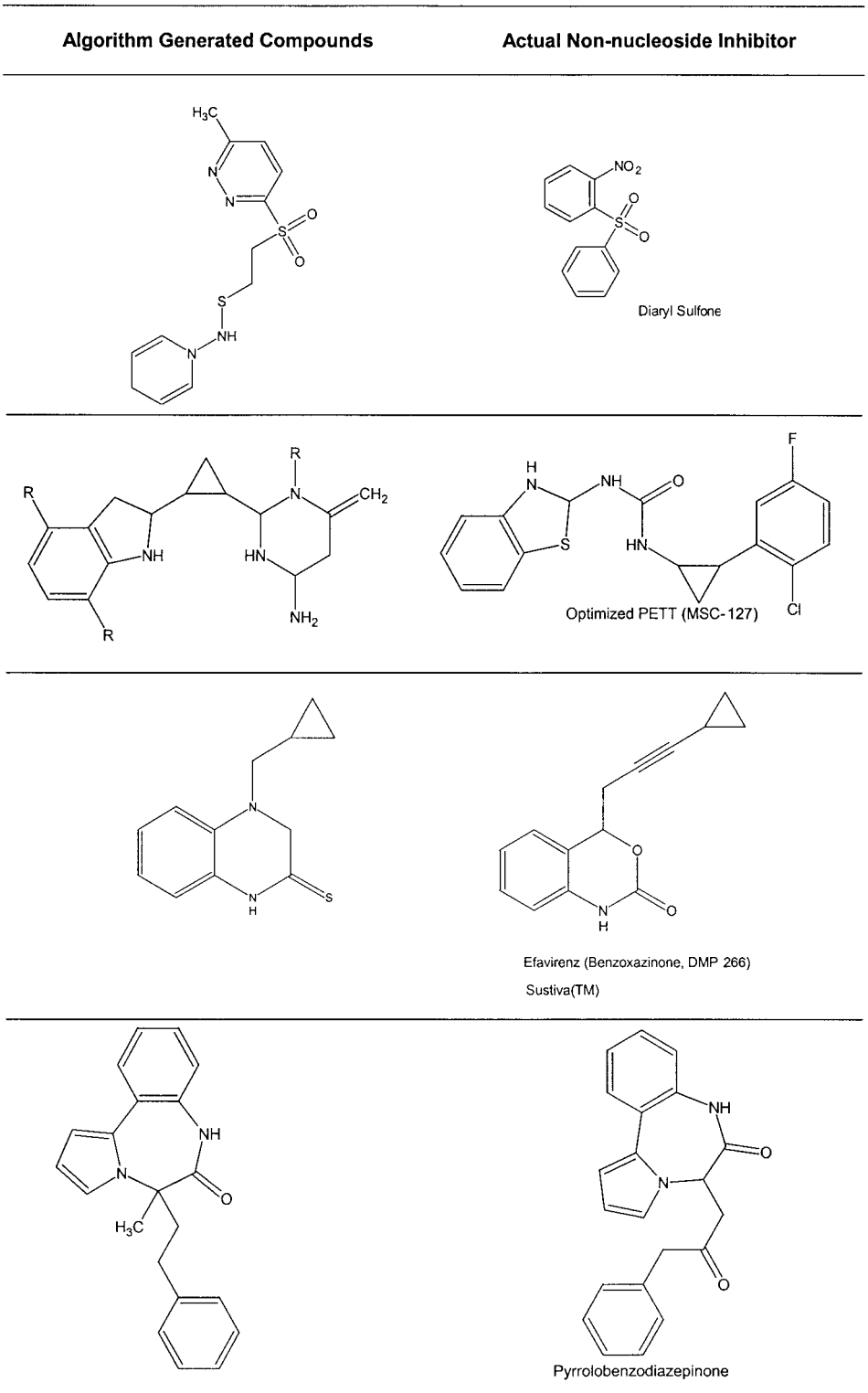

Figure 19. Three of the known NNI scaffolds were recovered by our algorithm. Of particular interest was the generation of a compound similar to MSC127 (second row). This compound was a result of an extensive synthetic optmization process that screened over 750 variations on the PETT scaffold.

that adding diversity back to the population during a run of the ADAPT program can help alleviate this by moving the algorithm to a random, but not too distant place in the search space. Repeatedly optimizing from these semi-diverse points can lead to generating good scoring compounds faster than starting repeatedly from a completely diverse population. Clearly there may be an optimal amount of diversity which should be added to provide a good restarting point-too little diversity puts the algorithm back where it started and too much diversity forces the algorithm to weed through many poor scoring structures. This optimal amount of diversity may be system invariant and finding it is a consideration in future work.

Measuring the success of any de novo design program is difficult, primarily because the number of unique compounds that can be generated is much larger than the number of compounds for which we have experimental binding data. One possible measure of success is to see how easily an algorithm can reproduce a known ligand which has been shown to bind experimentally. In this case the ADAPT program performed poorly in the cathepsin D system, and has only limited success with HIV-1 RT and DHFR where the chemical search space was deliberately much less restricted. In the DHFR experiments, however, ADAPT produced compounds with fitness better than the known inhibitor, indicating that our failure to reproduce the known inhibitor was at least partly the result of the fitness function not having the known inhibitor as a minimum. Another possible measure of success for a de novo design program is whether it can reproduce the trends seen in known ligands. In this case the ADAPT program does fairly well in identifying the 'fused-ring plus chain' motif seen in DHFR ligands and the 'butterfly' motif seen non-nucleoside HIV-1 RT inhibitors, and does very well in identifying the side chain fragments most likely to produce good inhibitors in the cathepsin D combinatorial chemistry system. The difference in ability between finding known ligands and finding trends appears to be mostly the result of the DOCK score being an inexact measure of true binding affinity.

Ideally we would like to be able to measure the success of a de novo design algorithm by being able to accurately (and rapidly) compute the binding affinities of compounds produced by the algorithm and simply analyzing their distribution. Binding affinity is often not the only quality criterion of the user, however, and success is likely to be measured differently under different conditions – there is no definitive magic formula of properties for good drug candidates. Nevertheless, genetic algorithms are quite attractive in their abilities to construct diverse solutions to problems with large search spaces and many constraints. As ligand affinity scoring methods (and those which consider other important properties in the drug development process) improve, genetic algorithms (or hybrid variants) for structure based de novo design will become more valuable and subject to greater scrutiny as strictly optimization methods in comparison to other de novo design algorithms.

Acknowledgements

We thank Geoff Skillman, Connie Oshiro, Jim Arnold, and Ken Foreman for helpful discussions. This work was funded by NIH Biotechnology Training Grant GM08388 (SP) and NIH General Medical Sciences Grant GM31497 (IDK and JJH).

References

1. Walters, P.W., Stahl, M. and Murcko, M.A., Drug Des. Today, 3 (1998) 160.
2. Goodford, P.J., Med. Chem., 28 (1985) 849.
3. Miranker, A. and Karplus, M., Proteins, 11 (1991) 29.
4. Eisen, M.B., Wiley, D.C., Karplus, M. and Hubbard, R.E., Proteins, 19 (1994) 199.
5. Miranker, A. and Karplus, M., Proteins, 23 (1995) 472.
6. Pearlman, D.A. and Murcko, M.A., J. Med. Chem., 39 (1996) 1651.
7. Bohm, H.-J., J. Comput. Aid. Mol. Des., 6 (1992) 61.
8. Lauri, G. and Bartlett, P., J. Comput. Aid. Mol. Des., 8 (1994) 51.
9. Roe, D.C. and Kuntz, I.D., J. Comput. Aid. Mol. Des., 9 (1995) 269.
10. Lewis, R.A. and Leach, A.R., J. Comput. Aid. Mol. Des., 8 (1994) 467.
11. DeWitte, R. and Shakhnovich, E., J. Am. Chem. Soc., 118 (1996) 11733.
12. Bohacek, R.S. and McMartin, C., J. Am. Chem. Soc., 116 (1994) 5560.
13. Rotstein, S.H. and Murcko, M.A., J. Comput. Aid. Mol. Des., 7 (1993) 23.
14. Rotstein, S.H. and Murcko, M.A., J. Med. Chem., 36 (1993) 1700.
15. Moon, J.B. and Howe, W.J., Proteins, 11 (1991) 314.
16. Joseph-McCarthy, D., Pharmacol. Therapeut., 84 (1999) 179.
17. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeny, P., J. Adv. Drug Del. Rev., 23 (1997) 3.
18. Schneider, G., Lee, M.-L., Stahl, M. and Schneider, P., J. Comput. Aid. Mol. Des., 14 (2000) 487.
19. Westhead, D.R., Clark, D.E., Frenkel, D., Li, J., Murray, C.W., Robson, B. and Waszkowycz, B., J. Comput. Aid. Mol. Des., 9 (1995) 139.
20. Glen, R.C. and Payne, A.W.R., J. Comput. Aid. Mol. Des., 9 (1995) 181.

21. Douguet, D., Thoreau, E. and Grassy, G., J. Comput. Aid. Mol. Des., 14 (2000) 449.
22. Devillers, J., J. Chem. Inf. Comput. Sci., 36 (1996) 1061.
23. Weber, L., Wallbaum, S., Broger, C. and Gubernator, K., Angew. Chem. Int. Ed. Engl., 34 (1995) 2280.
24. Blaney, J.M., Dixon, J.S. and Weininger, D., in Weininger, D. (Ed.), Molecular Graphics Society Meeting on Binding Sites: Characterising and Satisfying Steric and Chemical Restraints, York, U.K., 1993.
25. Sheridan, R.P. and Kearsley, S.K., J. Chem. Inf. Comput. Sci., 35 (1995) 310.
26. Sheridan, R.P., SanFeliciano, S.G. and Kearsley, S.K., J. Mol. Graph. Model., 18 (2000) 320.
27. Holland, J.H., Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975.
28. Holland, J.H., Adaptation in Natural and Artificial Systems, 2nd ed., MIT Press, Cambridge, MA, 1992.
29. Darwin, C., The Origin of Species. Dent Gordon, London, 1973.
30. Weininger, D.J., Chem. Inf. Comput. Sci., 30 (1990) 237.
31. Daylight Toolkit, v. 4.6, Daylight Chemical Information Systems. Inc., Santa Fe, NM, USA.
32. Ewing, T.J. A. and Kuntz, I.D., J. Comp. Chem., 18 (1997) 1175.
33. CONCORD, Tripos, Inc., St. Louis, MO.
34. SYBYL, v6.5, Tripos, Inc., St. Louis, MO.
35. Kick, E.K., Roe, D.C., Skillman, G.A., Liu, G., Ewing, T.J.A., Sun, Y., Kuntz, I.D. and Ellman, J.A., Chem. Biol., 4 (1997) 297.
36. Queener, S.F., Bartlett, M.S., Jay, M.A., Durkin, M.M. and Smith, J.W. Antimicrobiol. Agents Chemother., 31 (1987) 1323.
37. Bertino, J.R., J. Clin. Pharmacol., 30 (1990) 291.
38. Blaney, J.M., Hansch, C., Silipo, C. and Vittoria, A., Chem. Rev., 84 (1984) 333.
39. Verkhivker, G.M., Rejto, P.A., Bouzida, D., Arthurs, S., Colson, A.B., Freer, S.T., Gehlhaar, D.K., Larson, V., Luty, B.A., Marrone, T. and Rose, P., J. Mol. Recog., 12 (1999) 371.
40. De Clerq, E., Antiviral Res., 38 (1998) 153.
41. Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P.A., Hughes, P.A. and Arnold, E., Nat. Struct. Biol., 2 (1995) 407.
42. Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R., J. Mol. Graph., 6 (1988) 13.
43. Artico, M., Il Farmaco, 51 (1996) 305.
44. Goldberg, D.E., in Schaffer, J.D. (Ed.), Third International Conference on Genetic Algorithms. Morgan Kaufmann, Fairfax, VA, 1989, pp. 70.
45. Grefenstette, J., J. IEEE Transactions Systems, Man, and Cybernetics, 16 (1986) 122.
46. Schaffer, J.D., Caruana, R.A., Eshelman, L.J. and Das, R., in Schaffer, J.D. (Ed.), Third International Conference on Genetic Algorithms. Morgan Kaufmann, Fairfax, VA, 1989, pp. 51.
47. Caflisch, A. and Karplus, M., Persp. Drug Disc. Design, 3 (1995) 51.
48. Zou, X. Q., Sun, Y. and Kuntz, I.D., J. Am. Chem. Soc., 121 (1999) 8033.