Design of Libraries To Explore Receptor Sites

Catherine M. Murray[†] and Stephen J. Cato*[,‡]

Chemical Design Ltd., Roundway House, Cromwell Park, Chipping Norton, UK OX7 5SR, and
Chemical Design Inc., 8380 Miramar Mall, Ste 224, San Diego, California 92121-2550

Received June 6, 1998

Despite rapid progress in both combinatorial chemistry and high-throughput screening, the number of molecules that could potentially be made and tested for biological activity still far exceeds the capacity for synthesis or screening. Consequently, it is potentially valuable to select and synthesize sublibraries that contain rationally selected subsets. When the structure of the protein receptor site is known, this may be used to impose restrictions of the selection on molecules. This paper describes a method for rapid analysis of large virtual libraries to select a subset that can exhibit at least one conformer which will interact strongly with the receptor and fit within the receptor site.

INTRODUCTION

One of the important challenges of computer-aided drug design is the study of the mode of binding of known ligands into a protein binding site and to identify the pharmacophores involved. This can be achieved through various commercially available programs such as LUDI,[1] AUTODOCK,[2] and FlexiDock.[3] These programs have been designed to evaluate the best possible binding geometries during the lead optimization process in drug design. They rely on force fields for calculation of the energy associated with the binding

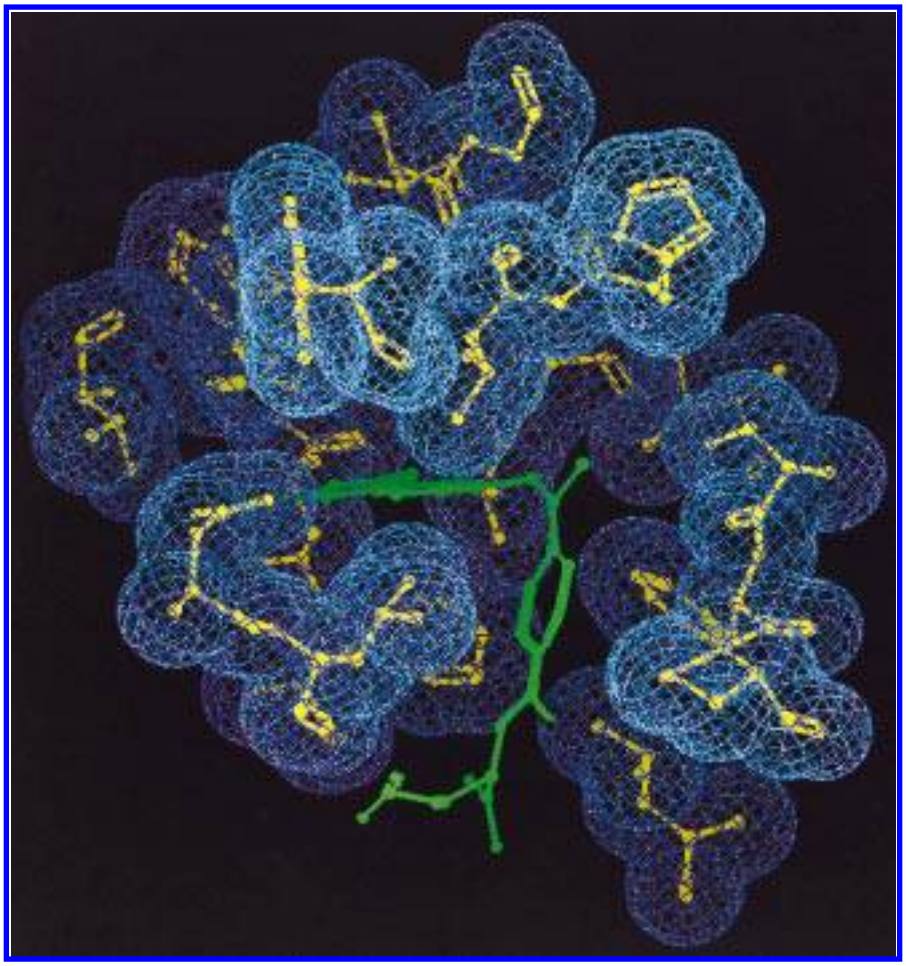

Figure 1. Methotrexate docked into a crystal structure of DHFR.

10.1021/ci980104h CCC: $18.00 © 1999 American Chemical Society
Published on Web 01/05/1999

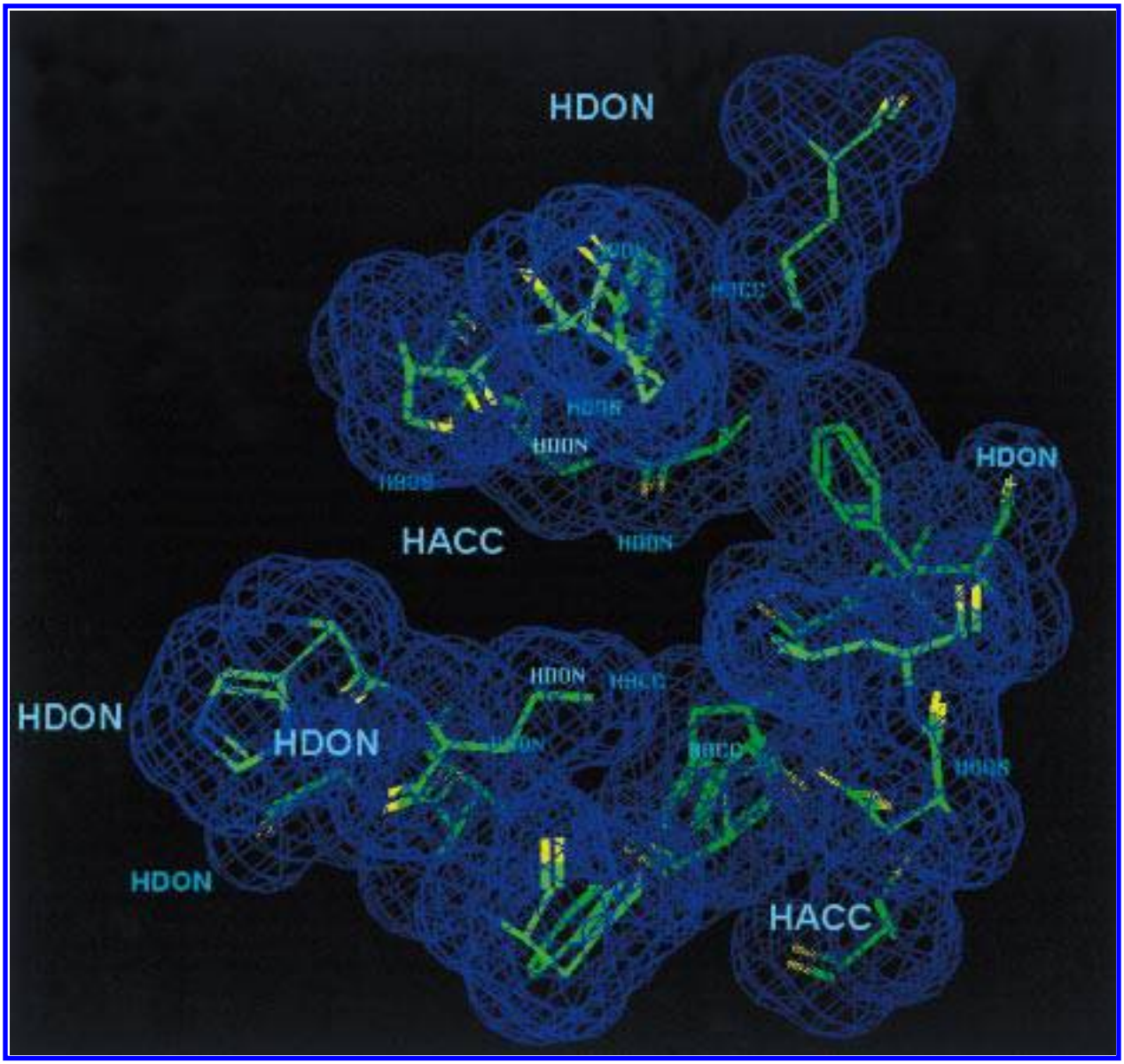

Figure 2. Receptor site of DHFR with pharmacophoric centers identified.

action and aim to help selection of the best ligand based on binding strength.

However, if the aim of a project is to produce a focused library for primary or secondary screening purposes a different approach should be employed, one that is designed specifically for the requirements of lead generation. Chemical Design (CDL) has previously introduced Receptor Screening[4] to achieve this purpose. This successful technique identifies the possible binding modes for ligands within the receptor cavity through the identification of pharmacophoric centers complementary in character to the centers found on the surface of the receptor site itself and the automatic consideration of all possible three- or four-center pharmacophores created therefrom. It then uses CDL's pharmacophore technology to screen a database of possible ligands providing the user with a set of potentially active compounds. This technique is very thorough in its approach and can return all conformations of all suitable ligands that have been selected by all possible pharmacophores.

During primary screening, it is often the case that you wish to select for screening all ligand molecules that fit into the receptor site without regard to how many possible pharmacophores caused them to be selected or in which conformations they were chosen. This can speed up the selection-process significantly. Design in receptor (DiR), CDL's new technological advancement in this area, developed in conjunction with several major pharmaceutical companies, addresses these issues. Through user selection and editing, a more refined set of pharmacophores can be produced without loss of any information that describes the volume and shape of the receptor site. This, combined with a new faster algorithm for searching, allows fast identification of potentially active compounds, allowing larger virtual libraries to be searched in an acceptable timeframe.

METHODOLOGY

A protein structure, derived from crystallography or homology modeling, together with any information about the receptor site that is available is the starting point for either receptor screening or design in receptor.

[†] Chemical Design Ltd.
[‡] Chemical Design Inc.

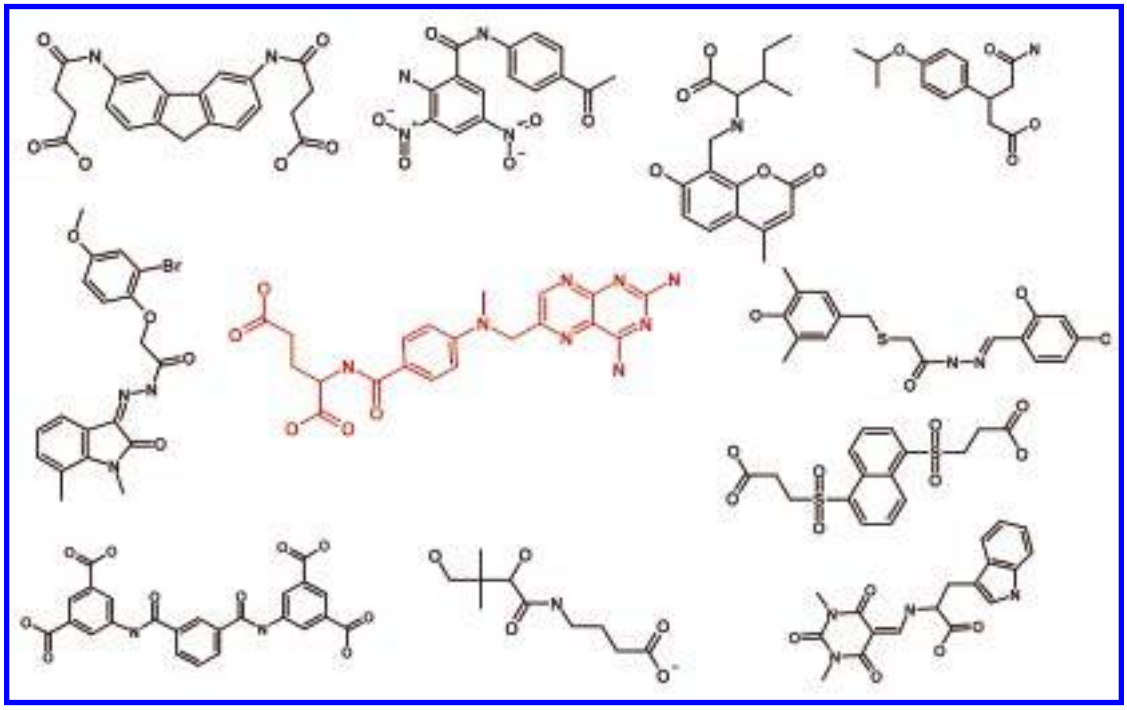

Figure 3. The ten compounds chosen as cluster means compared to methotrexate (center).

The technique requires that the receptor cavity be defined as a "selection" (a collection of atoms) within the protein structure. This can be achieved by choosing all residues within a user-defined distance from any atom of a known ligand bound into the active site or a known active residue. Best results are obtained when enough residues are included in this selection to give the receptor site shape and volume, enabling a check that the ligands chosen can physically fit into the cavity.

The pharmacophoric centers on the surface of the cavity are automatically identified, and the complements, required in the ligand for binding to them, are generated from among the seven pharmacophoric center types that CDL provides by default (acid, base, lipophile, hydrogen bond donor, hydrogen bond acceptor, positive charge center, and aromatic ring). In DiR, any knowledge that the user may have about the preferred binding geometry may be incorporated at this point by use of the editing facility to remove unwanted centers, move existing centers, or add new centers. For example, if it is known that a particular point of the receptorsite has potential to flex significantly during binding, the positions of centers can be adjusted to allow for this. Likewise, if a particular atom such as a hydroxyl oxygen could potentially behave as more than one center type, but is known to only behave as a hydrogen bond donor for a specific protein, then the definition of that oxygen can be changed to reflect this specific information. This ability to edit the set of centers allows better results to be obtained from approximate structures such as those obtained from homology modeling. If the aim of the project is to find new leads that bind in the same way as a known ligand, then any centers complementary to those not present in that ligand can be removed.

Once the user has defined the desired complementary pharmacophoric centers, the production of the pharmacophores derived from them proceeds. All possible pharmacophores from the available centers are calculated automatically. The centers can be combined as three-center pharmacophores or four-center pharmacophores. The extra center involved in the latter gives extra information about the binding criteria required, and, because the points are no longer necessarily planar, the four-center pharmacophore better expresses the shape and volume required to fill the receptor site.

The database containing potential ligands is then searched. The new fast algorithm generates each possible conformation only once and then compares it against all query pharmacophores. In comparison to receptor screening, this represents a reversal of the two nested loops involved. In DiR loopingover the molecular conformations becomes the outer loop, looping over the query pharmacophores the inner one. Typically, if a match is found, the process proceeds to the next compound, since in a screening selection tool there is no need to know how many, or indeed which, conformations caused the compound to be selected.

EXPERIMENTAL AND RESULTS

The Brookhaven[6,7] Database file of DHFR (code 3DFR[8] with methotrexate bound into the receptor site was used, as shown in Figure 1. The active site residues were identified, as indicated in the header of the pdb file. A cavity was constructed by combining the residues in the previously defined sites, SITE_NMR, SITE_NND, SITE_MPT, SITE_MNM, SITE_MAB, and SITE_MGL. This was easily achieved because of the automated manner in which Chem-X identifies the residues associated with active sites as detailed in the pdb header file and makes these available within the Chem-X environment. The protein structure was saved into a Chem-X database, along with the combined site, now designated as SITE_CDL. This combined site contained 19 residues, namely 4LEU, 5TRP, 6ALA, 13ILE, 14GLY, 18HIS, 19LEU, 21TRP, 26ASP, 27LEU, 30PHE, 45THR, 48SER, 49PHE, 50PRO, 54LEU, 97ALA, 98GLY, and 116THR.

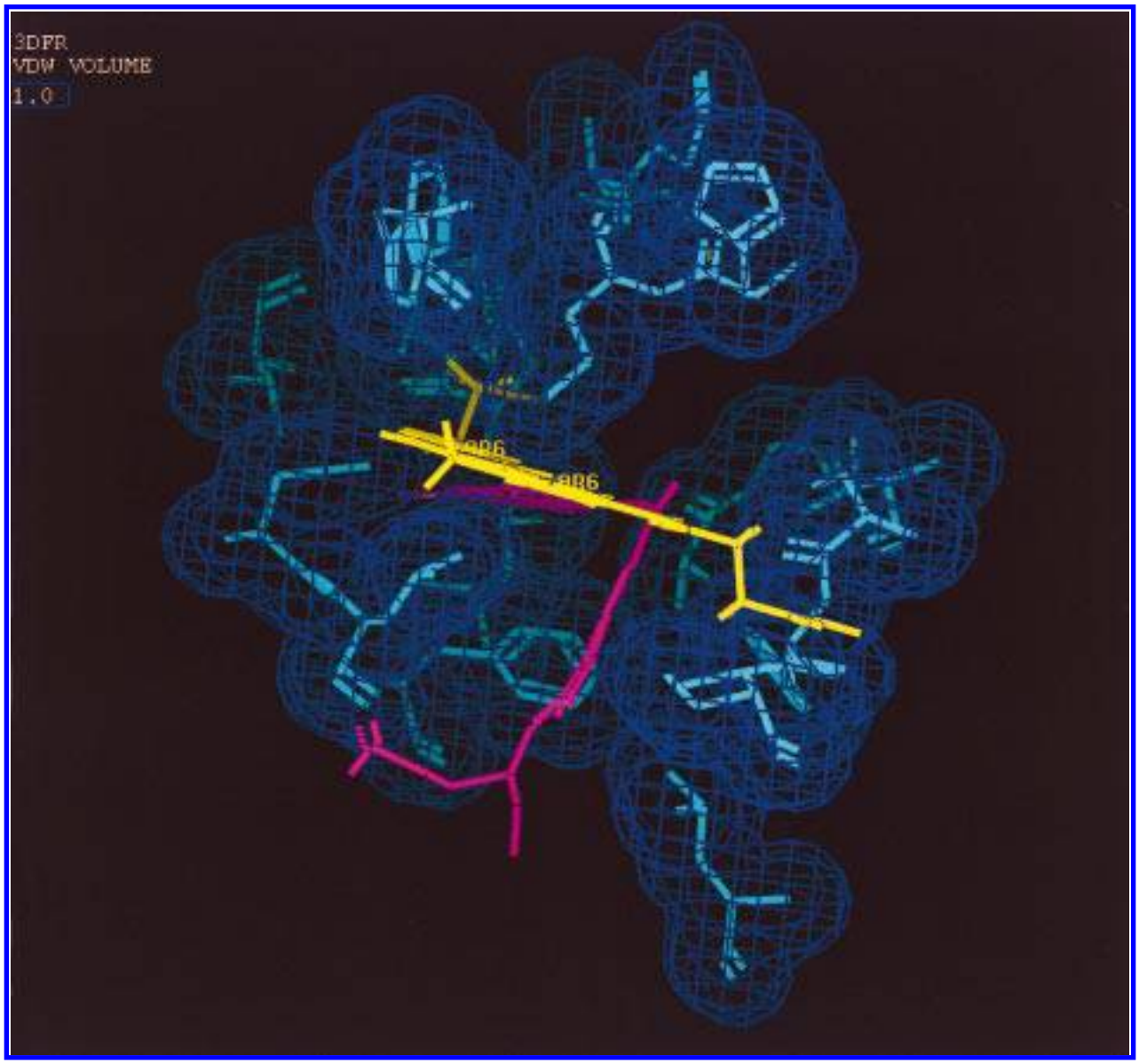

Figure 4. Overlay of one selected compound (224 286) (compared with methotrexate) in the binding site of DHFR.

Table 1. Results of the Clustering of the 225 Matches Obtained from the ChemScreen Database

cluster ID	mean structure	no. in cluster	max. separation
1	224 286	51	5.16
2	181 815	39	4.22
3	225 481	35	4.24
4	235 416	29	4.35
5	232 890	17	3.86
6	233 895	16	3.76
7	241 008	15	5.13
8	230 172	10	3.92
9	229 199	7	4.03
10	243 650	6	4.44

Parameters were set to ensure that all the lipophillic and aromatic sites were correctly identified. The methotrexate-ligand was included in the display, and all the complementary pharmacophoric centers were identified. For the purposes of clarity, Figure 2 shows the binding site with all the pharmacophoric centers and without the methotrexate ligand. Chem-X automatically checks that any complementary centers generated are indeed within the receptor cavity and available to bind.

From the displayed choice of centers, 13 that were directly adjacent to the methotrexate molecule were retained, and the rest removed through the editing facility. The ChemScreen[9] database supplied by ChemBridge Corporation was then searched for compounds which bind in a similar way to methotrexate. From the 13 centers, 167 three-center pharmacophores were developed. The DiR search was implemented including a volume map for the protein cavity. DiR was set to accept one conformation of each compound identified by any one pharmacophore. This produced a set of 225 matches that would be expected to fit inside, and bind with, the receptor site. This selection required 60 h to search the approximately 11 000 structures in the database running on a Alpha processor from Digital Equipment Corp.[10] Generally, speed improvements by a factor of up to 7 were observed compared to the earlier receptor screening technology.

These matches were then clustered using an 85% similarity index. They fall into 10 clusters, from which the mean compounds were carried forward. These results are illustrated

in Table 1 and Figure 3. One of these compounds (reference number 224 286) is shown bound (yellow) into the receptor site overlaid onto the methotrexate molecule (red) in Figure 4.

CONCLUSIONS

Use of receptor site information to focus libraries can provide a valuable means of compound selection for the purposes of primary screening. Computationally testing the molecules allows the user to determine whether they physically fit into the protein receptor site and also if it is possible for them to exist in a conformation that allows a minimum of three essential interactions to be available for binding purposes.

The new design in receptor development from chemical design allows this process, sometimes considered "virtual screening" to take place based on either 3D information about the structure of the protein receptor site or on the pharmacophore centers available on a known active ligand. There have been significant changes in the speed of searching obtained through a change in the order of the nested loops of the search algorithm.

REFERENCES AND NOTES

(1) Bohm, H. J. *J. Computer-Aided Mol. Design* **1996**, *10*, 265−272.
(2) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. *J. Computer-Aided Mol. Design* **1996**, *10*, 293−304.
(3) FlexiDock is part of the Sybyl module Biopolymer from Tripos Associates Inc., 1699 South Hanley Road, Suite 303, St. Louis, MO, 63144-2913, USA.
(4) Receptor Screening is available through Chem-X from Chemical Design, Roundway House, Cromwell Business Park, Chipping Norton, OXON OX7 5SR, U.K.
(5) Benkovic, S. J.; Fierke, C. A.; Naylor, A. M. *Science* **1988**, *239*, 1105−1109.
(6) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. Protein Data Bank in *Crystallographic Databases-Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; pp 107−132.
(7) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a Computer-based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(8) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. Crystal Structures of Escherichia Coli Lactobacillus Casei Dihydrofolate refined at 1.7 Angstroms resolution. I. General Features and Binding of Methotrexate. *J. Biol. Chem.* **1982**, *257*, 13650.
(9) ChemScreen, Diverse Screening Database from ChemBridge Corporation, 16981 Via Tazon, Ste G, San Diego, CA, 92127.
(10) Digital Equipment Corporation, Corporate Headquarters, 111 Powdermill Road, Maynard, MA 01754−1418.

CI980104H