Design and Characterization of a Functional Library for NMR Screening Against Novel Protein Targets

Kelly A. Mercier, Katherine Germer and Robert Powers*

Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Abstract: In the past few years, NMR has been extensively utilized as a screening tool for drug discovery using various types of compound libraries. The designs of NMR specific chemical libraries that utilize a fragment-based approach based on drug-like characteristics have been previously reported. In this article, a new type of compound library will be described that focuses on aiding in the functional annotation of novel proteins that have been identified from various ongoing genomics efforts. The NMR functional chemical library is comprised of small molecules with known biological activity such as: co-factors, inhibitors, metabolites and substrates. This functional library was developed through an extensive manual effort of mining several databases based on known ligand interactions with protein systems. In order to increase the efficiency of screening the NMR functional library, the compounds are screened as mixtures of 3-4 compounds that avoids the need to deconvolute positive hits by maintaining a unique NMR resonance and function for each compound in the mixture. The functional library has been used in the identification of general biological function of hypothetical proteins identified from the Protein Structure Initiative.

Keywords: NMR functional library, FAST-NMR, NMR high-throughput screen, protein-ligand binding, protein structure initiative, chemical library design.

INTRODUCTION

High-throughput screening (HTS) is an essential component of the drug discovery process where the success of HTS is fundamentally dependent on the composition of chemical libraries [1-5]. Significant effort by the pharmaceutical industry has been devoted to the accumulation and design of these libraries that may contain hundreds of thousands of compounds. Historically, the compounds that comprise these libraries may come from multiple sources such as the synthetic effort of chemists, through acquisition of chemical libraries, from natural sources [6, 7] and from combinatorial chemistry efforts [8-10]. The advent of automated technologies has allowed for these large chemical libraries to be screened in a reasonable amount of time. More recently, there has been an increase in effort to control the composition of the screening libraries by maintaining structural diversity and "drug-like" qualities [11-17]. There has also been a great deal of interest in virtual screening prior to any HTS or combinatorial chemistry efforts to further guide the design of the screening library [18-22]. NMR has evolved as an invaluable resource in drug discovery screening either as a validation of HTS hits or through complimentary screening efforts [23-29]. Similar efforts have been devoted to develop fragment based libraries that capture drug-like characteristics where the size of the chemical libraries are amenable to NMR-based screens [27, 30, 31].

In general, the receptor of interest in an HTS or NMR-based screen is already well characterized before a drug discovery effort is initiated. Nevertheless, pharmaceutical companies are constantly searching for new protein targets that may lead to the development of novel therapeutics or alleviate adverse side-effects [32]. The success of various genomic projects has identified a wealth of potential therapeutic targets [20, 33]. From the human genome alone, 30,000 – 90,000 proteins have been predicted [34, 35]. Unfortunately, only about 50% of these hypothetical proteins can be annotated with a molecular function based on sequence homology alone [36-38]. Traditional biochemical methods for determining protein function are highly specific and may take years to develop [39]. Recently, computational methods have been developed to assign biological functions to hypothetical proteins based on sequence, structure and predicted ligand interactions [33, 40-43]. While these computational predictions are efficient for genomic-scale analysis, experimental results are still essential for validating function. Additionally, the Protein Structure Initiative (PSI) is supporting the development of high-throughput technologies to expedite the process of determining a protein structure by NMR and X-ray, where the availability of the structure is expected to assist in the functional annotation of hypothetical proteins. Our analysis of the literature suggests that ~60% of the protein structures being determined by PSI are novel folds that provide little to no insight into the protein's biological function, which is not surprising given the general PSI target selection process [44].

NMR-based screening methodologies can be applied to assign a biological function to these novel proteins identified from PSI. The Functional Annotation Screening Technology using NMR (FAST-NMR) annotates the general biological function of novel proteins through the structural and sequence analysis of protein-ligand interactions. By experimentally identifying ligands with a known biological function that bind to proteins of unknown function, by identifying the protein's active site from this interaction, determining a corresponding co-structure of this protein-ligand complex, and identifying a protein of known function

*Address correspondence to this author at the University of Nebraska Lincoln, Department of Chemistry, 722 Hamilton Hall, Lincoln, NE 68588-0304, USA; Tel: (402) 472-3039; Fax: (402) 472-2044;
E-mail: rpowers3@unl.edu

1386-2073/06 $50.00+.00

that shares an active-site with a similar sequence and structure, it will be possible to readily infer a general biological function for these novel proteins. At the core of the FAST-NMR methodology is the functional compound library that has been developed to probe the function of hypothetical proteins evaluated by PSI. This library is comprised of functional ligands such as co-factors, inhibitors, metabolites and substrates that are used to infer a biological function through a binding interaction. The chemical compounds that comprise the NMR functional library were drawn from databases such as: BRENDA [45], the cabinet [46], the KEGG database [47], MetaCyc Metabolic Pathway Database [48], Protein Database [49], and the WOrld of Molecular BioAcTivity (WOMBAT) database [50]. The NMR functional library has been compiled and used to screen novel proteins identified by PSI for functional annotation. This paper describes the design, composition, development and maintenance of the NMR functional chemical library.

MATERIALS AND METHODS

Chemical Library

All compounds for the NMR functional chemical library were purchased from Sigma Aldrich (St. Louis, MO). Compounds are selected based on a number of criteria which include the existence of biological activity, existence of a co-structure in the PDB, likelihood of aqueous solubility ($\geq$100 μM), purity ($\geq$ 90%), commercial availability ($\geq$ 5 mgs) and cost ($32 average to add a compound to the library). The goal of the chemical library is to cover the diversity of protein function. Thus, new compounds are identified and added to the library based on unique interactions with protein functional classes that are not currently represented in the library. The compounds are dissolved in either D_2O or D_6-DMSO at a 20 mM concentration and are stored in 96 well plates in a -80°C dessicator. It is imperative that the plates be stored in a dessicator, since D_6-DMSO is very hydroscopic and water absorption may result in significant changes in volumes and concentrations. Similarly, a frost-free freezer will slowly evaporate samples dissolved in D_2O, changing the concentration and potentially causing the sample to precipitate. The library composition, mixture assignments and compound classification are listed in supplemental table 1S.

NMR Data Collection and Analysis

The NMR samples contained 100 μM of compound, 5% DMSO and 20 mM of d-Bis Tris buffer at pH 7.0. When screening the library against a protein target, the NMR sample also contains 25 μM of protein. The 1D ^{1}H NMR spectra were collected on a Bruker 500 MHz Avance spectrometer equipped with a triple-resonance, Z-axis gradient cryoprobe, a BACS-120 sample changer and Icon NMR software for automated data collection. ^{1}H NMR spectra were collected with 256 transients at 298 K with solvent presaturation of the residual HDO, a sweep-width of 6009 Hz and 32K data points and a total acquisition time of 11 min. Screening a protein target against the 113 mixtures by 1D ^{1}H NMR requires ~20.7 h of instrument time while using 12.7 mg of unlabeled SAV1430, a 9kDa protein.

The 1D ^{1}H line-broadening NMR experiment is most suited for the screening of large number of compounds using minimal resources while providing binding information for a wide-range of affinities. A protein-ligand binding interaction is identified by a change in line-width or complete disappearance of the ligands ^{1}H NMR resonances in the presence of the protein. This occurs because of the large molecular-weight difference between the protein and ligand and the corresponding difference in the T_2 relaxation time and NMR line-widths. Essentially, a bound ligand acquires the broad NMR line-widths of the large MW protein. The 1D ^{1}H line-broadening NMR experiment identifies a protein-ligand binding interaction for K_D's < ~20 mM [51, 52]. There is no effective limit for tight binders ($\leq$ nM) since a change in NMR peak intensity would simply be proportional to the amount of added protein. The addition of 25 μM of protein to 100 μM of ligand with a $K_D \leq$ nM would result in at least a 25% decrease in the intensity of the free ligand's NMR spectrum, which is easily observable by NMR for identifying a binding interaction.

The NMR spectra were processed automatically using a macro in the ACD/1D NMR manager (Advanced Chemistry Development, Inc., Toronto, Ontario). The NMR data was Fourier transformed, zero-filled, phased, and baseline corrected. Each spectrum was referenced with the (2, 2, 3, 3-D_4) trimethyl-3 propionic acid (TMSP) peak set to 0.0 ppm. S/N $\geq$ 4 in the NMR spectrum was required to keep a compound in the chemical library.

NMR Functional Library Database

A ChemFinder database (Cambridge Soft, Cambridge, MA) with extensive structural searching and similarity capabilities was developed to maintain, utilize and track each of the compounds in the NMR Functional chemical library. The database contains information about the compound's structure, experimental NMR spectrum, biological activity, known protein targets, structures and binding pockets. The database also contains practical information such as plate and well location, Sigma-Aldrich catalogue number and price. The ChemFinder database is interfaced with ACD/1D NMR manager (Advanced Chemistry Development, Toronto, Ontario) and RasMol [53] to display NMR spectra and protein-ligand structures, respectively. The database will be accessible through our web-site http://bionmr-c1.unl.edu/.

RESULTS AND DISCUSSION

Design of the NMR Functional Chemical Library

Prior efforts that have described the design of a compound library for HTS or NMR screening have primarily focused on drug-like features as a selection criteria. By analyzing known drugs, a number of characteristics have been identified that generally correlate structural features with drug activity. These include "Lipinski's Rule of Five" that relate molecular-weight, cLogP, number of hydrogen bond donors and acceptors with drug activity [16, 17]. This analysis has been expanded to include features such as the number of rings, rotatable bonds and the number of heteroatoms [54]. Intrinsic to this approach to library design is the ability to use structural descriptors to predict the likelihood that a compound may be a drug. Therefore, it is relatively straightforward to develop a software algorithm to

analyze a virtual database of millions of compounds to identify a subset of chemicals consistent with the identified drug-like characteristics [55, 56]. The limited accuracy of these predictions are easily counterbalanced by a large library size along with the expected goal of only identifying a few compounds as drug leads for the project to be successful [57]. Conversely, the design of a chemical library with the goal of aiding in the functional annotation of a protein is fundamentally distinct from a library designed for drug discovery. The NMR functional library expands on the library design proposed by Hajduk *et al.* for drug discovery efforts targeting hypothetical proteins [58]. The manuscript by Hajduk *et al. briefly* describes assembling a list of 160 compounds corresponding to substrates, cofactors, and other compounds known to bind proteins, where the compounds are screened as *singletons* using a 2D ^{1}H-^{15}N HSQC NMR experiment. Compounds that exhibit binding are then used as initial leads in a drug discovery effort.

Similarly, essential to the design of the NMR functional library is the knowledge that each compound in the library has a *demonstrated* biological activity. Simply relying on algorithms to predict the likelihood that a compound may bind a protein and exhibit a biological activity would have minimal value in assigning function to a hypothetical or novel protein. Effectively, the goal of the NMR functional library is to correlate the known biological activity of a ligand to the function of the hypothetical or novel protein it binds. There is no software available that will mine chemical databases and select compounds based on experimentally observed biological activities. Thus, assembling the NMR functional library required a large amount of manual research. The compounds for the NMR functional chemical library were primarily selected based on known functional characteristics. Compounds were added to the library by drawing from databases that contain functional information for ligands such as the PDB, KEGG, MetaCyc, BRENDA, the cabinet, and WOBAT.

The NMR functional library needs to maintain diversity, similar in concept to the structural diversity that is important to the design of drug discovery library. In the case of the NMR functional library, diversity is based on the functional activity of the compound as opposed to covering different regions in structure space. Specifically, having the library contain different chemical classes of protease or kinase inhibitors provides minimal value compared to having a variety of compounds that all bind to a distinct protein family. In addition to identifying if a compound has a known biological activity it is equally important to determine if the activity of the compound covers unique protein functions relative to the remainder of the library. As an example of functional diversity, consider three members of the NMR functional library: ATP, tyramine and picrotin. ATP is a known substrate for ~22 functional classes of enzymes from ATPase to synapsin. Similarly, the amino acid tyramine is known to bind to amine oxidase, tyramine N-methyltransferase, tyramine N-feruloyltransferase, aralkylamine dehydrogenase, tyrosine decarboxylase, monophenol monooxygenase, and aromatic-L-amino-acid decarboxylase. Picrotin, a substrate and potent antagonist of GABA receptors, blocks responses of receptors composed of either α_1 or α_3-subunits. The simple combination of these three compounds will identify a wide-range of protein function through a positive binding interaction with a novel protein. Additionally, if two or more compounds do exhibit an overlap in protein binding targets, confirmation that a novel protein binds all of these ligands may further aid in the refinement of the protein's functional annotation.

A second important factor that contributed to the selection of compounds for the NMR functional library was predicted aqueous solubility. Since protein binding is initially being monitored by observing changes in the ligand NMR spectrum, a compound solubility of ~100 μM is desirable for rapid acquisition of NMR data. While a number of software programs are available for predicting water solubility [59], our experience has indicated that these algorithms tend to be relatively unreliable when assembling compounds with the high solubility required for NMR. Instead, we relied on our own chemical intuition to select compounds. Since the aqueous solubility of a number of compounds increased significantly with the addition of 5% DMSO, we were cautious about *a priori* eliminating compounds without experimental validation of low solubility. This was achieved by collecting an NMR spectrum under standard buffer conditions for each proposed compound in the library.

Another important consideration that contributed to the selection of compounds for the NMR functional library was the presence of an NMR or X-ray structure in the PDB [60] for the ligand complexed to a protein. This is based on a secondary goal in the design of the NMR functional library, the use of the sequence and structural characteristics of a ligand defined binding site to annotate the function of a hypothetical or novel protein [61]. Clearly, correlating compounds in the NMR functional library with experimental protein-ligand co-structures contributes to this process.

Composition of the NMR Functional Chemical Library

General functional categories that were initially targeted included: amino acids, carbohydrates, co-factors, hormones, inhibitors, known drugs, lipids, metabolites, neurotransmitters, nucleotides, substrates and vitamins. Fig. **1** outlines the general composition of the library by assigning compounds into these general categories; it should be noted that some compounds may be classified into more than one group. Supplemental Table 1S provides an exhaustive list of the compounds, mixture definitions, and functional classifications for each compound. The chemical properties for the compounds that comprise the NMR functional library are also diverse and consistent with typical property values observed in chemical libraries for drug discovery. The majority of compounds have a molecular-weight below 500 Da, calculated LogP below 5, total number of heteroatoms (O, N, S) below 10 and a number of rings below 5 [62]. It should be noted, that while most of the compounds fall within the drug-like characteristics used to filter compounds for a library designed for drug discovery, there are also compounds that fall outside this range. This illustrates a typical concern with the design of a drug screening library. An absolute filter may exclude interesting and novel compounds that are viable drug candidates that do not follow historical trends. Distribution charts listing standard chemical properties of the compounds in the NMR functional library are illustrated in Fig. **2**.

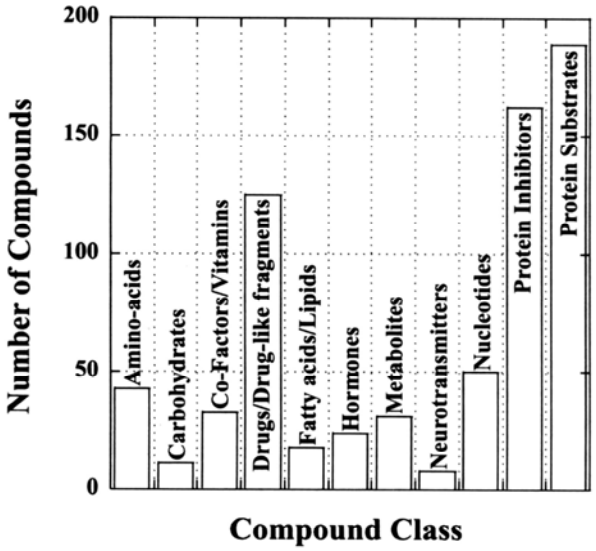

Fig. (1). General classification of the current composition of the functional chemical library. Because of overlapping functionality, numerous compounds are included in multiple chemical classes. Compounds that either bind DNA or are incorporated into DNA are grouped together.

To date, the NMR functional chemical library consists of 414 compounds. Due to issues with solubility, poor 1D ^{1}H NMR spectral quality and long term stability, ~70 compounds have been eliminated from the library. Compounds that have been withdrawn from the library due to low solubility include pyridoxal phosphate, dichloromethylenediphosphonic acid disodium salt, and DL-erythro-Dihydrosphingosine. These three compounds, among others, were not soluble enough to achieve a S/N of 4 in a reference NMR spectrum. A second set of compounds were removed from the library because of long term storage issues that resulted in precipitation or degradation. This included compounds such as: myricetin, Boc-L-phenylalaninol and O-(carboxymethyl)hydroxylamine hemihydrochloride.

5% DMSO is added to the NMR samples to assist with the solubility of the more hydrophobic compounds. There are a limited number of cost-effective deuterated buffers available for a high-throughput NMR-based assay. D_{19}-Bis Tris has an appropriate buffering capacity near pH 7 and was chosen to minimize any potential binding competition with the chemical library. Although phosphate buffers are typically used for NMR studies, phosphate at millimolar concentrations may effectively compete with a number of functional ligands in the library that contains a phosphate group (ATP, O-phospho-tyrosine, FAD, NADPH, pyridoxal phosphate, etc.). While it is still plausible that D_{19}-Bis Tris may still compete with some of the compounds in the library, the negative impact will be significantly less than a phosphate buffer.

To minimize the NMR instrument time and the amount of protein needed for screening the entire library by NMR, the compounds are screened as mixtures [63]. The NMR functional library is currently composed of 113 mixtures. By manually designing these mixtures based on functional diversity, there is a decrease in the likelihood that more than

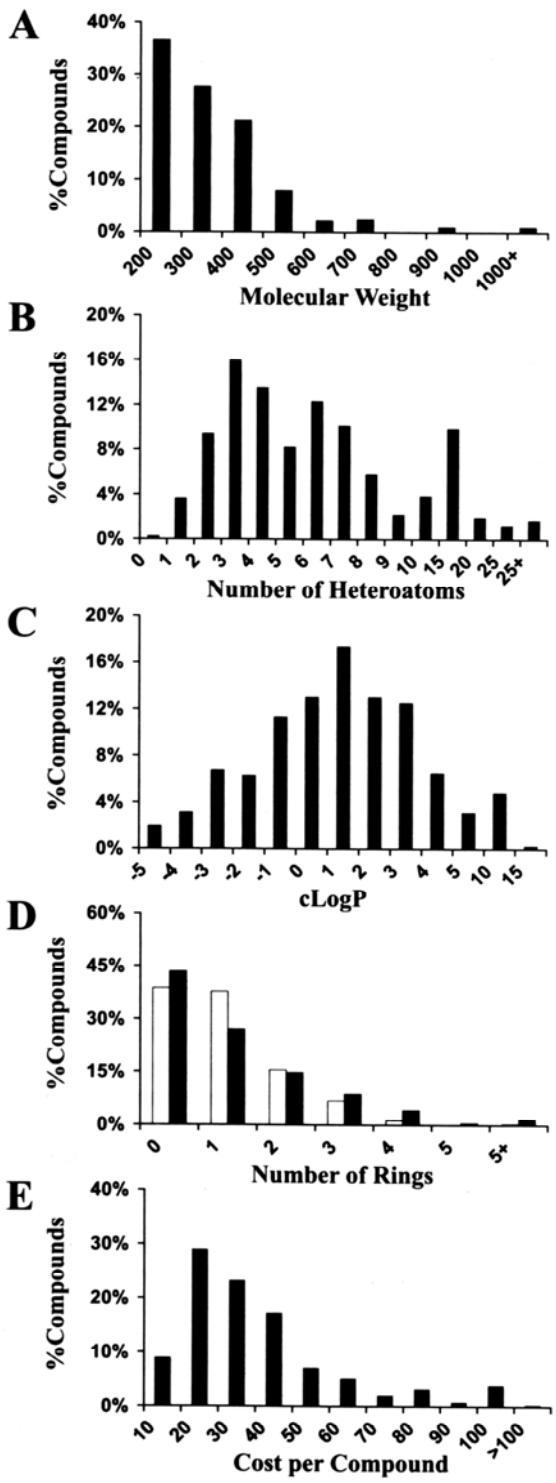

Fig. (2). Binned distributions of the typical chemical properties of the compounds in the functional library: molecular weight (A), number of heteroatoms (B), clogP (C), number of rings (D), where white bars are aromatic rings and black bars are aliphatic rings, and cost (E). The upper range is listed for each binned region. The average molecular weight is 296.6 ± 172.8, the average number of heteroatoms is 6.3 ± 4.8, the average number of aromatic ring structures is 1.1 ± 1.1, the average number of aliphatic rings is 0.9 ± 1.1, the average clogP is 0.47 ± 3.08 and the average cost per compound is $32.38 ± $24.60.

one compound in a mixture will have an interaction with the protein target. Each mixture contains 3-4 compounds and was also manually designed to minimize any overlap in NMR resonances. Thus, deconvolution is avoided by requiring that each compound in the mixture contains a unique NMR resonance that identifies the compound. This is an important and significant difference in the design of the functional NMR library compared to the library proposed by Hajduk *et al.* [58]. Screening of the NMR functional library greatly improves throughput and minimizes resources by an order of magnitude compared to the Hajduk library design, where compounds are screened as singletons using a 2D ^{1}H-^{15}N HSQC NMR experiment.

The structure and NMR spectrum for each compound aided in identifying appropriate combinations of 3-4 compounds. Nevertheless, trial and error was required to finalize the compound mixtures, where an NMR spectrum was used to experimentally validate the mixtures. A reference NMR spectrum was collected for each mixture to ensure that no interactions between the compounds occurred and to verify that no changes in chemical shift occurred. Effectively, the NMR spectrum of validated mixtures is simply equal to the sum of the individual NMR spectra for each compound in the mixture. An NMR reference spectrum for a typical mixture is shown in Fig. **3**.

Compounds are still actively being identified and added to the library, where a goal of ~1000 compounds has been initially set as a practical limit for efficiently screening a hypothetical or novel protein by NMR. New compounds are being selected based on a number of criteria which include the existence of biological activity, existence of a co-structure in the PDB, likelihood of aqueous solubility (≥100 μM), purity (≥ 90%), commercial availability (≥ 5mg) and cost ($32 average to add a compound to the library). As new compounds are added to the library, new mixtures are designed and reference NMR spectra are collected for both the individual compounds and the mixtures. The reference spectra provides NMR assignments for the compounds, while verifying compound solubility, compatibility, consistency of the NMR spectra between the mixture and

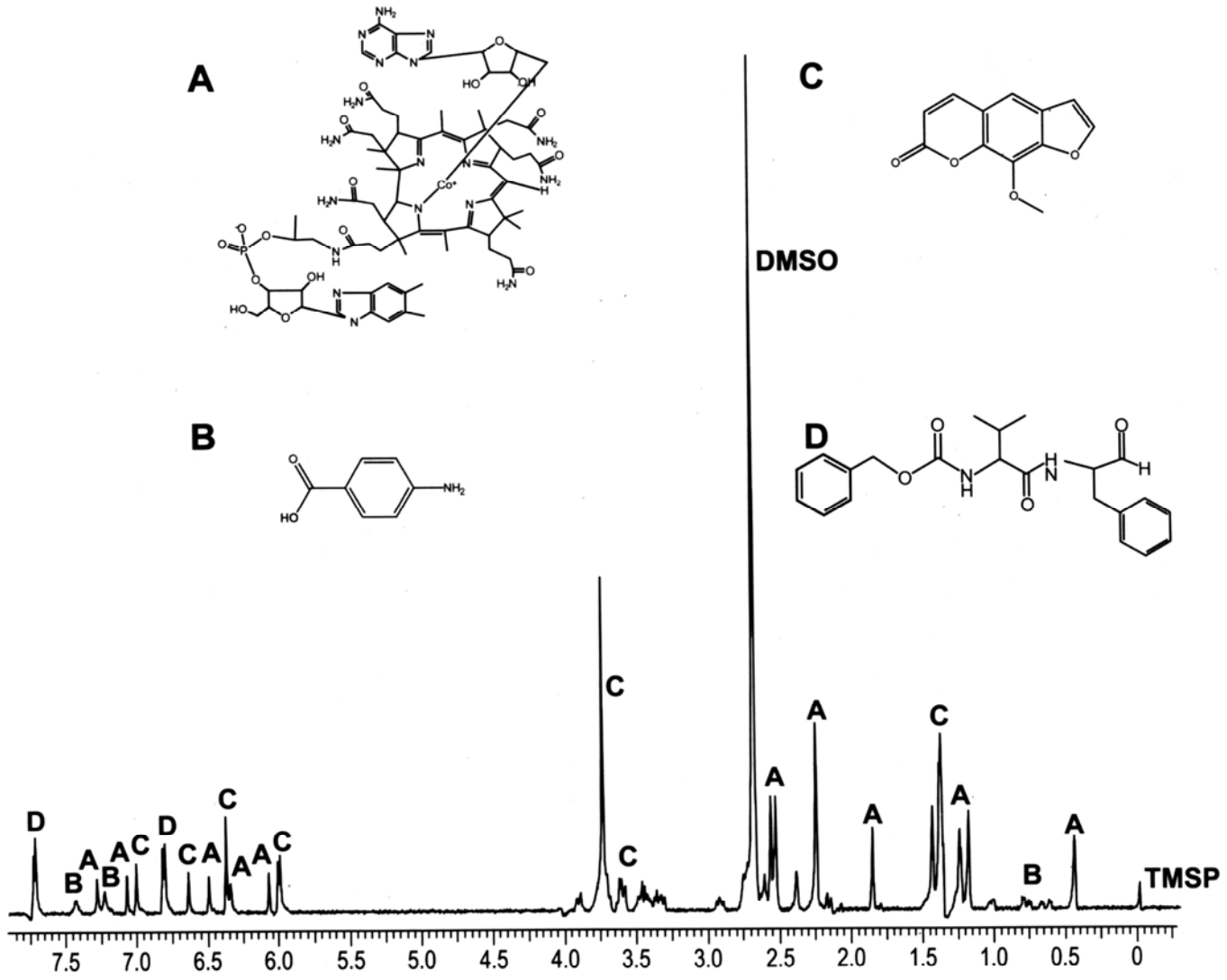

Fig. (3). A typical 1D ^{1}H reference spectrum of a mixture of small molecules from the NMR functional library used for FAST-NMR. Compound A is deoxyadenosyl cobalamin, compound B is 4-aminobenzoic acid, compound C is 8-methoxypsoralen, and compound D is carbobenzoxy-valinyl-phenylalaninal.

individual compounds, and the presence of distinct NMR resonances attributed to each compound in the mixture to avoid deconvolution.

NMR Functional Chemical Library Database

A ChemFinder database has been developed to maintain, utilize and track the progress of the compound library. ChemFinder has extensive structure, sub-structure and similarity search capabilities. This will enable valuable searches on related chemical structures, where an exact match with the compound in the chemical library may not exist. The database contains standard chemical information such as the structure, molecular weight, IUPAC name, PDB HET group dictionary and practical information such as the Sigma Aldrich catalog number and cost, solubility information, storage requirements, and location in the 96 well plates. The database contains a description of the biological activity of the ligand and a list of proteins the ligand binds. The reference 1D ^{1}H spectrum of each compound is linked by ACD/1D NMR manager. An important feature of the database design is a direct link from the PDB records associated with the ligand and a graphical display of the complex using RasMol, which is accessed through a simple display structure button. The graphical display is automatically focused on the ligand-defined active-site and highlights active-site residues. This enables easy visual comparison of protein-ligand co-structures identified by FAST-NMR with comparable structures in the PDB. The protein name, PDB identification number, and literature reference of the structure are also linked to the graphical display of the protein-ligand co-structure. A screenshot of the NMR functional chemical library database is shown in Fig. **4**.

FAST-NMR Utilization of the NMR Functional Chemical Library

To date, the NMR functional chemical library, as part of FAST-NMR, has been used to screen three hypothetical proteins identified from Protein Structure Initiative efforts: *Staphylococcus aureus* protein SAV1430, *Vibrio cholerae* protein VC0424, and *Pseudomonas aeruginosa* protein PA1324. The observed hit rates of the NMR functional chemical library for these proteins were 5.1%, 7.5%, and 4.8%, respectively. These hit rates are consistent with typical hit rates observed with NMR-based assays using directed or focused fragment chemical libraries [25, 64]. The NMR Functional chemical library has similarities to the SHAPES library, in that all members of the library have a predisposition to bind proteins. The SHAPES library was designed to contain small, structurally diverse compounds that are water soluble and correspond to *fragments* or *molecular frameworks* of known drugs [65]. This was accomplished by classifying compounds in the comprehensive medicinal chemistry (CMC) database (MDL Information Systems, San Leandro, CA) into frameworks and sidechains [65]. These frameworks and sidechains were used to search the available chemical directory (ACD) to assemble a library of a few hundred compounds. Screening the SHAPES library against therapeutic protein targets resulted in significantly higher hit rates (0.7-7.7% or higher) compared to random libraries (0.1-0.5%) [66]. Thus, the NMR functional chemical library is comparable in

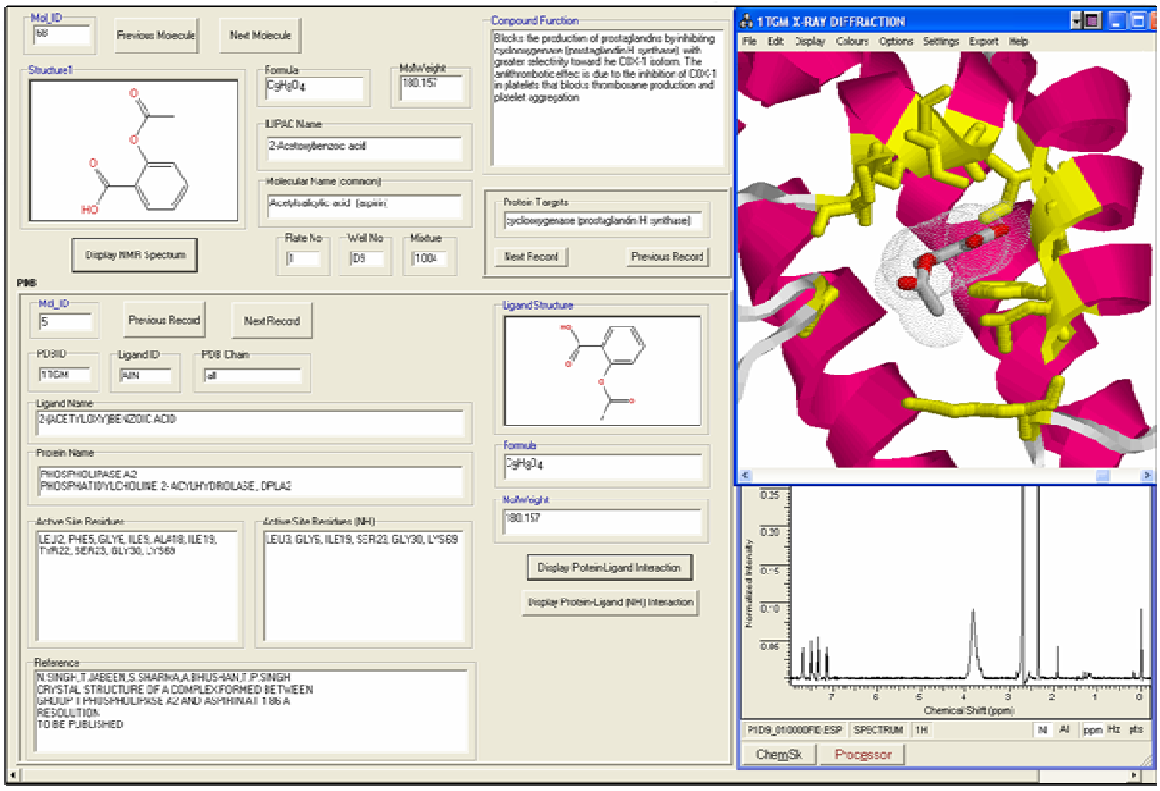

Fig. (4). A screenshot of the ChemFinder database that links the NMR functional library to a RasMol view of the protein-ligand co-structure and the experimental NMR spectrum.

application to a biased library where every compound has an experimentally observed binding affinity to a specific class of proteins. Fig. **5** outlines the hit rates and number of compounds from the functional library that exhibited binding against the three hypothetical proteins.

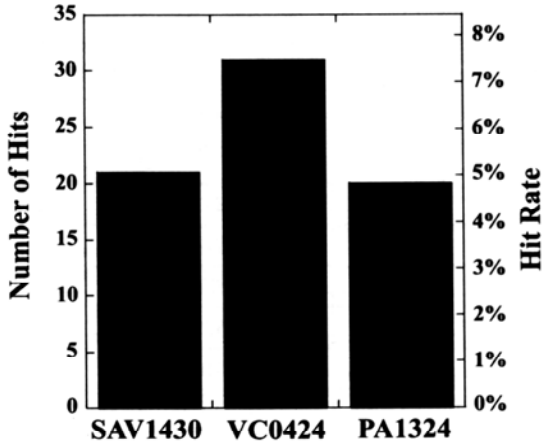

Fig. (5). The hit rate and number of compounds from the NMR functional library that exhibited binding against hypothetical proteins SAV1430 from *S. aureus*, VC0424 from *V. cholerae*, and PA1324 from *P. aeruginosa*.

The three hypothetical proteins bind a unique set of compounds, which suggests a distinct functional activity for each protein. Among the 21 compounds that bind SAV1430, the primary functional ligand was identified as O-phospho-L-tyrosine. The other 20 compounds were structurally similar to O-phospho-L-tyrosine with a heterocyclic structure and a charged substitutent (2-amino-4-methylphenol, nicotinic acid and acetylsalicylic acid). Similarly, screening PA1324 against the functional NMR library identified suramin as the primary functional ligand. Although suramin was clearly the tightest binder, the other ligands had a similar aromatic ring structure and negatively charged (sulfate, phosphate) substituents. Conversely, the initial functional screen of VC0424 yielded no consensus structural pattern for the ligands that bound VC0424. It was subsequently determined that VC0424 was not stable and probably partially unfolded under the standard FAST-NMR buffer conditions. This result suggests that the ligands that bound VC0424 were not related to its functional activity.

In addition to providing information regarding the biological function of a novel protein, an NMR-based screen using the NMR functional library will also provide a reasonable starting point for a structure-based approach to drug design [58]. The compounds that comprise the NMR functional library exhibit a number of properties that are desirable as initial chemical leads. The compounds are commercially available, economical, highly water soluble, exhibit generally specific biological activity and would be bioavailable since they correspond to natural cell metabolites or known protein inhibitors or drugs. While the compounds in the NMR functional screen may not be themselves viable drugs, they would provide an initial framework to design a directed combinatorial library for further screening.

CONCLUSION

The NMR functional library is the cornerstone of the FAST-NMR method that has been developed to screen hypothetical proteins from the PSI to annotate a general biological function using NMR-based screening methods. The design of the NMR functional library was based on identifying compounds with known biological functions, such as co-factors, protein inhibitors, and metabolites. Other important factors that contribute to including a compound in the library are the availability of an NMR or X-ray structure in the PDB, aqueous solubility and a quality NMR spectrum. To improve the efficiency of screening the functional library by NMR, the compounds have been assembled into mixtures containing 3-4 compounds. NMR reference spectra were used to validate the mixture composition by insuring that the compounds do not interact with each other and that each compound has a unique NMR resonance to eliminate the need for deconvolution. The NMR functional library has been assembled manually through the use of several ligand databases and maintained using our own ChemFinder chemical database that includes experimental NMR spectra and protein-ligand co-structures for each compound in the library.

ACKNOWLEDGEMENTS

This project could not have been completed without the generous protein donations from Dr. Guy Montelione, Dr. John Cort, and Dr. Theresa Ramelot from the Northeast Structural Genomics Consortium (NESG). This work was supported by grants from the Protein Structure Initiative of the National Institutes of Health (P50 GM62413), the Nebraska Tobacco Settlement Biomedical Research Development Funds and Nebraska EPSCoR. The research was performed in facilities renovated with support from NIH (RR015468-01).

SUPPLEMENTAL MATERIAL

Table **1** contains a list of compounds currently present in the NMR functional library. The compounds are listed based on their NMR mixture composition and contain functional classification.

ABBREVIATIONS

FAST-NMR	=	Functional annotation screening technology using NMR
NMR	=	Nuclear magnetic resonance
HTS	=	High-throughput screening
DMSO	=	Dimethyl sulfoxide
D_2O	=	Deuterium oxide
TMSP	=	(2, 2, 3, 3-d4) trimethyl-3 propionic acid
1D	=	One-dimensional
PSI	=	Protein structure initiative
PDB	=	Protein data bank
WOMBAT	=	World of molecular bioactivity
ATP	=	Adenosine triphosphate
GABA receptor	=	γ-aminobutyric acid receptor
FAD	=	Flavin-adenine dinucleotide
NADPH	=	Nicotinamide adenine dinucleotide phosphate

Table 1. Design and Characterization of a Functional Library for NMR Screening Against Novel Protein Targets (K.A. Mercier, K. Germer and R. Powers)

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Adenosine 5'-triphosphate disodium salt (ATP)	1001										X		X		179.2	$26.70	8	1.4	2	1
Cinoxacin						X						X			333.4	$32.40	7	-1.4	1	2
Cycloheximide				X											99.17	$17.40	6	-3.6	0	0
Sodium glycocholate hydrate (Bile salt)				X											86.03	$26.40	4	-1.7	0	0
Cytidine 5'-triphosphate disodium salt (CTP)	1002										X		X		479.6	$99.60	6	3.0	1	3
Cytochalasin B									X						179.2	$10.60	6	-3.1	0	0
Doxycycline hyclate						X						X			230.3	$8.90	3	2.8	2	0
L-Glutathione reduced				X											155.2	$10.10	5	-3.7	0	1
Colchicine	1003					X									251.2	$35.00	8	1.9	0	8
Nalidixic acid						X						X			311.4	$73.30	4	0.7	1	4
Tetrahydropteridine				X											182.2	$19.90	6	0.3	0	2
Acetylsalicylic acid (aspirin)	1004					X						X	X		162.3	$8.45	3	-3.3	0	0
Biotin (vitamin H)				X									X	X	424.5	$19.90	7	4.5	3	3
Thymidine 5'-triphosphate sodium salt (TTP)				X									X		478.1	$50.50	19	-3.8	0	2
Methoxatin (PQQ)	1005			X											166.2	$25.60	11	-0.7	0	2
Ouabain								X				X			387.4	$16.20	9	2.1	1	3
Praziquantel						X									360.4	$8.00	5	1.4	0	4
Uridine 5'-triphosphate trisodium salt dihydrate (UTP)											X		X		481.1	$25.05	20	-4.6	0	2
Adenosine 3',5'-cyclic monophosphate (cyclic AMP)	1006										X		X		347.2	$25.70	13	-3.2	1	2
Ethosuximide						X									165.2	$14.80	3	0.9	1	0
L-Leucine (Leu)		X											X		256.3	$14.60	3	3.2	2	0
Neostigmine						X						X			429.4	$144.60	13	-2.5	0	2
Benzamidine hydrochloride hydrate	1007											X	X		122.1	$13.70	4	5.1	0	1
Guanosine 3',5'-cyclic monophosphate sodium salt (cyclic GMP)											X		X		519.2	$29.05	22	-5.5	0	2
L-Isoleucine (Ile)		X													102.1	$21.80	3	-0.9	0	0
N-Phospho-Ile(O-ethyl)-Tyr(O-benzyl)-Gly dipotassium salt												X			116.2	$28.20	1	5.1	0	0
4-Aminobenzoic acid (PABA, vitamin H1)	1008			X									X	X	137.1	$7.35	3	1.0	1	0
8-Methoxypsoralen					X										336.4	$17.80	6	1.7	1	0
Deoxyadenosyl cobalamin (Vitamin B12)				X									X	X	218.3	$43.40	5	-3.1	0	1
MDL 28170												X			167.3	$27.10	1	2.8	0	2
Cantharidin	1009											X			217.3	$90.90	5	0.9	0	1
Coenzyme A sodium salt hydrate				X									X		399.4	$11.95	7	1.2	1	3
L-Arginine (Arg)		X											X		133.1	$13.60	4	-7.0	3	2
L-Carnitine inner salt (vitamin Bt)				X									X	X	147.1	$8.80	5	-2.7	0	0

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Calmidazolium chloride	1010											X			348.4	$55.00	6	0.9	2	3
L-Lysine (Lys)		X											X		149.2	$9.10	4	-1.7	0	0
Pantothenate (vitamin B5)				X									X	X	333.4	$31.75	7	-0.7	1	2
Acetamide	1011											X	X		222.3	$23.00	1	4.3	1	0
b-Nicotinamide adenine dinucleotide phosphate (NADP+,NADPH)				X									X		251.3	$12.30	4	2.4	1	1
L-Histidine (His)		X											X		234.3	$44.10	3	2.0	1	0
S-(5'-Adenosyl)-L-methionine chloride		X		X											237.2	$48.20	8	1.8	0	2
b-Nicotinamide adenine dinucleotide (NAD+, NADH)	1012			X									X		665.4	$50.50	23	-5.0	1	4
Chymostatin												X			238.3	$37.60	7	0.0	0	1
L-Aspartic acid (Asp)		X											X		179.6	$27.20	4	-7.0	0	0
O-Phospho-L-serine		X													261.2	$25.20	8	-3.6	1	0
L-Alanine (Ala)	1013	X											X		174.2	$12.00	5	-3.4	0	0
O-Phospho-L-tyrosine		X													330.4	$12.80	7	3.6	3	0
Thiamine hydrochloride (Vitamin B1)				X										X	425.3	$9.20	15	-5.0	1	1
Aminophylline hydrate	1014					X						X			291.4	$23.10	1	5.4	2	1
Phenylmethanesulfonyl fluoride (PMSF)												X			164.2	$10.40	3	-3.1	1	0
Thiamine pyrophosphate (TPP)				X									X		425.3	$10.50	15	-6.4	1	1
4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF)	1015											X			203.2	$25.35	5	1.6	1	0
erythro-9-(2-Hydroxy-3-nonyl)adenine hydrochloride												X			275.4	$39.00	4	3.1	1	0
Tetrahydrofolate (Folic acid, vitamin M)				X									X	X	430.4	$8.95	12	-3.5	1	2
4-Chloromercuribenzoic acid	1016											X			357.2	$12.95	8	-3.7	1	0
6-Hydroxydopamine hydrochloride (6-hydroxyDOPA)				X											202.3	$31.90	3	1.6	1	2
Ceftriaxone sodium salt						X						X			554.6	$38.90	13	-2.7	0	3
L-Tyrosine (Tyr)		X											X		399.4	$44.00	6	0.3	1	3
Allopurinol	1017											X			151.3	$17.00	5	2.9	1	1
L-2-Aminoadipic acid												X	X		89.09	$7.45	16	-1.4	2	1
L-Tryptophan (Trp)		X											X		181.2	$9.10	26	-3.9	1	0
L-Serine (Ser)	1018	X											X		119.1	$9.70	4	-2.4	0	0
Methyl 6,7-dimethoxy-4-ethyl-b-carboline-3-carboxylate						X						X			289.3	$32.10	10	-1.7	0	2
Pepstatin A												X			354.4	$11.40	6	-2.8	3	1
Pyridoxine (Vitamin B6)				X									X	X	172.1	$37.40	7	-3.0	0	0
Daphnetin; 7,8-Dihydroxycoumarin	1019											X	X		527.5	$29.30	11	0.9	2	3
Idazoxan hydrochloride												X			357.8	$33.20	5	-0.3	1	1
Lorglumide sodium salt												X			115.1	$9.60	3	-2.4	0	1
L-Threonine (Thr)		X											X		204.2	$11.40	4	-1.6	1	1
Bepridil hydrochloride	1020											X			308.4	$21.00	6	-0.3	1	0

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Flavin adenine dinucleotide (FAD,FADH2)				X									X		276.2	$21.80	7	0.6	1	0
L-Glutamic acid (Glu)		X											X		146.1	$38.30	5	-3.4	0	0
Suramin						X									1291	$54.70	35	0.0	8	0
4-Aminobenzamidine dihydrochloride	1021											X			135.2	$5.00	3	-0.5	1	0
Indomethacin						X						X			348.2	$14.70	13	-4.2	0	2
L-Glutamine (Gln)		X											X		307.3	$14.20	10	-3.1	0	0
N-Phenylanthranilic acid												X			551.5	$87.55	14	-0.9	2	0
Bay 11-7085	1022											X			121.1	$18.10	3	-0.5	0	0
L-Methionine (Met)		X											X		321.2	$22.85	6	4.8	2	1
Nicotinic acid (Niacin,vitamin B)				X									X	X	346.3	$31.80	8	3.1	1	1
Penicillin G potassium salt (benzyl penicillin)						X							X		238.3	$39.40	2	0.0	1	0
Estradiol	1023					X		X					X		288.4	$29.00	3	3.2	1	3
L-Proline (Pro)		X											X		105.1	$12.80	4	-2.7	0	0
Ranolazine dihydrochloride												X			197.2	$73.40	3	1.6	1	2
Rifampicin						X									340.4	$50.25	7	0.1	2	1
Diethylenetriaminepentaacetic acid	1024											X			268.4	$47.00	2	3.7	2	3
Puromycin dihydrochloride hydrate					X	X							X		247.1	$26.45	8	-1.7	1	0
Ribavirin						X						X			822.9	$21.76	16	0.0	3	3
trans-4-Hydroxy-L-proline		X											X		131.1	$11.30	4	-1.9	0	1
(±)-Epinephrine	1025							X		X			X		183.2	$31.85	6	-2.1	2	0
2,3,5,6, Tetramethyl-P-Benzoquinone (Duroquinone)													X		164.2	$21.35	2	2.3	0	1
Acetylthiocholine chloride													X		265.4	$19.34	8	3.8	0	0
rac-Glycerol 3-phosphate disodium salt							X						X		316.4	$17.50	8	0.6	0	1
4-Methylpyrazole hydrochloride	1026											X			82.1	$19.95	2	0.5	0	1
a-Methyl-DL-tyrosine		X										X			240.3	$14.70	8	0.0	0	2
Chelidamic acid												X			323.1	$24.20	3	2.7	3	0
Lonidamine						X						X			457.4	$56.70	7	5.7	1	0
6-Methoxy-1,2,3,4-tetrahydro-9H-pyrido[3,4-b]indole	1027											X			276.1	$51.80	11	-4.2	0	0
g-Aminobutyric acid (GABA)						X						X			270.2	$63.60	5	2.4	2	1
Naloxone hydrochloride dihydrate						X						X			230.3	$33.30	3	2.8	2	1
Picotamide												X			310.3	$79.00	7	-1.5	0	4
(-)-Tetramisole hydrochloride	1028					X						X			204.3	$21.65	3	1.8	1	2
3'-Azido-3'-deoxythymidine						X						X			94.11	$16.40	2	2.4	3	1
Atropine sulfate salt hydrate												X	X		282.3	$29.16	7	-0.3	1	1
Captopril						X						X			236.3	$33.40	3	2.4	2	1
1-Carbamoylmethyl-3-carbamoyl-1,4-dihydropyrimidine	1029					X							X		184.2	$47.70	1	1.1	3	1
Chlorzoxazone						X						X			104.2	$9.90	2	-4.4	0	0

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
DL-p-Chlorophenylalanine						X						X			253.3	$71.35	5	1.7	1	0
Melatonin								X				X	X		172.2	$13.30	2	2.5	1	1
(±)-2,3-Dichloro-a-methylbenzylamine hydrochloride	1030											X			190.1	$48.50	3	2.7	1	0
(±)-Thalidomide						X						X			258.2	$39.10	6	0.5	1	2
7,7-Dimethyl-(5Z,8Z)-eicosadienoic acid												X			328.4	$38.50	4	-1.4	0	0
Genistein						X						X	X		352.8	$42.00	6	1.2	1	1
(1R,2S)-(-)-Ephedrine	1031							X					X		165.2	$29.10	2	0.9	1	0
3,5-Dinitrocatechol												X			433	$27.80	6	-0.3	1	0
Choline bromide													X		607.7	$68.05	13	0.1	2	1
D-(+)-Maltose monohydrate			X										X		253.2	$30.10	9	-3.8	1	1
(±)-Norepinephrine (+)-bitartrate salt	1032									X			X		169.2	$44.25	4	-1.0	1	0
3-Amino-2,3-dihydrobenzoic acid hydrochloride												X			202.1	$63.60	8	-5.7	1	1
Methiothepin mesylate salt												X			454.4	$18.30	13	-0.5	3	3
Sucrose			X										X		342.3	$9.90	11	-3.1	0	2
3-Indoleacetic acid	1033							X				X			175.2	$26.65	6	1.4	1	0
3-Isobutyl-1-methylxanthine												X			222.2	$19.10	6	1.4	0	2
Acetylcholine						X				X			X		180.2	$5.72	4	1.0	1	0
Alaproclate hydrochloride												X			374.5	$79.70	5	0.4	0	5
Carbetapentane citrate salt	1034											X			260.3	$24.70	3	3.2	2	1
D-(+)-Glucose			X										X		342.3	$9.80	11	-4.4	0	2
Dopamine hydrochloride										X			X		543.5	$22.80	12	0.4	2	3
N-Acetyl-L-tryptophan 3,5-bis(trifluoromethyl)benzyl ester												X			245.3	$13.60	3	-1.0	0	1
3-Indoleacetic acid	1035						X					X			175.2	$26.65	6	1.4	1	0
Lipoic acid							X						X		131.2	$9.10	3	-1.8	0	0
Sodium DL-lactate													X		487.6	$14.45	8	-0.5	0	4
(±)-a-Lipoamide	1036						X						X		205.3	$10.20	4	1.5	0	1
3-Aminopropionitrile fumarate salt												X	X		143.2	$122.85	3	-1.9	0	1
D-Sorbitol			X			X						X	X		274.2	$72.60	5	0.2	3	3
Hydrocortisone (cortisol)						X		X					X		206.3	$206.28	2	3.7	1	0
3-Indoleacetic acid	1037						X					X			175.2	$15.55	6	1.4	1	0
Caffeine						X						X			652.3	$78.00	9	5.4	4	1
Phenylbutazone												X			174.2	$13.58	4	1.8	1	0
a-Lactose monohydrate	1038		X										X		255.7	$33.90	4	3.1	1	0
Betaine hydrochloride		X							X			X	X		330.2	$11.30	10	-0.1	0	2
Progesterone								X					X		116.1	$25.10	4	0.0	0	1
Serotonin hydrochloride								X		X			X		482.4	$10.70	1	2.2	0	5
3-(1-Naphthyl)-D-alanine	1039	X													215.3	$50.60	3	-0.4	2	0

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Histamine		X								X			X		182.2	$8.30	4	0.6	1	0
Sodium citrate tribasic dihydrate													X		211.1	$14.60	9	-1.7	0	0
(-)-Cotinine	1040								X						176.2	$52.55	3	-0.3	1	1
Porphobilinogen													X		312.4	$17.25	6	3.4	1	3
Sodium succinate dibasic							X						X		157.4	$27.52	5	2.8	1	1
Vasopressin						X		X		X					1084	$123.80	29	0.0	2	2
1-Octen-3-ol	1041						X						X		128.2	$20.50	1	2.4	0	0
N-Vanillylnonanamide													X		185.1	$10.85	8	-3.7	0	0
Phospho(enol)pyruvate potassium salt													X		184.2	$28.00	6	0.1	0	0
12-Hydroxydodecanoic acid	1042						X								216.3	$23.30	6	3.2	0	0
Adenine (Vitamine B4, 6-Aminopurine)				X							X		X	X	329.2	$17.20	12	-2.5	1	3
Sodium pyruvate													X		138.1	$0.00	3	0.3	1	0
D-(+)-Gluconic acid-lactone	1043		X										X		180.2	$15.00	6	-2.2	0	1
D-Panthenol (R)-(+)-2,4-Dihydroxy-N-(3-hydroxypropyl)-3,3-dimethylbutyramide)				X									X	X	182.2	$14.80	6	-2.0	0	0
Eserine (Physostigmine)						X						X			272.4	$20.40	2	3.8	1	3
Menadione (2-Methyl-1,4-naphthoquinone, vitamin k3)				X										X	447.6	$21.30	6	-1.7	1	1
Lithium acetoacetate	1044						X								138	$55.80	6	-1.3	0	0
N-Acetyl-L-tryptophanamide													X		309.3	$24.90	6	-2.0	3	1
Novobiocin sodium salt												X	X		213.2	$25.90	3	4.3	2	0
3,5-Diiodo-L-tyrosine dihydrate	1045											X	X		122.2	$13.50	2	2.4	2	0
3-Iodo-L-tyrosine		X										X	X		307.1	$22.60	3	-1.2	1	1
Cyclo(His-Pro)		X												X	234.3	$51.00	6	-1.9	0	3
(±)-Carnitine chloride	1046						X						X		176.2	$8.41	4	-4.8	0	0
N-p-Tosyl-L-phenylalanine chloromethyl ketone												X			624.6	$44.70	13	-3.5	2	1
Theophylline (2,6-dihydroxy-1,3dimethylpurine						X						X			182.2	$19.90	6	-0.8	0	2
1-(4-Chlorobenzyl)-5-methoxy-2-methylindole-3-acetic acid	1047											X			343.8	$79.60	4	1.8	0	0
Buspirone hydrochloride						X							X		194.2	$9.10	6	0.0	0	2
4-Deoxypyridoxine hydrochloride	1048			X										X	153.2	$38.00	3	0.7	1	0
5-Bromo-4-chloro-3-indolylb-D-galactopyranoside			X										X		408.6	$38.40	4	5.3	3	2
6,7-Dimethyl-5,6,7,8-tetrahydropterine hydrochloride				X											195.2	$10.80	6	2.9	1	1
Guanosine 5'-triphosphate sodium salt hydrate (GTP)											X		X		291.4	$77.40	5	1.6	2	1
Boc-L-phenylalaninol	1049	X													294.4	$27.10	3	1.4	1	0
Ellipticine					X	X						X			277.4	$110.50	6	2.0	1	1
Sepiapterin				X									X		176.2	$19.80	3	0.8	1	1

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Mitoxantrone dihydrochloride	1050				X	X									243.2	$33.70	6	6.4	1	2
N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide													X		177.2	$12.00	2	3.0	1	0
N-tert-Butyl-a-phenylnitrone												X	X		241.3	$67.70	2	-0.6	0	0
Thioctic acid				X									X	X	206.3	$13.60	4	2.4	0	1
Aztreonam	1051					X									249.3	$77.15	4	2.3	1	0
Chelerythrine chloride												X			183.1	$63.90	6	-1.6	0	1
D-(+)-Neopterin				X									X		335.5	$17.70	6	2.7	2	0
[Hydroxy(tosyloxy)iodo]benzene	1052												X		390.2	$23.00	4	2.2	2	0
2-Chloro-5,5-dimethyl-1,3-cyclohexanedione												X			174.6	$42.00	3	1.2	0	1
Phosphoramidon disodium salt												X			376.4	$48.75	7	1.0	2	0
Xanthtoxin L-Methionine sulfoximine												X			180.2	$40.30	10	-0.5	2	3
5,5-Diphenylhydantoin	1053					X									252.3	$17.00	18	-1.7	2	0
L-Leucine-b-naphthylamide		X											X		149.2	$9.10	4	-3.2	0	0
Sodium creatine phosphate dibasic tetrahydrate													X		90.08	$35.25	3	-0.7	0	0
Warfarin						X									308.3	$12.00	10	1.7	2	2
Adenosine 5'-monophosphoric acid (AMP)	1054										X		X		505.2	$26.80	18	-4.6	1	2
Ethidium Bromide					X										141.2	$16.75	3	0.4	0	1
Sodium oxamate												X	X		88.06	$26.00	3	-0.9	0	0
1-Methylimidazole	1055	X										X			82.1	$18.60	2	1.3	0	0
2-Fluoroaniline													X		111.1	$13.30	3	-3.4	1	1
3,4-Dimethylaniline													X		621.9	$35.30	9	0.5	3	1
Dansylcadaverine												X			308.4	$19.10	7	2.2	2	0
1-Aminocyclopropanecarboxylic acid	1056												X		101.1	$38.50	5	-1.1	2	0
Pentaethylene glycol							X								278.3	$29.30	7	0.1	0	2
Phenylpyruvic acid									X				X		168	$85.40	7	-1.0	0	0
Tyramine		X											X		137.2	$7.60	7	-1.5	4	1
2-Chloro-1,3,2-dioxaphospholane	1057						X								387.4	$16.20	9	2.1	1	3
Acridine Orange base					X								X		225.2	$44.95	2	2.8	0	2
Oxacillin						X									261.2	$22.70	10	0.2	0	2
Tubercidin					X	X					X				266.3	$23.10	3	1.4	1	0
3-Aminopyridine	1058										X		X		186.2	$30.75	6	-0.2	0	0
Lithium potassium acetyl phosphate													X		131.2	$8.80	3	-1.7	0	0
S-(-)-Carbidopa						X						X			418.9	$34.50	12	-1.8	0	3
Triethylenetetramine						X									146.2	$14.50	4	-2.4	0	0
2-Amino-5-methylthiazole	1059												X		114.2	$17.90	3	0.7	0	1
3,4-Dimethylphenol													X		121.2	$16.50	4	-1.3	2	0
Bisbenzimide H 33258					X										356.1	$17.30	7	2.6	2	2
Cytidine 5'-monophosphate (CMP)											X		X		481.1	$27.10	20	-4.8	0	2
2-Amino-4-methylphenol	1060												X		123.2	$17.40	2	0.7	1	0

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
2-Aminopyridine											X				94.11	$10.80	2	0.3	2	0
Inosine 5'-monophosphate disodium salt											X		X		286.2	$26.70	1	9.4	3	1
2-Aminophenol	1061												X		109.1	$17.60	12	1.1	1	0
2-Ethylimidazole		X											X		96.13	$15.60	1	0.8	1	0
Kaempferol												X	X		189.2	$9.10	2	6.8	2	2
2-Ketoglutaric acid	1062	X											X		146.1	$16.70	4	-1.6	0	0
2-Methylimidazole		X											X		82.1	$6.70	2	-0.8	1	1
Bromophenol Blue													X		371.5	$36.30	7	1.2	1	3
Uridine 5'-monophosphate											X		X		324.2	$65.20	2	12.5	0	0
2'-Deoxyadenosine 5'-monophosphate (dAMP)	1063										X		X		331.2	$15.90	12	-2.7	1	2
Cordycepin					X						X				466.4	$27.80	10	2.1	4	0
Netropsin dihydrochloride hydrate					X										123.1	$10.50	3	0.8	1	0
2-Aminofluorene	1064				X										181.2	$35.00	1	2.9	2	1
2'-Deoxyadenosine 3'-monophosphate sodium salt											X				315.2	$72.80	11	-2.0	1	2
Didecyldimethylammonium bromide							X								393.4	$31.00	13	-2.1	0	0
2-Cyclohexen-1-one	1065												X		96.13	$19.60	7	-1.8	0	4
Cyclohexylamine													X		261.1	$27.80	7	0.8	0	1
Picrotin												X			414.4	$15.90	8	3.7	2	3
(-)-Bilobalide	1066											X			310.30	$75.80	7	-1.0	0	4
Metergoline						X						X			356.6	$51.60	4	5.3	2	2
Pseudothiohydantoin											X		X		186.2	$28.90	3	-2.2	1	2
Ureidosuccinic acid		X													176.1	$6.95	4	-0.4	0	1
2-Aminothiazol	1067	X											X		100.1	$15.90	2	0.1	0	0
2-Hydroxyethyl disulfide													X		154.3	$35.80	7	-1.0	0	0
Androsterone								X					X		188.2	$7.10	2	4.4	1	2
Dimethyl 2-oxoglutarate		X											X		224.2	$25.10	7	2.1	3	1
1-Methylhistamine dihydrochloride	1068								X						125.2	$72.70	3	-1.0	0	1
2',3'-Dideoxyadenosine											X	X			235.2	$48.90	3	0.5	0	4
DL-Thiorphan						X						X			301.4	$39.60	4	2.4	1	0
2,3-diphosphoglycerate pentasodium salt	1069											X			261	$25.05	12	-5.1	0	0
Hydralazine hydrochloride												X			297.7	$38.80	4	-0.7	1	0
N-Propyl-1,3-propanediamine							X								351.9	$14.30	13	2.9	0	1
Pyridoxal 5'-phosphate				X									X		169.2	$16.60	4	-0.8	1	0
(-)-Perillic acid	1070											X	X		166.22	$46.10	2	2.8	0	1
Benzamide												X	X		120.2	$21.25	2	0.7	1	0
D(+)-Galactosamine hydrochloride			X										X		178.1	$9.50	6	1.4	0	0
3-Quinolinecarboxylic acid	1071					X									173.2	$26.40	3	-0.4	3	1
Aldosterone								X					X		287.2	$21.94	3	-0.1	2	0
Alizarin Yellow R													X		136.1	$24.60	5	0.6	1	1

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Ampicillin						X						X			290.4	$6.95	8	0.1	0	0
Clofibrate	1072					X						X			230.1	$30.60	5	1.7	1	1
Furosemide						X						X			103.1	$11.20	3	-2.8	0	0
Ibuprofen						X									204.2	$34.25	4	1.8	1	2
Naproxen						X						X			223.3	$29.85	4	2.8	1	0
Bromocresol Green	1073												X		670	$11.40	7	1.9	0	0
Digitoxin			X			X						X			174.2	$14.30	2	3.6	0	1
Methyl 4-hydroxyphenylacetate													X		314.3	$59.60	6	3.1	2	1
Phenol Red													X		308.4	$19.95	4	2.4	2	1
6,9-Diamino-2-ethoxyacridine-DL-lactate monohydrate	1074					X									253.3	$21.40	4	3.4	3	0
Ciprofloxacin						X						X			277.2	$19.30	5	2.4	1	0
Cyclo(His-Pro)		X							X						281.3	$39.70	5	-0.3	0	2
N-Acetylneuraminic acid			X										X		291.4	$50.20	5	1.9	1	0
3-Chlorophenol	1075												X		268.3	$32.50	9	0.0	0	2
6,9-Diamino-2-ethoxyacridine-DL-lactate monohydrate						X									169.2	$51.00	4	-0.6	1	0
Orange II sodium salt													X		598.7	$23.20	12	-3.6	0	6
Phosphocholine chloride calcium salt tetrahydrate							X						X		543.5	$130.65	14	1.5	1	3
3-Chlorophenol	1076												X		128.6	$11.60	2	0.3	3	0
Acetazolamide						X						X			146.2	$8.14	3	-3.5	0	0
Benzylamine													X		366.5	$85.60	3	6.0	2	1
Methotrexate				X											454.4	$32.90	13	-0.5	3	3
Diethylstilbestrol	1077					X		X							761	$23.50	11	0.0	0	8
Doxorubicin hydrochloride					X	X							X		444.4	$18.80	10	-0.5	1	3
Psoralen					X	X									321.4	$47.20	5	4.7	2	2
Tolbutamide						X							X		270.4	$20.60	13	-0.5	0	6
(±)-Verapamil hydrochloride	1078					X						X	X		491.1	$54.10	4	0.6	1	1
Diclofenac sodium salt						X						X			326.6	$38.40	1	4.3	0	0
Estriol								X	X				X		314.4	$10.60	-	-0.7	-	-
Amoxicillin	1079					X						X			365.4	$14.00	2	-1.8	1	0
Bestatin hydrochloride												X			308.4	$21.00	3	-2.9	0	0
Dansylglycin						X							X		178.1	$66.60	4	1.2	1	1
Diaminobiotin									X						296.2	$34.90	5	4.6	2	0
5-Hydroxyindole-3-acetic acid	1080								X						191.2	$15.00	3	2.2	1	0
Benzoic acid													X		107.2	$8.10	1	3.6	1	0
H-7 dihydrochloride												X			211.7	$18.40	3	1.4	1	0
Resveratrol												X	X		244.2	$18.10	9	-2.9	0	2
1,3-Dimethyluric acid	1081								X		X				196.2	$12.90	6	-0.3	1	0
Chlortetracycline hydrochloride						X									169.6	$11.90	5	1.7	1	1

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Flutamide						X						X			294.3	$25.00	7	0.7	1	1
Kynurenic acid													X		161.2	$16.95	5	-2.5	0	0
(-)-Arctigenin	1082											X			358.39	$38.00	6	1.5	2	1
5-Phenylvaleric acid													X		178.2	$19.30	3	2.4	0	3
Agmatine sulfate salt		X								X			X		342.3	$12.15	7	1.0	1	2
Diminazene aceturate					X	X									311.4	$19.50	8	-0.9	1	1
Podophyllotoxin	1083					X									225.2	$157.60	5	-2.8	0	1
Roscovitine												X			226.2	$47.00	6	2.0	1	0
Silibinin													X		138	$58.40	3	-0.5	0	0
tert-Butyl carbazate		X				X									132.2	$14.40	4	0.3	0	0
4-Hydroxy-3-methoxyphenylglycol sulfate potassium salt	1084								X						264.3	$40.60	8	-1.5	1	0
Carisoprodol						X			X						427.5	$48.90	12	-1.2	0	3
Ethylenediaminetetraacetic acid												X			783.5	$32.10	26	-3.4	2	4
Mycophenolic acid						X									387.5	$34.80	7	2.6	0	0
5a-Androstane-3,17-dione	1085							X					X		302.5	$29.10	2	3.7	0	4
Lumicolchicine									X						382.5	$102.20	6	3.6	2	0
PTH-tryptophan		X													471.5	$39.20	12	0.0	2	2
Sodium cacodylate trihydrate													X		258.1	$12.24	10	-2.4	0	0
2-Fluoro-2-deoxy-D-glucose 6-phosphate barium salt	1086								X		X				262.1	$29.40	6	1.6	0	0
AY 9944												X			391.4	$52.50	4	6.4	2	1
Cimetidine						X						X	X		252.3	$37.60	12	-1.2	0	2
N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide															293.4	$26.05	3	-0.4	0	1
6-Phosphogluconic acid trisodium salt	1087						X						X		336.6	$55.95	2	9.2	0	0
Haloperidol metabolite I									X						111.1	$18.37	3	-1.0	0	1
Nitrendipine						X						X			612.6	$22.20	13	2.9	2	2
Resorufin acetate													X		228.2	$57.00	3	-0.5	3	0
(±)-Camphor	1088					X							X		152.2	$16.30	3	7.9	0	0
Amantadine hydrochloride						X									195.2	$24.10	4	-1.7	1	0
Oxolinic acid						X						X			194.2	$37.40	5	-0.3	1	0
Tyrphostin 25												X			202.2	$44.50	2	2.8	0	0
(±)-Propranolol hydrochloride	1089					X							X		259.3	$21.60	10	-1.9	0	3
Adrenochrome								X	X						312.3	$26.90	7	0.3	1	0
Astemizole						X						X	X		266.3	$23.70	7	6.1	1	0
Homovanillic acid									X				X		160.2	$28.10	4	1.0	2	0
Cefoxitin	1090					X						X			554.6	$38.90	18	-1.6	0	4
Cephalexin hydrate						X									348.4	$87.20	5	-3.5	4	1
Cyproheptadine hydrochloride						X						X			323.2	$21.50	12	-3.3	0	1

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
N-(1H-Benzotriazol-1-ylphenylmethyl)benzamide						X									184.2	$70.70	4	1.9	0	1
1-Octanol	1091						X					X	X		130.2	$20.00	8	2.6	1	0
Aflatoxin B1													X		130.2	$30.20	4	-1.8	0	0
Daunorubicin hydrochloride						X						X	X		1580	$16.25	37	0.0	2	7
p-Aminohippuric acid		X													218.2	$15.00	6	-4.5	0	0
(-)-Bicuculline methiodide	1092											X			382.39	$31.40	7	4.1	4	2
3-Hydroxy-DL-kynurenine									X						224.2	$38.70	23	0.0	3	1
7-Deazaguanine											X	X			216.2	$21.10	5	-1.1	0	1
trans-4-(Aminomethyl)cyclohexanecarboxylic acid		X										X			157.2	$25.10	3	-1.8	0	1
5-Fluoro-5'-deoxyuridine	1093				X						X	X			246.2	$34.50	6	1.0	2	0
5-Hydroxyisophthalic acid						X									182.1	$14.00	11	0.9	1	1
Chloramphenicol						X						X	X		274.8	$50.40	7	3.0	1	0
Dipyrone						X						X			198.2	$9.60	7	2.7	3	0
3-Hydroxy-4-methoxyphenethylamine hydrochloride	1094								X						167.2	$37.80	3	0.6	1	0
DL-4-Hydroxy-3-methoxymandelic acid									X						199.6	$24.80	3	-0.8	1	0
Mecamylamine hydrochloride						X						X			232.3	$41.60	4	1.0	1	1
Nifedipine						X						X			348.4	$34.50	7	3.4	1	1
3-Indoleacrylic acid	1095											X			187.2	$35.90	2	2.8	0	2
Clindamycin hydrochloride						X						X			242.7	$28.30	6	-2.6	1	0
Sulindac Sulfide									X			X			340.4	$31.80	4	-1.8	2	1
Taxifolin												X	X		304.3	$36.00	7	0.9	2	1
4-Hydroxytamoxifen	1096											X	X		387.5	$44.80	1	6.2	3	0
N-Acetylprocainamide hydrochloride									X				X		309.4	$42.00	7	3.4	1	4
Tamoxifen						X						X	X		371.5	$43.30	2	6.8	3	0
2,6-Diisopropylphenol	1097					X						X	X		178.3	$15.70	1	3.9	1	0
4-Guanidinobutyric acid		X							X			X			159.2	$27.40	6	-4.1	1	0
Tyrphostin 1												X			184.2	$28.80	3	-1.8	1	0
Atenolol	1098					X						X			288.4	$23.70	8	-1.3	0	1
Methotrexate hydrate						X						X			330.2	$163.10	10	1.7	1	2
Nadolol						X						X			232.2	$14.80	5	-1.1	1	1
Procaine hydrochloride						X						X			314.5	$13.20	2	3.8	0	4
Clonidine hydrochloride	1099					X							X		783.5	$45.15	28	-4.9	0	3
Prednisolone						X	X	X				X	X		236.3	$16.40	4	2.5	1	0
Ranitidine hydrochloride						X						X			413.5	$90.00	4	1.2	2	1
Timolol maleate salt						X									316.4	$35.20	8	1.6	0	2
(−)-Scopolamine hydrobromide trihydrate	1100					X						X			303.4	$34.90	5	0.3	1	3
Camptothecin												X			196.2	$35.40	12	-3.6	2	0

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Myristic acid							X								328.4	$30.80	5	3.7	3	1
Acycloguanosine	1101										X	X	X		225.2	$49.30	7	2.6	1	0
Antipyrine						X									1300	$39.00	28	0.2	1	6
Ceftriaxone sodium salt						X						X			347.4	$38.30	8	-1.8	1	2
Cimetidine						X						X	X		262.2	$43.50	7	1.7	1	2
2,3,5-Triiodobenzoic acid	1102												X		499.8	$29.40	10	-2.1	0	2
3,3',5-Triiodothyroacetic acid								X					X		621.9	$35.30	0	0.0	0	1
Clenbuterol hydrochloride						X		X					X		423	$38.60	8	2.6	1	1
Tetracycline						X						X			446.5	$10.80	10	-1.5	1	3
4-Androstene-3,17-dione	1103							X					X		286.4	$17.10	2	2.8	0	4
Carbamazepine						X							X		190.2	$15.50	2	2.9	1	1
Amitriptyline hydrochloride	1104					X						X			365.4	$14.00	2	0.0	1	0
Cromolyn sodium salt						X									234.3	$51.00	6	-1.9	0	3
Sodium salicylate						X							X		116.1	$13.50	4	-6.5	0	0
2,3,5-Triiodobenzoic acid	1105												X		499.8	$29.40	10	-2.1	0	2
Amoxicillin						X						X			347.4	$36.70	7	-0.3	1	2
Dobutamine hydrochloride													X		153.2	$18.50	3	0.2	1	0
Pentoxifylline						X						X			685.9	$101.75	14	0.0	0	0
3-Methoxytyramine hydrochloride	1106								X				X		167.2	$31.10	3	0.6	1	0
5-Fluorocytosine					X						X				127.1	$61.80	3	0.7	1	0
Acycloguanosine											X	X	X		135.1	$23.50	5	-0.2	1	1
Aquocobalamin (Vitamin B12a)				X										X	458.6	$31.00	7	4.5	0	2
Bithionol	1107					X									265.4	$13.95	6	-4.7	1	4
Chlorpropamide						X		X							492.9	$20.00	11	0.5	1	3
Cyclophosphamide monohydrate					X	X									295.5	$27.80	1	6.0	1	3
Griseofulvin						X									345.2	$24.25	13	-3.5	0	4
(+)-Pseudoephedrine hydrochloride	1108					X		X							165.2	$30.20	2	0.9	1	0
Lidocaine hydrochloride						X						X			206.3	$9.90	7	1.8	3	1
Minoxidil sulfate salt						X									444.5	$23.70	10	1.1	2	1
(±)-Metoprolol (+)-tartrate salt	1109														267.4	$13.80	4	1.5	1	0
Hydrochlorothiazide												X			362.5	$13.70	5	1.7	0	4
N,N'-Dicyano-2,5-dimethylbenzoquinone-diimine						X									221.2	$22.60	8	4.3	3	0
Sulfaphenazole												X			314.4	$19.90	7	6.0	2	1
2-Acetamidophenol	1110								X						151.2	$21.80	9	-1.1	0	3
5-Azacytidine											X	X			244.2	$0.00	2	0.5	3	0
Brefeldin A						X			X						698	$9.25	6	3.3	1	1
Ebselen												X			246.3	$25.30	4	0.2	2	1
Acebutolol hydrochloride	1111					X						X			59.07	$19.95	3	0.4	0	1

(Table 1) contd.....

	Mixture	Amino Acid	Carbohydrates	Co-factor	DNA/RNA	Drug	Fatty Acid/Lipid	Hormone	Metabolite	Neurotransmitter	Neucleotide	Protein Inhibitor	Substrate	Vitamin	MW	Cost	Heteratoms	clogP	Aromatic Rings	Aliphatic Rings
Molsidomine						X									320.3	$38.20	6	4.6	1	0
Myriocin							X					X			214.3	$25.00	2	5.6	0	0
N-Acetyl-D-galactosamine			X										X		472.4	$45.30	11	4.4	2	1
1-Phenyl-1-cyclopropanecarboxylic acid	1112	X													162.2	$10.20	8	-2.9	2	0
2-Pyridylacetic acid hydrochloride									X		X				137.1	$29.00	3	-0.1	1	0
5-Aminosalicylic acid												X	X		153.1	$12.80	4	1.1	1	0
1,7-Dimethylxanthine	1113								X				X		180.2	$30.40	6	2.3	1	1
Dimethyl 4-methoxyisophthalate									X						281.3	$36.70	7	0.9	2	0
Ethyl 3-pyridylacetate						X					X				292.2	$18.60	10	-1.9	0	0

REFERENCES

[1] Fernandes, P.B. *Curr. Opin. Chem. Biol.*, **1998,** *2*(5), 597.

[2] Gonzalez, J.E.; Negulescu, P.A. *Curr. Opin. Biotech.*, **1998,** *9*(6), 624.

[3] Kenny, B.A.; Bushfield, M.; Parry-Smith, D.J.; Fogarty, S.; Treherne, J.M. *Prog. Drug Res.*, **1998,** *51*, 245.

[4] Oldenburg, K.R. *Annu. Rep. Med. Chem.*, **1998,** *33*, 301.

[5] Armstrong, J.W. *Am. Biotech. Lab.*, **1999,** *17*(5), 26.

[6] Koch, M.A.; Waldmann, H. *Method Principle Med. Chem.*, **2005,** *22*, 377.

[7] Messer, R.; Fuhrer, C.A.; Haener, R. *Curr. Opin. Chem. Biol.*, **2005,** *9*(3), 259.

[8] Golebiowski, A.; Klopfenstein, S.R.; Portlock, D.E. *Curr. Opin. Chem. Biol.*, **2003,** *7*(3), 308.

[9] Nilakantan, R.; Nunn, D.S. *Drug Discov. Today*, **2003,** *8*(15), 668.

[10] Geysen, H.M.; Schoenen, F.; Wagner, D.; Wagner, R. *Nat. Rev. Drug Discov.*, **2003,** *2*(3), 222.

[11] Lewis, R.A.; Pickett, S.D.; Clark, D.E. *Rev. Comp. Chem.*, **2000,** *16*, 1.

[12] Xue, L.; Bajorath, J. *Comb. Chem. High Throughput Screen.*, **2000,** *3*, (5), 363.

[13] Xu, J.; Stevenson, J. *J. Chem. Inf. Comp. Sci.*, **2000,** *40*(5), 1177.

[14] Willett, P. *Curr. Opin. Biotechnol.*, **2000,** *11*(1), 85.

[15] Spellmeyer, D.C.; Grootenhuis, P.D.J. *Annu. Rep. Med. Chem.*, **1999,** *34*, 287.

[16] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug Deliv. Rev.*, **1997,** *23*(1-3), 3.

[17] Lipinski, C.A. *J. Pharmacol. Toxicol. Method*, **2001,** *44*(1), 235.

[18] Gillet, V.J. *Method Mol. Biol.*, **2004,** *275*, 335.

[19] Brown, R.D.; Hassan, M.; Waldman, M. *J. Mol. Graph. Model.*, **2000,** *18*(4/5), 427.

[20] Fischer, P.M. *Curr. Med. Chem.*, **2004,** *11*(12), 1563.

[21] Matter, H.; Baringhaus, K.-H.; Naumann, T.; Klabunde, T.; Pirard, B. *Comb. Chem. High Throughput Screen.*, **2001,** *4*, 453.

[22] Stenberg, P.; Bergstrom, C.A.S.; Kthman, K.; Arthusson, P. *Clin. Pharmacokinet.*, **2002,** *41*(11), 877.

[23] Zartler, E.R.; Hanson, J.; Jones, B.E.; Kline, A.D.; Martin, G.; Mo, H.; Shapiro, M.J.; Wang, R.; Wu, H.; Yan, J. *J. Am. Chem. Soc.*, **2003,** *125*(36), 10941.

[24] Fejzo, J.; Lepre, C.; Xie, X. *Curr. Top. Med. Chem.*, **2003,** *3*, 81.

[25] Lepre, C.A.; Peng, J.; Fejzo, J.; Abdul-Manan, N.; Pcoas, J.; Jacobs, M.; Xie, X.; Moore, J.M. *Comb. Chem. High Throughput Screen.*, **2002,** *5*, 583.

[26] Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R.E. *J. Chem. Inf. Comp. Sci.*, **2004,** *44*(6), 2157.

[27] Jacoby, E.; Davies, J.; Blommers, M.J.J. *Curr. Top. Med. Chem.*, **2003,** *3*(1), 11.

[28] Ferna´ndez, C.S.; Jahnke, W. *Drug Discov. Today Tech.*, **2004,** *3*(1), 277.

[29] Powers, R. *J. Struct. Funct. Genomics*, **2002,** *2*, 113.

[30] Lepre, C.A.; Peng, J.; Fejzo, J.; Abdul-Manan, N.; Pocas, J.; Jacobs, M.; Xie, X.; Moore, J.M. *Comb. Chem. High Throughput Screen.*, **2002,** *5*,(8), 583.

[31] Stockman, B.J.; Dalvit, C. *Prog. NMR*, **2002,** *41*(3-4), 187.

[32] Cragg, G.M.; Newman, D.J. *PPSE*, **2002,** *47*, 285.

[33] Parsons, L.; Orban, J. *Curr. Opin. Drug Discov.*, **2004,** *7*(1), 62.

[34] Venter, C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; Gocayne, J.D.; Amanatides, P.; Ballew, R.M.; Huson, D.H.; Wortman, J.R.; Zhang, Q.; Kodira, C.D.; Zheng, X.H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P.D.; Zhang, J.; Miklos, G.L.G.; Nelson, C.; Broder, S.; Clark, A.G.; Nadeau, J.; McKusick, V.A.; Zinder, N.; Levine, A.J.; Roberts, R.J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A.E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T.J.; Higgins, M.E.; Ji, R.-R.; Ke, Z.; Ketchum, K.A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G.V.; Milshina, N.; Moore, H.M.; Naik, A.K.; Narayan, V.A.; Neelam, B.; Nusskern, D.; Rusch, D.B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.Y.; Wang, A.; Wang, X.; Wang, J.; Wei, M.-H.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.C.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M.L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y.-H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N.N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J.F.; Guigó, R.; Campbell, M.J.; Sjolander, K.V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert,

R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y.-H.; Coyne, M.; Dahlke, C.; Mays, A. D.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. *Science*, **2001**, *291*(5507), 1304.

[35] Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J.P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J.C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R.H.; Wilson, R. K.; Hillier, L.W.; McPherson, J.D.; Marra, M.A.; Mardis, E.R.; Fulton, L.A.; Chinwalla, A.T.; Pepin, K.H.; Gish, W.R.; Chissoe, S.L.; Wendl, M.C.; Delehaunty, K.D.; Miner, T.L.; Delehaunty, A.; Kramer, J.B.; Cook, L.L.; Fulton, R.S.; Johnson, D.L.; Minx, P.J.; Clifton, S.W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J.F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R.A.; Muzny, D.M.; Scherer, S.E.; Bouck, J.B.; Sodergren, E.J.; Worley, K.C.; Rives, C.M.; Gorrell, J.H.; Metzker, M.L.; Naylor, S.L.; Kucherlapati, R.S.; Nelson, D.L.; Weinstock, G.M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D.R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H.M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R.W.; Federspiel, N.A.; Abola, A.P.; Proctor, M.J.; Myers, R.M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D.R.; Olson, M.V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G.A.; Athanasiou, M.; Schultz, R.; Roe, B.A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W.R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J.A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D.G.; Burge, C.B.; Cerutti, L.; Chen, H.C.; Church, D.; Clamp, M.; Copley, R.R.; Doerks, T.; Eddy, S.R.; Eichler, E.E.; Furey, T.S.; Galagan, J.; Gilbert, J.G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L.S.; Jones, T.A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W.J.; Kitts, P.; Koonin, E.V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T.M.; McLysaght, A.; Mikkelsen, T.; Moran, J.V.; Mulder, N.; Pollara, V.J.; Ponting, C.P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A.F.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y.I.; Wolfe, K.H.; Yang, S.P.; Yeh, R.F.; Collins, F.; Guyer, M.S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K.A.; Patrinos, A.; Morgan, M.J.; Szustakowki, J.; de Jong, P.; Catanese, J.J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y.J.; Consortium, I.H.G.S. *Nature*, **2001**, *409*, 860.

[36] Pouliot, Y.; Gao, J.; Su, Q.J.; Liu, G.G.; Ling, X.B. *Genome Res.*, **2001**, *11*(10), 1766.

[37] Brenner, S.E. *Nat. Struct. Biol.*, **2000**, *7*(Suppl.), 967.

[38] Jensen, L.J.; Gupta, R.; Staerfeldt, H.H.; Brunak, S. *Bioinformatics*, **2003**, *19*(5), 635.

[39] Crawford, I.P. *Ann. Rev. Microbiol.*, **1989**, *43*, 567.

[40] Whisstock, J.C.; Lesk, A.M. *Quart. Rev. Biophys.*, **2003**, *36*(3), 307.

[41] Chakrabarti, R.; Klibanov, A.M.; Friesner, R.A. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*(29), 10153.

[42] Baxter, S.M.; Fetrow, J.S. *Curr. Opin. Drug Discov.*, **2001**, *4*(3), 291.

[43] Lan, N.; Jansen, R.; Gerstein, M. *Proc. IEEE*, **2002**, *90*(12), 1848.

[44] Brenner, S.E. *Nat. Struct. Biol.*, **2000**, *7*(Suppl.), 967.

[45] Schomburg, I.; Chang, A.; Ebeling, C.; Gremse, M.; Heldt, C.; Huhn, G.; Schomburg, D. *Nucleic Acids Res.*, **2004**, *32*, D431.

[46] Povolna, V.; Dixon, S.; Weininger, D. *Meth. Principle Med. Chem.*, **2005**, *23*, 241.

[47] Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. *Nucleic Acids Res.*, **2004**, *32*(Database), D277.

[48] Zhang, P.; Foerster, H.; Tissier, C.P.; Mueller, L.; Paley, S.; Karp, P.D.; Rhee, S.Y. *Plant Phys.*, **2005**, *138*(1), 27.

[49] Berman, H.; Henrich, K.; Nakamura, H. *Nat. Struct. Biol.*, **2003**, *10*(12), 980.

[50] Olah, M.; Mracec, M.; Ostopovici, L.; Rad, R.; Bora, A.; Hadaruga, N.; Olah, I.; Banda, M.; Simon, Z.; Mracec, M.; Oprea, T.I. *Meth. Principle Med. Chem.*, **2005**, *23*, 223.

[51] Fielding, L. *Curr. Top. Med. Chem. (Hilversum, Netherlands)*, **2003**, *3*(1), 39.

[52] Lepre, C.A.; Moore, J.M.; Peng, J.W. *Chem. Rev. (Washington, DC, United States)*, **2004**, *104*(8), 3641.

[53] Pembroke, J.T. *Biochem. Mol. Biol. Ed.*, **2000**, *28*(6), 297.

[54] Zheng, S.; Luo, X.; Chen, G.; Zhu, W.; Shen, J.; Chen, K.; Jiang, H. *J. Chem. Inf. Model.*, **2005**, *45*(4), 856.

[55] Gangal, R. *Molecules*, **2002**, *7*(8), 657.

[56] Menard, P.R.; Mason, J.S.; Morize, I.; Bauerschmidt, S. *J. Chem. Inf. Comp. Sci.*, **1998**, *38*(6), 1204.

[57] Tong, W.; Xie, Q.; Hong, H.; Shi, L.; Fang, H.; Perkins, R. *Environ. Health Persp.*, **2004**, *112*(12), 1249.

[58] Hajduk, P.; Betz, S.F.; Mack, J.; Ruan, X.; Towne, D.L.; Lerner, C.G.; Beutel, B.A.; Fesik, S.W. *J. Biomol. Scr.*, **2002**, *7*(5), 429.

[59] Eros, D.; Keri, G.; Kovesdi, I.; Szantai-Kis, C.; Meszaros, G.; Orfi, L. *Mini Rev. Med. Chem.*, **2004**, *4*(2), 167.

[60] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic Acids Res.*, **2000**, *28*(1), 235.

[61] Powers, R.; Copeland, J.C.; Germer, K.; Mercier, K.A.; Ramanathan, V.; Revesz, P. *PROTEINS: Struct. Funct. Bioinformatics*, **2006**, *in press*.

[62] Feher, M.; Schmidt Jonathan, M. *J. Chem. Inf. Comp. Sci.*, **2003**, *43*(1), 218.

[63] Mercier, K.A.; Powers, R. *J. Biomol. NMR*, **2005**, *31*(3), 243.

[64] Jahnke, W.; Widmer, H. *Cell Mol. Life Sci.*, **2004**, *61*(5), 580.

[65] Lepre, C.A.; Peng, J.; Fejzo, J.; Abdul-Manan, N.; Pocas, J.; Jacobs, M.; Xie, X.; Moore, J.M. *Comb. Chem. High Throughput Screen.*, **2002**, *5*(8), 583.

[66] Dove, A.; Marshall, A. *Nat. Biotech.*, **1999**, *17*(9), 859.

Received: October 18, 2005 **Revised: May 5, 2006** **Accepted: May 17, 2006**