# Leveraging structural approaches: applications of NMR-based screening and X-ray crystallography for inhibitor design

**Jonathan Moore,* Norzehan Abdul-Manan, Jasna Fejzo, Marc Jacobs, Christopher Lepre, Jeffrey Peng and Xiaoling Xie**

*Vertex Pharmaceuticals Incorporated, USA.*
*E-mail: moore@vpharm.com*

In the last several years, NMR strategies in drug discovery have evolved from a primarily structural focus to a set of technologies that are non-structural in nature but that have a much greater impact on the identification and optimization of real drug leads. NMR-based screening methods, such as the SHAPES strategy, help rapidly identify good starting points for drug design in a relatively high throughput implementation. The SHAPES method uses simple NMR techniques to detect binding of a limited, but diverse library of low molecular weight, soluble compounds to a potential drug target. SHAPES library compounds are derived largely from molecular frameworks most commonly found in known therapeutic agents. The NMR experiments used in these protocols are based on the well-known NMR techniques, and may be applied to targets with no limitation on molecular weight and no requirement for isotope labeling. Following screening, SHAPES hits may be used to guide virtual screening, synthesis of combinatorial libraries, and bias the first compounds that undergo high throughput screening. Integration of the SHAPES strategy with iterative X-ray crystallographic structure determination can be very useful in deriving an initial structural pharmacophore model and achieving significant *in* vitro potency in a short time frame. In this lecture, we will provide examples of how the combination of NMR SHAPES screening, virtual screening, molecular modeling and X-ray crystallography has led to novel drug scaffolds in several drug discovery programs: JNK3 MAP kinase and the fatty acid binding protein, aP2.

**Keywords: NMR-based screening; X-ray crystallography; drug design.**

## 1. Introduction

The SHAPES strategy is a process developed in our laboratory that combines nuclear magnetic resonance (NMR) screening of small drug-like molecules with other methods (such as enzymatic assays, conventional high-throughput screening (HTS), biophysical methods, combinatorial chemistry, X-ray crystallography, and molecular modeling) in a directed search to find new (or improve existing) leads for medicinal chemistry. Previous publications have presented the experimental details (Fejzo et al. (1999); Peng et al. (2001)) and described the design of the SHAPES screening libraries (Lepre (2001); Lepre (2002)). In this paper, we focus upon recent applications of the SHAPES strategy at Vertex, illustrating several different implementations of the method.

## 2. The SHAPES strategy – a synopsis

### 2.1. Origins of the SHAPES strategy

The principal assumption of the SHAPES screening strategy is that information regarding molecules that bind to a target weakly may be used to direct the design of more potent inhibitors of that target. An additional premise is that compounds containing molecular fragments found in marketed drugs are more likely to possess good clinical properties than random molecules. The idea of generating a NMR screening library from scaffolds found in known drugs was derived from a modeling study carried out at by Bemis and Murcko (Bemis and Murcko (1996)). In this study, 5120 drugs with a wide spectrum of therapeutic indications were selected from the Comprehensive Medicinal Chemistry (CMC) database and computationally reduced to graph frameworks (i.e., unadorned rings and linkers). The authors found that only 32 simple frameworks could describe approximately half of the drugs in the CMC database. When atom type and bond order were included, only 41 "complex" frameworks, or scaffolds, described about a quarter of the known drugs. This suggested that a relatively small number of molecular "shapes" could be used as a universal screening library for a wide variety of therapeutic targets.

### 2.2. Library design

Based on our assumption that commonly recurring drug scaffolds are predisposed to having favorable clinical properties such as good oral bioavailability, low toxicity, and metabolic stability, the original SHAPES library (Fejzo, et al. (1999)) was comprised of compounds containing the most common scaffolds found in known drugs, combined with commonly occurring drug side chains (Bemis and Murcko (1999)). Subsequently, the library has been expanded to incorporate drug-like and "lead-like" (Teague et al. (1999)) compounds that contain linkages amenable to combinatorial chemistry, are targeted against targets from a particular gene family, or exhibit diverse pharmacophores (Lepre (2001); Lepre (2002))].

The SHAPES library has been designed to satisfy stringent criteria for compound solubility, drug-likeness (or lead-likeness), and synthetic accessibility. This process has been previously described in detail (Lepre (2001); Lepre (2002)). Starting from a database of over 1 million commercially available compounds, computational filters were applied to remove compounds predicted to have poor solubilities, poor physicochemical properties (using "Rule of 5" (Lipinski et al. (1997)) and other criteria (Lepre (2001))), and possessing undesirable functional groups (Walters and Murcko (2000); Walters et al. (1998)). Candidates surviving the filtering process were then clustered using the Jarvis-Patrick method (Jarvis and Patrick (1973)), and individual compounds were picked manually from the centroids of the clusters. Mixtures of compounds were designed so as to minimize resonance overlap and avoid interactions between the components.

### 2.3. NMR methods

Because the molecules in the SHAPES screening library are typically smaller (< 350 MW) than most compounds screened by HTS, they are expected to have lower average affinities. In order to afford the broadest possible detection range, we employ NMR-based methods that are capable of detecting molecules that bind with dissociation constants ranging from low micromolar (and lower for some methods (Dalvit et al. (2002a); Dalvit et al. (2002b)) to millimolar in a single experiment. Although protein-based detection methods (Pellecchia et al. (2002); Shuker et al. (1996)) are very useful for mapping the binding sites of hits on protein targets, ligand-detected methods for primary screening consume less reagents, do not require isotopic labeling, and may be used for many types of targets with no limit to molecular mass. One preferred method for detection of ligand binding is saturation transfer difference spectroscopy (Mayer and Meyer (1999)), which offers high sensitivity, low protein consumption, and the ability to tune the

  

detection range of the experiment (Peng, et al. (2001)). SHAPES screening hits are validated by detection of inhibition (e.g., by enzymatic assay), competition with a known inhibitor (e.g., by NMR displacement experiment or calorimetry), or localization to the active site (e.g., by NMR chemical shift mapping experiments or crystallographic soaking studies)..

#### 2.4. SHAPES implementation

SHAPES screening has been implemented at Vertex at several stages of project development: pre-HTS, post-HTS, and during lead optimization. It is most commonly applied at the pre-HTS stage, when a new protein first becomes available and the HTS screen is still under development. Because NMR screening requires no assay development and only a few milligrams of protein, it is possible to find SHAPES hits, then purchase analogs or synthesize a combinatorial library around these hits quickly enough to be included in an initial HTS run.

Validated hits from pre-HTS SHAPES screens are most often followed up by purchasing analogs, as has been described elsewhere (Fejzo (2002); Lepre (2001)). Retrospective analysis of HTS screening results has shown that compounds containing SHAPES scaffolds (those shown to bind by NMR) typically have hit rates about ten-fold higher than general screening compounds that do not contain those scaffolds (Fejzo (2002)). Higher enhancements are possible by designing combinatorial libraries around SHAPES scaffolds (Fejzo (2002)). The observation of such enhancements when screening follow-up libraries based on SHAPES hits – more examples of which are given below – supports the assumption that binding information from compounds with weak affinity may lead to design of significantly (e.g. several orders of magnitude) more potent inhibitors.

## 3. Specific applications

### 3.1. Jnk3 MAP kinase

JNKs (c-Jun N-terminal kinases) are stress-activated serine/threonine kinases that phosphorylate c-Jun (reviewed by (Davis (2000))). Apoptotic neurons exhibit increased expression of the Jnk3 isoform, marking it as a potential target for the treatment of stroke and Parkinson's disease (Bruckner et al. (2001); Xia et al. (2001)). Efforts to design inhibitors of Jnk3 have targeted the ATP-binding site, which is located near the activation loop in a cleft between the C- and N-terminal domains.

Jnk3 was one of the first targets screened by the SHAPES method, using an initial version of the SHAPES library that only contained about 100 compounds (Fejzo (2002)). This screen yielded 17 hits with very weak affinities (estimated to be in the millimolar range), 13 of which were confirmed to target the ATP site since their binding could be blocked by 5'-adensosine-4-fluorosulfonyl benzoic acid (FSBA), a covalently bound ATP analog.

Attempts to soak the SHAPES hits into crystals of apo-Jnk3 for crystallographic studies were unsuccessful (presumably due to their low affinity). In the absence of information about the bound orientations of the hits, two approaches were taken to design follow-up libraries using commercially available compounds. First, approximately 200 analogs were chosen for purchase by using substructure-matching and similarity-based computational searches of the Available Chemicals Directory (ACD) database (1997, MDL Information Systems), as illustrated for two scaffolds in Figure 1.

Second, compounds were selected using a virtual screening procedure. Initially, multiple conformations of the original 17 SHAPES screening hits were computationally docked into the ATP site of the Jnk3 crystal structure. Of these, only four compounds produced distinct, energetically reasonable bound orientations. A substructure search of the ACD database was then carried out using the scaffolds from the four preferred fragments, followed by filtering for drug-like character (Walters and Murcko (2000); Walters, et al. (1998)), to give a starting set of 1647 analogs. Approximately 10,000 energy-minimized conformers of these compounds were then docked into the ATP site using an in-house genetic-algorithm-based program (Charifson et al. (1999)). The best 171 conformers from the consensus (Charifson, et al. (1999)) of four scoring functions (PLP, ChemScore, DOCK NRG, Volume Overlap) were then manually reduced to 94 top picks.

The approximately 300 follow-up compounds selected using the above substructure searching and virtual screening procedures were screened at 30 μM in an enzymatic assay, yielding 8 inhibitors with potency better than 20 μM. This hit rate was approximately tenfold higher than was observed in the HTS screen of Jnk3 using random compounds. It is noteworthy that none of the inhibitor classes derived from the SHAPES hits were found in the the initial HTS screen of Jnk3, even though compounds containing those scaffolds were present in the HTS screening library. A likely explanation for this observation is that the compounds representing those scaffolds in the HTS library are more highly elaborated, which reduces the likelihood of binding. The unelaborated SHAPES scaffolds are more likely to bind, albeit with affinities too weak to detect reliably by enzymatic methods. These observations reinforce the notion that NMR screens of weakly binding scaffolds can be used to direct HTS to novel classes of compounds that might otherwise be missed.

The SHAPES leads were followed up by medicinal chemistry after efforts to develop the leads from the original HTS screen were unsuccessful. Several hundred compounds were prepared (mostly in combinatorial libraries) based on four SHAPES-derived leads, producing three sub-micromolar classes of compounds, the thiazoles, uracils and isoxazoles, shown in Figure 2. X-ray structures of examples from each class of compounds complexed to Jnk3 (Figure 3) allowed a pharmacophore model to be developed, which suggested clear avenues for improving the potency of these inhibitors. By incorporating substituents to access all available binding pockets in the active site, a number compounds were synthesized which were able to inhibit Jnk3 activity in the range < 10 nM. Ultimately, crystallographic structures of over sixty different compounds bound to Jnk3 were solved. These data, besides being useful for design of Jnk3 inhibitors, also provide insight that has been leveraged across numerous other potential targets in the kinase gene family.

### 3.2. Adipocyte lipid-binding protein

Intracellular fatty acid binding proteins (FABPs) are small (14-16 kDa) proteins that are responsible for lipid trafficking, mobilization, and metabolism in a variety of tissues. Adipocyte lipid-binding protein (ALBP, or aP2) is phosphorylated in response to insulin and appears to regulate both insulin resistance and hyperinsulinemia, making it a possible therapeutic target for the treatment of type II diabetes (Hotamisligil et al. (1996); Scheja et al. (1999); Uysal et al. (2000)). In addition, ALBP-deficient hypercholesterolemic mice are protected against atherosclerosis, suggesting it as a target for treating that disease (Layne et al. (2001); Makowski et al. (2001)).

In the aP2 project, initial SHAPES screening was undertaken to identify ligands for the active site that could be used for lead generation. Using the prototype 100-compound SHAPES library, thirteen hits were found that bound to ALBP. The ligands were subsequently validated as active-site ligands by fluorimetric assay as well as chemical shift mapping studies using $^{1}H$ and $^{15}N$ chemical

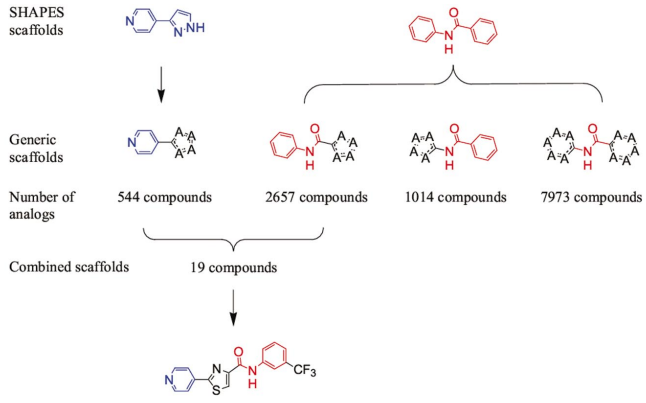


**Figure 1**

Scaffold-based search for commercially available analogs of Jnk3 SHAPES screening hits. Scaffolds from two of the original hits (top) were used to search the ACD database using the generic substructures shown in the second row. In the substructures, "A" denotes any atom and the dashed lines represent single or double bonds. Replacement of the pyrazole and phenyl rings with generic five and six-membered rings produced the number of analogs listed in the third row. Nineteen compounds contained analogs of both of the original SHAPES scaffolds, one of which (bottom) was found to have a $K_I$ of 13 μM. (Reproduced with permission from Fejzo, et. al. (2002)).

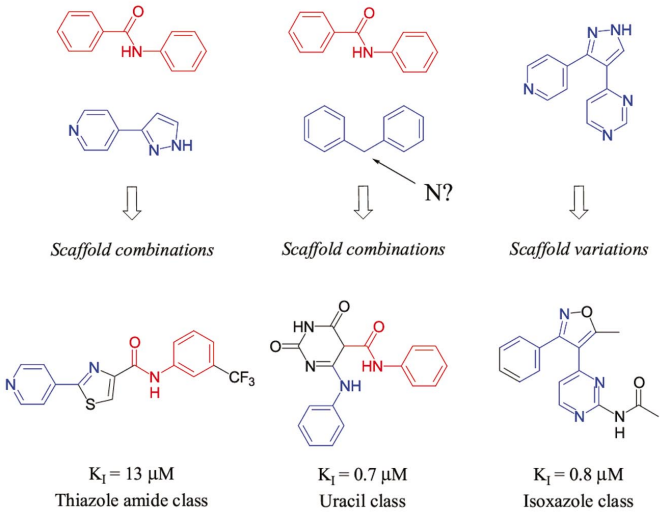


**Figure 2**

Evolution of SHAPES screening hits for Jnk3 into lead classes. Scaffolds from the primary SHAPES screening hits are shown in the top row. Follow-up compounds were selected from commercial sources based on combinations or variations of the active scaffolds. The resulting leads are shown at the bottom. The thiazole amide class (left) was derived from a combination of bisaryl amide and a variant of the pyridyl pyrazole scaffold. The uracil class (center) combined bisaryl amide and a variant of the bisphenyl methane scaffold. The isoxazole class (right) was based on variation of the 4,5-bisaryl pyrazole scaffold. Optimization of the isoxazole class led to multiple compounds with $K_I$ < 10 nM. (adapted from Fejzo et al . (2002)). An elaboration strategy was used, in that preference was given to compounds having additional functional groups capable of interacting with the protein and increasing potency. In addition, a fragment combination strategy was used, in which compounds were selected that contained two of the scaffolds found in the SHAPES hits (Figure 1).

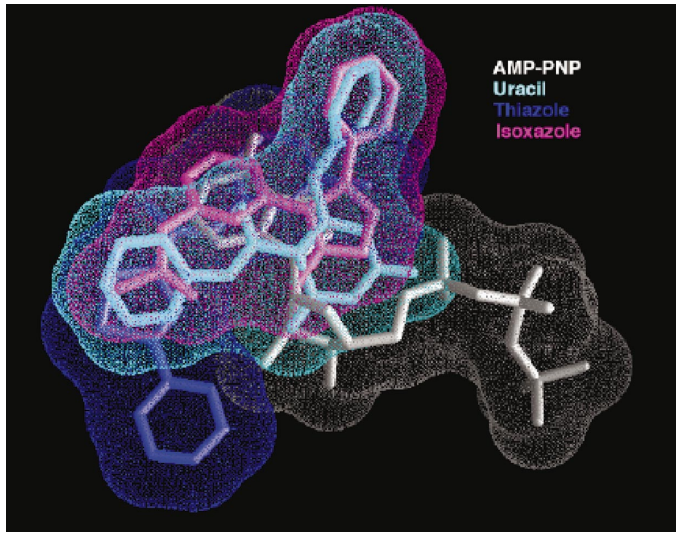


**Figure 3**

Superposition of the X-ray structures of the thiazole (blue), uracil (cyan) and isoxazole (magenta) compounds of Figure (2) with AMP-PNP (white) in the active site of Jnk3 (X. Xie, unpublished results). The comparison of these structures was used to extend the isoxazole compound to fill subsites occupied by the thiazole and uracil compounds (e.g., the lower left pocket), leading to increased potency.

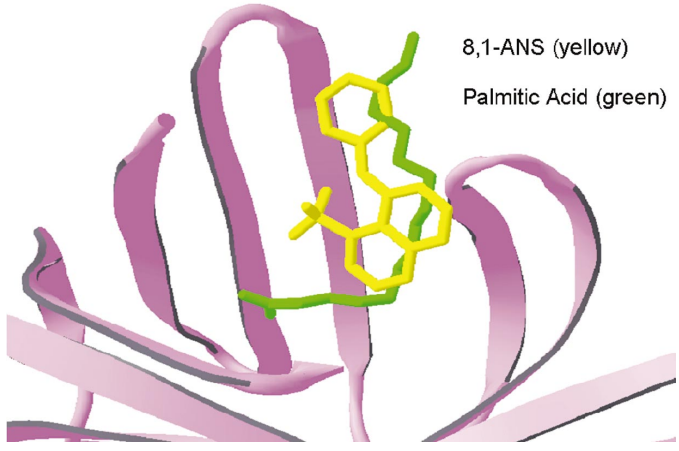


**Figure 4**

Cutaway view of the ALBP active site showing the bound structures of palmitic acid (green) (LaLonde et al. (1994)) and 1,8-ANS (yellow) (M. Jacobs, unpublished results). Binding of palmitic acid is stabilized by hydrogen bonds between the carboxylate and the sidechains of R126 and Y128, while 1,8-ANS employs a water-mediated hydrogen bond to R126.

shift perturbation experiments. The binding affinities of the validated hits ranged from 0.3 μM to 800 μM. One of these, 1-anilinonapthlalene-8-sulfonic acid (1,8-ANS), was found to have an apparent dissociation constant of 0.4 μM. 1,8-ANS binds to the active site of ALBP (see Figure 4), and because this probe has previously been used to assay lipid-binding proteins (Kane and Bernlohr (1996)), is a suitable molecule for the development of fluorescence-based competition assays directed at ALBP.

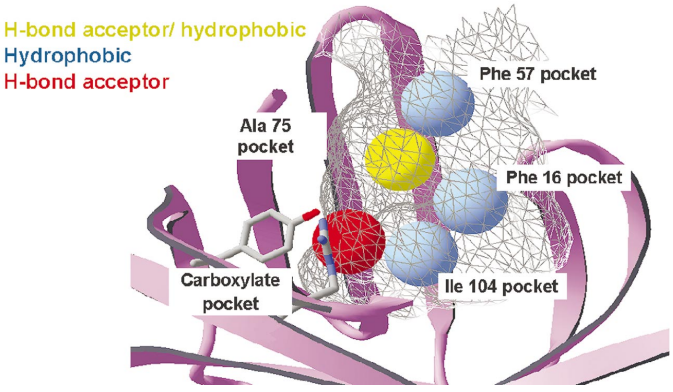


**Figure 5**

Pharmacophore model for ALBP derived from ligand binding and crystallographic studies.

In order to identify more potent binders of aP2, two of the primary SHAPES hits were successfully soaked into crystals of apo-ALBP and the structures of the complexes solved by X-ray crystallography. This information, along with molecular models of other hits, was used to identify possible sites of interaction with the protein. A set of twelve commercially available analogs (chosen to have high similarity to the primary SHAPES hits) was screened in order to explore variations at positions predicted to make favorable protein contacts. All of the analogs were active, and the bound structures of several were subsequently determined by X-ray crystallography. Based on this information, a second-generation follow-up library of 134 commercially available compounds was screened by calorimetry, yielding nine leads with low micromolar and nanomolar affinities, including one compound class with <100 nM affinity.

The fatty acid binding proteins are a unique class of targets, in that unlike most enzyme active sites, there is considerable space in the fatty acid binding pocket that can accomodate a variety of fatty acid substrates as well as bound water molecules. Understanding the binding modes of different fatty acids and ligand classes by solving multiple crystallographic structures allows one to derive a pharmocophore model describing key interactions in the binding pocket that one may access to achieve higher affinity ligands. A pharmacophore model for aP2 based on eight different crystallographic structures is shown in Figure 5.

This application demonstrates that using NMR-based screening, computational methods, and X-ray crystallography, it is possible to rapidly converge upon multiple sub-micromolar lead classes (in this case, requiring the testing of less than 250 compounds), using modest resources and without employing a HTS assay. The information derived from this work serves as an excellent starting point for a structure-based drug design program for this unique class of drug target.

### 4. Conclusions

NMR SHAPES screening can be an extremely useful tool for lead generation. Even in the absence of HTS hits, SHAPES screening can identify chemically accessible, drug-like fragments that can be rapidly translated into potent lead compounds. As illustrated in the examples, the unique ligand-binding data derived from NMR SHAPES screening and close integration with other enabling technologies (including x-ray crystallography, molecular modeling, calorimetry, and other biophysical techniques) can greatly accelerate discovery of lead molecules as well as aid in optimization of existing compounds.

### References

Bemis, G. W., & Murcko, M. A. (1996). *J. Med. Chem.* **39,** 2887-93.
Bemis, G. W., & Murcko, M. A. (1999). *J. Med. Chem.* **42,** 5095-9.
Bruckner, S. R., Tammariello, S. P., Kuan, C. Y., Flavell, R. A., Rakic, P., & Estus, S. (2001). *J. Neurochem.* **78,** 298-303.
Charifson, P. S., Corkery, J. J., Murcko, M. A., & Walters, W. P. (1999). *J Med Chem* **42,** 5100-9.
Dalvit, C., Fasolini, M., Flocco, M., Knapp, S., Pevarello, P., & Veronesi, M. (2002a). *J Med Chem* **45,** 2610-4.
Dalvit, C., Flocco, M., Knapp, S., Mostardini, M., Perego, R., Stockman, B. J., Veronesi, M., & Varasi, M. (2002b). *J Am Chem Soc* **124,** 7702-9.
Davis, R. J. (2000). *Cell* **103,** 239-52.
Fejzo, J., Lepre, C. A., Peng, J. W., Bemis, G. W., Ajay, Murcko, M. A., & Moore, J. M. (1999). *Chem. Biol.* **6,** 755-69.
Fejzo, J., Lepre, C., Xie, X. (2002). *Current Topics in Medicinal Chemistry* **2,** 1349-1364.
Hotamisligil, G. S., Johnson, R. S., Distel, R. J., Ellis, R., Papaioannou, V. E., & Spiegelman, B. M. (1996). *Science* **274,** 1377-9.
Jarvis, R. A., & Patrick, E. A. (1973). *IEEE Transactions on Computers* **C-22,** 1025-1034.
Kane, C. D., & Bernlohr, D. A. (1996). *Anal Biochem* **233,** 197-204.
LaLonde, J. M., Bernlohr, D. A., & Banaszak, L. J. (1994). *Biochemistry* **33,** 4885-4895.
Layne, M. D., Patel, A., Chen, Y. H., Rebel, V. I., Carvajal, I. M., Pellacani, A., Ith, B., Zhao, D., Schreiber, B. M., Yet, S. F., Lee, M. E., Storch, J., & Perrella, M. A. (2001). *Faseb J* **15,** 2733-5.
Lepre, C. A. (2001). *Drug Discovery Today* **6,** 133-140.
Lepre, C. A. (2002). *In* "BioNMR Techniques in Drug Research" (O. Zerbe, ed.), pp. 1349-1364. Wiley-VCH, Weinheim.
Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeny, P. J. (1997). *Advanced Drug Delivery Reviews* **23,** 3-25.
Makowski, L., Boord, J. B., Maeda, K., Babaev, V. R., Uysal, K. T., Morgan, M. A., Parker, R. A., Suttles, J., Fazio, S., Hotamisligil, G. S., & Linton, M. F. (2001). *Nat Med* **7,** 699-705.
Mayer, M., & Meyer, B. (1999). *Angew. Chem. Int. Ed.* **38,** 1784-1788.
Pellecchia, M., Meininger, D., Dong, Q., Chang, E., Jack, R., & Sem, D. S. (2002). *J Biomol NMR* **22,** 165-73.
Peng, J. W., Lepre, C. A., Fejzo, J., Abdul-Manan, N., & Moore, J. M. (2001). *Methods Enzymol* **338,** 202-30.
Scheja, L., Makowski, L., Uysal, K. T., Wiesbrock, S. M., Shimshek, D. R., Meyers, D. S., Morgan, M., Parker, R. A., & Hotamisligil, G. S. (1999). *Diabetes* **48,** 1987-94.
Shuker, S. B., Hajduk, P. J., Meadows, R. P., & Fesik, S. W. (1996). *Science* **274,** 1531-4.
Teague, S. J., Davis, A. M., Leeson, P. D., & Oprea, T. (1999). *Angew. Chem. Int. Ed.* **38,** 3743-3747.
Uysal, K. T., Scheja, L., Wiesbrock, S. M., Bonner-Weir, S., & Hotamisligil, G. S. (2000). *Endocrinology* **141,** 3388-96.
Walters, W. P., & Murcko, M. A. (2000). *In* "Virtual screening for bioactive molecules" (H.-J. Böhm and G. Schneider, eds.), Vol. 10, pp. 15-32. Wiley-VCH, New York.
Walters, W. P., Stahl, M. T., & Murcko, M. A. (1998). *Drug. Discov. Today* **3,** 160-178.
Xia, X. G., Harding, T., Weller, M., Bieneman, A., Uney, J. B., & Schulz, J. B. (2001). *Proc. Natl. Acad. Sci. USA* **98,** 10433-10438.