Journal of Cellular Biochemistry Supplement 37:99–105 (2001)

Object-Oriented Approach to Drug Design Enabled by NMR SOLVE: First Real-Time Structural Tool for Characterizing Protein–Ligand Interactions

Daniel S. Sem,* Lin Yu, Stephen M. Coutts, and Rick Jack

Triad Therapeutics, Inc., 5820 Nancy Ridge Drive, San Diego, California

Abstract As a result of genomics efforts, the number of protein drug targets is expected to increase by an order of magnitude. Functional genomics efforts are identifying these targets, while structural genomics efforts are determining structures for many of them. However, there is a significant gap in going from structural information for a protein target to a high affinity ($K_d < 100$ nM) inhibitor, and the problem is multiplied by the sheer number of new targets now available. nature frequently designs proteins in classes that are related by the reuse, through gene duplication events, of cofactor binding domains. This reuse of functional domains is an efficient way to build related proteins in that it is object-oriented. There is a growing realization that the most efficient drug design strategies for attacking the mass of targets coming from genomics efforts will be systems-based approaches that attack groups of related proteins in parallel. We propose that the most effective drug design strategy will be one that parallels the object-oriented manner by which nature designed the gene families themselves. IOPE (Integrated Object-Oriented PharmacoEngineering) is such an approach. It is a three-step technology to build focused combinatorial libraries of potential inhibitors for major families and sub-families of enzymes, using cogent NMR data derived from representatives of these protein families. The NMR SOLVE (Structurally Oriented Library Valency Engineering) data used to design these libraries are gathered in days, and data can be obtained for large proteins (> 170 kDa). Furthermore, the process is fully object-oriented in that once a given bi-ligand is identified for a target, potency is retained if different cofactor mimics are swapped. This gives the drug design process maximum flexibility, allowing for the more facile transition from in vitro potency to in vivo efficacy. J. Cell. Biochem. Suppl. 37: 99–105, 2001.

Key words: enzyme mechanomics; IOPE; NMR SOLVE; object-oriented drug design; structural genomics

There is currently a great need for an increase in the efficiency of the drug discovery process, since the cost is becoming prohibitively large, with the process taking > 10 years and costing several hundred million dollars on average for one drug to reach the market [Mathieu, 2000]. Now that genomes have been sequenced for human [International Human Genome Sequencing Consortium, 2001; Venter, 2001] as well as numerous pathogenic microbes [Cole, 1998; Alm, 1999; Tettelin, 2001], there is the expectation that functional genomics efforts will translate this information into thousands of new drug targets [Drews, 1998]. However, given that all previous drug discovery efforts have been fully occupied in the pursuit of only 500 targets, it is not clear that simply being presented with more targets will increase the efficiency of drug discovery, unless new approaches to target-driven drug discovery are developed.

OBJECT-ORIENTED DRUG DISCOVERY PARALLELS NATURE

Systems-based approaches to solving problems efficiently rely upon the grouping together of related tasks, and solving them in parallel [Senge et al., 1994]. Such a strategy may be applied to drug discovery in the post-genomic era by grouping proteins in classes that are related by binding sites, and pursuing inhibitor design for the whole class in parallel (Fig. 1). A class of proteins related by binding site properties is termed a pharmacofamily. A drug design strategy is needed that can efficiently attack proteins one pharmacofamilly at a time. In designing such a strategy, it is

*Correspondence to: Daniel S. Sem, Triad Therapeutics, Inc., 5820 Nancy Ridge Drive, San Diego, CA 92121.
E-mail: dsem@triadt.com

Received 13 September 2001; Accepted 14 September 2001

DOI 10.1002/jcb.10070

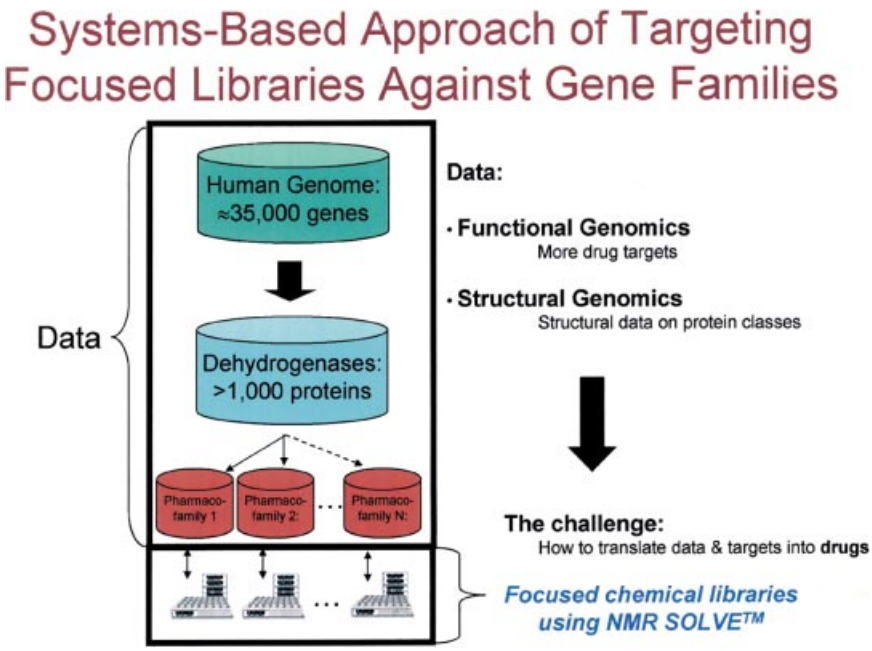

Fig. 1. Systems-based approach to designing drugs involves attacking entire pharmacofamilies of related proteins at once. The post-genomic challenge requires a novel strategy to drug design that makes more efficient use of drug target information.

illustrative to consider how nature has designed the gene-family or pharmacofamily. In general, nature does not design proteins de novo, but rather uses gene duplication events, whereby modules/domains with reusable functions are attached to other domains which confer a new, extended function. For example, the Rossman folds of dehydrogenases all bind NAD(P)(H) (a reusable function), and this domain is attached to different substrate domains to create novel functions such as NAD(P)H-based reduction of (a) dihydrodipicolinate, (b) aldehydes, and (c) pyruvate (Fig. 2). The concept of streamlining development efforts by reusing modules with conserved functions and attributes is called an object-oriented design strategy. Object-oriented approaches are well established, and have revolutionized the software industry. We propose that the most efficient systems-based approach to drug discovery will parallel nature's object-oriented design strategy, and that the science of object-oriented modeling [Booch et al., 1999] will find molecular parallels in the pharmaceutical industry. Indeed, from an operations perspective, object-oriented technologies are easily managed and industrialized, as long as the business and management strategy, as well as informatic infrastructure, mirrors the object-oriented problem [Taylor, 1995; Booch, 1996]. Thus, an object-oriented strategy lends itself well to efficient industrialization.

IOPE AS THE OPTIMAL OBJECT-ORIENTED DRUG DESIGN STRATEGY

IOPE (Integrated Object-Oriented Pharmacoengineering) is a technology, whereby

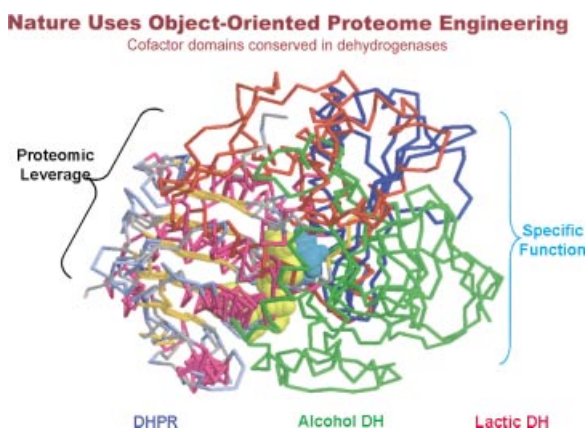

Fig. 2. The reuse of cofactor binding domains through gene duplication events suggests that nature used object-oriented strategies to efficiently create gene families of proteins within proteomes. Shown are the overlaid backbone traces for dihydrodipicolinate reductase (blue), alcohol dehydrogenase (green), and lactic acid dehydrogenase (red). The conserved Rossman fold is shown on the left, while the substrate binding domain is on the right.

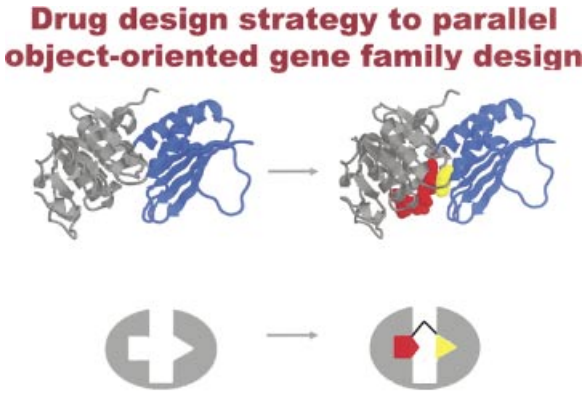

Fig. 3. A drug design strategy that parallels nature's object-oriented strategy for creating bi-ligand enzymes involves the use of bi-ligand inhibitors that are themselves object-oriented. That is, one half of the bi-ligand would correspond to a molecule that binds in the conserved cofactor binding domain. The top structures represent a dehydrogenase without (left) and with (right) cofactor (red), and substrate (yellow) bound. The bottom cartoon represents the same protein with a bi-ligand inhibitor that spans the common cofactor and variable substrate sites.

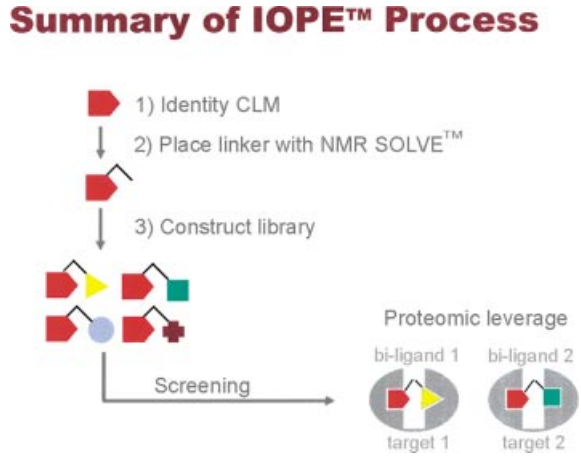

Fig. 4. Summary of the IOPE process. In step 1, a cofactor mimic (CLM) is identified. In step 2, the ideal location for placing a linker is identified using NMR SOLVE as a real-time structural probe. In step 3, a bi-ligand library is constructed whereby $< 10{,}000$ diversity elements are added to the end of the linker, such that they are directed into the substrate pocket. This library is then screened as a rich source of nM inhibitors for enzymes in the pharmacofamily that bind the starting CLM (in orange). Since the CLM is chosen to bind hundreds of proteins in a proteome, the library provides proteomic leverage to drug discovery.

collections of bi-ligand inhibitors are tailored to pharmacofamilies of proteins in such a way that parallels the object-oriented manner by which pharmacofamilies have been produced by nature (Fig. 3). IOPE is a three step process (Fig. 4) that starts with the identification of small molecules called common ligand mimics (CLMs) that bind with modest affinity to the cofactor domains of most members of a pharmacofamily. These CLM molecules bind in the NAD(P)(H) sites of dehydrogenases. Just as cofactor domains are conserved in the pharmacofamily of proteins, so too is the CLM molecular module/object conserved throughout a bi-ligand inhibitor collection where multiple diversity elements called specificity ligands (SLs) are added to the CLM (Fig. 4).

In the second step of IOPE, a structural tool called NMR SOLVE [Pellecchia et al., 2001] is used to determine whether the CLM is proximal to the substrate binding site (the SL site), and to determine what portion of the CLM is most proximal. It is here that a linker is added such that the linker is geometrically directed into the SL site. Just as the geometric relationship between cofactor and substrate sites is conserved in bi-ligand enzymes, so is the relationship between CLM and SL, as dictated by linker placement (Figs. 3 and 4).

In the third step of IOPE, a library of $< 10{,}000$ diversity elements is added to the end of the linker, as guided by NMR SOLVE (Fig. 4). Because of the powerful thermodynamic driving force of the chelate effect (10^8-fold effect on K_d [Page and Jencks, 1971]) coupled with the universal binding affinity of the CLM ($K_d = 10$–100 μM), addition of only a modest binding SL fragment should produce a nanomolar bi-ligand inhibitor. This process increases the likelihood of producing nM inhibitors with specificity for any member of a pharmacofamily. This is thus an optimal application of a systems-based approach to drug discovery. These SL fragments will bind in the specificity pocket, so that all additional binding energy beyond that generated by interactions with the CLM, is now in the form of specificity. This strategy has been used at Triad to produce specific nM inhibitors (50–200 nM) for three different dehydrogenases, starting with CLMs with K_d values in the 100 μM range. Further reinforcing the object-oriented nature of this process, a bi-ligand CLM-SL combination was modified by swapping four different CLMs with a consistent SL, while still retaining potent binding. Thus, the drug design process not only provides proteomic leverage, it is object-oriented in that CLMs have conserved properties, allowing them to be freely interchanged. This gives further flexibility to the drug design process by allowing one to move from potent in vitro binding to in vivo efficacy by trying those CLMs that empirically give best efficacy.

STRUCTURAL CHARACTERIZATION OF PROTEIN–LIGAND INTERACTIONS

Although it is conceivable that construction of large combinatorial libraries (> 1 million compounds) combined with high throughput screening technology, could produce nM inhibitors for many targets, such a strategy is expensive and unreliable. To increase the likelihood of discovering potent inhibitors through a bi-ligand approach, we sought a strategy to effectively focus libraries by building off of weak binding inhibitors. This is embodied in the second step of IOPE, which requires a structural tool to direct the bi-ligand library into the specificity pocket. We make a case that a novel methodology called NMR SOLVE [Pellecchia et al., 2001] is the best tool for this purpose, since it provides real-time structural information on protein–ligand interactions. However, since the two tools currently used for structural characterization are NMR and X-ray crystallography, we will briefly compare these methods.

Complementarity of NMR and Crystallography

X-ray crystallography is the optimal method for determining structures of large crystallizable proteins. However, NMR is advantageous not only for determining structures of non-crystallizable proteins, but also for extending initial crystal structure data when studying protein–ligand interactions. NMR data can be obtained more quickly than X-ray crystallography data if methods that focus on binding sites are used [Pellecchia et al., 2001], and the structure obtained will more realistically represent the relevant protein–ligand interactions. The latter is true because NMR methods reflect the time averaged state of the protein–ligand complex in solution, whereas the crystal structure reflects that single state which could be crystallized. While often there is little distinction, there are many exceptions. Some notable examples which have been recently reviewed [Sem and Pellecchia, 2001] include thymidylate synthase [Stout et al., 1999], fatty acid binding protein [Hodsdon and Cistola, 1997a; Hodsdon and Cistola, 1997b], hammerhead ribozyme [Torres and Bruice, 2000], and MotA [Li et al., 2001]. These studies demonstrate that once a crystal structure is available, further characterization of protein–ligand interactions by solution-state NMR is advantageous. Given that NMR experiments such as NMR SOLVE exist to gather this information on proteins as large as 170 kDa in a matter of minutes to days, we think that this is now a viable experimental option.

Structure-Based Screening

Most NMR-based screening methods only identify which ligands bind. However, some of the more powerful NMR screening approaches actually identify where a ligand binds. Such methods truly exploit the power of NMR. The "SAR by NMR" screening procedure [Hajduk et al., 1999] is a method whereby (a) a protein is isotopically labeled with ^{15}N, (b) all proton and ^{15}N resonances are assigned to specific residues in the protein, (c) a three-dimensional (3-D) structure of the protein is determined by NMR, and (d) screening is done by comparing 2-D [^{1}H-^{15}N] spectra in the presence of some novel ligand to that for a reference protein in the absence of any ligand. When a ligand binds in the active site, cross-peaks corresponding to binding site residues change in chemical shift, indicating it is likely that the ligand binds in that binding site. Although this method can identify ligands and binding sites, it is limited to small proteins (< 30 kDa) that can be isotopically labeled and expressed in large quantities.

To overcome this molecular weight limitation, we recently developed an alternative strategy for structure-based screening that provides structural information useful in drug design, even for proteins too large for complete assignments (Fig. 5). Using the NMR SOLVE method [Pellecchia et al., 2001], we measure reference spectra with a structurally characterized reference ligand in "Step 1" to identify cross-peaks corresponding to key binding site residues (Fig. 5). In "Step 2", we perturb these nuclei, whether on the protein or another ligand, by a method such as irradiation in the presence of a novel ligand or pool of ligands (Fig. 5). This perturbation is transferred to the ligand that binds, thereby, not only identifying which ligand binds, but also where it binds in the binding site relative to the reference ligands. A preliminary version of this structural information can be gathered in as little as 2 min/sample, using protein sample concentrations as little as 10 μM. Full structural characterization of binding mode and orientation is then determined in follow-up experiments in 1–3 days. NMR SOLVE has been used on monomeric

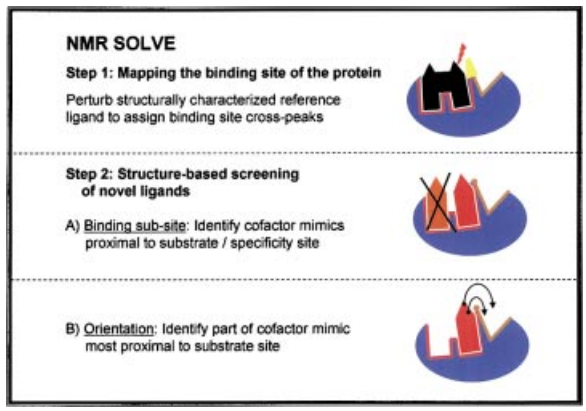

Fig. 5. Drug discovery using NMR SOLVE. NMR SOLVE guides the design of focused combinatorial libraries that are useful against hundreds of proteins related by the binding of a common ligand, such as a cofactor. In step 1, the binding site of a protein is mapped relative to reference ligands such as the natural cofactor (NAD(P)(H) for dehydrogenases). Having correlated NMR data (cross-peaks) with binding site atoms, this information is then used in follow-up experiments (step 2) to determine what part of a novel small molecule is binding closest to the substrate (SL) site. It is here that a linker is attached for construction of a bi-ligand library (Fig. 4). This type of information is obtained in days for proteins even as large as 170 kDa.

proteins as large as 71 kDa, and tetrameric proteins as large as 174 kDa [Pellecchia et al., 2001].

Structure-Based Drug Design With NMR

NMR can be used beyond ligand- and structure-based screening in drug discovery. In fact, "SAR by NMR" has been used for the actual design of potent bi-ligand inhibitors [Shuker et al., 1996]. Ligands are identified that bind in each of two adjacent sites. Then, complete 3-D structures of the protein target in complex with the two bound ligands are determined. If the ligands are close to each other and can be chemically linked such that the binding interactions of neither ligand is overly disrupted, there is a tremendous increase in binding affinity due to the chelate effect. This results from the fact that one ligand rather than two is binding to the receptor, thus avoiding the loss of three degrees of rotational and translational entropy. Thus, "SAR by NMR" is a powerful way to design inhibitors for individual protein targets, when the protein is small enough (< 30 kDa) to do a complete structure with NMR.

Structure-Based Focusing of Combinatorial Libraries

By contrast to the one target at a time approach of "SAR by NMR", NMR SOLVE provides the structural information needed to design combinatorial libraries of bi-ligands in a systems-based strategy targeting an entire pharmacofamily at a time. The method is used to identify inhibitors that bind properly in the conserved site of the pharmacofamily, such as the NAD(P)H site of dehydrogenases or the ATP site of kinases. Screening is performed as described above for compounds that bind in the conserved site. Structural data are obtained that identify where the small molecule binds and what part of it is most proximal to the adjacent substrate site. These experiments are fast (minutes-days), can be performed on very large proteins (170 kDa), and do not require the time consuming step of assignment or complete structure determination. Using this structural information, we produce a geometrically directed combinatorial library of bi-ligands, wherein several thousand random small molecule fragments are attached to the first ligand at the site most proximal to the second binding site (Fig. 5). This type of library is thus a rich source of bi-ligand inhibitors for the hundreds of proteins that bind the same cofactor in the first site. As such, it is a parallel, systems-based approach to drug discovery that relies on the fact that large classes of proteins exist with adjacent sites, one of which is conserved throughout the family.

CONCLUSIONS AND FUTURE DIRECTIONS

Now that the sequencing of the human and other genomes [Tettelin, 2001; Venter, 2001] has

been completed, we are faced with a post-genomic quandary, namely how can sequences of thousands of potential drug targets be most efficiently translated into new drugs (Fig. 1)? Realizing that nature is object-oriented (Fig. 2), it becomes clear that the best drug design strategies will also be object-oriented (Fig. 3). In the case of bi-ligand enzymes, the object-oriented drug design strategy that most naturally parallels nature is IOPE (Fig. 4). Since IOPE attacks entire gene families of proteins related by binding sites (pharmacofamilies), it is a systems-based strategy with tremendous proteomic leverage. The structure-based focusing of libraries of compounds towards pharmacofamilies requires a structural tool that is fast and can deal well with the flexible and dynamic nature of proteins. NMR is best suited for this purpose.

The role of NMR in drug discovery and development has traditionally been in the early stages of protein target structural characterization, and in analytical characterization of small molecule leads and natural products. NMR is now playing a central role in the screening process with the advent of numerous methods and hardware improvements. NMR methods are increasingly finding utility in the actual drug design process, most productively when used in combination with X-ray crystallographic data on reference protein structures. NMR is being used to leverage genomic information in parallel drug design, as a rapid structural tool for building combinatorial libraries tailored to large pharmacofamilies. Its role in drug discovery is only now becoming significant and clearly defined. Its strengths are that it is (a) fast, (b) noninvasive, and (c) a sensitive source of information on structure and dynamics. In terms of protein structure studies, it should be viewed as complementary to, not competitive with X-ray crystallography. NMR will likely become the method of choice for the rapid structural characterization of protein–ligand interactions for proteins, after reference crystal structures have been obtained. This will be because protein NMR can provide structural information that is (a) obtained faster, (b) more realistic (reflecting the population of binding modes), and (c) more biologically relevant when proteins are studied within cells [Serber et al., 2001]. Because of recent advances like NMR SOLVE, it is now possible to obtain structural information about protein–ligand interactions in a truly high-throughput manner. This speed will also make protein NMR a key contributor to the medicinal chemistry SAR process. Such rapid structural characterization of protein–ligand interactions will also allow for the study of binding properties across large classes and subclasses of proteins (pharmacofamilies) in the new field of enzyme mechanomics [Sem and Pellecchia, 2001].

REFERENCES

Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. Nature 397:719.

Booch G. 1996. Object solutions, managing the object-oriented project. Menlo Park, CA: Addison-Wesley.

Booch G, Rumbaugh J, Jacobson I. 1999. The unified modeling language user guide. Menlo Park, CA: Addison-Wesley.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeir K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544.

Drews J. 1998. In: Search of tomorrow's medicines. New York: Springer.

Hajduk PJ, Gerfin T, Boehlen JM, Haberli M, Marek D, Fesik SW. 1999. High-throughput nuclear magnetic resonance-based screening. J Med Chem 42:2315–2317.

Hodsdon ME, Cistola DP. 1997a. Discrete disorder in the nuclear magnetic resonance structure of apo intestinal fatty acid binding protein: Implications for the mechanism of ligand entry. Biochemistry 36:1450–1460.

Hodsdon ME, Cistola DP. 1997b. Ligand binding alters the backbone mobility of intestinal fatty acid binding protein as monitored by ^{15}N NMR relaxation and ^{1}H exchange. Biochemistry 36:2278–2290.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Li N, Zhang W, White SW, Kriwacki RW. 2001. Solution structure of the transcriptional activation domain of the bacteriophage T4 protein, MotA. Biochemistry 40:4293–4302.

Mathieu MP, editor. 2000. Parexel's Pharmaceutical R&D Statistical Sourcebook 2000. Chicago: Parexel International Corporation.

Page MI, Jencks WP. 1971. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. Proc Natl Acad Sci USA 68:1678–1683.

Pellecchia M, Meininger D, Jack R, Sem D. 2001. (submitted).

Sem DS, Pellecchia MA. 2001. NMR in the acceleration of drug discovery. Curr Opin Drug Disc Dev 4:479–492.

Senge PM, Kleiner A, Roberts C, Ross RB, Smith BJ. 1994. The fifth discipline fieldbook: Strategies and tools for building a learning organization. New York: Doubleday.

Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dotsch V. 2001. High-resolution macromolecular NMR spectroscopy inside living cells. J Am Chem Soc 123:2446–2447.

Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. 1996. Discovering high affinity ligands for proteins: SAR by NMR. Science 274:1531–1534.

Stout TJ, Tondi D, Rinaldi M, Barlocco D, Pecorari P, Santi DV, Kuntz ID, Stroud RM, Shoichet BK, Costi MP. 1999. Structure-based design of inhibitors for bacterial thymidylate synthase. Biochemistry 38:1607–1617.

Taylor D. 1995. Business engineering with object technology: New York: John Wiley & Sons.

Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS, Gwinn M, Kolonay JF, Nelson WC, Peterson JD, Umayam LA, White O, Salzberg SL, Lewis MR, Radune D, Holtzapple E, Khouri H, Wolf AM, Utterback TR, Hansen CL, McDonald LA, Feldblyum TV, Angiuoli S, Dickinson T, Hickey EK, Holt IE, Loftus BJ, Yang F, Smith HO, Venter JC, Dougherty BA, Morrison DA, Hollingshead SK, Fraser CM. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293:498–506.

Torres RA, Bruice TC. 2000. The mechanism of phosphodiester hydrolysis: Near in-line attack conformations in the hammerhead ribozyme. J Am Chem Soc 122:781–791.

Venter JC, et al. 2001. The sequence of the human genome. Science 291:1304–1351.