Journal of Cellular Biochemistry Supplement 37:1–6 (2001)

Molecular Recognition: Identifying Compounds and Their Targets

Prabhavathi B. Fernandes*

Ricerca, LLC, 7528 Auburn Road, Concord, Ohio

Abstract As a result of gene sequencing and proteomic efforts, thousands of new genes and proteins are now available as potential drug targets. The milieu of these proteins is complex and interactive; thousands of proteins activate, inhibit, and control each other's actions. The effect of blocking or activating a protein in a cell is far-reaching, and can affect whole, as well as adjacent pathways. This network of pathways is dynamic and a cellular response can change depending on the stimulus. In this section, the identification and role of individual proteins within the context of networked pathways, and the regulation of the activity of these proteins is discussed. Diverse chemical libraries, combinatorial libraries, natural products, as well as unnatural natural products that are derived from combinatorial biology (Chiu [2001] Proc. Natl. Acad. Sci. USA. 98:8548–8553), provide the chemical diversity in the search for new drugs to block new targets. Identifying new compounds that can become drugs is a long, expensive, and arduous task and potential targets must be carefully defined so as not to waste valuable resources. Equally important is the selection of compounds to be future drug candidates. Target selectivity in no way guarantees clinical efficacy, as the compound must meet pharmaceutical requirements, such as solubility, absorption, tissue distribution, and lack of toxicity. Thus matching biological diversity with chemical diversity involves something more than tight interactions, it involves interactions of the compounds with a variety host factors that can modulate its activity. J. Cell. Biochem. Suppl. 37: 1–6, 2001.

Key words: target validation; genomics; proteomics; leads; drug targets

During the last two decades, drug discovery has moved from simple single target identification and screening for inhibitors in complex cellular, and in many instances whole animal systems, towards isolated target systems [Fernandes, 2001]. This change was brought about by the ability to clone and express proteins and other potential drug targets that could be associated with disease. The generation of genome sequences and resolution of all cellular proteins has made drug discovery even more complicated and difficult. The drug discoverer is faced with the validity of the protein as a drug target, its function in normal vs. diseased tissue, as well as a suitable assay that can be used to screen thousands of compounds. Lastly, the need for animal models to validate the targets and determine the activity is still needed to give assurance of potential human use prior to clinical development.

Identifying New Drug Targets in the Proteomic World

Until the dawn of biotechnology, target proteins were isolated from blood and tissues by bioassay-guided fractionation. The isolation and identification of disease-causing and disease-associated genes and proteins were simplified by gene cloning and protein expression. Large-scale identification of proteins associated with disease is now operational [Hatak et al., 2001]. However, validation of the new protein as a drug target still remains an arduous task, the final proof often emerging after the development of a drug (Fig. 1).

It is essential to understand the environment of the cell. Localizing the protein in or on a particular cell type can be a useful exercise because the function of the tissue in which it resides may be known. For example, receptors localized in the hypothalamus have been thought to be linked to satiety and feeding

*Correspondence to: Prabhavathi B. Fernandes, Ph.D., Ricerca, LLC, 7528 Auburn Road, Concord, OH 44077.
E-mail: prabha@fernandes-domain.com

Received 17 September 2001; Accepted 17 September 2001

DOI 10.1002/jcb.10060

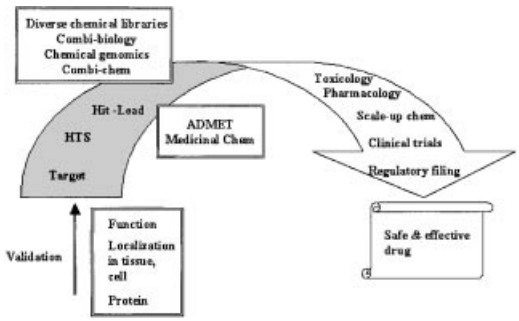

Fig. 1. The yellow brick road from proteomics to drugs.

[Durkin et al., 2000; Funahashi et al., 2000; White et al., 2000]. Furthermore, certain proteins are selectively expressed in diseased tissue. The expression of new proteins relative to those found in the same tissue when normal, does not validate it as a drug target because it may be the effect, rather than the cause of the disease. The new protein may be in a pathway resulting from the activation of another protein. However, inhibition of the new protein may help in modulating the disease or inhibiting the diseases progression. Thus, localization of receptors and other proteins in diseased tissue relative to normal tissue could be useful in identifying drug targets [Yeung et al., 1999; Zhang et al., 2001]. Expression of receptors in tissue that usually do not express these receptors could result in coupling and activation of pathways that could be harmful [Storm and Khawaja, 1999; Nasman et al., 2001]. Differential protein expression could be even more useful than differential gene expression in diseased and normal tissues because the former includes post-translational modification as well the stable expression of the protein itself [Black, 2000; Ahn and Resing, 2001].

Another means of identifying potential drug targets is to look at individuals and families with diseases of interest and identify genes that are linked to the disease process. Many diseases, such as hypertension and heart disease, are polygenic in origin, and therefore, complicate this analysis. Animal models of disease, such as the ob mouse, nacroleptic dogs [Lin et al., 1999; White et al., 2000] can also be used to identify potential disease genes. Model organisms, such as the worm, *Coenorhabditis elegans* and the Zebrafish [Walhout and Vidal, 2001; Wakamatsu et al., 2001], can also be used to identify genes and proteins of interest.

Family studies are used to identify genes that are associated with diseases. Gene mutations that are linked with the actual disease gene are separated by careful genetic linkage studies. In extremes of the expression of a characteristic phenotype for example hyperlipidemia, an analysis of cholesterol metabolism may be used to identify and study genes controlling specific pathways [Brooks-Wilson et al., 1999]. Proteins that are expressed differentially in diseased tissues can be drug targets. It is important to differentiate those proteins that occur as the effect of the disease vs. the proteins that are involved with causing the disease. In some situations, proteins that are differently expressed according to the stage of the disease, may actually be good drug targets. Blocking these targets may halt disease progression; e.g., halting the accumulation of the β-amyloid peptide in the brain of Alzheimer's disease patients. [Ghosh et al., 2001].

When chemical genomics are used, validation of the target can simulate the drug discovery process itself. In this method, compound libraries are designed and tested on cellular systems to obtain phenotypic effects of interest [Fernandes, 2000; Stockwell, 2000]. Although the specific target for the compound of interest may be not be easily known, the compound when optimized, already achieves the desired phenotype in cellular systems and may be closer to becoming a drug than when cell-free screening systems are used.

Identifying Circuitry

In addition to intracellular networks of proteins that communicate through pathways, cells and tissues communicate with each other to construct a coordinated whole animal system. This is especially true in the nervous system that communicates through neural connections as well as by neurotransmitters that can work distally. Tracking of signaling through nerve fibers has been elegantly done through the use of pseudorabies virus [Card, 2001; DeFalco et al., 2001], as well as through excitation of receptors using indicators that show changes in intracellular calcium concentrations. In vivo imagining of protein interactions can also be achieved by expressing proteins of interest fused to mutant Green Fluorescent Proteins (GFP) in vivo. The interaction of two different mutant GFP results in fluorescence resonance

energy transfer (FRET) that can be visualized in vivo [Thompson, 2001].

One Protein in Its Life Can Play Many Parts

Proteins expressed in normal and diseased tissue can be identified by traditional methods of purification, cloning, and gene expression. They can also be identified by 2-dimensional polyacrylamide gel electrophoresis and mass spectrometry. These methods can be used to identify new proteins and protein expressed in diseased tissue. However, neither the function of these proteins nor their utility as a drug target is identified by these methods. Just as an actor who behaves differently in different roles, a protein placed in different environments may present different personalities. Therefore, the protein must be first studied in reference to cell function and later within the context of the cell environment. The defined protein has many choices. It can choose to be activated, or inhibited; it can choose different partner proteins to interact with, it may choose to activate or inhibit another protein, it could choose to destroy another protein, it could choose to activate other proteins in its environment, activation could result in cell division, or it choose to destroy its environment, i.e., the host cell in which it resides [Downward, 2001]. A mutation in the protein can decrease or enhance any or all of these activities. When choosing a protein as drug target, it is necessary to identify the select environment, or set of conditions, under which it needs to be inhibited or activated. Cross-talk between receptors could silence other receptors, thus activity against one receptor could change the activity of another receptor [Tonra et al., 1999]. The effect of a compound on the protein is judged by the sum of the effect on all of the interactions of the protein within the cell. Thus, the actual environment in which each protein is expressed and activated within a cell is important. For example, in bone marrow cells, erythropoietin binds to the erythropoietin receptor and stimulates the production of red blood cells. While in the brain, erythropoietin can also bind to erythropoietin receptor and signal through the JAK/STAT pathway to phosphorylate Ikkinase β and the release of the transcription factor, NF-κB [Digicaylioglu and Lipton, 2001]. Activation of NF-κB in nerve cells protects them from apoptosis.

Cell Surface Proteins

New proteins that are on the cell surface may be useful as targets for immunotherapy. Blocking of the antigens may eliminate the cells or somehow block their activity, for example, adhesion molecules, angiogenesis receptors. Additionally, secreted proteins may themselves be therapeutic in instances where the diseases is the result of, or results in, a decrease in the amount of the secreted protein. Well-known examples of secreted proteins that are in themselves therapeutics are erythropoietin and insulin.

Protein–Protein Networks

As mentioned in the introduction, in order for a protein to perform its function it has to interact with other proteins. The yeast two-hybrid system has been useful in identifying interacting proteins [Fernandes, 2001; Ito et al., 2001] as well as for mapping the total cellular interactions in cells [Vidal, 2001]. There are several variations of the yeast two-hybrid system, each with described advantages of showing selectivity and sensitivity [Fernandes, 2001]. The yeast two-hybrid system is faced with some disadvantages as the organism grows slowly. False results may arise if there is a native protein in the wild type that is similar to the protein of interest. The proteins are expressed in the nucleus of the yeast cell and thus, cellular proteins expressed in the two-hybrid system will have an unnatural localization. Thus other systems, such as the bacterial two-hybrid system that allows simple manipulation as well as an alternative to the yeast two-hybrid system has been described. Mammalian cell systems that identify protein–protein interactions and protein–protein interactions in whole cell systems are also known. Recently, a method for determining protein–protein interaction using dihydrofolate reductase enzyme reconstitution was reported [Remy and Michnick, 2001], and this method could be useful for functional annotation of new proteins. The Cre-lox system for knocking out genes in adult animals has provided a useful tool for studying the effect of gene deletions or mutations in adult animals [Luo et al., 2001]. The in vivo environment is especially important while studying the nervous system, as this is a delicate system of circuitry that must be understood in its totality. Various methods of in vivo image to determine

network connections within the nervous system and localization of receptors and to understand their interactions have been described. Labeling with variants of GFP have made in vivo imaging possible, as no extraneous reagents need be added. In situ imaging of the human brain is being used to decipher intricate connections that occur and form during various stimuli [Neet and Campenot, 2001].

The Other World—Non-Protein Targets or Unproteomics

Cell surface carbohydrate receptors as well as DNA, and DNA–RNA or RNA–protein interactions are also drug targets. The complexity of sugars makes them more difficult to address with small molecules [Williams and Davies, 2001]. DNA or RNA containing targets must be very selective as any nonspecific effects may have disastrous effects [Gerber and Keller, 2001]. Introducing unnatural sugar variants that intercept the sialic acid biosynthetic pathway may be used for the validation of carbohydrate targets [Lemieux and Bertozzi, 2001].

Identifying Inhibitors by High Throughput Screening

In days before the generation and screening of large chemical libraries, the activity of each compound tested was examined for its cellular and in vivo activity. Many classes of drugs that are in clinical use today were identified by pharmacological screening in animals. With the advent of cloning and production of proteins, high throughput screening of chemical libraries has become the norm. However, the challenge is to look beyond the most potent inhibitors to find those inhibitors that are selective and present properties that could be the best choice to further optimize and develop [Fernandes, 1998, 2000; Oldenburg et al., 2001]. Potent inhibitors from screens are frequently non-selective. The problem is in sifting through many potent inhibitors that could mask the profile of activity of less potent compounds. Allosteric inhibitors are generally weaker inhibitors. In examining interacting surfaces of proteins, it may not be possible to block large binding surfaces. In such instances, it may be possible to identify those proteins that bind at a site distal to the binding site in such a manner as to alter the structure of the binding site [Tachedjian et al., 2001]. Such allosteric inhibitors, found in assays designed to measure enzyme activity, may not be as potent as active site inhibitors.

Chemical Diversity

Traditionally, drug discovery was focused on microbial and plant extracts to identify new drug candidates. In the late 1980s, chemical collections of pharmaceutical companies were screened for activity in a variety of screens. This was labeled the "synthetic compound collection," and had been synthesized as part of making analogs for drug discovery programs at these companies. Success from screening these libraries led to the demand for larger compound collections, and the designing and building of combinatorial libraries. Each library generated in a combinatorial fashion has yielded thousands of compounds for screening, but has not yielded the chemical diversity for identifying new drugs for new targets. Therefore, combinatorial chemistry is now commonly used to generate smaller and more focused libraries consisting of 50–100 molecules instead of hundreds of thousands compounds. The rich source of natural products is still being explored for new agents and more recently, anticancer compounds such as discodermalide and epothilones have been identified as clinical candidates [Altmann, 2001]. However, the long and difficult task of isolating new molecules, often found in trace amounts, from natural source has led to the creation of gene libraries from antibiotic producing organisms and the "reconstruction" of new compounds using defined synthetic pathways [Chiu et al., 2001]. This technique, called combinatorial biology, promises to be the source of new drugs that may be too complicated for the bench chemist to synthesize.

Screening hundreds of molecules against hundreds of targets is expensive and therefore, many methods to miniaturize the screen are available [Oldenburg et al., 2001]. Data interpretation and data analysis is still the critical factor for these high throughput and ultra-high throughput screens. "Hits" with lesser activity, but perhaps with interesting binding and inhibitory characteristics can be masked by those "hits" that are stronger and perhaps more traditional in their active site inhibitory process.

Diminishing Returns

Considerable amount of time and effort is spent on improving old products that have

proven clinical efficacy. Improvements have yielded promising new drugs with extended patent lives. One strategy has been to develop single isomer drugs in the hope of extending patent life, and developing a drug that is at least as effective as the proven racemate form. In some instances, the active enantiomer is not necessarily safer than the mixture, of active and inactive forms, as has been shown for fluoxetine or Prozac [Thayer, 2000; Choi et al., 2001]. Removal of the "L" form of fluoxetine, which is not active, did not make a safer product. Another strategy has been to clone and characterize subtypes of receptors and enzymes that could be used also screen for more selective leads. However, as in the case of identifying active isomers of drugs, from the point of view of the target, extreme target selectivity may not always be a guarantee for improving an existing drug product. For example, the COX-1 and COX-2 are targets for aspirin, and COX-2 inhibitors were designed with the interest in developing a more selective pain killer that did not have gastric side effects. Indeed, more potent and selective pain killers without gastric side effects were made. However, patients treated with COX-2 inhibitors have shown higher level of atherosclerosis with resulting cardiac effects [Boers, 2001]. This is because COX-1 inhibition, as a side effect of inhibiting COX-2, provided the anti-inflammatory properties that gives aspirin its beneficial vascular effect. Thus, target selectivity and compound purity can be stretched to an extent that could be detrimental.

CONCLUSIONS AND DIRECTIONS

Even after gaining specificity for the isolated target, and proving activity in the whole animal, a new drug is faced with the natural mechanism that is used by the human body to eliminate toxic and foreign substances [Hodgson, 2001]. Because many patients are being treated with more than one drug it is important to bear in mind drug–drug interactions and the effect of one drug on the metabolism of another. The chemist is faced with designing a molecule that can interact with the target, but escape the metabolizing and excretory systems for a desired period of time. This challenge, which straddles science and the arts, is faced daily in the biopharmaceutical industry.

REFERENCES

Ahn NG, Resing KA. 2001. Towards the phosphoproteome. Nat Biotech 19:317–318.

Altmann K. 2001. Microtubule-stabilizing agents: A growing class of important anticancer drugs. Curr Opin Chem Biol 5:424–431.

Black DL. 2000. Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology. Cell 103:367–370.

Boers M. 2001. NSAIDS and selective COX-2 inhibitors: Competition between gastroprotection and cardioprotection. Lancet 357:1222–1223.

Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, van Dam M, Yu L, Brewer C, Collins JA, Molhuizen HO, Loubser O, Ouelette BF, Fichter K, Ashbourne-Excoffon KJ, Sensen CW, Scherer S, Mott S, Denis M, Martindale D, Frohlich J, Morgan K, Koop B, Pimstone S, Kastelein JJ, Hayden MR, Genest J, Hayden MA. 1999. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. Nat Genet 22:336–345.

Chiu HT, Hubbard BK, Shah AN, Eide J, Fredenburg RA, Walsh CT, Khosla C. 2001. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. Proc Natl Acad Sci USA 98:8548–8553.

Choi BH, Choi J, Yoon SH, Rhie D, Min DS, Jo Y, Kim M, Hahn SJ. 2001. Effects of norfluoxetine, the major metabolite of fluoxetine, on the cloned neuronal potassium channel Kv3.1. Neuropharmacol 41:443–453.

Curd JP. 2001. Pseudorabies virus neuroinvasiveness: A window into the functioning organization of the brain. Adv Virus Res 56:39–71.

DeFalco J, Tomishima M, Liu H, Zhao C, Cai X, Marth JD, Enquist L, Friedman JM. 2001. Virus-assisted mapping of neural inputs to a feeding center in the hypothalamus. Science 291:2608–2613.

Digicaylioglu M, Lipton SA. 2001. Erythropoietin-mediated neuroprotection involves cross-talk between Jak2 and NF-κB signaling cascades. Nature 412:641–647.

Downward J. 2001. The ins and outs of signaling. Nature 411:759–762.

Durkin MM, Walker MW, Smith KE, Gustafson EL, Gerald C, Branchek TA. 2000. Expression of a novel neuropeptide Y receptor subtype involved in food intake: An in situ hybridization study of Y5 mRNA distribution in rat brain. Exp Neurol 165:90–100.

Fernandes PB. 1998. Technological advances in high throughput screening. Curr Opin Chem Biol 2:597–603.

Fernandes PB. 2000. The ATCG of drug discovery. Curr Opin Mol Therapeutics 2:624–632.

Fernandes PB. 2001. Microbe based screening systems. In: Seethala R, Fernandes PB, editors. Handbook of drug screening. New York: Marcel Dekker, Inc. p 129–152.

Funahashi H, Hori T, Shimoda Y, Mizushima H, Ryushi T, Katoh S, Shioda S. 2000. Morphological evidence for neural interactions between leptin and orexin in the hypothalamus. Regul Pept 92:31–35.

Gerber AP, Keller W. 2001. RNA editing by base deamination: More enzymes more targets, new mysteries. Trends Biochem Sci 26:376–384.

Ghosh AK, Bilcer G, Harwood C, Kawahama R, Shin D, Hussain KA, Hong L, Loy JA, Nguyen C, Koelsch G, Ermolieff J, Tang J. 2001. Structure-based design: Potent

inhibitors of human brain Memapsin 2 (beta-secretase). J Med Chem 44:2865–2868.
Hatak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, Haroutunian V, Fienberg AA. 2001. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proc Natl Acad Sci USA 98:4746–4751.
Hodgson J. 2001. ADMET-turning chemicals into drugs. Nat Biotech 19:722–726.
Ito T, Tomoko C, Ozawa R, Yoshida M, Hattori M, Sakaki Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98:4569–4574.
Lemieux GA, Bertozzi CR. 2001. Modulating cell surface immunoreactivity by metabolic induction unnatural carbohydrate antigens. Chem Biol 8:265–275.
Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qiu X, de Jong PJ, Nishino S, Mignot E. 1999. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. Cell 98:365–376.
Luo JL, Yang Q, Tong WM, Hergenhahn M, Wang ZQ, Hollstein M. 2001. Knock-in mice with a chimeric human/murine p53 gene develop normally and show wild-type p53 responses to DNA damaging agents: A new biomedical research tool. Oncogene 20:320–328.
Nasman J, Kukkonen JP, Ammoun S, Akerman KE. 2001. Role of G-protein availability in differential signaling by alpha 2-adrenoceptors. Biochem Pharmacol 62:913–922.
Neet KE, Campenot RB. 2001. Receptor binding, internalization, and retrograde transport of neurotrophic factors. Cell Mol Life Sci 58:1021–1035.
Oldenburg KR, Kariv I, Zhang J-H, Chung TDY, Lin S. 2001. Assay miniaturization: Developing new technologies and assay formats. In: Seethala R, Fernandes PB, editors. Handbook of drug screening, New York: Marcel Dekker, Inc. p 525–562.
Remy I, Michnick SW. 2001. Visualization of biochemical networks in living cells. Proc Natl Acad Sci USA 98: 7678–7683.
Stockwell BR. 2000. Frontiers in chemical genetics. Trends Biotechnol 18:449–455.
Storm SM, Khawaja XZ. 1999. Probing for drug-induced multiplex signal transduction pathways using high resolution two-dimensional gel electrophoresis: Application to beta-adrenoceptor stimulation in the rat C6 glioma cell. Brain Res Mol Brain Res 71:50–60.
Tachedjian G, Orlova M, Sarafianos SG, Arnold E, Goff SP. 2001. Nonnucleoside reverse transcriptase inhibitors are chemical enhancers of dimerization of the HIV type 1 reverse transcriptase. Proc Natl Acad Sci USA 98:6991–6992.
Thayer A. 2000. Eli Lilly pulls the plug on Prozac isomer drug. Chem Eng News 78:8.
Thompson WJ. 2001. Seeing is believing. GFP transgenics illuminate synapse elimination. Neuron 31:341–342.
Tonra JR, Ono M, Liu X, Garcia K, Jackson C, Yancopoulos GD, Weigand SJ, Wong V. 1999. Brain-derived neurotrophic factor improves blood glucose control and alleviates fasting hyperglycemia in C57BLKS-Lepr(db)/lepr(db) mice. Diabetes 48:588–594.
Vidal M. 2001. A biological atlas of functional maps. Cell 104:333–339.
Wakamatsu Y, Pristyazhnyuk S, Kinoshita M, Tanaka M, Ozato K. 2001. The see-through medaka: A fish model that is transparent throughout life. Proc Natl Acad Sci USA 98:10046–10050.
Walhout AJ, Vidal M. 2001. Protein interaction maps for model organisms. Nat Rev Mol Cell Biol 2:55–62.
White DW, Zhou J, Stricker-Krongrad A, Ge P, Morgenstern JP, Dembski M, Tartaglia LA. 2000. Identification of leptin-induced transcripts in the mouse hypothalamus. Diabetes 49:1443–1450.
Williams SJ, Davies GJ. 2001. Protein–carbohydrate interactions: Learning lessons from nature. Trends Biotechnol 19:356–362.
Yeung G, Mulero JJ, Berntsen RP, Loeb DB, Drmanac R, Ford JE. 1999. Cloning of a novel epidermal growth factor repeat containing gene EGFL6: Expressed in tumor and fetal tissues. Genomics 62:2.
Zhang H, Yu C-Y, Singer B, Xiong M. 2001. Recursive partitioning for tumor classification with gene expression micorarray data. Proc Natl Acad Sci USA 98:6730–6735.