CHAPTER 1

Drug Discovery: New Compounds

Tully Speaker

Then and Now

Discussion of the generation of new drug compounds by the pharmaceutical industry must consider the topics of drug-receptor modeling, high-throughput screening, and candidate selection. These topics have driven most of the industry over the last two decades. Advances in these areas have developed remarkable, nay, amazing, new science and technology. But the dearth of new products now in most pipelines calls into question the corresponding allocation of effort and expenditures that have compounded by about 13% per year since 1970.[1,2] Yet it is entirely possible the near future will see the pipelines quickly fill. In light of that possibility, it is useful to consider how drugs have historically come about, how the new methods arrived, and what they do.

Long ago plant extracts were the sources of formulations effective in treating a limited number of serious conditions, for instance, opium tincture for pain, digitalis fluid extract for dropsy. The conditions were physiologically and/or pathologically defined, and formulations were developed from folklore with the likelihood of few hits and uncounted numbers of complex misses. In the 19th century, use of plant extracts as drugs began to give way to treatment with specific chemicals separated from the extracts. New chemicals patterned on isolated natural substances became drugs. Then, in the 1840s, volatile chemicals that came into use as general anesthetics made surgery less utterly barbaric. Gradually matching drug to the condition requiring treatment became more nearly possible. Clinical medicine gradually defined itself.

From the 1920s through the 1980s, rapid advances in chemistry, biology, and pharmacology were coupled with increasing reliance on careful clinical research. This coupling produced very effective, if not miraculous, drugs to treat infectious, inflammatory, cardiovascular, psychiatric, respiratory, and invasive diseases, even if the actual target for a drug or its mechanism of action were ill-defined. Retrospective consideration of the current best-selling 100 drugs shows the drug targets were selected on the basis of convincing published biological research extending to human studies. Analogies with bullets that magically found their targets or with keys to hidden locks romanticized the wide endeavor.[3]

■ Drug and Receptor

Drug-receptor modeling may be considered an outgrowth of the 19th- and 20th-century lock and key analogy as applied to drug and receptor, imaginative as that analogy may be. It is that outgrowth and much more. Existence of a number of what might be called keys was demonstrable. Administration of a specific drug in a defined dose could reasonably be relied upon to produce an effect more or less specific to that drug, for example, sleep, analgesia, wakefulness, emesis. Further, isolation and identification of specific physiologic neurotransmitters, acetylcholine by Loewi and Navratil[4] and norepinephrine by von Euler,[5] gave reason to believe a mechanistic description of drugs was feasible. Presumably locks that accepted these neurotransmitters existed.

Proof of the existence of drug receptors emerged only in the latter half of the 20th century. Demonstration of drug binding to specific cells and subcellular fractions was greatly facilitated by the availability of radiolabeled drug substances. Development of methods for the culture of mammalian cell lines and demonstration of specific drug–protein interactions served to show the locks of the lock and key theory really do exist. For example, the several types of steroid hormones independently bind with a specific cytoplasmic receptor protein. This binding causes the receptor to change shape and to dissociate from a preexisting intracellular complex with another protein. Dissociated, the receptor proteins diffuse from the cell cytoplasm into the nucleus where they associate with DNA to initiate protein formation.[6] Similarly, aspirin and other nonsteroidal anti-inflammatory drugs interact with cyclooxygenase enzymes, crippling them and preventing the enzymes from converting arachidonic acid to inflammatory and other mediators.[7]

■ Modeling Molecules

From the earliest days of chemistry, isolation of a substance in crystalline form was considered a requisite to recognizing and identifying a new solid chemical substance. Liquid substances were, if at all possible, converted to solid derivatives to demonstrate their unique identities. The wet, messy complexity of living things only very slowly gave way to recognition that order existed in the complexity, that chains of linked chemical reactions involving specific substances occurred in orderly processes.

In the 1920s, William Bragg began to study the paths of X-rays from essentially point sources as the rays passed through or were scattered by pure crystalline solids. He showed that such crystals allowed narrow beams of X-rays to travel uninterrupted along some paths relative to the crystal axes and were scattered by beams traveling along other paths in a manner analogous to lines of sight in an orchard planted in a regular grid pattern. From this he was able to infer and calculate the relative positions of atoms within the crystals studied. However, the manual measurements and calculations were very laborious.

Over the intervening decades, X-ray crystallography grew in sophistication and complexity so that by the 1960s Max Perutz had determined the crystal structure of myoglobin, a protein with, at that time, a staggering number of atoms in each molecule.[8] By the year 2000, both private and public funding supported X-ray crystallographic determination of structures of proteins from all the families for which amino acid sequences had been established. X-ray crystallography of proteins has benefitted immensely from development of highly sophisticated computing capability and automated equipment that: (1) serially adjusts the angle at which the X-ray beam impinges on the crystal being studied; (2) records the intensity and emergent angle of the beam; (3) calculates the position of the atoms in the crystal which scatter the beam; and (4) with these data map the three-dimensional structure of one or more of the molecules that comprise the crystal.

■ Picturing Molecules

Quite independently, beginning in the 1970s, computer programs able to represent the two- and later the three-dimensional structures of chemical substances were developed. These programs were based on the known composition and reactivity of individual substances, together with their NMR spectra and X-ray crystallographic structures. As the capability of these programs increased, representations of individual molecules could be moved about on the screen at will, rotation of functional groups about a bond could be displayed, and the interaction of pairs of compounds could be visualized.

Ideally, there is one more step before proceeding with drug-receptor modeling in silico. The drug is allowed to interact with and bind to the protein that acts as receptor for it, and then, it is hoped, the pair will crystallize in the bound state. If this in fact occurs, a new X-ray crystallographic study may show, atom for atom, both the drug and receptor in their bound relation to one another. It is more likely that the drug (key) being studied is not a perfect fit to the receptor (lock) and even more likely that the drug–protein complex will not crystallize.

It is possible to display the three-dimensional structure of the receptor protein and the putative drug on the same screen. A number of commercially available or proprietary programs allow the drug image to be nudged near the protein, to estimate the goodness of fit and to optimize it by repositioning the drug and/or by flexing the structures of drug and receptor. And, of course, if such an interaction can be studied with one putative drug, the same evaluation may be carried out with representations of other candidate drug molecules to identify a best fit.

It sounds simple, but programs to represent a structure in two or three dimensions quickly grow complex as more atoms are added and as the flexibility of a structure is displayed. Add a protein with 20,000 to 50,000 atoms and the demands on computer size and time become proportionally greater and therefore more expensive. It is possible to trade off either the sizes of the molecules to be visualized or the goodness of fit, but neither is a property willingly lost. Drug-receptor modeling is not limited to displaying the physical size and structure of molecules. Many programs allow use of colors and color intensities to represent the properties of parts of molecules, such as relative acidity or basicity, electron density, charge distribution, hydrophobicity or hydrophilicity, and solvent accessible surface. Massive computing power and machine time are simply expensive, as are molecular modeling programs capable of handling large molecules with high accuracy and speed.

This is not a text on computer graphics, but the elements of computer representation of molecules may be summarized briefly. The very smallest parts of an image, often referred to as *primitives*, are points, polygons, and vectors. Modeling programs combine these primitives into objects with a specific shape and coordinates in *n*-dimensional space. These coordinates thus are elements in *n*-dimensional mathematical matrices, constructs that computer modeling programs can very efficiently manipulate.

The objects are then transformed in object space into atoms and bonds, transformation being a mathematical operation that allows all the coordinates of all the vertices in an object to be changed simultaneously, as when a whole molecule appears to rotate or, more selectively, when atoms linked by a covalent bond rotate relative to one another. In the context of drug-receptor modeling, objects may be atoms, bonds between atoms, or whole molecular structures of candidate drugs. Similarly, but on a larger scale, macromolecular objects may represent proteins, nucleic acids, or membranes with which a candidate drug object may interact.

■ Picturing Interacting Molecules

Interaction between two objects represents a next level of graphic complexity. It requires that the object spaces of individual molecules be transformed into a world coordinate system. In the world coordinate system, matrices of individual "objects" are reassigned positions in a larger common matrix so that interacting molecules may be positioned relative to one another in the larger matrix. Few people can think in the *n*-dimensional space of the larger matrix, so in almost all modeling programs the *n*-dimensional matrix is mapped to represent a two-dimensional display space.

Further transformations are applied to allow the person using the program to select the apparent orientation of the images (turned left, right, or upside down) in "viewer" space and to add Western artistic perspective so that an apparently more distant "object" or parts of it may appear proportionately smaller than those that are closer. In commercially available programs, the transformations are carried out without intervention or (usually) awareness of them by the person using the program. Using the best of these programs is akin to watching a video cartoon or playing a computer game. Commercial products are expensive because they enlist high talent and funding.

Selection of Preferred Models

The purpose of drug-receptor modeling is to find optimized or best fits between one or more similar drugs and a receptor. (Better fit implies better drug.) This requires calculation of the energy of association between candidate drugs and the receptor. Each candidate drug can, within limits, rotate, flex, stretch and/or compress all the bonds between its atoms. So can the receptor. Thus, the energy with which each conforms to the association varies.

The energy to be considered is the Gibbs free energy DG. *Free* here means available, not necessarily without cost. DG is defined as the difference between the enthalpy or heat content DH and the absolute temperature T of the system multiplied by the entropy DS. In equation form this is written as follows:

$$DG = DH - TDS$$

The heat content is the sum of the energy required to assemble a real compound from essentially infinitely separated atoms (its heat of formation) and the average kinetic energy of its molecules. The entropy of a system is a measure of its tendency to occupy all possible states. For molecular bonds this corresponds to all the energy stored in bonds, analogous in simple terms to the energy stored in springs connecting two atom models as the springs (bonds) are rotated, flexed, compressed, or stretched.

Additional contributions to the free energy of an interacting system arise from electrostatic attractions between oppositely charged portions of molecules and from van der Waal's attractions between any two atoms when they are in very close proximity to one another. For practical considerations in drug-receptor interactions, temperature is assumed constant or nearly so; humans are nearly constant temperature systems.

The sum of these types of potential energy in any specific instance is sometimes referred to as an *empirical force field*. It is considered empirical because the energy terms come from experimental data and basic quantum mechanical calculations. Adapting empirical force field calculations to biological systems has led to gradual improvements and to an approximate consensus now applied in popular programs, notably CHARMM[9] and AMBER.[10]

The whole point of performing these calculations is to estimate those geometries of interacting molecules that reduce the energy of the pair. In the real world, molecules preferentially move to structural forms that reduce their energy. So, too, in silico. A computer program can align the image of a drug molecule in approximation to its presumed receptor and then repetitively adjust the alignment so that oppositely charged atoms more closely approach one another; pairs of electron-rich atoms stretch their bonds to share electron-poor hydrogen atoms with one another to form hydrogen bonds, hydrophobic sections of drugs twist enough to match up more fully with hydrophobic patches of the receptor surface; and, importantly, the drug molecule snuggles deeper into a groove or depression on the receptor surface. At each stage, the energy of the resulting alignment is estimated until, by sequential iterations, a minimum energy is found. The fit is optimized.

Evaluation of fit by energy calculation is central to current efforts to discover new medicinal agents. When, as is usually the case, more than one candidate drug structure is considered, the process is repeated for each to find the drug-receptor pair with the lowest interaction energy, the pair that is very likely to be the most active when tested in vivo. The process is not simple but once set in motion is more effective, simpler, quicker, and less costly than animal studies.

■ The Need for Crystals

The mathematical selection process leads to the one or the few candidate drug computer images of the set available for trial that optimally bind with the computer image of the receptor protein in silico. The process relies on the availability of a real crystal of the receptor protein with well-defined facets and sharp edges. The crystal need not weigh more than the ink needed to print this sentence, but the X-ray diffraction pattern generated from this target receptor protein is the basis for the seemingly three-dimensional computer image of the protein. It may seem a small matter, but the work of crystallizing a useful quantity of each new receptor protein in a reasonable amount of time has been and continues to be a major bottleneck in drug discovery.

At least partial resolution of this bottleneck emerged from a separate series of experiments directed to resolution of the structure of nucleic acids.

■ The Double Helix and All That

As many literate adults know, in 1953 Watson and Crick[11] won the race with Pauling to generate X-ray crystallographic data about deoxyribonucleic acid (DNA) and to interpret the findings to yield new knowledge. They found: (1) DNA exists primarily in the form of helical double strands; (2) the strands consist of chain-like structures in which the sugar deoxyribose and phosphate groups alternate; (3) one of four cyclic nitrogenous bases is attached to each sugar link; (4) the nitrogenous bases of the strands' hydrogen bond to one another in very specific pairings (adenine with thymine, guanine with cytosine); (5) the specific pairings of the nitrogenous bases render the two strands of the double helix anti-parallel and complementary to one another; and (6) the helices have right-handed twists.

These bits of knowledge soon led to further understandings. The complementarity of pairs of DNA strands provides a mechanism for highly accurate replication and transmission of genetic information from one cell to its progeny. In this process, the enzyme DNA polymerase untwists the paired strands and, using each as a template, assembles its complementary strand. Genetic information encoded in DNA as sets (codons) of three consecutive nitrogenous bases is transcribed to corresponding codons in ribonucleic acid (RNA). It is then translated into successive amino acids of protein. These proteins determine the properties and activities of each cell. Some of these proteins act as receptors with which drug molecules interact to exert their effects.

In the 1970s, new types of enzymes were recognized. One type, *restriction enzymes*, breaks apart the chain of alternating ribose and phosphate units of DNA of any organism.[12] Each restriction enzyme acts at one link in a highly specific sequence of four to eight nitrogenous bases attached to a sugar-phosphate chain thousands of links long. The chopped bits are sometimes referred to as *restriction fragments*. Another new type of enzyme found in the 1970s, *DNA ligases*, is able to assemble and integrate restriction fragments into DNA strands and so to generate recombinant DNA.[13]

■ Determining Nucleotide Sequences

It helps, in interpreting X-ray crystal data, if the sequence of nucleotides in a sample of DNA is known. By 1975, Sanger et al.[14] had developed the first successful DNA sequencing method. The technique involves cutting each of multiple replicates of the DNA to be sequenced into restriction fragments and using a reserved fraction of the original DNA molecules as templates on which ligases assemble exactly matching strands.

The method cleverly attaches a different fluorescent label to a small fraction of each of the four nucleotides to be linked together by ligases. Linking is in the order dictated by the template, successively attaching nonlabeled nucleotides until, by chance, a labeled nucleotide is added to the chain. The label gums up the works because the bulk of the label cannot be accommodated by the ligase as it attempts to add another nucleotide. Chain building stops at the fluorescent label. What results is a soup containing newly made bits of DNA all of which started at the same point but with ends bearing differently fluorescent nucleotides. Separating the strands electrophoretically by size allows the sequence to be determined in reverse.

In 1986, Hood et al. described a method to attach different fluorescent compounds, each specific to one of the four nucleotide types, to an entire strand of DNA, thus allowing more rapid sequencing.[15] This improvement allowed successive bases to be identified more easily and served as the basis for current high-speed sequencing machines, that is, the laser detection of fluorescently tagged nucleotides as DNA strands flow through capillary tubes, a development that now enables arrays of high-throughput machines to sequence tens of thousands of nucleotides per day.[16,17]

Restriction fragments containing a few thousand base pairs are relatively easily replicated by polymerase chain reaction (PCR) and sequenced by automated procedures. In the 21st century, PCR instruments that automatically replicate DNA are available for modest sums and tens of commercial laboratories compete in offering sequencing services. However, PCR machines have been known to make random mistakes in replication of DNA strands.

■ From DNA to Protein

Recall that a crystal of receptor protein is needed to produce the X-ray pattern on which the computer-generated image is patterned. Converting strands of DNA into a protein crystal involves several steps. First, the DNA sequence of interest, usually a restriction fragment, must be inserted into a suitable recombinant protein-expression system and the resulting system employed to make useful amounts of protein. The choice of expression system is not trivial. Obviously, expressing a human DNA segment in a human cell system is very likely to produce a protein that is correctly folded and posttranslationally modified, for example, by correct attachment of methyl groups and sugars. Mammalian expression systems generally produce low yields, are complex, and are comparatively costly. For many years addition of fetal bovine serum to the growth medium of mammalian systems has been widely used to aid cell growth. However, it is possible and generally preferable to avoid adding complex mixtures of small peptides, growth factors, trace lipids, bovine viruses, and prions to cultures intended to yield therapeutic proteins.

But most strains of human cells, or of nonhuman mammals for that matter, eventually die, just as whole humans and other animals do. Having a cell strain die out raises hob with an experiment. It is easier to plan with immortal cells.

Mammalian cells able to live indefinitely in culture are referred to as a *cell line* or as *immortalized cells*. They are derived from tumors or other cells that have been transformed. They resemble tumor cells but perform most posttranslational modifications in the same way normal cells do. There are many readily available immortalized mammalian cell lines that have remained stable for decades, but other cell types also have advantages.

Generating substantial quantities of recombinant DNA is best done in replicating cells by taking advantage of small double-stranded circles of DNA, called *plasmids*, found in bacteria, yeasts, and a few higher organisms. Plasmids are not chromosomal DNA, but are replicated when a cell containing them replicates. Plasmids can be isolated from other cellular components. To make recombinant DNA, isolated plasmids are cut with the restriction enzyme used to generate the restriction fragment of interest. Then, with the aid of a ligase, the fragment of interest and the cut circle are linked to make a larger circle. The ligase must be chosen to act at the same nucleotide sequence as had been cut, both to make the fragment and to open the original plasmid circle.

The new recombinant plasmids are transplanted or *transfected* into intact cells, for example, *Escherichia coli*, by mixing the cells and plasmids and exposing the mix to high concentrations of certain divalent cations, such as calcium. The efficiency of such transfection is very low and it is often advantageous to link into the plasmid another restriction fragment, one conferring resistance to a specific antibiotic to aid selection of the cells of interest. The mix of cells, most normal and a few doubly recombinant, is then grown in culture medium containing the antibiotic to which the transfected cells are resistant. The antibiotic kills off normal nonresistant cells. Each colony of cells that arises from a single doubly recombinant transformed cell, a colony from one, is called a *clone*. A clone carrying the restriction fragment of interest can thus replicate easily and produce the protein wanted for crystallization experiments or other scientific and economic ends.

Alternatively, cut plasmids and restriction fragments are mixed and the mix is transplanted into yeast cells. Inside yeast cells, the cut plasmid and fragment are linked by yeast DNA repair enzymes to make larger plasmids. Expression systems utilizing yeasts such as *Saccharomyces cerevisiae* or *Pichia pastoris*[18] are perhaps the most frequently used.

In addition, and more amenable to adaptation to high-throughput processes, continuous flow systems providing cell-free protein have been described. These offer the advantages of stable genetic sequences, most posttranslational modifications of mammalian systems, and minimal background protein secretion.[19,20]

Numerous other procedures for generating transfected cells may also be employed. Insect cell-baculovirus expression systems are not costly and can generate proteins in posttranslationally modified soluble form, but yields vary from one protein to the next. More important, although insect-derived proteins are close to, they are usually not identical in posttranslational modification to the corresponding human proteins.

■ Converting DNA Sequences to Genomes

The ability to generate substantial amounts of each of the many restriction fragments available from the DNA has allowed sequencing the entire genetic assembly (the genomes), of many types of organism (*E. coli*), microbes that live in the intestines (*Drosophila melanogaster*), fruit flies, and, in 2001, *Homo sapiens* (people).[21–23]

■ Structural Genomics

Structural genomics research applied to the human and other genome sequences allows identification of a huge number of proteins capable of serving as drug targets. In many instances, stretches of DNA sequence correspond to the known structure of specific proteins; in many others the genetic information represents proteins with unknown structure or function.

As data on the human genome have been acquired, it has become evident that the DNA sequences of people, although very much alike at the large scale, vary from what may be considered a consensus genome at numerous points (or else, for example, we would all have the same color of eyes or hair or all have the same allergies, and so forth). Most frequently, these variations take the form of single-nucleotide polymorphisms (SNPs), which generally do not result in readily distinguishable phenotypic differences, such as eye color or allergies. They are surprisingly frequent as well, occurring about once in every thousand consecutive nucleotides.[24]

SNPs almost always take the form of biallelic polymorphisms in which one purine is replaced by the other (adenosine/guanosine interchange) or one pyrimidine substitutes for the other (cytidine/thymidine interchange). Theoretically, any of the base pairings could substitute for another. These differences help allow differentiation/identification of individuals on the basis of their DNA, a method of some forensic interest. Of broader importance, the presence of a specific SNP or set of SNPs has been linked to human disorders. But sequencing accurate enough to provide better than 99.99% assurance of an individual human's genome is expensive. Stimulus to develop the needed speed and analytic accuracy is provided, at least in part, by a recent National Institutes of Health grant program to support workers attempting to sequence a complete mammalian genome at a cost below $100,000 and eventually $1000. One may expect some employers and insurance companies to encourage this work.

Regrettably, there are as yet very few associations of a specific gene with a specific disease. It is technically possible to examine each gene in a person to find whether or not it is relevant to a specific disease by comparison with the genes of large well-defined sets of people with that disease and a comparable well-characterized set of control individuals. Accurate determination of an individual's genome and accurate comparison of sequences with millions upon millions of nucleotides in nearly identical strings of genomic data from the sets of disease and control populations are required. That can be done.

Goldstein et al.[25] retrospectively examined some 42 sequence variants of genes for which response to a drug had been identified at least twice. They reported that half of these coded for the drug target protein or a metabolic pathway associated with the target. As a result it is fair to say there are data supporting the notion that genetically linked targets can serve as guides to a drug target. It is also fair to say genetic variants may provide targets for new drugs, but also to recognize genetic susceptibility does not necessarily identify individuals who will develop and need treatment for the disease flagged by the variant.

Other studies have associated diseases with variations in gene structure more specifically. By 2002, research in the fields of genetic epidemiology and statistical genetics led to high-throughput searches for genome–disease correlations and estimation of their statistical significance.[26–28]

It is not unreasonable to consider also that expression of the same SNP in a small fraction of the population may be the basis of an infrequent side effect in those taking a candidate drug in a phase 1 or later trial. When multiple variants of a gene appear to militate toward expression of the disease/condition, the group is commonly referred to as *susceptibility genes*. Practically speaking, the multiplicity inherent in a set of susceptibility genes does not render them efficient targets for single drugs nor for high throughput screening procedures to select candidate magic bullets, magic shotgun shot perhaps.

The enormous effort and expense of enlisting clinicians, defining uniform criteria for assessment and diagnosis, assembling libraries comprising clinical data on cohorts of sick people and healthy controls, collecting their DNA and individual genome data, codifying their responses to marketed drugs and drugs in phases 2 though 4, gathering informed consent agreements, and integrating these multidimensional data into a coherent assemblage are daunting. An example is a project begun in 1997 by GlaxoSmithKline in which, by 2005, some 80,000 patients and controls had been enrolled.[29]

■ Genomes to Proteomes

Conversion of the sequence information to an understanding of the posttranslational modification and folding processes represents another major step toward generating a receptor model. Messenger RNA mediates the translation of the complete DNA sequence to the corresponding sequence of amino acids linked as peptides and in so doing edits out *introns*, large noncoding parts of the DNA sequence. The resulting proteins may differ substantially from what was writ as the inheritance from the dividing parent cell: repetitive sequences are omitted as are introns, noncoding sequences scattered throughout the genome.

The edited sequence of all the DNA in a parent cell is generally referred to as the *expressed sequence* and the set of corresponding protein structures is identified as the *proteome*. In 2005, the National Institutes of Health announced awards of approximately $300,000,000 in a continuation of the Protein Structure Initiative, an effort to determine the proteomes of many organisms. The project, initiated in 2000, had by 2005 deciphered more than 1000 protein structures from genomic data and was expected to add another 5000 proteins to the database by 2010.

■ Purifying Proteins and Growing Crystals

Having generated the receptor(s) of interest, a next major step in growing protein crystals is developing a protein purification system. Highly purified protein is essential to produce a useful crystal because the presence of impurities impedes growth of a single large crystal and fosters growth of many small ones. Chromatographic separation techniques provide excellent separations and allow isolation of individual proteins.

However, relatively small differences between similar proteins significantly alter chromatographic behavior and could complicate or perhaps frustrate attempts to automate the separation process as a set of parallel systems. It is now commonplace to include in the expression system a handle or tag that allows facile identification and separation of the desired protein. Such tags include maltose-binding protein or other large proteins with strong affinities, but these necessitate cleavage of the tag and another separation step. A widely used small tag is a sequence of six histidines at the amino end of a protein. The tag latches onto nickel complexing polymer films or to beads that can be separated magnetically from a complex mix. Separation of the hexamer from the nickel with an imidazole reagent is straightforward.

There are several methods to induce a solution of a protein to form single large (a millimeter or more in length) crystals of protein. The protein is almost always in scarce supply, so small volumes of protein solution are the rule. One attempts to dissolve as much protein in as little solvent as possible, centrifuging to separate (and recover) undissolved material. The solvent is almost always water, but it may be a more complex system containing three or more other components in addition to the anticipated ligand: a co-solvent such as dimethylsulfoxide, low concentrations of highly soluble buffer and other salts, and a reducing agent to protect against air oxidation of the anticipated ligand.

All these methods involve gradually increasing the concentration of dissolved protein to a value above its solubility. Obviously, one might add more protein and stir the mix to dissolve the added bit. But stirring or other agitation favors formation of multiple small crystals, not the single crystal wanted. One may slowly cool a warm solution of protein and so at some temperature exceed the solubility of the protein. That must be done slowly to avoid setting up convection currents that disturb the mix. The most gentle and most likely to succeed method is vapor diffusion. Controlled evaporation of water in a closed chamber containing a nonvolatile desiccant gradually concentrates a protein solution. Alternatively, exposing the aqueous protein solution to a water-miscible volatile may allow the volatile to equilibrate with the water and so effectively reduce the amount of water available to dissolve the protein, in effect concentrating it. This latter method also requires careful thermal control.

Membrane proteins pose considerable challenges. Their cellular location in vivo puts them in extensive contact with both cytoplasmic and lipid layers; thus these proteins have both hydrophilic and lipophilic surfaces that do not readily stack one above another to form a crystal lattice. Use of a small co-solute that forms a lipidic cubic phase, for example, glyceryl monostearate, has allowed crystallization of a limited number of membrane proteins in the past decade.[30]

■ High-Throughput Systems for Growing Crystals

Optimizing conditions for growth of a single-protein crystal involves independently varying the types and concentrations of the several components of the mixture in which the protein is dissolved. Such optimization may best be pursued with the aid of any of several commercially available robotic systems. Typically, these robotic systems employ industry standardized plastic plates in which multiple flat bottom wells are molded in uniformly positioned sets of 96, 385, or 1536. These plates were initially developed for use in manual microtiter assays of microbial growth, enzymatic activity, immunoprecipitation, and so on.

By about 1990, clinical and industrial needs for performance of large numbers of repetitive tests led to development of automated instruments able to deliver small liquid volumes and to measure absorbance or emission of light by material in the wells.

The need for additional capabilities in conjunction with the human and other genome projects led to use of multiple quantitative additions of smaller volumes and to realization of throughput speeds of more than 50,000 samples per day. Increasing automation of delivery of crystallization solution components has been matched by successive reductions in the finished volumes of crystallization experiments. In 2006, the volume of mixed fluids needed in one well of a crystallization experiment was less than 100 microliters.

Detection of crystallization in exactly positioned liquids resting on the optically flat bottoms of experimental wells is monitored using microscopes fitted with digital image recorders. Such monitoring is as quick as taking a digital photograph. But crystal nucleation does not occur at some predictable time, and crystal growth to a discernible size is not an instantaneous process. Therefore, crystallization experiments must be monitored by capturing and analyzing a series of consecutive images. It is possible to screen many variables (co-solvent, salts, antioxidant, etc.) in parallel and so to identify the conditions likely to provide X-ray diffraction-quality crystals.

Once a set or a few sets of conditions that yield single crystals of receptor protein have been experimentally identified, it is reasonable to invest more protein in growing larger crystals for X-ray study. At least one milligram-sized crystal of the native protein is currently needed for crystallographic characterization. If several crystals of the protein are available, it is of interest to soak at least one briefly in a solution of the candidate ligand expected (hoped) to bind to it. If the crystal binds with the ligand, the resulting diffraction pattern will differ from its previous pattern, indicate binding, and may define the position at which the ligand binds to the protein.[31]

In some instances, brief soaking in ligand solution results in cracking of the crystal. This is usually taken as evidence that the binding interaction is very strong and that binding induces substantial conformational changes in the protein. These findings are useful guides in determining the extent to which other candidate congeners are likely to fit the binding site on the protein molecule.

In some rare cases, it is possible to cocrystallize the protein and a candidate ligand.[32] In such an instance, there is little doubt about the goodness of fit of the candidate ligand to the binding site on the protein.

■ New and Novel Proteins

The availability of nearly complete human genome sequence information (Build 35 is about 99% complete in 2006[33]) and methods for isolating corresponding individual proteins of unknown structure or function have resulted in recognition of some 2000 to 3000 genes and their corresponding proteins as possible drug targets. The consensus druggable protein appears to be one that presents folds in which drug-like chemicals can fit and that displays functional moieties with which those chemicals may interact. Proteins lacking such properties may have interesting features and/or generate important biological responses but probably are not pharmaceutically accessible in the near term.[34,35]

This abundance of druggable genes has been a mixed blessing for the pharmaceutical industry and has prompted solving of X-ray crystal structures in as rapid a manner as possible. The intent is to categorize the function of a new protein by comparison with protein structures of known function. High-throughput crystal growth and X-ray analysis are essential to rapid screening and development of new drug entities targeting new proteins. Newly recognized protein targets call for synthesis of many candidate drugs and efficient means to screen those many candidates for drug-like activity at receptors whose function is not yet known. All this costs a lot.

Moreover, identifying the function of a newly recognized protein may not produce a therapeutic effect or one that is clinically successful. It is tempting to reason one gene Æ one protein–Æ one target, but many receptors comprise heteromeric assemblies of subunits encoded by distinct genes. (For example, for the pharmacologist an ion channel in an intact animal may behave as a single target but a candidate drug binding to an isolated protein of the channel may not elicit a recognizable response.)

Certainly many successful drugs act at multiple targets, accounting for at least some side effects, but the clinical utility of these drugs results more from the net effect than from single receptor specificity.[36]

■ Fitting Drug to Receptor in Silico and Culling Misfits

When the structure of a drug binding site is known, usually from studies involving an active compound, it is frequently of interest to consider analogs of the active compounds that may have greater bioactivity, an exploitable secondary activity, or fewer unwanted side effects. Virtual screening using rapid automated fitting of drug to receptor becomes an attractive tool. It may be used to select probably active compounds from libraries of structures, such as those in a corporate compound collection, or to assemble candidate structures from a virtual catalog of structural parts.

And virtual screening may be used concurrently for negative selection, ruling out toxicophores or compounds with poor water solubility or poor oral bioavailability. Lipinsky's rule of five, which eliminated compounds that violated one or more of the rules, may be considered an early virtual screening method.[37] Those rules amount to categorical tests predicting good oral absorption and/or permeation if the compound being considered has fewer than 5 hydrogen bond donors; fewer than 10 hydrogen bond acceptors; a molecular weight of less than 500; and a logarithm of the octanol/water partition coefficient (logP) of less than 5.

More recently, Clark and Picket[38] observed that oral bioavailability of absorbed molecules is minimal if their polar surface areas exceed 100 to 140 square angstrom units, and Veber[39] suggests oral absorption of a compound is maximal if it has seven rotatable bonds. Pursuing this set of ideas, Congreve et al.[40] developed a rule of three (or six, depending on what one counts) for synthons, the carbon, oxygen, nitrogen, sulfur, and other skeletons of molecular fragments such as chains of atoms, ring structures, and the like. They based the rule on electron density maps generated from X-ray crystallographic studies of weakly interacting small structures. The rule of three suggests selecting synthons in which molecular weight contribution is less than or equal to 300; the number of hydrogen bond donors is less than or equal to 3 and of hydrogen bond acceptors less than or equal to 3; the calculated logarithm of the octanol/water partition coefficient is less than or equal to 3; the number of rotatable bonds is less than or equal to 3; and the polar surface area is less than or equal to 60 square angstrom units.

More than 20 computer programs collectively offer many thousands of synthon fragment images that may be clicked together in silico to form candidate structures similar to the known actives. The synthon skeletons can be fleshed out with hydrogens by using the Sybyl program and then "3D-ized" with Concord and/or Corina programs. Additional programs rank similarities between the guide structures and the candidate analogs.[41]

Linking synthons together to represent structures likely both to have bioactivity and to be synthetically feasible allows in silico estimation of the ability of the constructs to bind to a receptor site, to dock. Docking programs seek to find the best fits of ligands and receptor sites. Extensively used docking programs include FlexX,[42] DOCK,[43] and GOLD.[44] In most docking programs, the small molecule can be modified to allow conformational changes, but the protein receptor is held rigid.

Several strategies are applied to match ligand and receptor. In FlexX, the Ogston three-point attachment concept[45] is brought to bear; successive triplets (triangles) of points on each putative ligand in its several conformations are matched to triads of all possible interaction sites on the receptor. The fits are scored and recorded with the conformations in a table of triplets of interaction sites. The DOCK algorithm has evolved over more than 20 years. In DOCK, the receptor is envisioned as a pocket in which spherical loci of possible ligand interaction are located. The "goodness" of fit of at least four atoms of a putative ligand (or its disembodied parts) into these interaction spheres is scored and recorded in successive iterations accommodating possible orientations and conformations of the ligand. The GOLD program utilizes parallel algorithms that allow full ligand and partial receptor flexibilities and iteratively searches for hydrogen bonding and other energy-minimizing interactions.

Docking programs differ, sometimes importantly, in how the score, an estimate of the change in energy of association between ligand and receptor (the binding affinity), is computed. Two broad types of scoring calculations are used. They are what might be called empirical and knowledge-based. Empirical systems rely on measures of binding constants in known ligand-receptor systems chosen to include several sets of ionic, hydrogen, and hydrophobic bondings and corresponding entropy values. Knowledge-based scoring systems are derived from the Boltzman law and are used to calculate energies between attracting and repelling atom pairs.

The systems differ from one another in the number of ligand-receptor pairs, reference states, ranges of interacting distance, and other variables used in computing the scores. Scoring programs in current use include ChemScore, GoldScore, and DrugScore. Neither docking nor scoring programs are perfect; human perception still works. Fiorino[46] recently described finding three additional highly active compounds patterned on a known active by in silico docking and scoring augmented by in-person viewing of the results to eliminate false-negative scores.

It is not enough to identify molecules that might bind to receptors and estimate the goodness of fit. The actual chemical substances need to be in hand for testing. The in silico ability to assess the match of a molecular structure of a candidate drug with that of a binding site on a protein relatively accurately and quickly has been paralleled by the ability to prepare extensive sets or libraries of chemical substances that share a common scaffolding and likelihood of exerting biological activity. Two main approaches to the generation of chemical libraries have been pursued, combinatorial and parallel chemical synthesis.

■ High-Throughput Chemistry: Combinatorial Synthesis

If a chemical substance is found to have some desirable/useful biological activity, for example, it reduces elevated blood pressure or stops an infection, the finder of that activity is likely to hope to capitalize on it not only to benefit all humankind, but particularly the finder. And the finder will no doubt hope to find other chemicals that might share that activity in more potent or less toxic form. The finder will want to prepare those other related chemicals and test their activities.

Preparing those other related chemicals has for decades relied on the one-by-one synthesis of new related compounds, *analogs*. That changed in 1982 when Árpád Furka described combinatorial chemistry, a method to prepare a multiplicity of closely related substances, for example, peptides, essentially simultaneously.[47,48]

The essence of Furka's approach was to generate mixtures of chains of a given number of amino acids in every possible sequence and submit the mixtures for screening tests. (It may be noted that shortly before Furka's work was recorded, reports appeared describing tripeptide pituitary hormones and the pentapeptide endogenous opioid peptides Leu-enkephalin and Met-enkephalin.[49])

The choice of a peptide chain as the basis of the combinatorial structure allowed use of amino acids for each link and thus the advantage of having the same functional group, amino or carboxyl, to react at each increment in chain length. To facilitate handling, one end of each peptide chain was fixed at the carboxyl group to resin beads, and additions to the chain were made by so-called solid state synthesis. The carboxyl groups were attached with a bond type that could be easily cleaved under specific conditions but that was otherwise stable. Additionally, side products of a coupling reaction could be removed by pouring off the solvent and rinsing the beads bearing their attached peptide chains with fresh solvent.

At the start, the activated resin was divided into N1 equal portions (where N1 was the number of different amino acids available to react at one end, the acidic end, of the anticipated chain). The amino group of each of the amino acids was protected by attachment of a small blocking group that could be removed without disturbing other bonds. Each type of amino-protected amino acid was then allowed to react at its acidic end with one of the resin bead portions. After reaction the beads with pendant amino acids were unblocked to generate aminoacyl modified resins. An aliquot of each resin sample was reserved for subsequent use.

The remaining portions of the aminoacyl resins were carefully mixed, divided into N2 equal portions, allowed to react separately with one of the N2 types of protected amino acids to generate dipeptides, and then unblocked. As before, aliquots of each resin sample were reserved for subsequent use and the mixtures of dipeptides on each resin sample were cleaved for use in bioactivity tests.

Again, the remaining portions of the now dipeptide-bearing resin were carefully mixed, divided into N3 equal portions, allowed to react separately with one of the N3 types of protected amino acids to generate tripeptides, and then unblocked. As before, aliquots of each resin sample were reserved for subsequent use and the mixtures of tripeptides on each resin sample were cleaved as before.

This process was repeated until the desired *n*-length of peptide chain was realized. The sum of all the peptides formed S is as follows:

$$\mathrm{S} = \mathrm{N1} \times \mathrm{N2} \times \mathrm{N3} \ldots \times \mathrm{N}n$$

a large number, but the number of coupling reactions C is comparatively small:

$$\mathrm{C} = \mathrm{N1} + \mathrm{N2} + \mathrm{N3} \ldots + \mathrm{N}n$$

After all the planned amino acid chains had been attached to the sets of resin beads, equal samples of each set of beads were mixed and the peptides were cleaved from the resins to provide material for bioactivity testing.

A Combinatorial Example

If 10 amino acids were employed at each of 5 consecutive steps, the process would generate 100,000 differently sequenced chains of 5 amino acids, a cornucopia of possible new drugs but in a mixture of enormous complexity. Larger sets of starting compounds would of course produce yet more complex mixtures, but orderly ones. The utility of combinatorial largess did not become evident until the mixture was screened for biological activity.

If the peptide mixtures produced by the last synthetic step showed bioactivity, isolation and identification of the active peptide(s) were the next order of business. Each of the samples of resin to which the final peptide links had been attached, the samples one peptide shorter, were separately treated to release their respective peptides. Then, each of the freed peptide mixes was screened for biological activity in as quantitative a manner as feasible to determine the most active peptide mix and to assess how activity varied with the terminal amino acids. Finding active mixes narrowed the field of search.

Next, each of the samples of resin to which the penultimate peptide links had been attached was separately treated to release the respective two-unit-shorter peptide mixes. Each of these two-unit-shorter peptide mixes was likewise screened for biological activity in as quantitative a manner as feasible to determine the most active shorter peptide mix and to see how activity varied with the amino acid sequences.

This process was repeated with resin samples bearing successively shorter peptide chains until an amino acid sequence without activity was encountered. Reasoning backward from the longest active peptide chain allowed identification of the sequence associated with the measured activity. Clearly, expenditure of considerable resources for biological screening was key to unraveling the complexity of the synthetic mixtures.

To confirm the inferences arising from an experimental sequence such as described previously, it is always necessary to make the inferred compounds and to show that the resulting substances have the expected biological activity.

The logic of combinatorial synthesis is not limited in its application to peptide substances. It may equally be applied to oligosaccharides, oligonucleotides, and sequential polycondensates. Recent development of readily cleaved and broadly applicable linker groups to join resin beads (or films) to the molecular scaffold on which candidate drugs are constructed has increased the range and utility of combinatorial synthesis. Further, combinatorial synthesis typically employs reaction mechanisms that may be applied to a wide range of related compounds under essentially identical conditions (eg, Mitsonobu reaction, an intermolecular dehydration reaction between alcohols and acidic components to give stereospecific products,[50] and Suzuki coupling,[51] joining of two aromatic nuclei through reaction of an arylboronic acid and an aryl halide).

■ High-Throughput Chemistry: Parallel Synthesis

Most chemical syntheses have classically been carried out in solution phase with each reaction mixture in a separate vessel. That has not usually been a problem. The difficulties in synthesis have typically been separating excess reactant, side products, and solvents to purify the desired product. Typically, this separation, the *workup*, has involved adding a second solvent immiscible with the first, more or less selectively partitioning excess reactant, side products, and desired product between the solvents, separating and collecting the product-rich solvent phase, and evaporating the solvent from a residue, mostly of desired product. Often, the desired product required a further chromatographic purification. In short, running the reaction has not been the problem, isolating the product has. The process did not readily yield large numbers of closely related candidate drugs and neither was it amenable to automation.

That changed with the commercial availability of high-quality polymer polystyrene resin beads carrying, initially, ion-exchange sulfonic acid or quaternary ammonium groups and later other reactive functionalities such as isocyanate or aldehyde. These beads, especially ion-exchange beads, prepackaged sterile in 5-,10-, 25-, and 100-mL hypodermic syringe tubes have greatly facilitated cleanup and solid phase extraction of urine samples in clinical laboratories. Such tubes are uniformly packed with resin or sorbent, typically occupying not more than half their volume, allowing all the sample to be added in one step. This application of tube processing was soon followed by its application to multiplexed parallel synthesis of small batches of compounds.[52–54]

The concepts involved are simple. In the simplest sort of application, a mixture from a completed small-scale reaction, for example, an amide synthesis mixture made using a slight excess of acidic reactant, is poured into a tube packed with a quaternary-ammonium fuctionalized resin bead. The excess acid is trapped by the resin and the product flows out dissolved in the initial reaction solvent. Any reaction solvent and contained product held up in the resin bed is eluted into clean receivers with an additional small volume of the initial reaction solvent. If 10 different acids are to be used in 10 similar reactions, 10 tubes and receivers are set up in parallel. If a hundred acids are to be used, a robotic system is desirable, but keeping track of the starting materials and labeling the receivers emerges as a task requiring attention.

Alternatively, a product may be held in a resin-packed tube while reaction solvent, excess reagent, and side products are rinsed away with more of the same solvent. Then, an appropriate solvent can elute the desired product. Obviously, separate tubes containing cation- and anion-exchange resins may be used in tandem, the first tube delivering into the second without intervention.

The variety of commercially available resin-supported functionalities useful as scavengers of unreacted components includes benzaldehyde for primary amines or hydrazines, benzylisocyanate for both primary and secondary amines and for alcohols, benzyltriethylammonium carbonate for carboxylic acids and phenols, triphenylphosphine for alkyl halides, and benzyltrisaminoethylamine for acid chlorides and isocyanates.

Variations in the packing of filtration tubes are useful and sometimes necessary. Polystyrene beads may disadvantageously swell in common halogenated solvents, so it may be useful to pack tubes with purified silica instead. Water and diluted acids or bases are easily retained by silica beds. Organic solvents added to such water-wetted columns readily equilibrate with the interstitial water, allowing selective partitioning of hydrophilic compounds into water retained in the silica bed and nearly complete elution of the organic phase.

Equipment for the separate steps in parallel synthesis protocols is available in almost all chemical synthesis laboratories but is not necessarily organized or integrated for parallel synthesis. Such equipment includes temperature-controlled heating blocks that accept multiple identical vessels, multiunit filtration racks, and nitrogen blow-down or heated centrifugal evaporators. Robotic systems offered by several manufacturers are adapted or adaptable to multisample processing.

Additionally, a number of liquid-handling systems may be adapted to parallel separation. Thus, it may be practical to generate many more candidate drugs a day using a parallel synthesis, or more accurately, a parallel separation approach. Refined automation of workup in place of the complex manipulations and high levels of eye–hand coordination that characterize classical product isolation are brought to bear in making parallel synthesis an economical way to prepare extensive libraries of candidate drugs. Robots are tireless.

Teflon 96-well plates in the format developed for microbiologic and enzymatic assays are of particular utility for high-throughput small-scale syntheses. Reaction mixtures in which precipitates form in one plate may be robotically pipetted to a second, filter-bottomed plate (eg, bottoms of glass fiber, microporous polypropylene, polyvinylidene fluoride, etc.) to capture the solid while the filtrates are collected in a third plate made in the same 96-well format.

■ Automated Cleanup

All the desired products from a group of parallel syntheses, even resin-based reactions, require workup to reduce impurities as much as practicable and, looking ahead, to lower the incidence of false positives when tested in biological systems. Selective resins might certainly be used to scavenge expected impurities and side products from reactions performed in 96-well plates. Resin beads, which quickly dry, develop high static charges, and scatter, are difficult to pack into even 10-mL tubes. Prepacked 96-well plates suited to parallel synthesis separations do not seem to be commercially available.

A high-throughput purification system adaptable to products of parallel syntheses is available, but there are none commercially available capable of purifying, quantitating, and tracking very large numbers of reaction mixtures, especially in the small volumes characteristic of the automatable 96-well format. Ideally, for a high-throughput system, each filtrate is separated/purified by high-performance chromatography before the individually collected fractions are characterized.

■ Product Characterization

Ultraviolet spectrophotometers able to utilize volumes of 1 μL became available in 2005 and with reference to a calibration file can estimate the amount of product(s) in a sample. These seem ideal for small-scale synthesis, but at this writing are not yet adapted to robotic sample handling or to automated data recording and tracking. Instead, it has become routine to characterize the structure and identity of major components in an aliquot of a chromatographic fraction with liquid chromatography/mass spectrometry or electrospray/mass spectroscopy.

The technique of ambient mass spectroscopy or desorption electrospray ionization (DESI) is a very recent innovation in structure determination applicable to high-throughput analysis of very small sample sizes. Very briefly, electrically charged solvent droplets are sprayed in ambient air onto a succession of dried samples arranged in an array on a substrate such as a microscope slide. The charged droplets cause ions from the sample surface to be released and these ions are swept into the vacuum interface of a mass spectrometer where the ions (and daughter ions in a tandem instrument) are rapidly and sensitively analyzed.[55,56]

In short order, high-throughput synthesis generates an enormous amount of data that must be linked unambiguously to each reaction mixture and anticipated candidate drug. Commercial systems capable of tracking huge volumes of data arising from synthesis, separation, identification, and quantitation of products in one 96-well plate per day do not seem to be available. Large pharma has assembled these for in-house use; Everett et al. have described one such as at Pfizer UK.[57] It is evident that many new series of chemical substances may be generated much more rapidly than before the adoption of high-throughput systems.

■ Structural Genomics Leads to Structural Proteomics

Structural genomics research applied to the human and other genome sequences allows identification of a huge number of proteins capable of serving as drug targets. In many instances, stretches of DNA sequence correspond to the known structure of specific proteins; in many others, the genetic information represents proteins with unknown structure or function.

The transition from structural genomics to structural proteomics calls on the traditional larger scale separation of proteins by chromatography or two-dimensional gel electrophoresis followed by mass spectroscopic identification of the proteins. An application of this information can provide a differential basis for comparison of expression during the maturation of an organism/culture or in the development of a disease. This traditional procedure provides a chance to monitor the associated proteins as the process unfolds. The information in an inventory of an organism's proteins and their functions is also useful in its own right.

■ Target Selection Through Structural Proteomics

Elucidation of the proteomes of many microorganisms has allowed assembly of this information in a series of flow diagrams separately displaying the individual metabolic pathways of each of numerous species of microbes. This information may be displayed as a diagram for each organism in which the sequences of enzymatic substrates and products are shown as a series of loci, each substance appearing only once. The enzymes catalyzing conversion of substrate to a first and then successive products are represented by arrows. Thus, the web-like diagrams show not only the metabolic routes but also metabolic junctions. It is immediately evident on examination of such diagrams that some few enzymes act as choke points because they are the only means by which a given organism can generate an essential product if it is not available in the environment.

These diagrams may be used in conjunction with information on the rates at which substrates are converted by enzymes to their respective products under various conditions. So informed, one can use these flow diagrams in various modeling approaches, notably incorporating constraint-based flux-balances, and successfully use these to calculate deletion phenotypes and either optimal or fatal growth conditions.

The constraint-based flux-balance approach has been applied to a number of microbes. Notable among these are *E. coli, Helicobacter pylori*, and *Saccharomyces cerevisiae*. The result is identification of a metabolic core of essential reactions for each organism, reactions that never stop, in any of thousands of simulated environments.

Core reactions appear to be evolutionarily conserved and always active. An analogous study of more than 700 *Salmonella enterica* enzymes in a smaller number of environments identified 15 absolutely essential enzyme reactions.

Noncore reactions are species specific and may be considered conditionally active, that is, functioning as may be advantageous depending on nutrient supply or balance.

Core reactions, the unconditionally essential reactions, serve as targets of antibiotics that interfere with bacterial metabolism, for example, trimethoprim, fosfomycin, cycloserine.[58–60] If the core reactions are considered choke points, they identify the microbial metabolic fluxes that are particularly vulnerable to antibiotics.

It follows that coupling knowledge of the proteome and corresponding metabolic fluxes of other microbes, such as the *Mycobacterium avium* complex, would allow an informed search for agents capable of blocking unconditionally essential pathways in such organisms, and thereby offer a means of blocking further emergence of drug-resistant strains.

However, as Rene Dubos[61] pointed out in 1942 and again in 1952, microbes inevitably develop drug resistance and the more quickly if overused, a caution almost universally unheeded by a pressured medical community with few alternatives.

■ High-Throughput Binding Studies

It must be self-evident that most biological functions of cells do not involve DNA itself but protein molecules ultimately derived from DNA. It follows that preparation of protein microarrays analogous to DNA microarrays would be useful in screening for biological activity. Microarrays of individual protein receptors printed on glass slides provide a convenient and automatable format with which to assess protein interaction with candidate ligands or other small proteins.[62] Printed fresh, they are excellent tools.

As an alternative to printing, protein microarrays have been prepared by transferring multiply protonated proteins selected from mixtures on the basis of mass and charge and gently depositing them on solid or liquid-coated surfaces. Mass spectra of proteins from the resulting arrays have been shown to match those of the authentic compounds, and the arrayed proteins retain their bioactivity.[63]

Unfortunately, protein array components have short working lifetimes. Even frozen in place, proteins have widely different stabilities, a weakness that limits the utility of such arrays. Neither is it feasible to dry the arrays; proteins dried in contact with vitreous surfaces are quickly denatured. Nonetheless, the idea of protein microarrays is highly attractive for high-throughput screening.

Just as a large number of essentially identical cells may be transfected with many replicates of the same plasmid to generate substantial quantities of one protein, it is possible to transfect a large number of essentially identical cells with a large number of different plasmids. Such a procedure can result in a large number of differently transfected cells, each expressing a different protein.

The process depends on the availability of copy DNA (cDNA) probes. These probes are specific nucleotide sequences that bind only to their complementary DNA to form duplex strands. cDNAs may be prepared using restriction enzymes as described earlier. Alternatively, cDNA may be prepared using *reverse transcriptases*, viral enzymes able to reverse transcribe single-stranded messenger RNA to the corresponding single-stranded DNA. This alternative process is sometimes called *retro-synthesis*.

cDNA is often radioactively or fluorescently labeled as an aid in detecting the probe when it is incorporated into a plasmid in an array. cDNA, however made, can be incorporated into plasmids, essentially as previously described. Microscope slides may be robotically printed with cDNA-containing plasmids suspended in gelatin solution to generate microarrays in a manner analogous to preparation of DNA microarrays. Gelling of the gelatin solution fixes the plasmids in place on the slide. The resulting plasmid microarrays can have densities of 5000 to 10,000 spots per slide. Thus, whole genomes may be represented on a small number of slides.

Adding a lipoidal transfection agent such as polyethylenimine converts the plasmid microarray to one of lipoidal DNA complexes. Adding a suspension of adherent mammalian cells on top of the spots quickly results in a new microarray in which each spot contains about 25 to 100 transfected cells. Each so transfected slide represents a living microarray. Because the cells are placed on the plasmids rather than the plasmids on the cells, this process is sometimes referred to as *reverse transfection* and the array is called a *transfected cell array* (TCA).

Because each of the few transfected cells in each spot expresses only one protein and the spots have been printed at discrete locations, the transfected cell array may be considered a kind of microarray displaying many different proteins. If the cells in the array are those of eukaryotic (nucleated cell) origin, posttranslational modification (eg, glycosylation) of the expressed proteins may be expected. Using several cell lines allows one to distinguish posttranslational modifications characteristic of the lines and possibly to recognize protein–protein interaction.[64,65]

■ Assemble a Mouse

Transgenic and/or knockout mice have traditionally been used in genetically based gain- or loss-of-function studies. Their use entails very high costs for commercial strains and/or lengthy periods of in-house animal care for home-grown strains. Additionally, using such designer mice is complicated by their similarity to normal mice; they generally do not display a readily evident alteration in phenotype that might be associated with certainty to the mutated gene.

Transfected cell arrays may also be utilized to reduce the costs and lengthy care needed to use such mice. Transfected cell arrays may be constructed to exhibit specific genetic RNA interference (RNAi), and thus to generate dozens of phenotypes in a single array. No one spot in the array is equivalent to a whole genetically modified mouse. The array may be thought analogous to those holiday toys for children that bear the label "Some assembly required."

RNA interference, first described in nematodes,[66] is initiated by the action of the enzyme Dicer on long double-stranded RNA (dsRNA) cutting it into bits called *short interfering duplex RNA* (siRNA). A resulting snippet of siRNA associates with certain cytoplasmic proteins to form an RNA-induced silencing complex (RISC) the anti-sense strand of which guides the RISC to corresponding messenger RNA. RISC acts on the mRNA, slicing it into smaller segments that are then readily lysed by nucleases in the cytoplasm. A knockout organism results.

Mammalian cells are more finicky; they do not respond as well to siRNA prepared from long dsRNA as do nonmammalian cells. dsRNA longer than 30 nucleotide triads (nt) stimulates a lethal interferon response. But, mammalian siRNA strands between 21 and 15 nt work well to silence gene expression and can either be transfected or introduced into cells as synthetic agents. In contrast to nonmammalian cells, neither replication of siRNA nor its interference with gene expression is heritable in mammalian cells; the siRNA effect is itself silenced as cells repeatedly divide and the siRNA becomes diluted in the growing culture.

This problem has been overcome by developing plasmid constructs that contain a so-called short hairpin RNA (shRNA) along with drug resistance coding sequences that allow stable transfection and the ease of antibiotic-based cell selection.[67–69] Furthermore the duration of siRNA effectiveness has been increased through use of retrovirus and lentevirus as expression vectors.[70]

Movement of siRNA techniques from multiwell screens to transfected cell arrays has been relatively slow and applied only to small sets of genes coding for specific proteins to date. However, it is reasonable to expect the technical advantages developed and refined for high-throughput DNA microarray systems will accrue to transfected cell arrays, for example, use of smaller volumes of reagents, robotic spot printers, automated array scanning systems, and commercial availability of siRNA libraries to make TCAs cost effective. In parallel, computational tools to identify the genomic targets of siRNAs have become available.[71–72]

An immediate advantage of cDNA spotted microarrays is that they may be stored for extended periods before use in a screen or to replicate/confirm a prior screen. Additionally, only a relatively few cells are used to prepare a microarray, in general providing economy of scale and an advantage if the biological system is in short supply. More generally, cell microarrays are well suited to high-throughput screening for specific activities, such as interference with kinase signaling pathways. The cDNAs are printed at specific locations on each slide so that the cDNA at a spot and its corresponding expressed protein are known; the target protein of an active drug can be traced and identified rapidly. Their corresponding amino acid sequences can thus be learned.

■ High-Throughput Screening: G-Protein-Coupled Receptors

Of the currently 100 best-selling drugs, about a quarter act through G-protein-coupled receptors (GPCR). These include opioid agonists to block pain, b2 adrenoceptor agonists to control asthma, histamine receptor antagonists to suppress peptic ulcers, and angiotensin receptor antagonists to reduce hypertension. Together they represent a large share of the world pharmaceutical market.

Yet these drugs target only about 30 of the approximately 750 receptors in the GPCR superfamily of the human genome. About 400 others are considered likely to be of interest as drug targets. Natural ligands are known for roughly half of these putative drug targets. These GPCRs (and other receptors) for which no ligand has been identified are often referred to as *orphan receptors* and appear to offer opportunities for development of new drugs.

On the basis of sequence homology, GPCRs are usually grouped in three sets. The largest set, Group A, contains receptors for catecholamines, chemokines, glycoproteins, lipids, neuropeptides, and nucleotides. Group B includes receptors for other peptide ligands, and Group C encompasses metabotropic receptors for ligands such as gamma-aminobutyric acid and calcium ion.

Structure of G-Protein-Coupled Receptors

All GPCRs are large membrane-bound filamentous proteins in which eight hydrophilic lengths alternate with seven hydrophobic segments. By convention, the hydrophobic segments are serially numbered starting with the segment nearest the very hydrophilic carboxylic acid end. The acid end and the first hydrophobic segment normally position themselves at the interface between the cytoplasm and the cell membrane. The acid end is in the cytoplasm, the hydrophobic segment in the oily membrane; the free energy change drives the system. The amino end extends into the extracellular medium.

In successive adjustments of free energy, the hydrophobic segments become aligned side by side spanning the oily membrane between the interfaces. The hydrophilic lengths form loops alternately projecting past the external interface into the surrounding medium or through the internal interface and into the cytoplasm. Tracing these in order, the hydrophilic length connecting segments 1 and 2 is in the surrounding medium, the one connecting segments 2 and 3 in the cytoplasm, the length between 3 and 4 in the surrounds, that between 4 and 5 again in the cytoplasm, and so on. The filamentous GPCR is said to assume a sinusoidal disposition, snaking back and forth through the cell membrane.

The arrangement of hydrophobic segments in the membrane is yet more complicated. They approximate a cylindrical shape in which the hydrophobic segments are wrapped around to form a barrel-like assembly. The segments are positioned like barrel staves, parallel to the cylindrical axis with segments 1 and 7 aligned side by side.

The whole assembly resembles a transmembranal pore in which the exposed terminal amino group and hydrophilic extracellular loops form a binding site for large ligands; smaller ligands bind deeper into the barrel structure. In the barrel-like structure, the terminal carboxyl group of every G-protein-coupled receptor protein is positioned near the intracellular loop between the sixth and seventh transmembranal strands.

The loop between segments 6 and 7 usually serves as the binding site for G proteins. There are three G-protein subunits, usually identified as a, b, and g, assembled as a heterotrimer. More than a dozen closely related types of G proteins have been recognized, each with quite specific activities.

A number of individual ligands bind at more than one type of G-protein-coupled receptor. Ligand binding causes subunits of the G-protein heterotrimer to dissociate and to activate nearby membrane-bound enzymes that generate so-called second messengers. One major type of thus activated enzymes catalyzes conversion of purine triphosphates into the corresponding cyclic purine monophosphate (cyclic adenylmonophosphate or cyclic guanylmonophosphate). Another is a family of phospholipases that release phospholipid-derived esters (diacylglycerol or inositol triphosphate). The second messengers in turn activate protein kinases that trigger a cellular response. For example, inositol triphosphate triggers release of calcium ion from bound intracellular stores. The calcium then binds to calmodulin and kinases activating yet another set of intracellular enzymes by phosphorylation.

It may be seen that taking into account the sheer number and variety of G-protein-coupled receptor proteins, the nearly promiscuous ligand binding at these receptors, the variety of G-protein types, and the range of second messengers, disentangling their mechanisms and functions allows wide scope for development of new drug entities.

But G-protein-coupled receptor processes are not haphazard. The structures of the receptor proteins are known or knowable and in the intricate ballet of second messengers the same steps are traced in each repeated performance.

An example of one strategy to identify ligands for orphan receptors is the linking of a restriction fragment bearing the genetic sequence for the orphan receptor into a plasmid and its transfection into cells. The resulting expression systems are then exposed to a set of compounds that in nature might serve as ligands for the orphan receptor. Binding of a candidate ligand to the GPCR is monitored by measuring second messenger–induced effects.

The expression system needs to be chosen with care. It should supply G proteins, membrane-bound enzymes activated by G proteins, and the wherewithal to generate second messengers. These needs may be met with any of several well-characterized cell lines, notably Chinese Hamster Ovary (CHO), Human Embryonic Kidney (HEK), *Xenopus laevis* oocytes, or *Saccharomyces cerevisiae*.

Monitoring Transfection

Further, in a well-controlled experiment seeking ligands for the orphan, it is necessary to be sure the plasmid has been expressed in all the cells to be used. It may happen that not all cells in a culture will be transfected. Failure to generate second messenger effects might not mean the absence of an effective ligand, but rather that the plasmid DNA was not transfected and expressed as a surface protein.

It is relatively simple to enlarge the plasmid slightly to carry a tag of convenience. The DNA sequence coding for a protein such as hemagglutinin can serve as such a tag. Linking it to the N-terminal sequence of the orphan receptor restriction fragment builds it into the plasmid. As cells carrying the plasmid grow and divide, the tag is displayed at the surface of the cell together with the amino group of the receptor when both proteins are expressed.

The presence of the tag, and thus of the GPCR too, may be established and assured by adding a fluorescently labeled hemagglutinin antibody to the cell culture. So labeled, transfected cells may be separated from nonfluorescent cells and collected with a fluorescence-activated cell sorter. The population of transfected cells may be grown to numbers needed for high-throughput screening in multiwell plates.

To screen for the ligand(s) binding to G-protein receptors one needs not only to generate the second messengers but also to have a way to recognize them should they appear. In a mammalian cell-based screen, there is almost always release of second messengers.

Some dyes, such as fura-2, fluoresce when they bind calcium. If such a dye in the form of its neutral lipophilic ester is added to a resting cell culture, the ester diffuses across plasma membranes and, once in the cytoplasm, is hydrolyzed to the corresponding acid. The resulting negatively charged acid now cannot diffuse back out of the cell. Though it does not fluoresce itself, it can form a stable fluorescent complex with calcium but not other ions. The change in cytoplasmic concentration of free calcium ion in response to ligand binding may thus be monitored by measuring fluorescence at a wavelength characteristic of the fura-2-calcium complex.[73,74]

Additional assays based on more specific imaging of the cell, of the GPCR, or of another interacting protein have been developed. These assays are based on movement of the protein within the cell or on a change in spectral properties resulting from the binding of ligand to receptor[75,76] and, because they provide more information, are often called *high-content assays*.

A green autofluorescent protein (GFP, derived from the jellyfish *Aequoria victoria*) when attached to the cytoplasmic acid end of a b2 adrenoceptor has been shown to move with the receptor from the cell surface. After binding the ligand, the receptor is internalized in an acidic endosome and subsequently recycled to the surface. Movement of the GFP-GPCR conjugate is evidence of ligand binding[77] and can be followed by pseudoconfocal imaging systems adapted to monitor multiwell plates.

If a laboratory can enlist the skills needed to culture and transfect melanophores, these cells can serve in a screen selecting ligands for orphan GPCRs. *Melanophores*, cells from the nearly pigment-free frog *Xenopus laevis*, contain melanosomes, intracellular organelles carrying the dark brown/black pigment melanin.

Binding of a candidate ligand to an orphan receptor expressed in a melanophore may traduce a signal that dissociates the a subunit of the guanine nucleotide binding protein from the heterotrimer. If the trimer is thus broken up, the enzymes adenylate cyclase and phospholipase C become activated and the second messengers cyclic adenosine monophosphate (cAMP) and diacylglycerol are formed. A melanocyte responds to cAMP by dispersing its melanosomes quite uniformly throughout the cell, quickly causing it to appear darker. Melanosomes in a cell aggregate if adenylate cyclase is inhibited and the cell thus appears to become lighter. The response is sensitive, occurs within minutes, and is readily monitored colorimetrically. Constituative activity serves as an indicator of successful transfection. *Xenopus* melanocytes are readily adaptable to growth in multiwell plates and high-throughput screening of orphan GPCRs for candidate ligands.

Complex as G-protein-coupled receptor screening assays may seem, they are readily adaptable to 96-well or denser plates and well within the capabilities of commercially available automated systems that can perform 100,000 screens per day. The strength of this approach derives from direct targeting of the disease phenotype as a mechanistic study using relevant cell models.

■ Validating Target Receptors

Identification of a class of ligands or a new ligand in a class that binds to an orphan GPCR is exciting. But it is important to connect (or to use the *mot de jour*, validate) the newly found receptor with a disease for which a marketable drug is needed. Receptor validation is a series of steps in which the cumulative weight of evidence is the relevant measure.

A first step is attempting to match the genomic sequence of the presumed orphan receptor with sequences of known receptor function, but matching, although a clue, does not guarantee the expected function. Similarly, finding a ligand that activates a receptor provides another item of evidence. Steps matching the orphan receptor to an endogenous ligand and in vitro to a receptor that varies expression in health and disease offer stronger validation. So too does matching the orphan to the expression of the receptor in healthy and diseased animal models and humans. Perhaps the last steps in validating a candidate receptor are refining the initial set of candidate ligands to an optimized candidate and to advance the ligand as a drug that is effective in clinical treatment of the disease. Validation that begins as high-throughput screening gradually changes to low-throughput clinical experiments extending over days to months and longer.

Validation is tricky business. For example, high-throughput screening coupled with an understanding of the global metabolic fluxes in an infectious organism may allow efficient design, synthesis, and validation of a new antimicrobial. It is not certain that these combined capabilities would today anticipate drug action on unintended targets, for example, aminoglycosides on the eighth cranial nerve and ability to hear.

It is immensely helpful, but not necessary, to understand the origins of a disease condition to find and validate drugs that benefit the patient. For example, though essential hypertension has been recognized for decades, its genetic or other origins remain uncertain. Useful drugs for its treatment appear to act by indirect targeting. Quick-paced technologies validated at each step have allowed development of effective therapies using angiotensin-converting enzyme inhibitors, calcium channel blockers, and a 1-adrenergic antagonists.

There is, however, another set of drug targets to which high-throughput methods do not now seem amenable. Many complex disease conditions, such as depression and schizophrenia, do not readily admit study of the phenotypic cell isolated from the intact organism. It is possible these and others may arise from single nucleotide polymorphisms.

Huge libraries have been developed by the pharmaceutical industry. These catalog synthons, chemical entities, X-ray diffraction and mass spectrometric data, ligand activities, known and orphan receptors, drug effects and side effects in many diseases, diagnostic criteria, clinical histories, treatment outcomes, patient-specific genomes, and patient-specific SNP cohorts. Holdings in these libraries have been correlated, mostly in the physical sciences. The computational power needed to integrate these entire libraries is enormous and the costs staggering. At successive stages validation will be both essential and challenging.

References

1. Booth JB, Zemmel R. Prospects for productivity. *Nat Rev Drug Discov*. 2004;3:451–456.
2. Chapman JT. Drug discovery—the leading edge. *Nature*. 2004;430:109.
3. Zambrowicz BP, Sands A. Knockouts model the 100 best drugs: will they model the next 100? *Nat Rev Drug Disc*. 2003;2:38–51.
4. Loewi O. Pflugers Arch. *Gesamte Physiol*. 1921;189: 239–242.
5. von Euler US. *Handbuch der Experimentellen Pharmakolgie*. Heidelberg, Germany: Springer-Verlag; 1946:186–230.
6. Chan L, O'Malley BW. Mechanism of action of sex hormones. *N Engl J Med*. 1976;294:1322–1328, 1372–1381, 1430–1437.
7. Vane JR. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nat New Biol*. 1971;231:232–235.
8. Perutz M. Early days of crystallography. *Meth Enzymol*. 1985;114:3–18.
9. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem*. 1983;4:187–217.
10. Weiner SJ, Kollman, PA, Nguyen DT, Sase DA. An all atom force field for simulations of proteins and nucleic acids. *J Comp Chem*. 1986;7:230–252.
11. Watson JD, Crick FH. A structure for deoxyribonucleic acid. *Nature*. 1953;421:397–398.
12. Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. *Nucleic Acid Res*. 2001;29:3705–3727.
13. Barany F. The ligase chain reaction in a PCR world. *PCR Methods Appl*. 1991;1:5–16.
14. Sanger F, et al. Nucleotide sequence of bacteriophage phiX 174. *J Mol Biol*. 1977;125:225–246.
15. Strauss EC, Kabori JA, Siu G, Hood LE. Specific primer directed sequencing. *Anal Biochem*. 1986;154:353–360.
16. Dovichi N. Development of a DNA sequencer (letter). *Science*. 1999;285:1016.
17. Zhou G, Kamahori MH, Okano K, Harada K, Kambara H. Miniaturized pyrosequencer for DNA analysis with capillaries to deliver deoxynucleotides. *Electrophoresis*. 2001;22:3497–3504.
18. Cregg JM, Cereghino JL, Shi J, Higgins DR. Recombinant protein expression in *Picia pastoris*. *Mol Biotechnol*. 2000;16:23–52.
19. Kigawa T, et al. Cell-free production and stable isotope labeling of milligram quantities of proteins. *FEBS Lett*. 1999;442:15–19.
20. Stewart L, Clark R, Behnke C. High throughput crystallization and structure determination. *Drug Disc Today*. 2002;7:187–196.
21. Venter JC, et al. The sequence of the human genome. *Science*. 2001;291:1304–1351.
22. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
23. Ewing B, Green P. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Gen*. 2000;25:232–234.
24. Brookes AJ. The essence of SNPs. *Gene*. 1999;234:177–186.
25. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. *Nat Rev Genet*. 2003;4:937–947.
26. Van Eerdewegh P, et al. Association of the ADAM32 gene with asthma and bronchial hyper responsiveness. *Nature*. 2002;418:426–430.

27. Giallourakis C, et al. IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease, and identification of a novel association with ulcerative colitis. *Am J Hum Genet*. 2003;73:205–211.
28. Martin E, et al. Association of single-nucleotide polymorphisms of the tau gene with late onset Parkinson disease. *JAMA*. 2004;286:2245–2250.
29. Roses AD, Burns DK, Chissoe S, Middleton L, St. Jean P. Disease specific target selection: a critical step down the right road. *Drug Disc Today*. 2005;10:177–189.
30. Landau EM, Rosenbusch JP. Lipidic cubic phases: a novel concept for crystallization of membrane proteins. *Proc Natl Acad Sci USA*. 1996;93:1452–1455.
31. Gill AG, et al. Identification of novel p38 alpha MAP kinase inhibitors using fragment-based lead generation. *J Med Chem*. 2005;48:414–426.
32. Card GL, et al. A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold based drug design. *Nat Biotechnol*. 2005;23:201–207.
33. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931–945.
34. Hopkins AL, Groom CL. The druggable genome. *Nat Rev Drug Disc*. 2000;1:727–730.
35. Orth AP, Batalov S, Perrone M, Chanda SK. The promise of genomics to identify novel therapeutic targets. *Expert Opin Ther Targets*. 2004;8:587–596.
36. Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Disc*. 2004;3:353–359.
37. Lipinsky CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997;23:3–25.
38. Clark DE, Picket SD. Computational methods for prediction of drug-likeness. *Drug Disc Today*. 2000;5:49–58.
39. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence oral bioavailability of drug candidates. *J Med Chem*. 2002;45:2615–2623.
40. Congreve M, Carr R, Murray C, Jhoti H. A "rule of three" for fragment-based lead discovery. *Drug Disc Today*. 2003;8:876–877.
41. Schneider J, Bohm H-J. Virtual screening and fast automated docking methods. *Drug Disc Today.* 2002;7:64–70.
42. http://www.ercim.org/publication/Ercim_News/enw29/kramer.html.
43. http://dock.compbio.ucsf.edu/44. http:// www.ccdc.cam.ac.uk/products/life_sciences/gold/45. Ogston AG. Specificity of the enzyme aconitase. *Nature*. 1951;167:693.
46. Forino JM, Jumg D, Easton JB, Houghton PJ, Pellechia M. Virtual docking approaches to protein kinase B inhibition. *J Med Chem*. 2005;48:2278–2281.
47. http://www.archiv bmn.com/sup/ddt/.CCHUN.pdf for the Hungarian original text.
48. Furka A, Sebestyen F, Asgedom M, Dibo G. General method for rapid synthesis of multi-component peptide mixtures. *Int J Peptide Protein Res*. 1991;37:487–493.
49. Hughes J, Smith TW, Kosterlitz HW, Fothergill LA, Morgan BA, Morris HR. Identification of two related pentapeptides from the brain with potent opiate agonist activity. *Nature*. 1975;258:577–580.
50. Mitsunobu O, Masaaki Y. Preparation of esters of carboxylic and phosphoric acid via quaternary phosphonium salts. *Bull Chem Soc Japan*. 1967;40:2380–2382.
51. Miyaura N, Yanagi T, Suzuki A. The palladium-catalyzed cross-coupling reaction of phenylboronic acid with haloarenes in the presence of bases. *Synth Commun*. 1981;11:513–519.

52. Shuker AJ, Siegel MG, Mathews DP, Weigel LO. The application of high throughput synthesis and purification to the preparation of ethanolamines. *Tetrahedron Letts*. 1997;38:6149–6152.
53. Gayo LM, Suto MJ. Ion-exchange resins for solution phase parallel synthesis of chemical libraries. *Tetrahedron Letts*. 1997;38:513–516.
54. Lawrence RM, Biller SA, Fryszman OM, Poss MA. Automated synthesis and purification of amides: exploitation of automated solid phase extraction in organic synthesis. *Synthesis*. 1997;5:553–558.
55. Takats Z, Wiseman JM, Gologan B, Cooks RG. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science*. 2004;306:471–473.
56. Cooks RG, Ouyang Z, Takats Z, Wiseman JM. Ambient mass spectrometry. *Science*. 2006;311:1566–1570.
57. Everett J, Gardner M, Pullon F, Smith GF, Snarey M, Terrett N. The application of non-combinatorial chemistry to lead discovery. *Drug Disc Today*. 2001;6:779–785.
58. Almaas E, Oltvai ZN, Barabasi A. The activity reaction core and plasticity of metabolic networks. *PLoS Comput Biol*. 2005;1(7):e68.
59. Almaas E, Kovacs B, Vicsec T, Oltvai ZN, Barabasi A. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*. 2004;427:839–843.
60. Becker D, Selbach M, Rollenhagen M, Ballmaier M, Meyer TF, Mann M, Bumann D. Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature*. 2006;440:303–307.
61. Dubos R, Dubos J. *The White Plague, Tuberculosis, Man and Society*. Boston: Little, Brown; 1952.
62. MacBeath G, Schreiber S. Printing proteins as microarrays for high throughput function determination. *Science*. 2000;289:1760–1763.
63. Outang Z, Takats Z, Blake TA, Gologan B, Guymon AJ, Wiseman JM, Oliver JC, Davisson VJ, Cooks RG. Preparing protein microarrays by soft landing of mass selected ions. *Science*. 2003;301:13251–13254.
64. Ziauddin J, Sabatini DM. Microarrays of cells expressing defined cDNAs. *Nature*. 2001;411:107–110.
65. Chang FH, Lee CH, Chen MT, Kuo CC, Chiang YL, Hang CY, Roffler S. Surfection, a new platform for transfected cell arrays. *Nucleic Acids Res*. 2004;32:e32.
66. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Melo CC. Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998;391:806–811.
67. Brummelkamp TR, et al. A system for stable expression of short interfering RNAs in mammalian cells. *Science*. 2002;296:550–553.
68. Paul CP, Good PD, Winer I, Engelka DR. Effective expression of small interfering RNA in human cells. *Nat Biotechnol*. 2002;20:4497–4500.
69. Paddison PJ, Caudy AA, Bernstein E, Hannon GJ, Conklin DS. Short hairpin RNAs (shRNAs) induce sequence specific silencing in mammalian cells. *Genes Dev*. 2002;16:948–958.
70. Scherr M. Modulation of gene expression by lentiviral-mediated delivery RNA interference. *Cell Cycle*. 2003;2:251–257.
71. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell*. 2003;115:787–798.
72. Rajewsky N, Socci ND. Computational identification of microRNA targets. *Dev Biol*. 2004;267:529–535.
73. Poenie M, Alderton J. Changes of free calcium levels with stages of the cell division cycle. *Nature*. 1985;315:147–149.
74. Szekeres PG. Functional assays for identifying ligands at orphan G-protein-coupled receptors. *Recept Channels*. 2002;8:297–308.

75. Conway BR, Demarest KT. The use of biosensors to study GPCR function: applications for high content screening. *Recept Channels*. 2002;8:321–341.

76. Milligan G. High-content assays for ligand regulation of G-protein-coupled receptors. *Drug Des Today*. 2003;8:579–584.

77. Barak LS, et al. Internal trafficking and surface mobility of a functionally intact b2-adrenergic receptor-green fluorescent conjugate. *Mol Pharmacol*. 1997;51:177–184.