Rediscovering the sweet spot in drug discovery

David Brown and Giulio Superti-Furga

Advances over the past decade in drug discovery technologies have not yet led to an increase in productivity. We analyzed the reasons that have led to this juncture and identify the selection of the right target and the right lead as crucial. New approaches are required to take full advantage of the genomics revolution. For targets, methods are becoming available for high-throughput proteome analysis and pathway characterization that synergize with studies of disease association and differential expression. For leads, methods are being developed that 'reverse' the high-throughput screening paradigm by mapping drugs and drug-like compounds back onto the proteome. The synergy between pathway mapping and compound mapping could allow the pharmaceutical and biotechnology industries to rediscover the sweet spot of research productivity.

David Brown
Cellzome AG
160 Centennial Avenue
Centennial Park
Elstree, Hertfordshire
WD6 3SH, UK
e-mail: dave.brown@cellzome.com
Giulio Superti-Furga
Meyerhofstrasse 1
69117 Heidelberg
Germany
e-mail: giulio.superti-furga@cellzome.com

▼ The pharmaceutical and biotechnology industries are in a strange and perhaps unexpected position. Until a decade ago the area enjoyed sustained growth and remarkable stability compared to many other industries. New medicines were discovered at a rate that supported double-digit growth rates broadly across the industry without resort to mergers and scientists enjoyed a high level of job security. The industry appeared to have found the formula for repeatedly hitting the 'sweet spot' of research productivity, commercial success and corporate growth [1].

That this is no longer the case appears puzzling after a particularly successful decade of advances in scientific understanding and a rapid rate of advance in technology development, the engines that drive the discovery of new medicines. Instead of the current uncertainties, we might have expected to be in a golden age of drug discovery. Yet pipelines are relatively thin, and drug launches have steadily fallen in recent years [2]. This appears to be more than a short-term issue, as evidenced by predictions from executives of major pharmaceutical companies that drug launches will remain flat through to 2008 at least (as discussed by several speakers at the conference '*Is there a best strategy for drug discovery?*' organized by the Society for Medicines Research at Imperial College in London, March 19, 2003). The only bright spot has been the successful introduction of a range of biological products that bypass the need for small-molecule lead discovery. One can only assume that (for small-molecule drug discovery) the new science and technology is taking longer than hoped to deliver new medicines. What is happening here? What are the prospects for rediscovering the sweet spot? And how long is it likely to take?

Evolution of the modern drug discovery process

Until about 30 years ago much of drug discovery proceeded by testing molecules against cell, tissue or whole animal systems [1]. Mechanistic assays were rare. A relatively small number of compounds were tested, a phenotypic response was measured and lead optimization often proceeded without any idea of the mechanistic target such that multiple structure-activity relationships (SARs) were being optimized in parallel. During the late 1960s and 1970s our understanding of human biology and biochemistry advanced rapidly and progressive 'drug hunters' began to identify mechanistic targets using the methods of classical pharmacology [1]. Chemistry and pharmacology proceeded in parallel, with chemical probes being used to identify receptor subtypes, for example, such that a lead molecule often emerged in parallel with the mechanistic target, and the target was often close to its natural state within a tissue preparation [3]. Whole classes of drugs still in use today emerged: α blockers, β blockers, histamine

1359-6446/03/$ – see front matter PII: S1359-6446(03)02902-7

antagonists, serotonin antagonists, among others. Then, with the advent of liquid-chromatography-based biochemical separations, followed by cloning technology during the 1980s, it became possible to obtain individual targets for screening or rational drug design either in isolation (e.g. for enzymes) or expressed in cells or membrane preparations (e.g. for ion channels or receptors). New classes of drugs were discovered, such as angiotensin converting enzyme (ACE) inhibitors, leukotriene antagonists, 3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase inhibitors, phosphodiesterase 5 (PDE5) inhibitors and many more. As a natural extension, the availability of these biochemical assays led to the establishment of high-throughput screening (HTS) technologies during the 1990s and a heavy focus on HTS as a primary generator of new leads. An HTS industry was born. The drug discovery process in its current form began to emerge, with target selection becoming increasingly separated from lead identification. The hungry screens demanded access to many compounds and so combinatorial and parallel synthesis technologies were developed and another mini-industry appeared. In recent years genome data has become available, and this is leading to further modification of the drug discovery process, driving another wave of mini-industries supplying the pharmaceutical industry with genome- and proteome technologies and databases. These new approaches have opened up new therapeutic opportunities, without speeding up the development of therapeutics.

A longer, more complicated drug discovery process

The cumulative attrition (failure) at each step of the drug discovery process is a real cause for concern. Rediscovery of the 'sweet spot' is dependent on achieving reduced attrition. Current evidence shows that this is unlikely to be achieved without rethinking the approach to drug discovery. Adding more steps and processes seems unlikely to help. Did we take a wrong direction some time in the past 10–15 years? Or do we just need to improve current methods? By examining the limitations of the current approach, from gene to target to lead to selection of a potential drug molecule, several bottlenecks can be identified.

The sequencing of the human genome was a significant landmark [4,5]. It was accompanied by predictions of miracle medicines, and considerably raised the expectations both of the public and of investors. However, experienced drug discoverers are aware that turning this data into information that is useful for drug discovery represents a great challenge; it is a long path from genome data to showing that a drug is safe and effective in clinical trials [6]. Sequence data alone tells us little about the complexities of gene expression and their influence on the (patho-) physiology of an organism. Genome data has marginally accelerated the pace of validated target discovery but in the short to medium term is perhaps likely to have more impact in the field of diagnostics [7]. It will also continue to drive the discovery of biological products, but why is the impact on oral small-molecule target and drug discovery slower than anticipated?

Human diseases are classified as either monogenic or polygenic. There are some 6000 monogenic diseases and these had been largely classified before the sequencing of the genome (GlaxoSmithKline Corporate Website, Genes and diseases, http://genetics.gsk.com/link.htm; [8]). Establishing a gene–disease link is considerably easier for monogenic diseases. In terms of incidence, cystic fibrosis claims more victims than any other monogenic disease, yet the patient population is not regarded as large enough by most pharmaceutical companies to give a satisfactory commercial return. Few of the major pharmaceutical companies are researching monogenic diseases.

Polygenic diseases are a greater burden, both in terms of the number of diseases and the number of patients afflicted by those diseases. Several genes might be implicated in the disease process and, furthermore, unraveling the influence of genetics is complicated by environmental factors. A range of methods is being used to identify genes involved in disease processes, ranging from the broad approach of population studies to hypothesis-driven studies of single genes in model systems [6,9]. When the probability of a genetic link is established there are several additional key steps before drug discovery can commence. These include identifying which protein is involved if the gene encodes several, determining whether upregulation or downregulation of function is required to treat the disease and deciding whether modulation of the function of the identified protein will provide effective therapy. There are several other key questions about the proposed approach. Will it lead to unacceptable side-effects? Will it be competitive with established therapies, if any exist? Would modulation of a compensatory pathway be a better mode of therapy? Is the target chemically tractable? Hurdles such as these can lead to a high rate of attrition and might explain why relatively few new targets have emerged so far from this approach.

Attrition data highlights the real needs

The effect of current limitations is graphically illustrated by benchmark attrition data from major pharmaceutical companies (Figure 1). In the Discovery phase, the cumulative attrition is ~80%; only 1 in 5 projects gets as far as selecting a compound for clinical trials. In the Development phase, cumulative attrition between selecting a compound

and marketing is >90%. Less than 1 in 10 projects survive through the clinical phases. Overall, through Discovery and Development, less than 1 in 50 projects get a drug to market. Why are these projects lost? Although attrition figures for any single phase need treating with caution (poor decision-making can pass a problem downstream and increase attrition at a later phase), meta-analysis of the data shows four main problems. Two of these are biology problems associated with target selection (poorly validated targets fail in Discovery or in Clinical Development); and two are chemistry problems associated with leads (failure in lead identification or optimization, or the leads prove to be toxic). What this means is that our ability to identify the best target in a disease pathway and our ability to find good chemical leads remain key limitations, and therefore limit our ability to rediscover the 'sweet spot'. Our analysis concurs with that of Jurgen Drews: 'The value of every technology that is offered by small biotech companies…must be assessed by asking…: how much does the new technology help to solve one of the two central problems: the identification and validation of a disease-specific target or the identification of a molecule that can modify this target in a way that makes therapeutic sense?' [2].

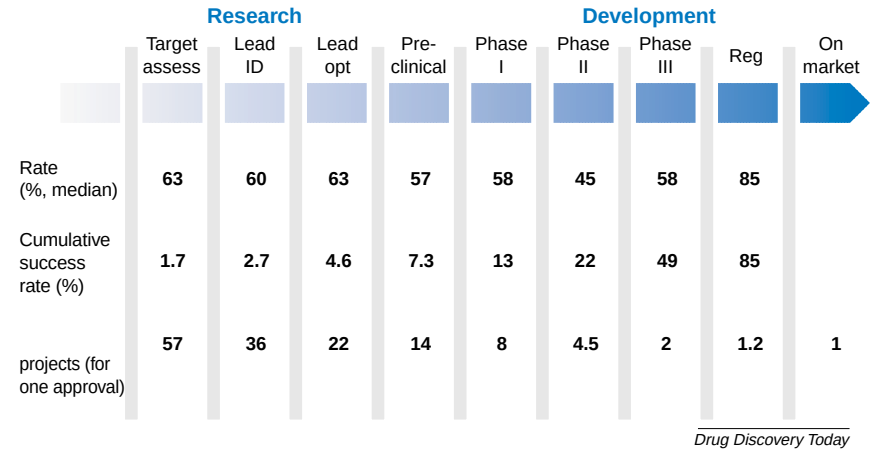

Figure 1. Success rates in Research and Development. The data represent the accumulated experience of several leading pharmaceutical companies. They were assembled by the Roche Strategy Unit, 2001 (courtesy of Paul Nakagaki and Mark Noguchi) and presented at the 'Discovery 2002' Conference, March 2002, Malaga, Spain, and at MipTec 2002, Basel, Switzerland, May 2002. Listed at the bottom is the number of projects that statistically are needed to get one drug on the market. Abbreviations: ID, identification; Opt, optimization; Reg, regulatory.

Breaking through the targets and leads bottleneck

There are unresolved technical issues in making use of the data provided by the mapping of the human genome. Although the initial steps (genome sequencing and bioinformatic annotation and searching) have been converted to high-throughput, most subsequent steps are still low-throughput. Once a disease link is established, one needs to understand the pathway involved with the protein that is encoded by the disease-linked gene; breakthroughs are needed in our ability to map the proteins suspected to be involved in disease pathways. Understanding the pathway can strengthen the hypothesis provided by the genetic link and, of equal importance, there might be alternative targets in the pathway with higher chemical tractability. There is also a need for better ways of demonstrating that modulation of putative targets does indeed have the desired effect on the disease phenotype. Moreover, our ability to find good chemical ligands as leads for optimization into drugs is quite limited for many of the potential target types [10].

Targets from mapping disease pathways

Biology relies on the concerted action of several molecular interactions of gene products and metabolites, operationally organized in so-called pathways (Figure 2) [11–13]. Impairment of pathway flow or connections through subtle perturbation of one or several individual pathway components can lead to pathology. The majority of targets of current therapeutics cluster in a limited number of these cellular pathways (Table 1). However, current appreciation of the 'wiring diagram' or 'molecular maps' of these pathways is scanty.

In principle, large-scale yeast two-hybrid approaches can produce such wiring diagrams. However, the technical limitations for this *ex vivo* genetic approach do not allow for efficient analysis of membrane proteins and transcription factors, which are often at the beginning and end of the signalling process, respectively, unless studied in non-physiological protein fragments. Moreover, because the analysis occurs in yeast cell nuclei and mostly monitors binary interactions, the very essence of signaling pathways (i.e. the physiological integration of posttranslational modification and of scaffolding/bridging molecules in human cells) is lost [14].

Recent developments in biochemical and mass spectrometry technologies have rendered the physical characterization of human cellular pathways feasible and more reliable [15,16,17]. It is possible to assign membership, position and a proposed role for each of the roughly 100–300 proteins responsible for a typical pathway [16–18]. Thus, there

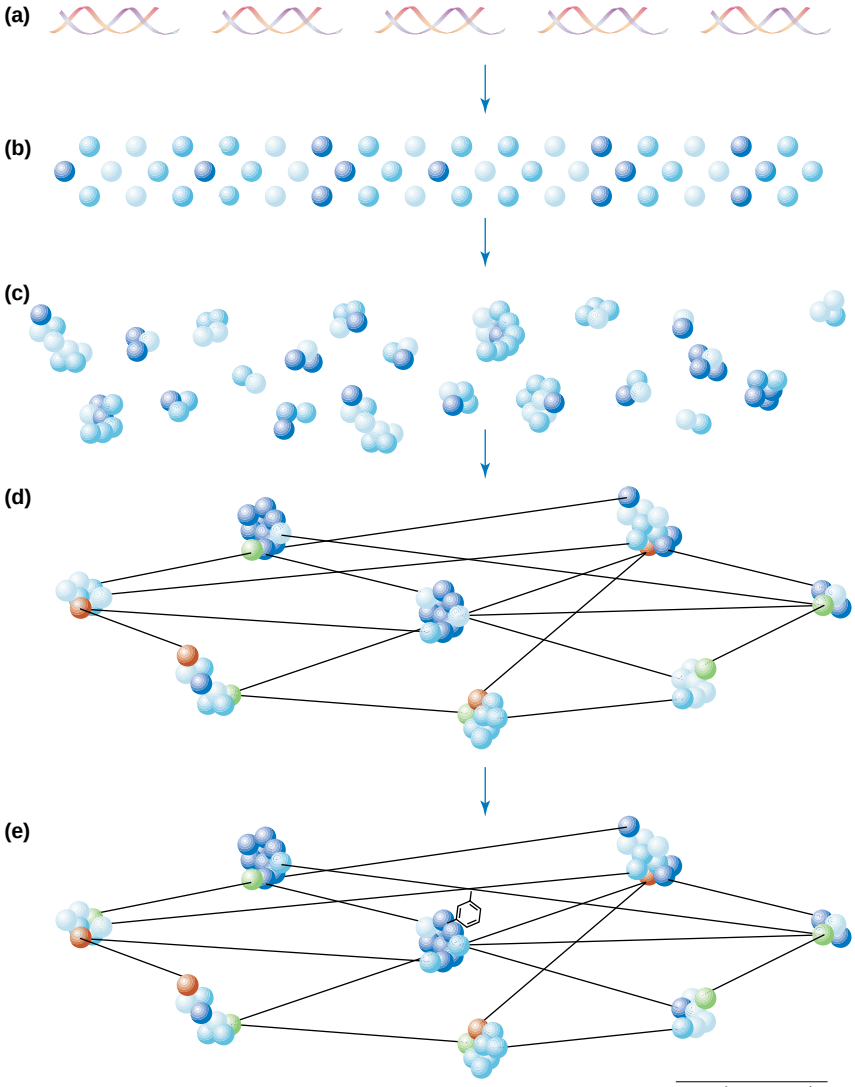

Figure 2. Genes to proteins to pathways to drug targets in pathways. Schematic view of the organization of biological material. Genes **(a)** give rise to individual gene products **(b)**, which assemble into protein complexes **(c)**, the basic units of biological activities. Through direct protein interactions, sharing of protein components or via metabolites, complexes are organized into pathways **(d)**. Pathways are responsible for more important cellular tasks. Drugs do not only affect individual proteins and complexes, but, through wiring, often affect whole pathways **(e)**.

is a new opportunity to elucidate selectively the pathways central to treatment of human diseases.

Pathway mapping

The propensity of proteins to assemble in order to achieve more complex tasks and increase localization and regulation efficiency can be exploited directly to isolate and characterize the physical assemblies that are the 'functional units' in the cell (Figure 2). Starting from known pathway components, expressed in retrievable forms in cells, it is possible to identify the immediate partners of these components. Scientists at Cellzome (http://www.cellzome.com) use tandem-affinity-purification (TAP) [15,17,19] coupled to mass spectrometry (Figure 3), whereas scientists at MDS Proteomics (http://www.mdsproteomics.com) for example, in a process called PathMap™, use an epitope-tagged protein [20]. In both cases, the components of the complex are identified by using mass spectrometry and bioinformatics [15–17]. Using the identified protein partners in turn and subjecting them to the same process confirms the association with the original components and also allows for the identification of other pathway components. This 'pathway expansion' continues until the newly identified components predominate and appear to pertain to a different, neighboring and/or intersecting pathway (see feedback loop in Figure 3). Statistical bioinformatics analysis, clustering algorithms, visualization tools and expert annotation deliver the map of the pathway, with boundaries and connections to other pathways. Compared with the large-scale proteome 'catalogues' this pathway-focused strategy appears to have several advantages. First, it allows an exploration of validated targets as starting points. Second, it allows the identification of non-abundant proteins. Third, it indicates a functional relationship with the starting points, through the 'guilt-by-association' concept [18]. However powerful these large-scale complex purification and mass spectrometry characterization approaches are, they provide poor quantitative information. Dedicated technologies relying on differential isotopic labeling of samples have been developed that address more quantitative aspects of protein complex composition [21,22].

The term 'pathway' implies motion along a path. Indeed, cellular proteins undergo protein complex assembly and disassembly in a dynamic fashion in accordance with cellular needs and the environment. It is therefore informative to study the same pathway under different experimental conditions, for example the presence or absence of drug or

Table 1. List of cellular pathways of potential interest to the pharmaceutical industry

Pathways	Targets and other prominent components	Disease
Apoptosis	cGMP phosphodiesterase, Bcl-2, p53, MDM2, Caspases (Caspase 3), IKK-α, -β, -γ	Cancer
PI3K signal transduction	PI3K, AKT/PKB, PDK, PTEN, mTOR, GSK3β	Cancer (e.g. lung), rheumatoid arthritis/ inflammation, respiratory disease
MAPK signaling	ERK pathway: ERK1, ERK2, EGFR (RTK), HER2 (RTK), B-RAF, MEK1, MEK2 JNK pathway: JNK1, JNKK p38 pathway: p38 α, β, γ, δ, MAPKAPK2	Inflammation, rheumatoid arthritis, cancer, and neurological diseases (Parkinson)
Angiogenesis	VEGF, VEGFR (RTK), FGF, FGFR (RTK), PDGF, PDGFR (RTK), FLT3 (RTK), PKC, CXCR2, MMP2, Met AP-2	Cancer
Tyrosine kinase (RTK) signaling	EphB3, c-Met, EGFR, HER2, VEGFR, FGFR, PDGFR, FLT3, Src, Lck, Abl	Cancer, diabetes, rheumatoid arthritis/inflammation
Protein prenylation	Protein farnesyltransferase α and β Rab geranylgeranyltransferase α and β	Cancer
Protein acetylation	HDAC1–11	Cancer, inflammation, neurodegenerative disease
Cell adhesion and motility	FAK, Cadherin 5, E-cadherin, Selectin P, LIMK1, LIMK2, p130CAS	Cancer, peripheral vascular disease
Integrin signaling	ILK, Integrin (α V/β 3, α 2/β 1, α 4/β 1)	Inflammation, cancer, peripheral vascular disease, respiratory disease
Cell cycle regulation and cellular life-span	CDK2, CDK4, CDC7L1, p16CDK, HSP90-α and -β, CDC37, Telomerase	Cancer, osteoarthritis
Hormone signaling	DPP4, CRHR1, Melanocortin 3 and 4 R, Neuropeptide Y receptor Y5, GPR24, ADCYAP1, ADCYAP1R1	Diabetes, depression, obesity, urogenital disease
Nuclear hormone receptor signaling	PPAR (PPARα, γ, δ), RAR (RAR γ), RXR, Progesterone receptor, Androgen receptor, Estrogen receptor, LXR-α, FXR	Lipid metabolism, inflammation, immunology, atherosclerosis, diabetes, cancer, Alzheimer's disease
Wnt pathway	Frizzled homolog 1, APC, β-catenin, axin, GSK3β, Bcl-9 TCF/LEF	Cancer, hair growth
Pro-inflammatory pathways	TNF-α, TNF-α receptor, TACE, IKK-α, -β, -γ, IL-1, IL-1R, IRAK1, IRAK2, IRAK3, IRAK4, IL-6, IL-6R, Toll-like receptors, BlyS (BAFF), CD40L, LT-β, GM-CSF	Rheumatoid arthritis and inflammatory disorders, autoimmune disorders, cancer
Anti-inflammatory pathways	IL-2, IL-2R, IL-4, IL-4R, IL-10, IL-10R	Inflammation, cancer
Chemokine signaling	CXCR2, CCR3, CCR4, CCR5, CCR6, ERK, AKT, p38	Cardiovascular disease, inflammation, infection, metabolic disease, respiratory disease, cancer
T- and B-cell signaling	TCR, calcineurin, IgE, FceRI, FceRII, kinases	Immune response, allergic disorders
IFN pathway	IFN-α/β, IFN-γ, JAK-STAT	Multiple sclerosis, antiviral immune response, cancer
Arachidonic acid metabolism	Lipoxygenases, prostaglandins, thromboxanes, COX leukotrienes, LT receptors	Hypertension, inflammation
Insulin signaling	Insulin, IGF, insulin receptor, IGF-1R, PTPN1	Diabetes, cancer
Glucose homeostasis	Glucagon, glucagon receptor, glucokinase, GFPT1, GFPT2, glycogen phosphorylase, glycogen synthase 1 and 2	Diabetes

Table continued on following page

Table 1. List of cellular pathways of potential interest to the pharmaceutical industry - continued

Pathways	Targets and other prominent components	Disease
Lipid homeostasis, cholesterol biosynthesis	HMG-CoA reductase, SCAP, SREBP, ACAT-1, ACAT-2, ACACA, ACACB, LXR-α, DGAT1, LDLR, LRP2, ABCA1	Hyperlipidemia, dyslipidemia, obesity, Alzheimer's disease, coronary heart disease, atherosclerotic plaque
Renin-angiotensin system	AGTR1, ACE	Hypertension
Amyloid precursor protein processing	APP, α-secretase, β-secretase, γ-secretase (presenilin, nicastrin, Aph-1, Pen-2), neprilysin, IDE, GSK3α	Alzheimer's disease
Tau phosphorylation	Tau, GSK3β, CDK5, MARK	Alzheimer's disease
Apolipoprotein metabolism	Apolipoprotein E, LDL Receptor	Alzheimer's disease
Expression and activation of neurotransmitter receptors	Neuregulin 1, ErbB4	Schizophrenia
Alpha-synuclein pathway	Synuclein, synphilin, parkin	Parkinson's disease
Nociception	μ opioid receptor, κ opioid receptor, vanilloid receptors	Acute pain
Cyclic nucleotide metabolism	PDE2, PDE3, PDE4A, B, D, PDE5A	Depression, asthma, cancer
Neurokinin signaling	Tachykinin receptor 1 and 2	Depression, urogenital disease
Neurotransmission	Neuropeptide signaling (orexin receptor 1 and 2), cannabinoid receptor 1, NAALAD2, type I s receptor, neurotensin receptor 1 and 2	Depression, Alzheimer's disease, anxiety, schizophrenia, obesity, urogenital disease
	Serotonergic neurotransmission: serotonin receptor 2C and 6	
	Inhibitory neurotransmission: GABA A receptor α 2, 3 and 5	
	Excitatory neurotransmission: metabotropic glutamate receptor 1, 2, and 5, folate hydrolase 1	

other stimulus, or whether the cell is in a healthy or pathological state, and so on. Often, it is the differential association of particular components that highlights a critical regulatory role and a possible interesting target.

Synergy with genetics and genomics

One of the most important advantages of the pathway focus is its synergy with other discovery strategies based on genetic and genomic data. From a genetic standpoint, disease often does not arise from a single atypical gene [9]. Instead, individual defects in several distinct genes can cause a similar pathology and the combination of particular alleles of these same genes might also lead to disease. Not surprisingly, the products of these genes are often found to contribute to the same pathway. There are numerous examples that illustrate this point, such as cancer susceptibility [23,24], blood clotting [25], insulin resistance [26], mitochondrial respiratory chain diseases [27], urea cycle defects [28] and cholesterol metabolism [28,29]. A full map of the pathway is a useful interpretation and predictive tool for human genetics data.

Disease-associated genes

Disease-associated genes that have a statistically weak association with a disease can be validated using pathway mapping. For example, armed with the knowledge that two gene products interact, the sub-threshold statistical significance of SNPs and haplotype association with a disease of the corresponding genes becomes revealing. Understanding the context can strengthen the hypothesis. In our own experience, we have found that confidence in the functional relevance of a newly predicted disease mechanism is significantly strengthened when more than one disease gene is connected in a pathway. Furthermore, genes identified through genetics either in human populations or in model organisms often predict targets that are

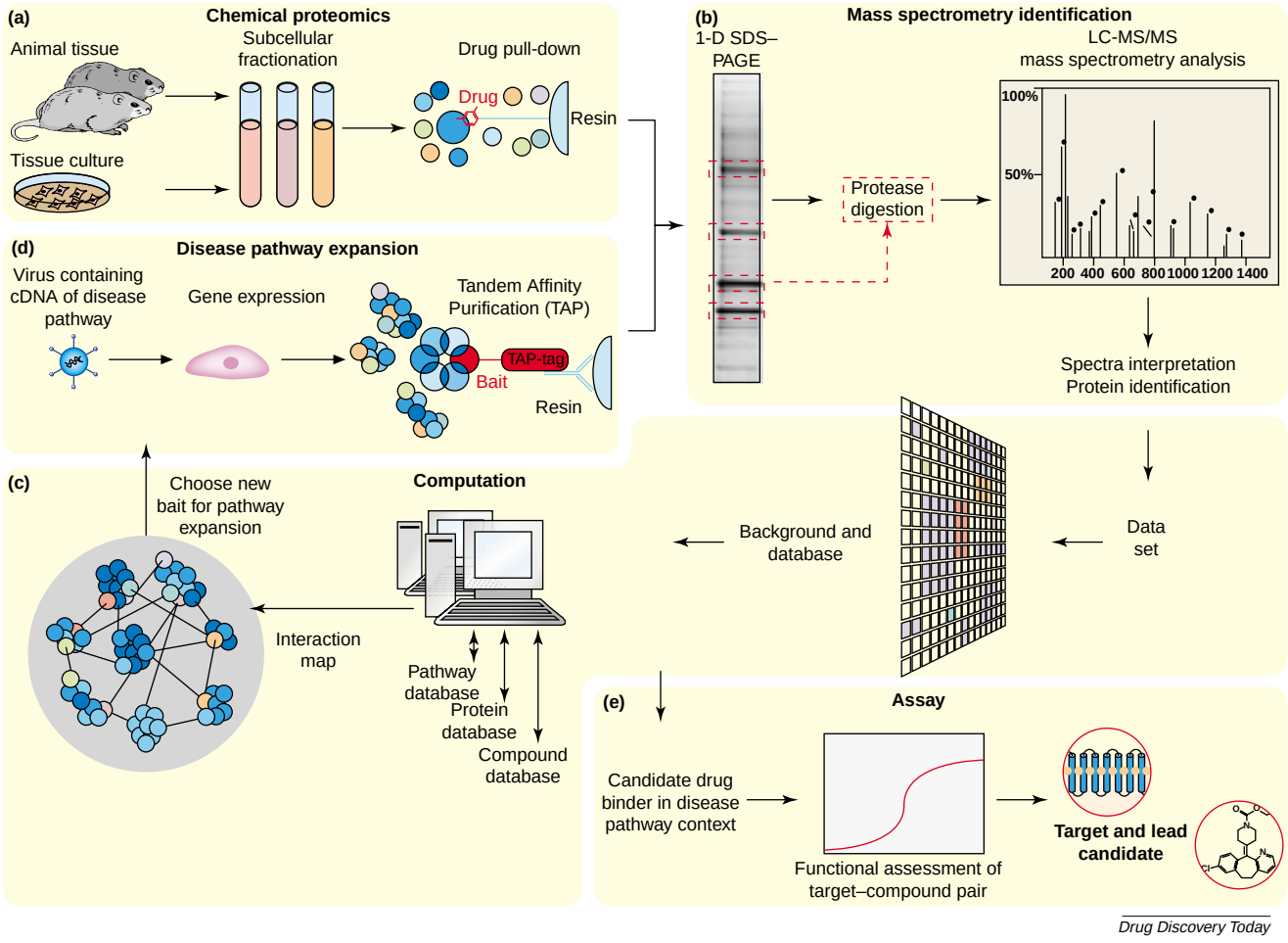

Figure 3. Functional and chemical proteomics discovery process. **(a)** Lysates are generated from tissues or cells. Subfractions are used to enrich for particular target types. Drugs or drug-like compounds are coupled to a matrix. The drug matrix is incubated with the lysate and washed. Specific binders are eluted with drug excess or by specific cleavage in the linker. **(b)** Proteins are separated by gel electrophoresis, excised and analyzed by liquid chromatography–tandem mass spectrometry. Spectra are obtained and analyzed. The proteins are identified through database searches. **(c)** Data are quality controlled and entered into the database. Filtering against cell-type and fraction-specific 'noise' and matrix contaminants allows identification of specific binding profile. The algorithms that allow the mining of the chemical proteomics database and automatic identification of specific binders candidates with the desired properties are crucial. Reiteration of the process leads to 'pathway expansion'. The protein–protein interaction database allows further insight into 'piggy-backing' components and positioning in disease pathways. **(e)** The candidate target is assessed with the drug in cellular and *in vitro* assays for potency and selectivity. **(d)** In the protein interaction assay, the cDNA encoding the protein of interest is fused to a Tandem-Affinity-Purification (TAP) cassette and expressed from a retroviral construct after virus generation and human cell infection. The protein is recovered along with its natural binding partners in a two-step purification procedure.

not chemically tractable. The elucidation of the pathways around the disease gene products allows the identification of alternative, tractable targets in the same pathway.

Differential gene expression

Differential expression patterns (protein or mRNA) in pathological versus normal cells or tissues are a symptomatic view of a disease and do not necessarily capture the underlying causal disease mechanism. The elucidation of molecular pathways associated with hits from differential display experiments allows an interpretation of the data at a functional level. The association of the observed hits with previously reported disease genes or known cellular pathways place the data into a functional disease context. Targets can then be validated and prioritized.

Pharmacologically validated pathways

Validated pathways are those for which drugs in clinical trials or on the market have already validated the original biological hypothesis, giving ultimate target validation, but for which the ideal small-molecule therapy is still lacking. For example, it might be that current therapy involves

parenteral administration of a biological. By focusing on these already validated pathways, one capitalizes on the work and experience gathered previously in the industry (Table 1). Mapping these pathways allows one to choose the most promising targets according to their position in the pathway, their possible cellular involvement in other pathways (potential side-effects), and their chemical tractability. Once all proteins in a major disease pathway have been mapped, one can parallel-process target validation and select for screening several targets in parallel from the same pathway to increase the probability of success.

Pathways so far

Perhaps the largest published study so far has involved mapping the complete proteome of baker's yeast (*Saccharomyces cerevisiae)*. Initial analysis of the yeast proteome focused on the gene products orthologous to human proteins [17] and those associated with signaling and response to DNA damage [20]. From this analysis and work at Cellzome, it appears that in excess of 85% of proteins are found in association with other proteins. Proteins are thus 'sociable' and exist mostly in complexes. Extrapolating the figures from the published data, we anticipate that the proteome might be organized into some 300 complexes, which can be assigned to around 50 pathways. Interestingly for drug discovery, a third of such pathways could contain targets for antifungal drugs. In view of the genome homology with key fungal pathogens, such studies might provide direction for target identification for new antifungals.

Signaling pathways have been studied for the identification of drug targets for many years [13]. The research community has even started alliances to integrate available data and coordinate efforts (Alliance for Cellular Signaling; http://www.signaling-gateway.org). The analysis of mammalian protein complexes, the building blocks of pathways, has been made easier by recent advances in protein complex purification and analysis by mass spectrometry [30,31]. Several dozens of human protein complexes have been characterized already (reviewed in Ref. [18]). Adaptation of the TAP method to mammalian cells [17,32,33] allowed the efficient characterization of complexes involved in cellular signaling. Ongoing work at Cellzome focuses on the elucidation of two pathways relevant to human disease – the tumor necrosis factor-α (TNF-α)-signaling pathway involved in inflammatory and growth-regulatory processes, and the amyloid precursor protein (APP)-processing pathway involved in Alzheimer's disease. So far, around 200 and 150 proteins have been mapped to the TNF-α and the APP pathways, respectively, with typical complex sizes of 5–10 protein constituents. It appears that several new proteins in the pathways are potentially chemically tractable. We propose that it will also be possible to pinpoint the interfaces to suspected intersecting pathways, such as apoptosis and stroke for the TNF-α pathway, and cholesterol metabolism for the APP pathway.

Leads

If pathway mapping can overcome critical barriers in target selection, what about the second major issue of lead generation? Pathway mapping in itself is a major asset in identifying chemically tractable targets. After a decade of high-throughput screening, a clearer picture has emerged of which target classes are more likely to give good leads and which classes remain essentially intractable to chemists [10]. Currently the target classes that we can approach with higher confidence of obtaining a lead include GPCR class I, particularly enzyme-class subtypes, ion channels and nuclear receptors [10,34]**.** By contrast, with most other target classes, the industry's experience has shown low success rates in finding good leads. Therefore, identifying all possible targets in a pathway through pathway mapping will lead to an increase in the number of potentially tractable targets in the three classes available to us. What is required in addition is a breakthrough in our ability to modulate those target classes that are still considered chemically intractable. Looking back over the past decade one has the impression of regular and rapid biological breakthroughs, but of chemistry drug design knowledge advancing at a slower pace. Consequently, there has been a rapid proliferation of potential targets, for many of which there is uncertainty over whether or not leads can be found. In an interesting twist to this problem, some companies are accepting that chemistry is limiting and are developing a 'chemo-centric' approach. They exploit genome and proteome knowledge based on an understanding of target subtype and/or ligand structural relationships and phylogenetic considerations. This entire concept, termed 'chemical genomics' can take several forms [35]. In one form, not reviewed here and exemplified by the discovery strategy of companies like Acadia (http://www.acadia-pharm.com) and Cellular Genomics (http://www.cellulargenomics.com), sophisticated cellular read-outs are coupled with HTS screening capability. Other approaches, exemplified by the initiatives of companies like Infinity (http://www.ipi.com) and Graffinity (http://www.graffinity.com), offer collections of diverse compounds on chips in parallel to biological material, typically purified proteins.

Another pragmatic, experimental approach takes existing drug-like molecules to 'screen' against the proteome. This 'chemical proteomics' approach differs from chemical genomics insofar as the tractability is not inferred computationally by phylogenetic considerations on the target

Table 2. List of companies active in the chemical proteomics and chemical genomics field[a]

Company/ group	Website	Pathway proteo-mics	Drug Pull-downs	Protein chip	Chemical chip	Technology description
Ambit	www.ambitbio.com		X			Phage-displayed proteins screened against immobilized drugs
Caprion	www.caprion.com	X				Subcellular proteomics
Cellzome	www.cellzome.com	X	X			Pathway proteomics and chemical proteomics
Curagen	www.curagen.com	X				Pathways by yeast two-hybrid
GPC	www.gpc-biotech.com	X		X		Yeast two-hybrid, dependent on compound: identifies targets
Graffinity	www.graffinity.com				X	Surface-plasmon resonance technology on chemical chip
Hybrigenics	www.hybrigenics.fr	X				Pathways by yeast two-hybrid
Infinity	www.ipi.com				X	Diversity-oriented synthesis chemical chips
MDS proteomics	www.mdsproteomics.com	X	X			Pathway mapping and chemical proteomics
Prolexys	www.myriad-proteomics .com	X				Yeast two-hybrid and mass spectrometry of complexes around recombinant baits
Protometrix	www.protometrix.com				X	Protein chip for function and protein–protein interaction
Reverse Proteomics Res.Institute	www.reprori.jp		X		X	Chemical proteomics and protein chips
Serenex	www.serenex.com		X			Isolate sub-proteomes with immobilized ligands
Zyomix	www.zyomix.com				X	Protein chip for function and protein–protein interaction

[a]Only companies using mass spectrometry, drug pull-downs or chemical chips are listed. A longer list that includes companies employing expression proteomics, structural genomics and those that use a purely computational approach or a screening platform is found in Ref. [43]

class. Instead, evidence for tractability is sought experimentally. This 'drug pull-down' method is essentially an affinity chromatography procedure followed by gel electrophoresis, mass spectrometry analysis (or, occasionally, microsequencing) and computational analysis [36–42]. It allows for the detection of a direct, physical interaction of proteins with compounds. The affinity matrix is incubated with lysates obtained from tissues or cells so that the physiological environment, the physiological partners and the post-translational modifications of the protein targets are maintained. The approach, in effect, reverses the HTS process, which traditionally throws many compounds at a single target. Instead, many proteins, with their modifications and their partners, see each chemical molecule. Individual companies, listed in Table 2, present different versions of the approach, replacing the screening of entire proteomes with the screening of collections of recombinant proteins.

A crucial factor in being able to exploit the technology effectively is the ability to evaluate the purification data statistically and mine the data from a bioinformatics and cheminformatics standpoint. Based on the proteins in the human genome bearing small-molecule-binding potential (using the PFAM and SMART algorithms) and experimental evidence from chemical proteomics, Cellzome's estimate of potentially tractable components is around 6000 proteins, although it should be noted that only a subset of these will be valid targets for disease modulation and drug discovery. Although it is impossible to know how many of these may have been detected already worldwide, initial efforts at Cellzome have already detected 600. Pulling together the need to find both targets and leads, it is notable that a small but interesting fraction are also among the already elucidated protein pathways, allowing immediate chemical assessment of the target–compound pair in the pathway context (Figure 4). A premier route might be

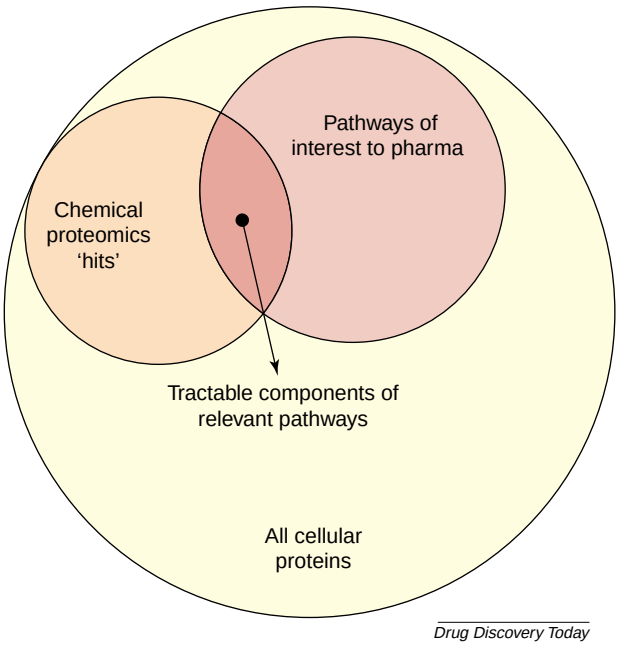

Figure 4. The 'sweet spot' of drug discovery. Of all the possible tens of thousands of proteins expressed in the human body, only several thousands, a fraction (pink circle), participate in the pathways on which present and future drug discovery efforts of pharmaceutical industries concentrate. Likewise, another small, but different fraction of the total proteome has small-molecule-binding properties (orange circle). These are the potential 'chemical proteomics targets' identified by 'drug pull-down' experiments. Leads for this subset can also be found by screening of focused libraries The sweet spot of drug discovery is represented by the several hundred 'druggable' proteins that are in the therapeutically valid pathways (mauve intersection).

through the use of existing efficacious drugs that have an unclear or ambiguous mechanism of action. The combined use of the technologies summarized here should allow the identification of the target protein(s) and the corresponding pathways. It should also be possible, based on information about the precise molecular action of the original drug, to initiate a new drug discovery program armed with candidate leads, a pharmacologically validated target and a defined biochemical read-out predictive of clinical efficacy.

Conclusions: focusing on the sweet spot

An efficient strategy to exploit knowledge of the human genome and recent proteomics technologies is to focus on the pharmacologically relevant part of the genome and/or proteome (i.e. targets that are valid, safe and chemically tractable). From all proteins and cellular pathways, there are likely to be only a limited number of targets on which drug discovery in pharmaceutical and biotech companies will converge. Typically, these cluster in pathways around known therapeutically valid targets or where there is sufficient genetic information and links to important diseases. A non-comprehensive list of 32 prominent pathways, with representative targets and drug development status, is listed in Table 1. From the technological breakthroughs described above, it should now be technically possible to fully characterize all components of key disease pathways, to elucidate the pharmacologically relevant part of the human proteome. Such an approach would achieve maximal exploitation through the screening of focused libraries against the chemically tractable members of the pathways. Likewise, in a second approach, it should be possible to capitalize on accumulated medicinal chemical wisdom, focusing on compounds with drug-like properties, and use chemical proteomics to identify target and lead in a single step. In essence, this would represent a return (at a higher level of scientific development) to the paradigm that was so successful in the past when lead and target identification were conducted in a single step. We anticipate that, using these approaches, the industry can rediscover the postgenomic 'sweet spot' of drug discovery.

Acknowledgements

We are indebted to Heinz Ruffner for expert compilation of Table 1. We wish to thank Gitte Neubauer, Gerard Drewes and Heinz Ruffner for critical reading of the manuscript, Monika Blank for expert editorial assistance and Frank Weisbrodt for the graphics.

References

1 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
2 Drews, J. (2003) Strategic trends in the drug industry. *Drug Discov. Today* 8, 411–420
3 Maehle, A.H. *et al.* (2002) The emergence of the drug receptor theory. *Nat. Rev. Drug Discov.* 1, 637–641
4 Venter, J.C. (2001) The sequence of the human genome. *Science* 291, 1304–1351; Erratum in *Science* 292, 1838
5 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
6 Schadt, E.E. *et al.* (2003) A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochem. Soc. Trans.* 31, 437–443
7 Chanda, S.K. and Caldwell, J.S. (2003) Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov. Today* 8, 168–174
8 Hamosh, A. *et al.* (2002) Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 30, 52–55
9 Roses, A.D. (2000) Pharmacogenetics and the practice of medicine. *Nature* 405, 857–865
10 Bleicher, K.H. *et al.* (2003) Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discov.* 2, 369–378
11 Pawson, T. and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* 300, 445–452
12 Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934
13 Persidis, A. (1998) Signal Transduction as a drug-discovery platform. *Nat. Biotechnol.* 16, 1082–1083

14 Drewes, G. and Bouwmeester, T. (2003) Global approaches to protein-protein interactions. *Curr. Opin. Cell Biol.* 15, 199–205
15 Puig, O. *et al.* (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24, 218–229
16 Figeys, D. (2002) Proteomics approaches in drug discovery. *Anal. Chem.* 74, 412A–419A
17 Gavin, A.C. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147
18 Gavin, A.C. and Superti-Furga, G. (2003) Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.* 7, 21–27
19 Rigault, G. *et al.* (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032
20 Ho, Y. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183
21 Ranish, J.A. *et al.* (2003) The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* 33, 349–355
22 Blagoev, B. *et al.* (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat. Biotechnol.* 21, 315–318
23 Figeys, D. (2002) Proteomics approaches in drug discovery. *Anal. Chem.* 74, 412A–419A
24 D'Andrea, A.D. and Grompe, M. (2003) The Fanconi anaemia/BRCA pathway. *Nat. Rev. Cancer* 3, 23–34
25 Venkitaraman, A.R. (2003) A growing network of cancer-susceptibility genes. *New. Engl. J. Med.* 348, 1917–1919
26 Krem, M.M. and Cera, E.D. (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem. Sci.* 27, 67–74
27 Moller, D.E. (2001) New drug targets for type 2 diabetes and the metabolic syndrome. *Nature* 414, 821–827
28 DiMauro, S. and Scon, E.A. (2003) Mitochondrial respiratory-chain diseases. *New. Engl. J. Med.* 348, 2656–2668
29 Scriver, C. *et al.* (2001) *The Metabolic and Molecular Basis of Inherited Disease* (8th edn), McGraw-Hill
30 Blackstock, W.P. and Weir, M.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17, 121–127
31 Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422, 198–207
32 Chen, C.Y. *et al.* (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* 107, 451–464
33 Basu, S. *et al.* (2003) Akt phosphorylates the Yes-associated protein, YAP, to induce interaction with 14-3-3 and attenuation of p73-mediated apoptosis. *Mol. Cell* 11, 11–23
34 Auld, D.S. *et al.* (2002) Targeting signal transduction with large combinatorial collections. *Drug Discov. Today* 7, 1206–1213
35 Caron, P.R. *et al.* (2001) Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* 5, 464–470
36 Harding, M.W. *et al.* (1989) A receptor for the immuno-suppressant FK506 is a cis-trans peptidy-prolyl isomerase. *Nature* 341, 758–760
37 Oda, Y. *et al.* (2003) Quantitative chemical proteomics for identifying candidate drug targets. *Anal. Chem.* 75, 2159–2165
38 Graves, P.R. *et al.* (2002) Discovery of novel targets of quinoline drugs in the human purine binding proteome. *Mol. Pharmacol.* 62, 1364–1372
39 Ding, S. *et al.* (2003) Synthetic small molecules that control stem cell fate. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7632–7637
40 Leung, D. *et al.* (2003) Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat. Biotechnol.* 21, 687–691
41 Knockaert, M. *et al.* (2000) Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilized inhibitors. *Chem. Biol.* 7, 411–422
42 Taunton, J. *et al.* (1996) A mammalian histone deacetylase related to the yeast transcriptational regulator Rpd3p. *Science* 272, 371–372
43 Mitchell, P. (2003) In the pursuit of industrial proteomics. *Nat. Biotechnol.* 21, 233–237