In need of high-throughput behavioral systems

Dani Brunner, Eric Nestler and Emer Leahy

One of the current major bottlenecks in drug discovery is *in vivo* testing of candidate drugs in behavioral paradigms in normal or genetically altered mice. This testing is essential in discovering gene function and predicting potential efficacy of CNS drugs in humans. New efforts in the biotech community aim to alleviate this bottleneck by developing higher-throughput systems of behavioral, neurological and physiological analyses. Together with large pharmacological databases, equipped with state-of-the-art bioinformatic and/or data-mining algorithms, these systems will provide rapid and accurate indices of the therapeutic potential of novel drugs. By providing a substantial increase in the speed of behavioral testing, new high-throughput systems will facilitate current behavioral research with faster, more reliable approaches. Furthermore, screening whole drug-libraries and comparing the profiles of novel compounds to those of known compounds will faciliate the discovery of novel drugs. Target validation will also become more efficient with the fast characterization of novel mutant mice.

Dani Brunner
Psychogenics
4 Skyline Drive, Hawthorne
NY 10532, USA
and Dept of Psychiatry
Columbia University, New York
NY 10027-6902, USA
tel: +1 914 593 0640
fax: +1 914 593 0645
e-mail: daniela.brunner@psychogenics.com
Eric Nestler
Psychogenics
and Dept of Psychiatry
University of Texas
Southwestern Medical Center
Dallas, TX 75390-9070, USA
Emer Leahy
Psychogenics

▼ During the past decade, new enabling technologies in molecular biology, chemistry, automation and information technology have dramatically reshaped pharmaceutical and biological research. The completion of genome sequencing in humans and mice has opened new opportunities to study the relationship between gene expression and behavior (http://www.phrma.org/). However, while the function of many genes is being unraveled, producing many promising therapeutic targets, progress in understanding many of the common neurological and psychiatric disorders (schizophrenia, depression and anxiety, and so on) is still lagging behind.

New scientific advances and several converging factors – including the introduction of new diagnostics and therapies, increased life expectancy, and unmet medical and commercial needs – are major contributing factors to the acceleration of CNS drug discovery and development in recent years. As the CNS drug market is the leading recipient of pharmaceutical R&D budgets, there has been a surge in CNS drugs in the pharmaceutical pipeline [1].

The drug approval process, however, is still a very lengthy and costly exercise. Independently of the indication, it takes on-average 15 years and US$500 million for a single drug for a specific disorder to reach successful commercialization [1], with the major bottlenecks being the *in vivo* testing of lead compounds.

Discovery bottlenecks

CNS drug discovery differs from most other therapeutic areas because of the complex and multigenic nature of most psychiatric and neurological disorders [1]. Preclinical testing for most disorders, such as cancer, infectious diseases, and metabolic and cardiovascular disorders, is primarily done *in vitro*, or involves an autopsy of the animal to determine the impact of the compound. By contrast, CNS research is highly dependent on *in vivo* behavioral research, in which the live animal is subjected to a series of tests to determine the behavioral effect of altering the target gene or administering the compound. Behavioral tests are required for both gene target-identification and -validation, and compound-selection and -refinement. Behavioral data therefore are the main drivers of CNS drug discovery.

However, behavioral testing is currently a very slow, costly and sometimes subjective process, requiring a significant amount of behavioral research expertise. To date, new technologies developed to expedite the drug discovery process have not been applied successfully to behavioral research. While some areas of the drug discovery

1359-6446/02/$ – see front matter PII: S1359-6446(02)02423-6

process have become highly automated, most aspects of the process that involve *in vivo* behavioral testing on animal models still depend on slow traditional laboratory techniques. This bottleneck is especially pronounced for CNS preclinical drug testing, where live behavioral tests must be conducted for target-validation and lead-compound selection. A high-throughput, automated behavioral research platform would substantially alleviate this critical and costly bottleneck for the CNS drug market.

High-throughput *in vitro* technologies

Spurred by the demand for new drug candidates, automated high-throughput *in vitro* systems can produce and preliminarily screen hundreds-of-thousands of new compounds, resulting in a vastly greater number of new compounds that need further screening *in vivo*. Moreover, *in vitro* techniques are target-driven, and most of today's leads in CNS research are still based on old targets. A fast and efficient way to assess drug action *in vivo* could lead to the discovery of new drugs and truly novel targets.

Whereas *in vitro* high-throughput systems for the development of target-driven lead-compound selection have been in place for some time, this effort has been minimally successful in CNS research, probably because most CNS disorders are multigenic and interact in complex ways with the environment.

The need for a high-throughput *in vivo* discovery tool

Continuous research activities of pharmaceutical and biotechnology companies have enabled them to build massive compound libraries that hold >1 million compounds each. Screening every eligible compound with traditional behavioral research techniques is prohibitively expensive. With the development of high-throughput techniques, many more compounds can be tested for CNS efficacy.

The genomics revolution

In the past, CNS drug discovery focused on only dozens of 'druggable' targets. By contrast, the mapping of the human genome has produced thousands of potentially druggable targets (http://www.globaltechnoscan.com/5thSep-11thSep01/human_genome.htm). However, the discovery of these new gene targets has not as-yet resulted in approved CNS drug therapies; it will take time before the field of genomics yields tangible results. Meanwhile, pharmaceutical companies need to balance their dependence on new technologies with continued research on proven targets. This means continuing to test a variety of compounds with known and relatively conventional targets.

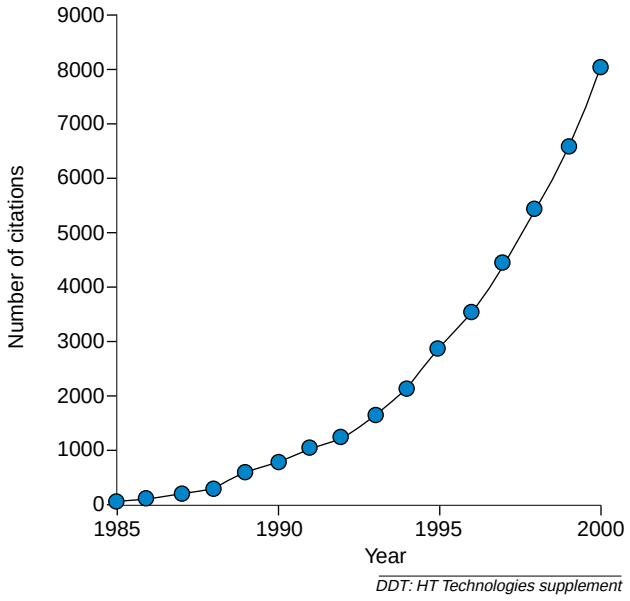

Figure 1. Number of citations per year of the terms 'knockout' and 'transgenics' in the scientific literature (Medline).

Phenotype–genotype correlation

Thousands of genetically manipulated animals are being generated in hundreds of different laboratories for many different purposes. Very frequently, there are unexpected consequences of the absence of the gene product, such as secondary adaptations, which confound or obscure the gene function of interest. For example, a gene involved in memory might result in abnormal sensory function. As a result, many tests for the assessment of memory might have to be ruled out if they depend on the sensory function affected. The difficulty here is that the laboratories developing these genetically manipulated animals rarely have the capacity to test for secondary adaptations and therefore most go unnoticed.

Functional genomics therefore requires a special need for comprehensive assessment of behavior. This will facilitate fast and accurate correlation of behavior, physiology and gene expression, and enables the investigator to rule out confounding variables as the cause of observed behavioral and physiological phenotypes.

The gene function and animal-model discovery bottleneck

New tools in molecular biology have led to an explosion of research using genetically manipulated mice and rats, either by targeted approaches or by applying ethylnitrosourea (ENU) to induce point mutations in sperm. A quick search through Medline for the terms 'knockout' and 'transgenic' shows a very clear exponential growth in the number of publications centered on genetically modified animals (Fig. 1).

In a way that parallels the problems faced by the pharmaceutical industry, assessment of the phenotype of these new mutants has a number of problems.

First, characterizing the phenotype of a new mutant requires much more than a simple behavioral or pharmacological test. For example, the phenotype of a gene knockout involved in processes of learning and memory might be obscured by deficits in sensory-motor function caused by loss of the gene product itself or by developmental compensations for the gene absence. A wide battery of tests is therefore required to assess multiple processes.

Examples of these test batteries are the SHIRPA protocol [2] and PsychoScreen™ (http://www.psychogenics.com/services/psychoscreen.htm). Although a first-pass battery can be rapid [an estimated 350 mice can be studied at the MRC Mammalian Genetics Unit (http://www.mgu.har.mrc.ac.uk) using the first-tier level of SHIRPA; 150 mice can be run per Research Associate in the neurological subtests of PsychoScreen™), a comprehensive assessment – from neurological assessment to cognitive function – takes so much time and effort that it can only be performed at a very low throughput. PsychoScreen™, which consists of ~12 highly automated tests, runs ~50 mice with two Research Associates in a full two-week period of data collection.

Second, in the area of ENU mutagenesis, the very large number of animals that are produced, and the cost of breeding them to produce new mouse lines, compound the problem. Many thousands of mice might need to be screened to detect dominant mutations, and as many as 25,000 might need to be screened for recessive mutations. Laboratories that focus on ENU mutagenesis, such as the Jackson Labs Neuroscience Mutagenesis Facility (http://www.jax.org/nmf/documents/protocols/genetics/phenotyping.html) are forced to implement high-throughput batteries for the assessment of physiology, behavior, and other endpoints, before they commit to breeding a given line. Similarly, quantitative trait-loci analysis involves behavioral phenotyping of a large number of cross-hybrid animals. By necessity, currently used batteries of the highest throughput are limited in their scope, particularly in relation to behavior.

Third, implementing any kind of large-scale HTS is costly, and only large laboratories with a particular interest in behavior and physiology can afford to run such programs.

Fourth, in the analysis of behavior, and especially of the effect of a genetic manipulation, a serious problem arises when many variables are collected and analyzed independently of each other: the more statistical analyses are conducted, the higher the occurrence of a false-positive or Type I error. If the level of Type I error is fixed at $\alpha = 0.05$ for each individual test, the overall probability of a Type I error for the complete study (experiment-wise Type I error probability) is much larger than $\alpha = 0.05$. In the characterization of mutants, this is a particularly critical issue because many tests are conducted to assess behavior, physiology and pharmacological response. A false-positive is a false-lead that could guide costly efforts in the wrong direction.

Standard behavioral techniques

Although great progress has been made in the development of techniques that permit objective and quantitative studies of behavior, these techniques involve specialized expertise and considerable labor. For example, the study of sensorimotor gating is one of the most promising techniques for the study of compromised-processing schizophrenia-like states that has been successfully applied to animal models [3]. However, the test requires considerable technical expertise; not many laboratories have this technology available for research. There are many automated tests that are easier to set-up and run, such as the measurement of exploratory activity in rodents or the prepulse inhibition of startle [4], but these provide only a partial assessment of a narrow behavioral repertoire.

Standard behavioral tests are therefore not only time-consuming but also provide a limited picture of the behavior of the animal and do not enable a comprehensive assessment of the test subject. Behavioral data are limited to what the scientific community considers a relevant behavioral measure, based on the way the tests are performed, and the context and time constraints of the testing. Although this is the correct approach in hypothesis-driven research as it addresses specific questions about function, a novel, faster and wide-based approach is required for the analysis of novel drugs, drug indications, or characterization of novel mutants.

Moreover, there is an intrinsic variability and subjectivity in the behavioral data generated in some of the standard tests. Behavioral data are collected using many different techniques; in some cases, by trained observers who use rating scales to assess drug-induced behavior. Although a trained observer can detect complex or subtle changes in behavior, the reliability of the data heavily depends on the expertise of the observer. There is also a high inter-observer variability. This method is obviously constrained by the short duration of the observation and by physical processes experienced by the observer (e.g. fatigue, boredom and distraction).

Standard behavioral tests collect only a small subset of ongoing behaviors. Devices are often used for measuring the activity of an animal, for example, by video, infrared light sensors, interruption of magnetic fields, or Doppler shifts. Although these devices permit prolonged observation of one aspect of the animal's activity, other concurrent behaviors are not captured. Standard behavioral tests mostly collect overt

Table 1. Comparison of available systems for behavioral and physiological assessment of small mammals

System	Type of product	Type of behavior												Type of measure	Type of database	Vendors
		Locomote	Eat/drink	Rear	Groom	Startle	Twitch/seizure	Freezing	Sleep	EEG	ECG	Integrated states	Other[a]			
Smart Cube	Hardware plus software	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	Automatic	Archive and data-mining	Psychogenics; http://www.psychogenics.com
Hypnion	Hardware plus software	✓	✓	×	×	×	×	×	✓	✓	×	×	×	Automatic	Archive and data-mining	Hypnion Science & Discovery; http://www.hypnion.com/discovery/score.htm
Activity monitors	Hardware plus software	✓	×	✓	×	×	×	×	×	×	×	×	×	Automatic	Archive	Actimetrics; http://www.actimetrics.com Columbus; http://www.colinst.com/index.shtml Coulborn; http://www.coulbourninst.com/ Med. Associates; http://www.med-associates.com/ SanDiego Instruments; http://www.sd-inst.com/
Freezing monitors	Software	✓	×	×	×	×	×	✓	×	×	×	×	×	Automatic	n.a.	Actimetrics Imetronic; http://www.imetronic.com San Diego Instruments Med. Associates
Columbus CLAMS	Hardware plus software	✓	✓	✓	×	×	×	×	×	×	×	×	×	Automatic	Outlier detection	Columbus Jackson; http://www.jax.org/nmf/documents/protocols/multiple_domain/goals.html.
Noldus Ethovision	Some hardware plus software	✓	×	✓	×	×	×	×	×	×	×	×	×	Automatic	n.a.	Noldus; http://www.noldus.com/
Noldus Observer	Some hardware plus software	✓	✓	✓	✓	×	×	×	×	×	×	×	×	Manual	n.a.	Noldus
Clever systems	Software	✓	✓	✓	×	×	×	×	×	×	×	×	×	Automatic	n.a.	CleverSys; NIH grants R43DA014889, R43MH058964, Liang, Yiqing
Stoelting Metabolic Cages	Hardware plus software	✓	✓	✓	×	×	×	×	×	×	×	×	×	Automatic	n.a.	Stoelting; http://www.stoeltingco.com/
Startle boxes	Hardware plus software	×	×	×	×	✓	×	×	×	×	×	×	×	Automatic	n.a.	Med. Associates SanDiego Instruments
Operant boxes	Hardware plus language	×	✓	×	×	×	×	×	×	×	×	×	×	Automatic	n.a.	Coulborn SanDiego Instruments
Telemetry	Hardware plus software	✓	×	×	×	×	×	×	×	×	✓	×	×	Automatic	n.a.	DataScience; http://www.datasci.com/References/miscref.html Minimitter; http://www.minimitter.com/

[a]Other includes: jumping; supported versus unsupported rearing; head versus anogenital licking and grooming; stereotyped versus non-stereotyped grooming; burying; digging; gait analysis; conditional aversive; and reward-motivated responses.
Abbreviations: ECG, electrocardiogram; EEG, electro-encephalogram; n.a., not applicable.

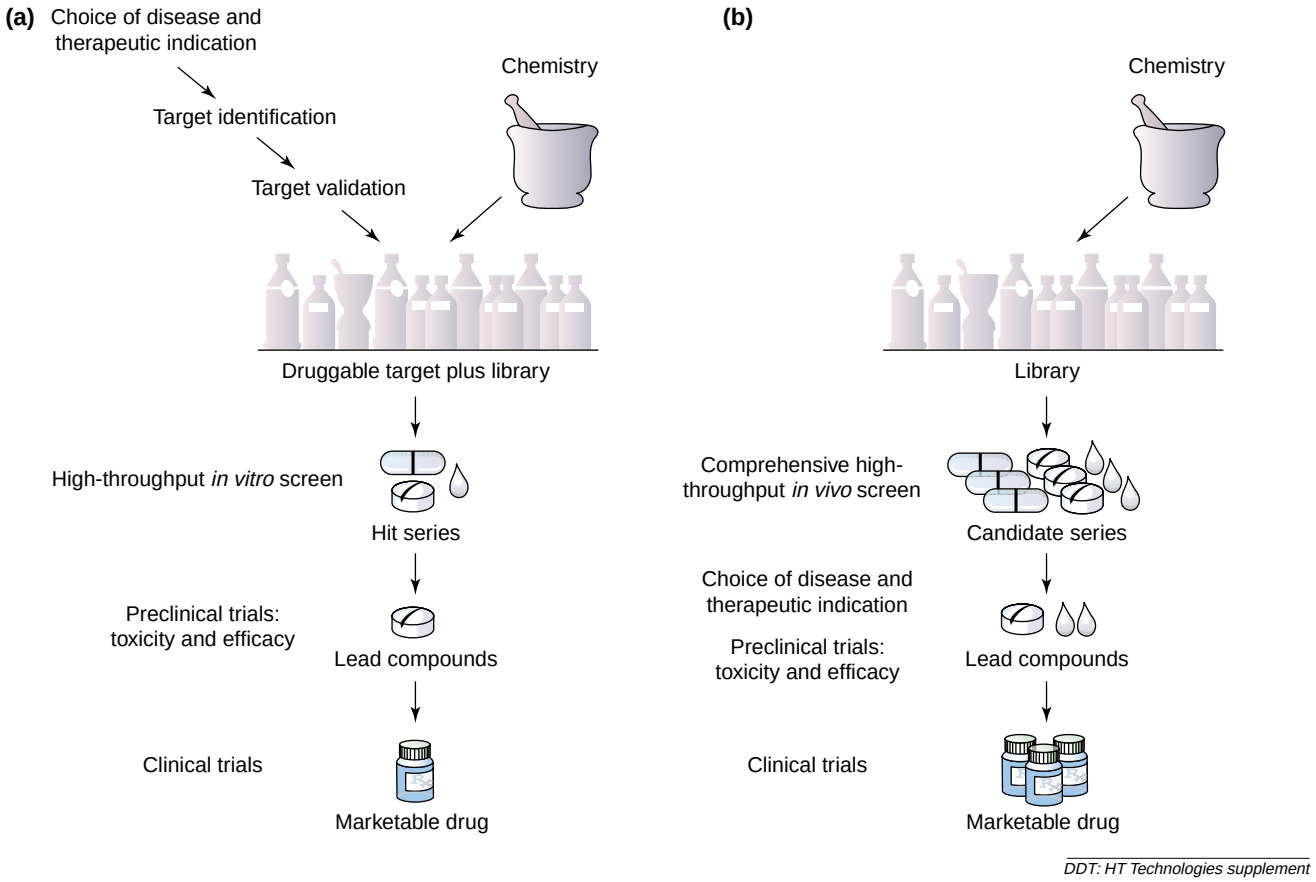

Figure 2. A comparison between the typical drug discovery process and a new discovery pathway based on a comprehensive library screening. (a) The standard process requires detailed validation of a target before candidate drugs for preclinical studies can be obtained from *in vitro* screens. (b) A novel pathway based on a comprehensive high-throughput *in vivo* screen enables the testing of compounds without regard for their mechanism-of-action. This approach will be of special interest for: (1) shelved compounds that failed because of lack of efficacy for a given indication; (2) compounds on the market that might have novel applications; (3) representative compounds of chemical libraries that await preclinical studies in a saturated pipeline; and (4) compounds generated through combinatorial chemistry for which targets have not yet been identified. As it enables assessment of drug action for many indications simultaneously, the system will result in a higher number of hits, with a concomitant reduction in the number of animals necessary for the development of a lead compound.

behavior. Most behavioral monitoring systems are limited because they only assess one dimension of the animal's visually detectable gross motion (e.g. locomotion and stereotyped motor behavior), and do not provide complex behavioral assessment or concurrent physiological or biomechanical measures.

Performing a battery of tests means either that the same animals are tested several times, with all the problems of carry-over effects and stress induced by such manipulations, or that a very large number of animals is needed.

High-throughput *in vivo* behavioral systems

Recent years have witnessed significant advances in improving the throughput of behavioral analyses. Numerous semi-automated systems have been developed, each with its own merits and limitations. For example, there are several options for the automated measurement of locomotor activity, freezing, or startle, but a system that enables measurement of these three behaviors simultaneously with others is lacking. Most current systems are not integrated in a larger database system for the analysis of behavior and physiology, although some private companies have moved in that direction in the area of electroencephalograms (EEGs) and sleep (Organon; http://www.organon.com/). For example, Hypnion (http://www.hypnion.com/discovery/score.htm), a company that focuses on drug discovery for sleep disorders, has developed SCORE-2000™, a high-throughput *in vivo* assay system that monitors EEG and other physiological and behavioral variables in rats. The system incorporates a large database that enables the prediction of drug efficacy and side-effects of existing and experimental compounds for sleep disorders by using bioinformatic tools.

SmartCube™ (Psychogenics), one of the most comprehensive approaches to-date for the analysis of behavior and physiology, works by compressing numerous tests into one, by gathering more comprehensive and detailed data into a large database, and by implementing sophisticated data-mining algorithms. By doing so, SmartCube™ should reduce the amount of handling and the number of animals required.

The system consists of two main components: SmartCube™, a proprietary behavioral data-capture system; and SmartBase™, an intelligent database that enables sophisticated statistical analysis and data-mining. It is able to monitor many behavioral and physiological phenomena based on standard criteria, including anxiety, activity patterns, avoidance, learning, memory and motivation, among others.

Table 1 shows a comparison of available systems detailing the type of behavior measured and the level of database archiving and data-mining capabilities.

A new solution to an old bottleneck

Breaking the bottleneck of *in vivo* behavioral models requires a system for capturing, collecting and analyzing behavioral changes resulting from manipulating genes or administering compounds. For example, CNS drug discovery would be revolutionized if it became possible to achieve a hundred-fold increase in throughput. Such a system would open up significant new opportunities for pharmaceutical and biotechnology companies by enabling them to cost-effectively screen compounds from their existing libraries for their potential impact on CNS disorders (Fig. 2).

The goal is not only to automate the traditional behavioral research process, but also to bring greater precision to behavior analysis. By automating and systematizing the way in which behavior is captured and collected, the ideal system would standardize the translation and interpretation of the visual data, enabling this critical information to be stored in a relational database for analysis (along with the non-visual data also captured during the screening). Furthermore, being able to store and compare the captured data in a standardized format would have the potential to populate an unparalleled database related to CNS functions and disorders. This would significantly increase accuracy, and facilitate interpretation by providing a strong foundation for comparative analysis.

A non-invasive automated system that can run more animals in less time, and can assess many processes at the same time, yields a compounded efficiency that not only brings a significant increase in throughput, but also a relief to the ethical issue of the use of animals in drug discovery.

Significantly enhancing the throughput of early behavioral studies opens up the opportunity to screen compounds *in vivo* in a non-hypothesis driven approach and to 'rule-in' as well as 'rule-out' a vastly larger number of compounds. Furthermore, many of these compounds will interact with novel targets, thereby resulting in a much more efficient and ultimately successful development of new medications with novel mechanisms-of-action.

Concluding remarks

It can be argued that high-throughput *in vivo* screening will not be achieved until an equally high-throughput drug metabolism and pharmacokinetic (DMPK) assessment system has been developed. However, screening representative compounds from libraries before seeking DMPK information implies that DMPK studies will be carried out only for those compounds that have demonstrated *in vivo* activity. If the *in vivo* assessment is fast, reliable and inexpensive, this new pathway becomes a real possibility.

References

1 Friedman, Billing, Ramsey and Co. (May, 2001) Healing the mind and brain: CNS European Survey. *FBR International Technology Research*, London, UK

2 Rogers, D.C. *et al.* (1997) Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. *Mamm. Genome* 8, 711–713

3 Light, G.A. and Braff, D.L. (1999) Human and animal studies of schizophrenia-related gating deficits. *Curr. Psychiatry Rep*. 1, 31–40

4 Varty, G.B. *et al.* (2000) Environmental enrichment and isolation rearing in the rat: effects on locomotor behavior and startle response plasticity. *Biological Psychiatry* 47, 864–873

5 DeAngelis, E.A. (1999) High-throughput phenotypic screening of the mouse electrocardiogram. *Circulation* 100, 1–276