Combinatorial chemistry. Facing the challenge of chemical genomics*

Ferenc Darvas[1,†], Gyorgy Dorman[1], Laszlo Urge[1], Istvan Szabo[1], Zsolt Ronai[2], and Maria Sasvari-Szekely[2]

[1]ComGenex Inc., H-1027 Budapest, Bem rkp. 33-34, Hungary; [2]Semmelweis University Medical School, Institute of Medical Chemistry, Molecular Biology and Pathobiochemistry, H-1088, Budapest, Puskin u. 9, Hungary

Abstract: In the age of high-throughput screening and combinatorial chemistry, the focus of drug discovery is to replace the sequential approach with the most effective parallel approach. By the completion of the human gene-map, understanding and healing a disease require the integration of genomics, proteomics, and, very recently, metabolomics with early utilization of diverse small-molecule libraries to create a more powerful "total" drug discovery approach.

In this post-genomic era, there is an enhanced demand for information-enriched combinatorial libraries which are high-quality, chemically and physiologically stable, diverse, and supported by measured and predicted data. Furthermore, specific marker libraries could be used for early functional profiling of the genome, proteome, and metabolome. In this new operating model, called "combinatorial chemical genomics", an optimal combination of the marker and high-quality libraries provides a novel synergy for the drug discovery process at a very early stage.

INTRODUCTION: NEW CHALLENGES IN THE POST-GENOMIC DRUG DISCOVERY

Combinatorial chemistry has evolved to provide novel chemical entities for screening by increasing both the number and diversity of the compounds supplied. On the other hand, in the post-genomic era 3000–10 000 new disease-related targets are expected to emerge. The increasing number of small molecules and the potential targets, however, does not automatically lead to higher efficiency of the whole drug discovery process.

The major task in the post-genomic discovery is to establish a synergy between the increasing number of targets and small molecules in many aspects. The new targets will provide novel therapeutic interventions of traditional and newly emerging diseases with more specific, efficient, and safer small molecules.

One of the new trends in the post-genomic drug discovery is that the chemical libraries can be implemented in the early phases of the genomics-based drug discovery scheme (Fig. 1). Chemical genomics is expected to accelerate the determination of the function and biological role of the gene, through small-molecule interactions with the protein encoded by the gene (this is often referred to as chemical proteomics). Small molecules provide multiple opportunities for "genome or proteome mining" [1,2]. Known robust inhibitors/agonists/antagonists are able to identify targets that belong to cer-

†Corresponding author

tain enzyme or receptor families. However, in the human genome we anticipate completely new disease-related molecular targets that do not fit into currently known families. Those proteins can be quickly identified primarily by novel combinatorial libraries (indicated as drug candidates) in biological media such as originating from diseased tissues or samples.

Figure 1 presents a new parallel discovery model [2,3] where combinatorial libraries are linked to target identification by joint application of marker (reporter) and quality libraries that simultaneously provide hit compounds. In further steps, the selective ligand-binding interaction validates the target, and inversely the binding to the target confirms the biological activity of the ligand.

The laborious target validation phase is a real bottleneck in post-genomic drug discovery and prevents the fast impact of genomics on the development of new drugs. *In vitr*o proteomics provides only indirect information about protein function, since gene transcription and translation are different in tissues, cells, and organs. That is the major challenge combinatorial chemistry faces and provides a solution through *chemical genomics*.

The ligand-driven approach shortens significantly the routine proteoma screening for identifying "druggable" targets that allow effective small-molecule therapeutic intervention and also provides help to *in vivo* functional genomics/proteomics. In the new, parallel operating model [3], ligand-driven target validation is simultaneous to lead optimization, which is carried out with a series of focused libraries.

One of the major responses to the post-genomic challenge [4] is the development of special libraries and their integrated use with diverse discovery libraries. Furthermore, the quality of the discovery libraries should be radically improved, and a rapid increase in productivity is needed in chemical synthesis as well as in high-throughput screening. Library members linked with special functionalities are able to mimic biological target validation steps using normal and diseased tissues, such as "knocking out" different functions at a cellular level and blocking signaling pathways by chemically interacting with the molecular target. Chemical genetics, using small molecules to selectively disrupt signals or modulate phenotypic changes through interaction with proteins, basically mimics the effect of mutations that lead to the development of disease states [5].

The whole parallel process requires integrated technologies with increased high-throughput from library generation to screening, which leads to an explosion of the data recorded. With the implemen-
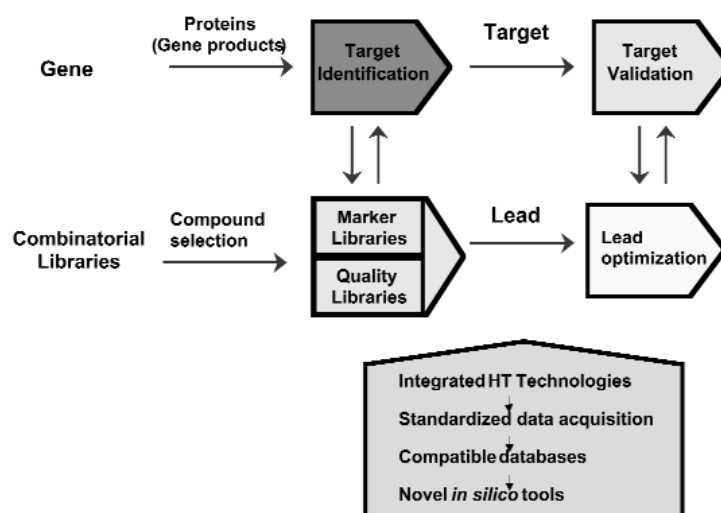

Fig. 1 New parallel discovery model.

tation of proper standardization of data acquisition, large compatible databases will be easily generated. They can simply reveal intrinsic and comprehensive new structure–property relationships that can be the starting point for novel *in silico* tools [6]. Chemoinformatics and bioinformatics further refine the above process, contributing to the quality of libraries and target functional identification as well as validation and optimization.

COMBINATORIAL SYNTHESIS IN THE POST-GENOMIC ERA

A major new trend in the post-genomic combinatorial library generation is the integration of different technologies and, in principle, finding the best choice for specific purposes.

Solution-, solid-, and fluorous-phase technologies show a markedly different productivity and development timeline (Fig. 2). Solution-phase combinatorial chemistry can apply directly to all the accumulated synthetic knowledge of organic chemistry without any extra effort. On the other hand, for solid-phase synthesis the solution-phase protocols will be adapted. The emerging fluorous-phase technology represents an interface between the above technologies, combining the advantages of both of them.

Combinatorial library design and virtual library generation combine a high level of organic chemistry knowledge with computer-assisted methods, including synthetic matrix planning, reagent database retrieval, and different, state-of-the-art selection methods.

In our practice, we use an in-house, established chemical reaction catalog with validated procedures, together with our "drug-like" multifunctional building block collection to define the optimal diversity building route (2–6 steps) in a cascading, tree-like manner (Fig. 3). This strategy is realized by ComGenex matrix technology (CMT) [7], utilizing manual and robotized parallel synthesis stations of different size that reflect the cascading diversity building approach.

CMT is practically a *sequential* split–pool technique where the intermediates are divided into small portions after each diversity-building synthetic step and reacted with a pool of different reagents in a parallel manner in isolated vessels. CMT is practically a technological line that also contains state-of-the-art selection methods (diversity, drug-likeness, and ADME prefiltering), reaction piloting, HT purification, and analytics.
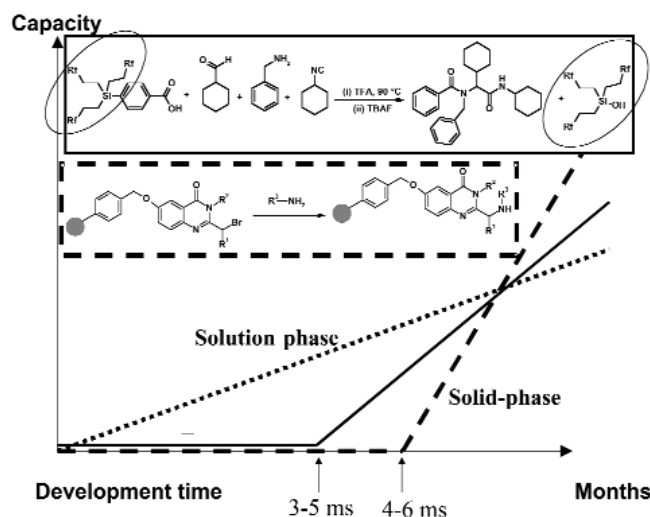

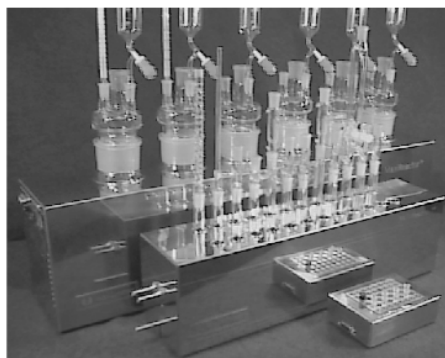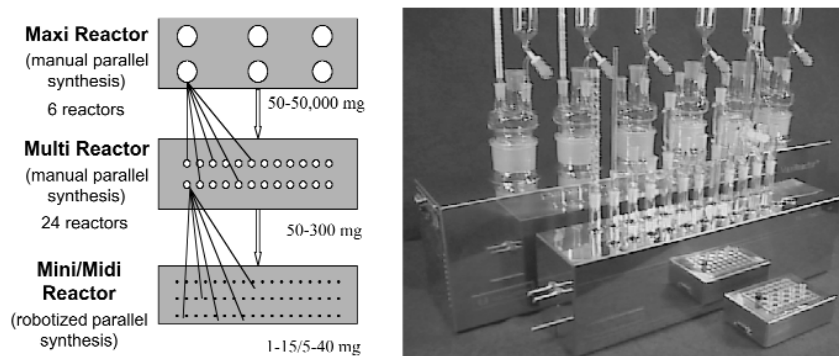**Fig. 2** Major synthetic approaches in combinatorial chemistry.

Fig. 3 CMT cascading matrix technology.

Table 1 Essential characterization of single compounds in drug discovery.

Data class	Measured parameters	Calculated from structure	Calculated from measured parameters
In silico predictions		• pK_a (acidic, basic) • log P, log D (at 3 pHs) • list of metabolites, fingerprint • toxicity in different classes (e.g., oncogenicity, etc.)	
Ex silico (measurement-supported predictions based on HPLC)	• retention time calibrated with reference compounds (CHI)		• solubility • cell permeability • oral absorption • log D confirmation at various pHs • BBB penetration
	• retention time measured at various thermal stress condition		• long-term stability • compound shelf-life
In vitro measurement	• *in vitro* metabolism (stability and fingerprint) • CYP induction/ inhibition, drug–drug interactions • cell toxicity		

INFORMATION-ENRICHED, HIGH-QUALITY DISCOVERY LIBRARIES

The chemical genomics approach requires higher standards of the discovery libraries. This new challenge can be handled at multiple levels. First, each member of a combinatorial library should be characterized according to a hierarchical data acquisition scheme. This scheme includes the following:

•   proper chemical characterization and quality control
•   proper estimation for chemical stability

- proper physicochemical characterization
- proper description of adsorption, distribution, metabolism, elimination, and toxicology (ADME/Tox) features
- *priori* knowledge of biological activity (expected activity classification)

Secondly, the compounds characterized in the above manner can be pooled into combinatorial libraries based on structural diversity, ADME admissibility, specific physicochemical parameter requirements, or according to target-related biological activities.

Most practically, the compound characterization should be performed before any synthetic efforts, however, the present structure–property relationships are still limited in certain areas and the accuracy of the *in silico* assessment needs significant improvement.

As a response to this challenge in the post-genomic combinatorial library generation we propose a good compound characterization practice (GCCP) [8], that integrates the measured, measurement-based predictive, and *in silico* predicted data (Table 1). While some of the properties are already evident and accepted for chemical characterization and quality control, more uniform characterizations are needed in order to meet the drug "developability" requirements and increase the success rate in the preclinical and clinical phases.

Estimation of ADME/Tox-related parameters

The *in vitro* nature of HTS techniques generates hits with often unfavorably low absorption and metabolic stability and fast clearance; in other words, with poor oral bioavailability. In the late 1990s, early assessment of the ADME/Tox properties of libraries was recognized [9] in order to identify and eliminate candidate molecules that are unlikely to survive later stages of discovery and development. Thus, a new slogan was put into operation: "fail fast, fail cheap".

In the efforts of "turning chemicals into drugs" [9], ADME-related physicochemical parameter limits are generally applied as Lipinski's rule (for Mwt, log P, log D, H-donors, and H-acceptors) [12] in the synthetic design phase, removing reagents contributing to the violation of the limit.

HPLC platforms that are primarily used for the high-throughput quality control of the compound libraries allow a different type of parameter prediction based on the measured retention times. Chromatographic hydrophobicity index (CHI) was introduced by Valkó at GlaxoSmithKline [10]. CHI values are used to estimate lipophilicity if the reversed-phase HPLC system is calibrated with a representative set of compounds. Valkó *et al.* further refined the model to predict cell permeability and oral absorption using measured CHI values together with calculated molar refractivity (cMR) as a size parameter [10].

Chemical stability and shelf-life estimation [11]

Archiving compound collections in solid state or in frozen dimethyl sulfoxide (DMSO) for repeated biological assays often leads to chemical degradation, thus, false-positive activity measurements. It is assumed that degradation processes are accelerated by elevating the storage temperature. The degradation rate at high temperature is characteristic of the structure, and it can be used to estimate the *shelf-life* under storage conditions at various temperatures by using the Arrhenius kinetic model [11].

The *shelf-life* of a compound, library, or repository according to our definition is a time interval that the species is expected to remain within an approved purity specification under the required storage conditions.

In vitro ADME/Tox measurements

For *in vitro* measurement of *ADME/Tox parameters*, cell-based HTS methods are utilized. The details are reported in a recent review [13].

NEW GENERATION OF COMBINATORIAL LIBRARIES FOR CHEMICAL GENOMICS/ PROTEOMICS

Compound pools characterized by high-value data allow a generation of quality libraries based on diversity or target-oriented selection. Maximizing the diversity of a compound library is the central concept in combinatorial chemistry, by removing structural redundancy and covering most of the property space. The approaches that address the issues of chemical diversity go beyond the focus of the present account [19].

On the other hand, quality libraries can be used early in the post-genomic discovery without any modification or together with the attachment of special functionalities in order to measure direct interaction with proteins as gene-products (Table 2). Simple affinity selection methods use label-free compound mixtures or microarrays; however, for identification of ligand target pairs, if they are not covalently attached, deconvolution or spatial decoding and further iterative refining steps are needed.

The major feature of *covalent bond-forming* libraries is a creation of the stable linkage between the small molecules and the biopolymers targeted by the ligand or substrate toward the active site. The stable covalent linkage survives proteolysis, chemical fragmentation, and sequencing; thus, the small molecule will stay trapped within the active site during the entire manipulation.

The covalent bond-forming libraries [22] have the following elements: ligand, chemical (electrophilic) or photoactivatable reactive group, reporter tag, and linker. The utility of electrophilic covalent probes is limited to specific enzyme classes [21] only, where mechanism-based inactivators and irreversible inhibitors have been developed over the years.

Photoactivatable marker libraries represent small-molecule libraries, where a photolabile crosslinking unit together with a reporter tag (fluorescent, biotin, or isotopic label) (Fig. 4) is attached to each compound at various sites around a core structure. In contrast to the electrophylic probes, they bind reversibly *in dark*, and only irradiation at defined wavelengths results in the covalent attachment [20].

Another advantage of photolabile ligands/substrates compared with the chemical probes is that they do not require nucleophilic functional groups within the binding region. Finally, but most importantly, irreversible inhibitor probes can conjugate only enzymes, while the photoprobes can react with virtually all the proteins.

In our practice, where ca. 10 000–25 000 compounds are synthesized each month, marker library generation is implemented into the general synthetic design and realization scheme directly utilizing intermediates as precursors. The marker unit is attached to a suitable (or tethered) functionality in a late step of the synthesis (Fig. 5).

Table 2 Special libraries for chemical genomics/proteomics.

| Library class	Library	Utility
Label-free libraries	Affinity microarrays [16]	Target identification
	Diverse natural product-like libraries [14].	Chemical genetics, induction of phenotypic changes
	Reference ligand libraries (biomolecules) [18]	Ligand identification for receptors (e.g., orphan receptors)
	Reference libraries of toxic compounds [17]	Toxicogenomics
Tagged libraries: *binding in a noncovalent manner*	Fluorogenic substrate libraries [15]	Target class identification/profiling
Tagged libraries: *binding in a covalent manner*	Irreversible inhibitor libraries [21]	Target class identification/profiling (serine/cystein proteases, thus far*)*
	Photoactivatable libraries [2]	Target (class) identification Target validation
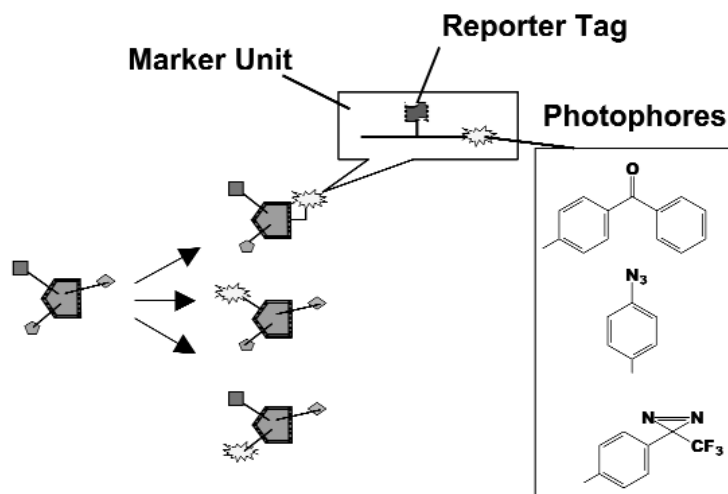
Fig. 4 Design of photomarker subsets for combinatorial libraries.

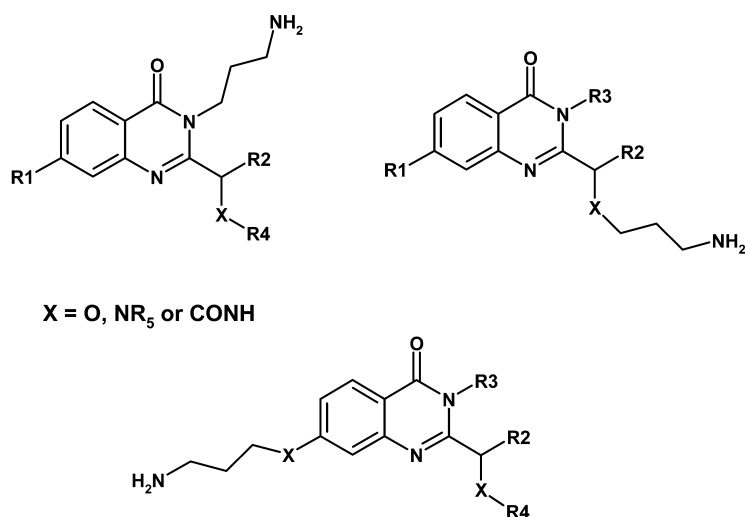**X = O, NR₅ or CONH**

X = O, NR$_5$ or CONH

Fig. 5 Tethered sublibraries around a quinazolone core.

The marker compounds constitute a sublibrary related to their label-free counterparts. The simultaneous production and the combinatorial design approach make the process highly efficient, unlike the traditional photoaffinity technique.

The marker libraries can be classified into diverse or focused subsets. For *target identification/validation*, a highly diverse sublibrary is used. The libraries contain a collection of known reference substrate/ligand photoprobes (functionality directed marker libraries) and/or a diverse photolabile subset of novel drug-like compounds (discovery marker library). Through competition experiments with label-free discovery library members, further hits can be identified directly from the crude tissue or proteome.

Cell-based assays often provide hits where the exact target is not known. Based on the hit structures, a *focused marker library* can be prepared where the photoactivatable groups are placed separately at the variable sites around the core (scaffold) structure. These libraries can be used to identify the intracellular molecular target.

In the next section, we discuss a specific example of how to generate a focused marker library based on a known antagonist structure.

EXAMPLE: INVESTIGATION OF THE POLYMORPHISM OF DOPAMINE D4 RECEPTORS BY CHEMICAL GENOMICS METHODS

After the completion of the Human Genome Project in the "post-genomic" era, genetic studies focus on the function and the variations of the genetic information. The variable part of the human genome is 0.1%, which adds up to 3 million base-pairs; these variations are mostly single nucleotide polymorphisms (SNP), but length polymorphisms (insertions, deletions, and repeat polymorphisms) are also described. A large number of these variations are thought to lack any function, being in the non-coding region. However, SNPs resulting in illnesses are also well known, they are usually referred to as point mutations and often cause the decreased or missing activity of an enzyme. On the other hand, length polymorphism can also be linked to altered intracellular signaling through different expression or modified binding affinity to certain chemical messengers, etc., leading to specific disorders.

Polymorphisms in the gene of the dopamine D4 receptor

Dopamine is one of the most important neurotransmitters of the central nervous system taking part in the regulation of movements, emotions, and the neuroendocrin system. Dopamine takes its effect on the dopamine receptors that belong to the G-protein coupled receptors possessing 7 transmembrane domains. The dopamine D4 (DRD4) and D2 (DRD2) receptors belong to the D2-type receptor family; they take their intracellular effect by decreasing the level of cAMP. D2 receptors can be found mostly in the striatum, nucleus accumbens, tuberculum olfactorium, and substantia nigra and play a significant role in the control of movements. DRD4 receptors cannot be found in the nigrostriatal pathway, but they are mainly in the limbic system (prefrontal cortex, amygdala, hypothalamus). This pattern of location of the two different receptors explains the features of the atypical antipsychotic drugs: they are specific antagonists of the dopamine D4 receptors, thus, they are efficient in the therapy of schisophrenia without strong extrapyramidal side effects.

Recently, the dopamine D4 receptor gene has become the focus of psychogenetic studies, since a large number of polymorphisms were described in both the coding and the 5′ untranslated region of the gene. The unique polymorph structure of the D4 receptor is responsible for the diversity of its involvement in the dopaminerg regulation. One of the most investigated polymorphisms is a variable number of tandem repeats in exon III, where the number of copies of a 48-bp-long sequence varies from two to ten [23]. This part of the gene codes the third cytoplasmatic loop of the receptor protein, which probably plays an important role in signal transferring being associated with the G-proteins. The assumed difference between the function of the proteins possessing long and short cytoplasmatic loop was demonstrated *in vitro*: the cAMP decreasing activity of the 2x or 4x form is twice higher than that of the 7x receptor. However, there is not a simple parallel relation between the number of copies and the function of the receptor: the 10x protein was described to have the highest efficiency [24].

Benjamin [25] and Ebstein [26] demonstrated an association between the DRD4 exon III VNTR (variable number of tandem repeats) and psychological traits, e.g., novelty-seeking personality trait, and numerous workgroups studied psychiatric disorders such as the development of attention deficit hyperactivity disorder (ADHD) [27].

Figure 6 shows the result of the genotyping of the DRD4 exon III 48 bp VNTR of a series of 8 samples. Based on the position of the primers used in our experiments, the expected size of the PCR
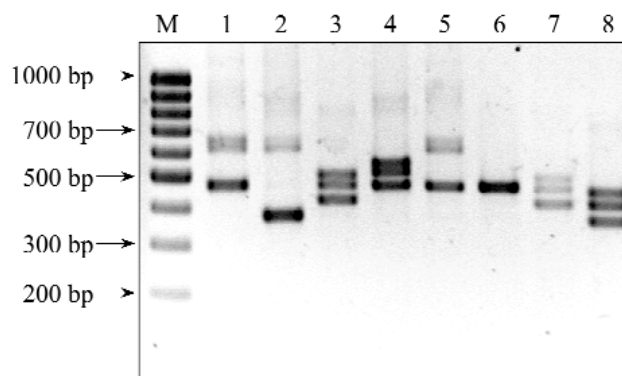
Fig. 6 Genotyping of the DRD4 exon III 48 bp VNTR.

product can be calculated using the $283 + 48n$ equation where n is the number of repeats. The most common form can be seen in case of sample 6: the single 475-bp-long fragment shows that this person possesses the 4 times repeated form on both alleles in homozygous form shortly referred to as "4x/4x". The other samples are heterozygous for the polymorphism, their genotypes are 1: "4x/7x", 2: "2x/7x", 3: "3x/4x", 4: "4x/5x", 5: "4x/7x", 7: "3x/4x", 8: "2x/3x", respectively. The third band with the smallest electrophoretic mobility in the case of the heterozygous samples is the so-called "heteroduplex", a double-stranded DNA formed by a shorter and a longer strand during the PCR. Although the size of this molecule is smaller than that of the long double-stranded DNA fragment, because of the irregular conformation it migrates slower than both the short and the long PCR-product. Lane *M* shows the separation of a 100-bp ladder.

Medical considerations

Exploration, diagnoses, and treatment of D2 and D4 receptor-linked disorders require selective ligands effective in the nanomolar range. Combinatorial chemistry provides HTS technique for rational ligand/drug design, allowing the synthesis of thousands of substances around a chemical core structure. *In silico* hit selection based on molecular fingerprint similarity to known D4 selective ligands resulted in the identification of nearly 1000 compounds from ComGenex's repository. Determination of *in vitro* hits by means of receptor binding studies allows lead optimization based on quantitative structure–activity relationship (QSAR). Ligands identified based on affinity and selectivity criteria should be used for mechanistic studies such as: (1) elucidation of the mechanism of action, e.g., the role of the activation of extracellular signal-regulated kinases (ERKs); (2) clarification of the importance of allosteric regulatory sites; (3) determination of ligand selectivity to polymorph variants and the therapeutic importance of this selectivity; (4) elucidation of the regulation of D4 receptor density through desensitization and down-regulation, which is not known at present. The above achievements should contribute to the better understanding of CNS function and medical control.

Optimized leads screened with ADME/Tox *in silico* prediction tools, MetabolExpert and HazardExpert [13], showing required "drug-likeness" should enter *in vivo* animal studies and clinical phase. Approved candidates become drugs, directly improving the quality of life for those suffering in dopamine-linked medical conditions such as Parkinson's disease, schizophrenia, social disorders, and enhancing generic well-being.

Photomarkers as probes

Dopamine D4 polymorphism can be studied with the previously described chemical genomics armory. As assumed, a photoactivatable D4 antagonist probes (Schemes 1–2) could covalently modify the receptor at the binding site, and the cross-linking pattern can be recorded. That pattern contains all the covalently modified amino acids in the sequence. Since the cAMP decreasing activity strongly depends on the repeated alleles, this length polymorphism very likely causes changes in the secondary structure of the biopolymer. Thus, D4-specific photoaffinity probes can lead to a specific cross-linking pattern related to the polymorphism. Through validating the pattern with a series of photophores (a focused library around the core of the antagonist structure) a fast and reliable tool could be provided. Furthermore, polymorph-selective antagonist variants can be easily developed using that ligand-driven
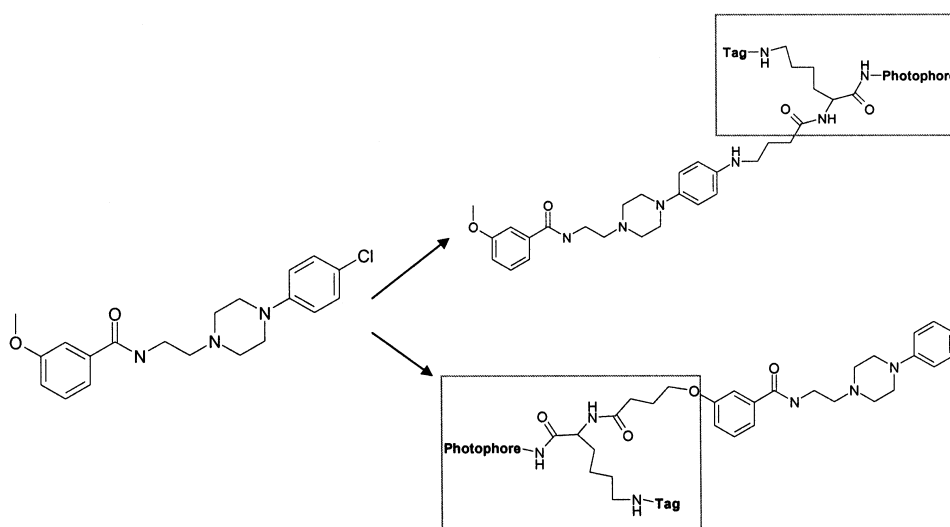
Scheme 1 Transformation of D4 receptor antagonists into "tagged" photoaffinity probes.
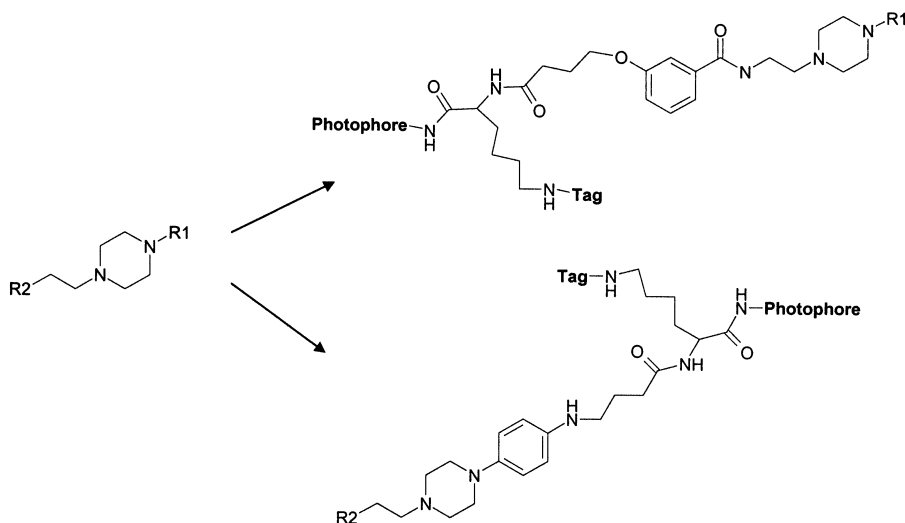
Scheme 2 Design of a photomarker library around the major structural element (core).

technology. Photocovalent irreversible probes should also be used for the determination of polymorph receptor subtype density in different brain regions, which is not mapped at present.

CONCLUSIONS AND OUTLOOK

The post-genomic era initiated a differential development on combinatorial chemistry. The quality requirements are much higher for the discovery libraries, while new generation of combinatorial libraries have to emerge to study the proteome/genome directly by affinity-directed methods.

The implementation of the chemical genomics/proteomics in the drug discovery pipeline leads to a *shortcut* through identifying hits and their potential target in just one single step. That new approach shortens significantly the routine proteoma screening for target identification and supports the *in vivo* functional genomics/proteomics technologies and the lead optimization.

The new parallel discovery model [2,3] requires rapid technological advancement: miniaturization and nanotechnology, which can produce millions of compounds per week that can be immobilized as small-molecule arrays on glass [14], or gold surface and surface plasmon resonance is used for detecting the binding.

Finally, in the near future, ultra-thin surfaces [28], microfluidic networks [29], and related chip-based microsystems [30] will allow synthesis, purification, and high-throughput screening in the same device, providing another milestone on the way to total integration.

ACKNOWLEDGMENTS

REFERENCES

1. T. A. J. Haystead. Proteome mining: exploiting serendipity in drug discovery. *Curr. Drug Discov.* **March**, 22–24 (2001).
2. G. Dormán, P. Krajcsi, F. Darvas. *Curr. Drug Discov.* October, 21–23 (2001).
3. S. Myers and A. Baker. *Nature Biotech.* **19**, 727–730 (2001).
4. D. S. Thorpe. *Combin. Chem. High-Through. Screen.* **3**, 421–436, (2000).
5. D. S. Tan, M. A. Foley, B. R. Stockwell, M. D. Shair, S. L. Schreiber. *J. Am. Chem. Soc.* **121**, 9073–9087 (1999).
6. G. C. Terstappen and A. Reggiani. *Trends Pharm. Sci.* **22**, 23–26 (2001).
7. F. Darvas and L. Kovács. In *High-Throughput Screening,* J. P. Devlin (Ed.), pp. 223–242, Marcel Dekker, New York (1997).
8. F. Darvas. "Drug Discovery by Information-Content Enriched Combinatorial Libraries: The GCCP Concept", Symposium on Molecular Informatics and Combinatorial Quantum Chemistry, Budapest, Hungary, February 19 (2001).
9. J. Hogson. *Nature Biotech.* **19**, 722–726 (2001).
10. K Valkó. Measurements and Predictions of Physicochemical Properties In: HT ADMETox estimation based on *in vitro* and *in silico* approaches, F. Darvas and G. Dormán (Eds.), Eaton Publishing (in press).
11. F. Darvas, G. Dormán, T. Karancsi, T. Nagy, I. Bágyi. "Estimation of stability and shelf-life for compounds and repositories (With combination of systematic discovery of new rearrangement pathways)". In *Combinatorial Chemistry,* R. Hanko and K. C. Nicolaou (Eds.), Wiley-VCH, New York (in press).
12. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).

13. I. Szabó, G. Dormán, F. Darvas, "*In vitro* metabolism screening", and P. Krajcsi and F. Darvas, "High-throughput in vitro toxicology". In *HT ADMETox Estimation Based on* In Vitro *and* In Silico *Approaches,* F. Darvas and G. Dormán (Eds.), Eaton Publishing (in press).
14. G. MacBeath, A. N. Koehler, S. L. Schreiber. *J. Am. Chem. Soc.* **121**, 7967–7968 (1999).
15. J. L. Harris, B. J. Backes, F. Leonetti, S. Mahrus, J. A. Ellman, C. S. Craik. *PNAS*, **97**, 7754–7759 (2000).
16. G. R. Lenz, H. M. Nash, S. Jindal. *Drug Discov. Today* **4**, 145–155 (2000).
17. B. R. Stockwell, *TIBTECH* **18**, 449–455 (2000).
18. M. Detheux. *Inn. Pharm. Techn.* 27–33 (2000).
19. P. Arya, D. T. H. Chou, M-G. Baek. *Angew. Chem., Int. Ed.* **40**, 339–346 (2001).
20. G. Dormán and G. D. Prestwich. *Trends Biotechnol.* **18**, 64–77 (2000).
21. D. Greenbaum, A. Baruch, L. Hayrapetian, Zs. Darula, K. Medzihradszky, M. Bogyo. *J. Cell. Proteomics* (in press).
22. B. J. Cravatt and E. J. Sorensen. *Curr. Opin. Chem. Biol.* **2**, 663–668 (2000).
23. J. B. Lichter, C. L. Barr, J. L. Kennedy, H. H. Van Tol, K. K. Kidd, K. J. Livak. *Hum. Mol. Genet.* **2**, 767–773 (1993).
24. V. Jovanovic, H. C. Guan, H. H. Van Tol. *Pharmacogenetics* **9**, 561–568 (1999).
25. J. Benjamin, L. Li, C. Patterson, B. D. Greenberg, D. L. Murphy, D. H. Hamer. *Nat. Genet.* **12**, 81–84 (1996).
26. R. P. Ebstein, O. Novick, R. Umansky, B. Priel, Y. Osher, D. Blaine, E. R. Bennett, L. Nemanov, M. Katz, R. H. Belmaker. *Nat. Genet.* **12**, 78–80 (1996).
27. S. V. Faraone, J. Biederman, B. Weiffenbach , T. Keith, M. P. Chu, A. Weaver, T. J. Spencer, T. E. Wilens, J. Frazier, M. Cleves, J. Sakai. *Am. J. Psychiatry* **156**, 768–770 (1999).
28. A. Guttman and Z. Rónai. *Electrophoresis* **21**, 3952–3964 (2000).
29. I. Gibbons. *Drug Discov. Today*. (Suppl. *High Throughput Screening*) **5**, S33–S37 (2000).
30. S. A. Sundberg, A. Chow, T. Nikiforov, H. G. Wada. *Drug Discov. Today* (Suppl. *High Throughput Screening*) **6**, S92–S103 (2001).