The Role of Automation in the Identification of New Bioactive Compounds

Pasqualina Liana Scognamiglio[1], Giuseppe Perretta[2] and Daniela Marasco[1]
[1]Department of Biological Sciences, University "Federico II" of Naples, Naples,
[2]Institute of Motors, CNR, Naples,
Italy

1. Introduction

Automation is nowadays implemented in many areas of the drug discovery process, from sample preparation through process development. High-Throughput Screening (HTS) is a well-established process for lead discovery that includes the synthesis and the activity screening of large chemical libraries against biological targets via the use of automated and miniaturized assays and large-scale data analysis. In recent years, high-throughput technologies for combinatorial and multiparallel chemical synthesis and automation technologies for isolation of natural products have tremendously increased the size and diversity of compound collections. The HTS process consists of multiple automated steps involving compound handling, liquid transfers and assay signal capture. Library screening has become an important source of hits for drug discovery programmes. Three main complementary methodologies are actually used: 1) *in silico* virtual screening of libraries to select small sets of compounds for biochemical assays, 2) fragment-based screening using high-throughput X-ray crystallography or NMR methods to discover relatively small related compounds able to bind the target with high efficiency and 3) HTS of either diverse chemical libraries or focused libraries tailored for specific gene families. Furthermore a variety of assay technologies continues to be developed for high-throughput screening; these include cell-based assays, surrogate systems using microbial cells and systems to measure nucleic acid-protein and receptor-ligand interactions. Modifications have been developed for in vitro homogeneous assays, such as time-resolved fluorescence, fluorescence polarization and the scintillation proximity. Innovations in engineering and chemistry have led to delivery systems and sensitive biosensors for Ultrahigh-Throughout Screening working in nanoliter and picoliter volumes. Spectroscopic methods are now sensitive to single molecule fluorescence. Technologies are being developed to identify new targets from genomic information in order to design the next generation of screenings. As HTS assay technologies, screening systems, and analytical instrumentation the interfacing of large compound libraries with sophisticated assay and detection platforms will greatly expand the capability to identify chemical probes for the vast untapped biology encoded by genomes.

As a consequence the growing demand for new, highly effective drugs is driven by the identification of novel targets derived from the human genome project and from the

understanding of complex protein-protein interactions that contribute to the onset and maintaining of pathological conditions. To illustrate the dynamics of quantitative and qualitative process approaches to accelerated drug development, a model pipeline is depicted in Fig. 1.

Fig. 1. Drug Discovery pipelines: **IND**: Investigational New Drug, **CTA**: Clinical Trial Application, **NDA**: New Drug Application, **MAA**: Marketing Authorization.

At the basis of HTS is the simultaneous employment of different sets of compounds that are rapidly screened for the identification of active components. This approach is nowadays regarded as a powerful tool for the discovery of new drug candidates, catalysts and materials. It is also largely utilized to improve the potency and/or the selectivity of existing active leads by producing analogues derived by systematic substitution or introduction of functional groups or by merging active scaffolds. In Combinatorial Chemistry (CC) the higher is the number of building blocks and the number of transforming synthetic steps, the more attainable molecules can be had. However, the higher is the number of attainable molecules, the higher should be the handling capacity of so many reagents and products; thus automation in CC plays a very critical role. Automation in chemical synthesis has been essentially developed around the solid phase method introduced by R. B. Merrifield for the synthesis of peptides, which, after its introduction, has been continuously improved in terms of solid supports, linkers, coupling chemistry, protecting groups and automation procedures, making it nowadays one of the most robust and well-established synthetic methods (Shin et al., 2005). For these reasons, the basic concepts of CC, such as compound libraries, molecular repertoires, chemical diversity and library complexity have been developed using peptides and later transferred to the preparation of libraries of small molecules and other oligomeric biomolecules (Houghten et al., 2000). The easiness of preparation, characterization and the robustness of the available chemistry provide high purity levels and the built-in code represented by their own sequences have promoted the employment of large but rationally encoded mixtures instead of single compounds, leading to the generation and manipulation of libraries composed of hundreds of thousands and even millions of different sequences. The broad complexity of mixture libraries has also led to the development of several screening procedures based on iterative or positional scanning deconvolution approaches for soluble libraries.

2. HTS

High-throughput screening (HTS) has achieved a dominant role in drug discovery over the past two decades. Its aim is to identify active compounds (hits) by screening large numbers of diverse chemical compounds against selected targets and/or cellular phenotypes. The HTS process consists of multiple automated steps involving compound handling, liquid transfers, and assay signal capture, all of which unavoidably contribute to systematic variation in the screening data. It represents the process of testing a large number of diverse chemical structures against disease targets to identify 'hits'.

Compared to traditional drug screening methods, HTS is characterized by its simplicity, rapidness, low cost, and high efficiency, taking the ligand-target interactions as the principle, as well as leading to a higher information harvest.

Independent of the precise nature of the applied screening technology, lead discovery efforts can always be analyzed and optimized along the same fundamental principles of performance management: time, costs, and quality of the process (Fig. 2).
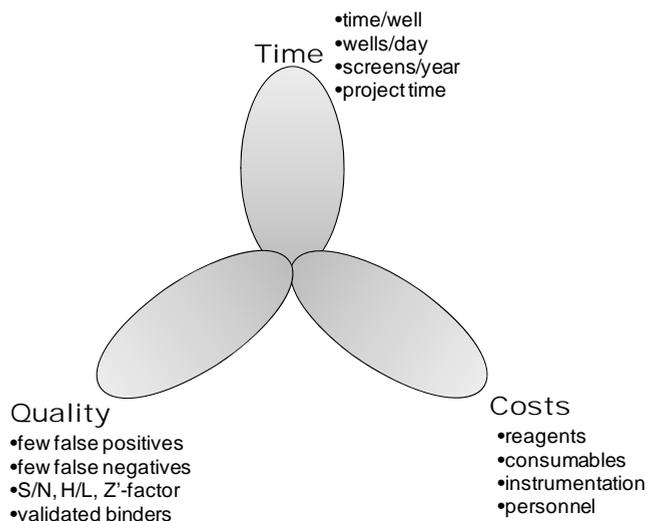

Fig. 2. HTS "magic triangle".

As a multidisciplinary field, HTS involves an automated operation-platform, highly sensitive testing system, specific screening model (in vitro), abundant component libraries, and a data acquisition and processing system. Several technologies such as fluorescence, nuclear-magnetic resonance, affinity chromatography, surface plasmon resonance and DNA microarray, are actually available, and the screening of more than 100 000 samples per day is already possible.

The data analysis challenge is to detect biologically active compounds from assay variability. Traditional plates, controls-based and non-controls-based statistical methods have been widely used for HTS data processing and active identification by both the pharmaceutical industry and academic sectors. Recently, the introduction of improved robust statistical methods has reduced the impact of systematic row/column effects in HTS data. In practice, no single method is the best hit detection method for every HTS data set. Nevertheless, to help in the selection of the most appropriate HTS data-processing and active identification methods a 3-step statistical decision methodology has been developed: Step 1) to determine the most appropriate HTS data-processing method and establish criteria for quality control review and active identification from the 3-day assay signal window and validation tests. Step 2) to perform a multilevel statistical and graphical review of the screening data to exclude data that fall outside the quality control criteria. Step 3) to apply the established active criterion to the quality-assured data to identify the active compounds.

The principles and methods of HTS find their application for screening of combinatorial chemistry, genomics, protein, and peptide libraries. For the success of any HTS assay or screening several steps -like target identification, reagent preparation, compound management, assay development and high-throughput library screening- should be carried out with utmost care and precision. Historically, the majority of all targets in HTS-based lead discovery fall into a rather small set of just a few target families (Table 1). Enzymes such as kinases, proteases, phosphatases, oxidoreductases, phosphodiesterases, and transferases comprise the majority of biochemical targets in today's lead discovery efforts. Among cell-based targets, many GPCRs (G-protein-coupled receptors, 7-transmembrane receptors), nuclear hormone receptors, and some types of voltage- and ligand-gated ion channels (e.g., Ca2+-channels) are very well suited for screening large compound collections. Despite the large number of human genes (>25,000) and the even larger number of gene variants and proteins (>100,000), the number of molecular targets with drugs approved against the target is still fairly limited (~350 targets). The explanations for this discrepancy can be due to some targets might not be feasible at all for modulation via low molecular weight compounds. Others, however, might simply not be approachable by current technologies and therefore constitute not only a great challenge but also a tremendous potential for future lead discovery. Among those, a large number of ion channels, transporters, and transmembrane receptors but also protein-protein, protein-DNA, and protein-RNA interactions, even RNA/DNA itself, might form innovative targets for modulation via low molecular weight compounds. With better tools with sufficient predictivity, one might be able to expand the current HTS portfolio into novel classes of pharmaceutical targets. Besides the obvious classes with targets on the cell surface, such as ion channels, transporters and receptors, we predict that modulation of intracellular pathways, particularly via protein-protein interactions, might have a great potential for pharmaceutical intervention. In this regard, some modern technologies such as subcellular imaging (High Content Screening [HCS]) will certainly enable novel and powerful approaches for innovative lead discovery.

Established Target Classes	Novel Target Classes
• Kinases/Phosphatases • Oxidoreductases • Transferases • Proteases • Nuclear Hormone Receptors • G-protein-coupled receptors, 7-transmembrane receptors (GPCRs) • Ion Channels (some) • Signaling Pathways (some)	• Ion Channels • Transporters • Transmembrane Receptors • Signaling Pathways • Protein-Protein Interactions • Protein-DNA Interactions • Protein-RNA Interactions • DNA/RNA

Table 1. Established and potential future target classes for HTS

HTS not only helps in drug discovery but it is also important in improving present drug moieties to optimize their activity. In past years many advances in science and technology and economic pressures have kept every researcher to develop speedy and precise drug discovery and screening technologies to tackle the ever increasing diseases and the many pathogens acquiring resistance to currently available drugs. This also applies to screening the ever increasing compound libraries waiting to be screened due to increase in the parallel and combinatorial chemical synthesis. Research is also carried out so to cut the drug development costs, so that industries keep abreast with ever increasing competition.

2.1 Screenings of libraries

The greatest potential of combinatorial chemistry is represented by the number and variety of screenable compounds. Major efforts of researchers in the last decades have been focused on the development of methodologies to further increase molecular diversity. An interesting approach is based on general reversible reactions that produce "dynamic mixtures": in these, reactants and products are present in thermodynamic equilibrium. The assayed biological system selects the best binding structures among the different mixture components, thus producing a shifts of the reaction equilibrium by subtracting the product. Importantly, dynamic libraries do not need a deconvolution step and ligands are selected directly in the reaction mixture. Although innovative and promising, dynamic combinatorial libraries have been so far limited only to a small number of reversible reactions and libraries of moderate size. Similarly, the methodology named "libraries from libraries" has represented an innovation compared to the traditional concepts of combinatorial chemistry. By this approach, combinatorial libraries of peptides are built on a solid phase and are subsequently modified in order to maximize the chemical diversity. Oxidations, reductions, alkylations and acylations can be performed, exponentially increasing the number of new compounds. Once synthesized, libraries are employed in screening processes to determine active components for a given target. The choice of the assay is of utmost importance to succeed with a screening program. Indeed, different assays can be chosen depending on whether binders for an unspecified site are searched for, or ligands with predetermined properties are needed. Competition assays ensure the selection of active molecules with a specificity for the interacting interface of the proteins employed as targets in the screening. Binding assays can also be performed "on bead", in this case the libraries have been prepared with the Mix and Split method, the use of bead bound molecules has the advantage of having a local high concentration of ligands (several picomoles on a very limited surface area). The assay principle is simple and is based on the interaction between molecules and a labelled target-any natural or artificial receptor, enzymes, antibodies, nucleic acids or even small molecules in solution. When the labelled target binds the bound molecule, the bead will also be labelled and therefore can be visualized/detected by chromogenic methods and using micromanipulators. The labeled bead can be isolated and the molecule/peptide microsequenced. Though the on-bead assay is faster and easier to perform homogeneous phase assays in solution it can provide more specific functional evaluations. However highly charged or very hydrophobic molecules, when locally highly concentrated, can result in a high number of false-positive hits. Partially releasable libraries, based on light sensitive chemical linkers have also been described. These linkers, being very resistant to cleavage under acidic conditions, allow the complete removal of amino acid side chain protections, while they can be easily cleaved by irradiation at a

defined wavelength. By tuning light intensity and duration, small amounts of peptides can be released in solutions, while the molecules bound to the bead can be microsequenced to reveal the molecule identity. The deconvolution of libraries bound to solid supports can be performed following different approaches. Peptide arrays can be distinguished on the basis of their preparations: some ones involve the immobilization of pre-synthesized peptide derivatives and others the *in situ* synthesis directly on the array surface. The pre-synthesis approach requires a chemoselective immobilization on solid supports that can provide a useful method for controlling the orientation and the density of the immobilized peptides.

A critical factor for the screening of compound arrays is the accessibility to the immobilized molecules by the target proteins. Several spacers between peptides and chip surface have been proposed, such as 11 mercaptoundecanoic acid, hydrophilic polyethylene glycol chain, dextran, bovine serum albumin or human leptin and water-compatible supramolecular hydrogels.

In situ parallel synthesis methods can provide cheaper and faster miniaturized spatially addressed peptide arrays. Two general approaches have been introduced for the preparation of peptide arrays on solid supports: the photolithographic and the SPOT synthesis. The photolithographic approach consisted in the translation of the solid phase method for peptide synthesis into the preparation of supported peptide arrays by mean of photolabile protecting groups that allow the synthesis to proceed only on defined surface spots that are illuminated by light at a given wavelength. In the SPOT synthesis, the peptides arrays are synthesized in a stepwise manner on a flat solid support, such as functionalized cellulose membrane, polypropylene and glass, following the standard Fmoc-based peptide chemistry. Each spot is thus considered as an independent microreactor: thereby the selection of the solid support is a key factor. The support has to meet the chemical and the biological requirements of the target and it determines the synthesis and screening methods. It also dictates the functionalization type and the insertion of spacers and/or linkers. Described supports include ester-derivatized planar supports, CAPE [celluloseamino-hydroxypropyl ether] membranes, amino-functionalized polypropylene membranes and glass surface. Libraries can thus be seen as a source of bioactive molecules that are selected on the basis of the biochemical properties pre-defined by appropriate assay settings. The more diverse will be the library, the higher will be the probability to select good "hits". The concept of diversity is therefore of utmost importance in choosing a library for a particular screening and good libraries must first fulfill the requirement of highest diversity instead of the highest complexity. Molecules with overlapping structures will contribute little or nothing at all to the overall probability to find out a positive hit. The diversity of a library is generally associated to its complexity, that in turn depends on the number of different components. However this is not always true: small but smart libraries can display a higher diversity than libraries with a huge number of components. Indeed the synthesis of random combinatorial libraries of peptides generates a large number of "quasi-duplicates" deriving from the strong similarity between several side chains. Using common amino acids, in L- or D- configuration, sequences where Glu is replaced by Asp, Leu by Ile or Val, Gln by Asn and so on, can display very similar properties. Such residues, although being different in their propensity to adopt secondary structures, can be considered almost equivalent in terms of intrinsic physico-chemical properties, as for example the capacity to establish external interactions or to fit in a given recognition site. In large libraries, the need to manage large arrays of tubes and codes can puzzle the way to the identification of lead

compounds and slows down the synthetic and deconvolution steps. To simplify the synthesis and deconvolution procedures without significantly affecting the probability to find out active peptides, we have reasoned on the general properties of L and D-amino acids, reaching a compromise between the need to maintain the highest possible diversity and that of reducing the number of building blocks. We have called these new libraries "Simplified Libraries", intending with this any new ensemble of possible sequences achievable with a reduced and non redundant set of amino acids. The set of "non redundant" different residues (12 instead of 20) is chosen following several simple rules. Firstly "quasi-identical" groups of amino acids are selected: Asp and Glu; Asn and Gln; Ile and Val; Met and Leu; Arg and Lys; Gly and Ala; Tyr and Trp; Ser and Thr. Then, from each group, one is selected trying to keep a good distribution of properties, including hydropathic properties, charges, pKa, aromaticity. The cysteine is converted to the stable acetamidomethyl-derivative to prevent polymerization and to increase the number of polar residues within the final set. The distribution of residue molecular weights is also an important parameter in cases of deconvolution of the library by mass spectrometry approaches. The molecular weights do not overlap, allowing unambiguous assignment of sequences by tandem mass spectrometry. Using this set of amino acids in the L- or D-configuration, several libraries have been designed and synthesized (Marasco et al, 2008).

2.2 Automated libraries characterization

A key element for a successful screening of peptide libraries is the preparation of good quality libraries. Unclear screening results are often ascribed to the random generation of large populations of impurities, therefore even for very large mixture libraries, gross analytical characterizations must be performed to assess the relative amounts and the distribution of experimental molecular weight of library components. While classical analytical methods, as for example LC-MS, can be in most cases enough for the characterization of single compounds, more sophisticated techniques are required for the analysis of complex mixtures. The recent progresses in parallel separation methods, such as orthogonal chromatography, associated to synergic combinations of different detectors of complementary selectivities, have allowed the development of high-throughput analytical technologies based on traditional methods as HPLC, capillary electrophoresis (CE), NMR, FTIR, LC-MS, evaporative light scattering (ELSD), chemiluminescent nitrogen (CLND). A list of useful techniques is reported in **Table 2**.

Mass spectrometry (MS) is the best method to assess compound identity and purity due to its sensitivity, speed and specificity. Electrospray ionization (ESI) coupled to quadrupole analyzers or Matrix Assisted Laser Desorption Ionization Time-Of-Flight (MALDI-TOF) are useful for peptide library analysis. LC-MS systems using ESI and different analyzers (single or triple quadrupoles, ion traps, TOF), offer the advantage of chromatographic separation prior to on-line mass analysis, thus enormously simplifying data interpretation. Single quadrupole instruments typically provide unit mass resolution and therefore do not resolve isobaric ions (i.e. two compounds with the same nominal mass but different exact mass), which are often encountered in combinatorial libraries. LC-MS systems equipped with ESI and TOF analyzers (ESI-TOF) or with combined quadrupole-TOF (Q-TOF) analyzers provide increased mass accuracies up to 5 ppm and great versatility for tandem mass analysis. Limitations of MS include the inability to distinguish isomers and, with the increase of potential molecular formulae, differences in masses are too close to measurement error.

Detection mode	Advantages	Disadvantages
UV-absorbance spectroscopy	• High sensibility and response linearity • Easy on-line integration	• Relative purity measurements • Compounds have to bear chromophores • Responses depende on extinction coefficients
Evaporative light scattering	• Response factors of diverse compounds • Easy on-line integration	• Relative purity measurements • Limited for compounds with MW>300 amu • Right response depends on geometry molecule
Chemiluminescent nitrogen	• Universal response to number of nitrogen structure • Quantitative purity (if used with an internal standards) • Easy on-line integration	• Compounds must contain nitrogen
MS	• Excellent sensibility • Exact mass identification • Easy on-line integration	• Response factors depend on compound ionizability and matrix effects • Limited resolution with compounds with a high MW
NMR	• Universal response indipendent from structure • Quantitative purity (if used with an internal standards) • Detailed structure information	• Complex data interpretation • UV-absorbance Peak assignement difficult in presence of impurities • Difficult on-line integration

Table 2. Rivelator Method for HTS Characterization Of Compound Libraries

In addition to standard HPLC columns systems for interfacing to MS, several alternative separation techniques or platforms have been investigated for high-throughput purity analysis. An interesting application regards imaging time-of-flight secondary ion mass spectrometry (TOF-SIMS) to perform high-throughput analysis of solid phase synthesized combinatorial libraries. Also NMR techniques have been extensively utilized for the

identification and quantification of combinatorial compounds. NMR has the capability of elucidating compound structure and it is a quantitative detector whose response factor is directly proportional to the number of nuclei for a given signal, independent of molecular structure. Recently advances in NMR probe design have enabled sampling from smaller sample volumes and have significantly improved sample throughput overcoming some limitations of traditional NMR for combinatorial analysis included its low sensitivity, low sample throughput and complexity of data interpretation. The characterization of large mixture libraries has also been performed by pool amino acid analysis and by pool N-terminal sequencing techniques. Such techniques, though not providing single compounds purity and identity, can offer a rough but useful indication of equimolarity and representativeness of components.

2.3 Assays preparation

Automation is an important element in HTS's usefulness. Typically, an integrated robot system consisting of one or more robots transports assay-microplates from one station to another for sample and reagent addition, mixing, incubation and finally readout or detection. A HTS system can usually prepare, incubate, and analyze many plates simultaneously, further speeding the data-collection process. Currently HTS robots are able to test up to 100,000 compounds per day. The key labware or testing vessel of HTS is the microtiter plate: a small container, usually disposable and made of plastic, that features a grid of small, open divots called wells. Modern microplates for HTS generally have either 384, 1536, or 3456 wells. These are all multiples of 96, reflecting the original 96 well microplate with 8 x 12 9mm spaced wells. Most of the wells contain experimentally useful matter, often an aqueous solution of dimethyl sulfoxide (DMSO) and some other chemical compounds, the latter of which is different for each well across the plate. The other wells may be empty, intended for use as optional experimental controls. A screening laboratory typically holds a library of stock plates, whose contents are carefully catalogued, and each of which may have been created by the same lab or obtained from a commercial source. These stock plates themselves are not directly used in experiments; instead, separate assay plates are created as needed. An assay plate is simply a copy of a stock plate, created by pipetteing a small amount of liquid (often measured in nanoliters) from the wells of a stock plate to the corresponding wells of a completely empty plate. To prepare an assay each well of the plate is filled with some quantity of a protein, or an animal embryo. After some incubation time has passed to allow the biological matter to absorb, bind to, or otherwise react (or fail to react) with the compounds in the wells, measurements are taken across all the plate's wells, either manually or by a machine. Manual measurements are often necessary when the researcher is using microscopy to (for example) seek changes or defects in embryonic development caused by the wells compounds, looking for effects that a computer could not easily determine by itself. Otherwise, a specialized automated analysis machine can run a number of experiments on the wells. In this case, the machine outputs the result of each experiment as a grid of numeric values, with each number mapping to the value obtained from a single well. A high-capacity analysis machine can measure dozens of plates in the space of a few minutes like this, generating thousands of experimental datapoints very quickly. Depending on the results of this first assay, the researcher can perform follow up assays within the same screen by "cherrypicking" liquid from the source wells that gave interesting results into new assay plates, and then re-running the experiment to collect

further data on this narrowed set, confirming and refining observations. Problems associated with screening robotics have included long design and implementation time, long manual to automated method transfer time, non-stable robotic operation and limited error recovery abilities. These problems can be attributed to robot integration architectures, poor software design and robot–workstation compatibility issues (e.g., microplate readers and liquid handlers). Traditionally, these integrated robot architectures have involved multiple layered computers, different operating systems, a single central robot servicing all peripheral devices, and the necessity of complex scheduling software to coordinate all of the above. Usually robot-centric HTS systems have a central robot with a gripper that can pick and place microplates around a platform. They typically process between 40 and 100 microplates in a single run (the duration of the run depends on the assay type). The screener loads the robotic platform with microplates and reagents at the beginning of the experiment and the assay is then processed unattended. Robotic HTS systems often have humidified CO_2 incubators and are enclosed for tissue culture work. Like in assembly-line manufacturing, microplates are passed down a line in serial fashion to consecutive processing modules. Each module has its own simple pick and place robotic arm (to pass plates to the next module) and microplate processing device. The trend towards assay miniaturization arose simultaneously with move towards automation from the direct need to reduce development cost. Although at present most HTS is still carried out in 96-well plate format, the move towards 384-well and higher density plate formats is well taking place. Instrumentation for accurate, low-volume dispensing into 384-well plates is commercially available, so they are sensitive plate-readers that accommodate this format. The combination of liquid handling automation and information management processes supports the entire compound management cycle from compound submission to delivery of assay ready compound plates. The compound management often consists in the organization of hundreds of thousands of compounds (small molecules, natural products and peptide libraries in different formats). Compounds are stored offline in an automated compound store in 96-well tube format and online in 384-well format for hitpicking operations, 1536-well for HTS operations. Often there are dedicated LC-MS systems with multimode mass spectrometry (ESI, APCI, ELSD) for quality control of incoming HTS compound libraries and hit confirmation efforts. This platform is designed for automated analysis and generates database-ready reports.

2.4 Screening assays

Assays are either heterogeneous or homogeneous. Heterogeneous assays are a bit complex, requiring additional steps like filtration, centrifugation etc. beyond the usual steps like fluid addition, incubation and reading. Homogeneous assays are simpler, consisting of the latter three usual steps - this may also be called a true homogeneous assay. However at times homogeneous assays could be complex due to the need for multiple addition and different incubation times. Though homogeneous assays are advantageous, many companies prefer to continue to use heterogeneous assay, eyeing their better precision over its counterpart, though it is true only in few number of cases. The driving force for use of homogeneous assays is the lower number of steps, which will help reduce assay cost. This simplicity may also reduce the robotic complexity requirement for automation. When performing HTS of free compounds in solution, automation, miniaturization and very sensitive detection methodologies are required. Different approaches such as ELISA, cell-based cytotoxic,

antimicrobial, radiometric and fluorescence-based assays, affinity chromatography methodologies can be used. Fluorescence-based techniques are likely to be among the most important detection approaches used for HTS due to their high sensitivity and amenability to automation, given the industry-wide drive to simplify, miniaturize, and speed up assays. Fluorescence resonance energy transfer (FRET), fluorescence polarization (FP), and fluorescence correlation spectroscopy (FCS) are indeed already broadly utilized for screenings of large collections of compounds and many optical readers capable of handling multi-well plates are commercially available. FRET is based on energy transfer between appropriate energy donor and acceptor molecules. It is typically used in protein-protein interaction studies where one protein partner is labelled with a fluorescence donor and the other one is labelled with an appropriate fluorescence acceptor molecule. The donor has the specific property of being excitable, emitting a fluorescence photon at a wavelength that is able to in turn excite the acceptor. When the two proteins interact and the system is excited at the donor excitation wavelength, a specific fluorescence emission at the acceptor emission wavelength is recorded. The energy transfer occurs only when the donor-acceptor pair is within a minimum distance, the Förster distance (the distance at which energy transfer efficiency is half-maximal), which is around 50 Å. When performing a screening assay, if a library component is able to disrupt the protein-protein interaction, the effect can be quantitatively measured by a reduction of the FRET effect. Similar assays can be performed using the Time Resolved FRET (TR-FRET) whereby the fluorescence of the acceptor molecule has a duration on the milliseconds time scale, allowing fluorescence measurement after a short time delay to remove interference by the excitation energy or by inhibitors. TR-FRET can be integrated on large time intervals to increase sensitivity. Typical donor-acceptor molecules are the Allophycocyanin-Europium chelates or the fluorescein - Terbium chelates. FRET can also be used to determine enzyme activity using internally quenched probes. In these systems short peptides corresponding to the sequence for a natural cleavage site of the enzyme is synthesized and labelled at opposite ends with appropriate donor and quencher pairs. Before cleavage, donor and quencher are very close and the effective fluorescence emitted is low; once the two parts drift apart, the fluorescent signal increases. This approach has been successfully utilized to screen the substrate specificity of an alkaline serine proteinase. FP experiments allow measurements of changes in the emitted light intensity of small labelled probes on binding to larger molecules. The sample is excited with polarized light and, when a binding equilibrium is established, the observed polarization of the emitted light increases. In FCS the main detected parameter is the spontaneous intensity fluctuations caused by the minute deviations of the small system from the thermal equilibrium. FCS is an emerging technique for HTS. In this case, measurements are carried out using confocal optics to provide the highly focused excitation light. It is used to monitor binding interactions as well as other molecular events. At times it was reported that assays for biological targets cannot be conveniently designed to fit with standard cellular or biochemical assay formats. For example, in the search for new antibacterial agents, genomic experiments have indicated a large number of proteins that are essential for the survival of the bacterium but whose function in the cell is unknown. In this case there is no known biological function that will allow the design a biochemical or cellular screen. To screen these types of target, an alternative to conventional chemical or cellular screenings may be used. One alternative screening approach that does not require knowledge or analysis of the biological function of the target of choice is direct measurement of compound interaction with protein. A range of techniques are available to measure the direct binding events such

as NMR, SPR and calorimetry. One advantage afforded by NMR is that it can provide direct information on the affinity of the screening compounds and the binding location of protein. The structure-activity relationship acquired from NMR analysis can sharpen the library design, which will be very important for the design of HTS experiments with well-defined drug candidates. Affinity chromatography used for library screening will provide information on the fundamental processes of drug action, such as absorption, distribution, excretion, and receptor activation; also the eluting curve can give directly the possibility of candidate drug. SPR can measure the quantity of a complex formed between two molecules in real-time without the need for fluorescent or radioisotopic labels. SPR is capable of characterizing unmodified biopharmaceuticals, studying the interaction of drug candidates with macromolecular targets, and identifying binding partners during ligand fishing experiments. They indeed allow real-time detection of soluble peptide binding to different biomolecules, like proteins, nucleic acids or sugars and also to cell membranes or whole cells. In addition, they allow the one-step measurement of kinetic rates and affinity of binding, thus providing an affinity ranking of molecules peptides.

2.5 Data analysis

In validating a typical HTS assay, unknown samples are assayed with reference controls. The sample signal refers to the measured signal for a given test compound. The negative control (usually referred to as background) refers to set of individual assays from control wells that give minimum signals. The positive control refers to the set of individual assay from control wells that give maximum signals. In validating the assay, it is critical to run several assay plates containing positive and negative control in order to assess reproducibility and signal variation at two extremes of the activity range. The positive and negative control data can then be used to calculate their means and standard deviations (SD). The difference between the mean of the positive controls and the mean of the negative controls defines the dynamic range of the assay signal. The variation in signal measurement for samples, positive control, and negative controls (i.e., SDs) may be different. The mean and SD of all the test samples are largely governed by the assay method and also by intrinsic properties of the compound library. Because the vast majority of compounds from an unbiased library have very low or no biological activity, the mean and SD of all the sample signals should be close to those of the positive controls for inhibition/antagonist type assays and near to those of the negative controls for activation/agonist types assays.

The Z-factor is a measure of statistical effect. It was proposed for use in HTS to judge whether the response in a particular assay is large enough to warrant further attention. The Z-factor is defined by four parameters: the means (μ) and standard deviations (σ) of both the positive (p) and negative (n) controls (μ_p, σ_p, and μ_n, σ_n). Given these values, the Z-factor is defined as:

$$\text{Z-factor} = 1 - \frac{3 \times \left(\sigma_p + \sigma_n \right)}{\left| \mu_p - \mu_n \right|}$$

Generally HTS suffers from two types of errors false positives and negatives. A poor candidate or an artifact gives an anomalously high signal, exceeding an established threshold. While a perfectly good candidate compound is not flagged as a hit, because it

gives an anomalously low signal. Moreover, a low degree of relevance of the test may induce a high failure rate of type. Much more attention is given to false-positive results than to false-negative results. Some of the false positives are promiscuous compounds that act non competitively and show little relationship between structure and function.

In HTS each biochemical experiment in a single well is analyzed by an automated device, typically a plate reader or other kind of detectors. The output of these instruments comes in different formats depending on the type of reader. Sometimes multiple readings are necessary, and the instrument itself may perform some initial calculation. These heterogeneous types of raw data are automatically fed into the data management software. In the next step raw data are translated in contextual information by calculating results. Data on percentage inhibition or percentage of control are normalized with values obtained from the high and low controls present in each plate. The values obtained depend on the method used for the normalization step (e.g. fitting algorithms used for dose-response curve) and have to be standardized for screens. All the plates that fail against one or more quality criteria are flagged and discarded. A final step in the process requires the experimenter to monitor visually the data that have been flagged, as a final quality check. This is a fundamental step to ensure the system has performed correctly. In addition to registering the test data, all relevant information about the assay has to be logged, e.g. the supplier of reagents, storage conditions, a detailed protocol, plate layout, and algorithms for the calculation of results. Each assay run is registered and its performance documented. HTS will initially deliver hits in targeted assays. Retrieval of these data has to be simple and the data must be exchangeable between different project teams to generate knowledge from the total mass of data.

3. Virtual screening

Even with HTS, the discovery of new lead compounds largely remains a matter of trial and error. Although the number of compounds that can be evaluated by HTS methods is apparently large, these numbers are small in comparison to the astronomical number of possible molecular structures that might represent potential drug-like molecules. Often, far more compounds exist or can be synthesized by combinatorial methods than can be reasonably and affordably evaluated by HTS. As the costs of computing decreases and as computational speeds increase, many researchers have directed efforts to develop computational methods to perform "virtual screens" of compounds. Thus since performing screens *in silico* can be faster and less expensive than HTS methods, virtual screening methods may provide the key to limit the number of compounds to be evaluated by HTS to a subset of molecules that are more likely to yield "hits" when screened. For the practical advantages of virtual screening to be realized, computational methods must excel in speed, cheapness, and accuracy. Striking the right balance of these criteria with existing tools presents a formidable challenge. An inspiring example study is related to the structure-based virtual screening applied to the selection of new thyroid hormone receptor antagonists when only a related receptor structure is available. Receptor-based virtual screening uses knowledge of the target protein structure to select candidate compounds with which it is likely to favorably interact (Schapira et al., 2003). Even when the structure of the target molecule is known, the ability to design a molecule to bind, inhibit, or activate a biomolecular target remains a daunting challenge. Although the fundamental goals of

screening methods are to identify those molecules with the proper complement of shape, hydrogen bonding, electrostatic and hydrophobic interactions for the target receptor, the complexity of the problem is in reality far greater. For example, the ligand and the receptor may exist in a different set of conformations when in free solution than when bound. The entropy of the unassociated ligand and receptor is generally higher than that of the complexes, and favorable interactions with water are lost upon binding. These energetic costs of association must be offset by the gain of favorable intermolecular protein–ligand interactions. The magnitude of the energetic costs and gains is typically much larger than their difference, and, therefore, potency is extremely difficult to predict even when relative errors are small. While several methods have been developed to more accurately predict the strength of molecular association events by accounting for entropic and solvation effects, these methods are costly in terms of computational time and are inappropriate for the virtual screening of large compound databases. The challenge in developing practical virtual screening methods is to develop an algorithm that is fast enough to rapidly evaluate potentially millions of compounds while maintaining sufficient accuracy to successfully identify a subset of compounds that is significantly enriched in hits. Accordingly, structure-based screening methods typically use a minimalist ''grid'' representation of the receptor properties and an empirical or semiempirically derived scoring function to estimate the potency of the bound complex (Schulz-Gasch & Stahl, 2003). Several programs now employ a range of scoring functions, but it is often difficult to assess their effectiveness on difficult ''real-world'' problems. Virtual screening based on receptor structure therefore has the distinct advantage of aiding the discovery of new antagonist structural classes or pharmacophores. First, and most importantly, from the many case studies published over the past decade, it has become evident that the applicability and thus the usefulness of a particular virtual screening method for a given drug discovery project depends on the macromolecular target being investigated. Thus it seems more appropriate to consider virtual screening from a problem-centric rather than a method centric perspective. Depending on what is already known about a target and its ligands, different approaches to virtual screening — and consequently different sets of methods — are preferred. In addition, some virtual screening methods that have been reported might be premature or simply not sufficiently accurate. Second, the perceived success or failure of virtual screening in a particular organization depends on the depth and mode of integration of virtual screening in the organization's hit identification process, and whether expectations are realistic. For example, an important factor for the chances of success of virtual screening could be whether hits with interesting characteristics, such as structural novelty and/or patentability, can be sufficiently nurtured by medicinal chemists to produce leads that can compete with those arising from HTS. Indeed, this is typically the scenario for hits arising from alternative lead discovery approaches such as fragment-based screening. Third, a major achievement of virtual screening so far has been to help eliminate the bulk of inactive compounds (negative design), rather than to actually select bioactive molecules for a given target (positive design). Although this statement is a simplification it highlights the extent of the challenges for developing improved virtual screening methods.

It will be important to systematically determine which ligand–receptor interactions are amenable to such an approach and which require other or additional features to be considered indeed, dynamic descriptions of molecules will have to replace our

predominantly static view of both targets and ligands. Molecular dynamics simulations can sample conformational ensembles of targets and ligands. However, some of the popular force-field approaches used to describe the energetics of molecular systems might be inadequate for drug design. Furthermore, although in general it might be more valuable to identify ligand chemotypes for which receptor–ligand complex formation is dominated by enthalpy changes rather than entropy changes, improvements are required to allow for a more accurate estimation of both enthalpic and entropic contributions (Freire, 2008). The thermodynamics of ligand–receptor interactions are commonly treated in a similar way as molecular reactions, and this may not always be appropriate. How can we reliably and efficiently predict that some protein–ligand interactions become stronger with increasing temperature, or identify the role of buried water on ligand binding? Questions like these must be answered by computational chemistry as the forces that govern ligand–receptor interactions are only understood at a rudimentary level: flexible fit phenomena, the role of water molecules, protonation states in proteinaceous environments, and the entropic and enthalpic contributions and compensations upon complex formation are not satisfactorily addressed by the existing virtual screening methods. In this respect, advanced and specialized computer hardware might enable extended (>100 ns) dynamics simulations of macromolecular targets and receptor–ligand complexes on a routine basis, which might help our understanding of allosteric effects and flexible fit phenomena of druglike ligands and effector molecules. As a consequence, modelling of dynamic molecular features (in contrast to static properties such as molecular mass, logP and other time-invariant molecular properties), which cannot be accurately achieved at present, could improve the accuracy of future predictions of novel bioactive compounds.

4. Fragment screening

Fragment-based drug discovery has proved too to be a very useful approach particularly in the hit-to-lead process, acting as a complementary tool to traditional HTS. Over the last ten years, fragment-based drug discovery has provided in excess of 50 examples of small molecule hits that have been successfully advanced to leads and therefore resulted in useful substrate for drug discovery programs. The unique feature of fragment-based drug discovery is the low molecular weight of the hit. It has the potential to supersede traditional HTS based drug discovery for molecular targets amenable to structure determination. This is because the chemical diversity coverage is better accomplished by a fragment collection of reasonable size than by larger HTS collections. Fragments represent smaller, less complex, molecules than either drug compounds or typical lead series compounds. It is now widely acknowledged within the pharmaceutical and biotech industries that weakly active fragment hit molecules can be efficiently optimised into lead compound series if structural insight is obtained at the outset for the binding interaction between each fragment hit and the target protein of interest. This is supported by recent reports of the progression into human clinical trials of drug molecules developed from weakly active fragment starting points (Jhoti et al., 2007). Fragment-based drug discovery can explore the drug-like chemical diversity space in an efficient and effective manner. Two key factors govern this approach, firstly the coverage of fragment chemical diversity space during the screening stage and then, that drug chemical diversity space is explored in an efficient iterative fashion during the optimisation stage as fewer combinations need to be evaluated than through a purely

random screening and undirected optimisation approach. For example, a fragment collection of 10 000 molecules may virtually represent the diversity of one billion molecules if one considers the combinatorial power of fragment merging or linking (e.g. by assuming two adjacent binding sites to which fragments bind and 10 different possibilities of fragment linking) but only a small part of the larger chemical space defined by fragment merging and linking needs to be explored in the structure- directed elaboration of fragments into leads. Employing fragment-based drug discovery a relatively small number of low molecular weight fragment molecules can provide a higher degree of sampling of the chemical diversity space for fragments than a very large number of higher molecular weight compounds is able to sample the respective chemical diversity space for drug-like compounds. Furthermore lower molecular weight molecules exhibit reduced complexity than the larger molecules in drug-like collections and it can be hypothesized a model to rationalise ligand–receptor interactions in the molecular recognition process. Accordingly the theoretical probability of a useful interaction falls dramatically with increasing molecular complexity of the ligand. The selection of the fragment screening method is of key importance since there are two general factors that have an impact for all methods. The first is sensitivity of the screening method and the second is throughput. Sensitive screening methods enable weakly active fragment molecules of lower molecular weights to be identified as hit compounds and so fragment libraries with a lower molecular weight range can be used. On the contrary the use of a low throughput screening method necessitates the use of a smaller fragment library with a concomitant sparser coverage of fragment chemical diversity space. Perhaps the most elegant method of fragment screening is by X-ray crystallography, in that it provides directly structural information on the interaction between fragment ligands and the protein target. However, owing to the method's low throughput, even when fully automated, the technique can only be effectively applied to targets for which a robust crystallographic system is available that allows soaking of preformed crystals with fragment cocktail mixtures of up to 10 compounds at high concentrations. This requirement imposes two key limitations: in the number of evaluable fragment compounds, typically limited to no more than 1000 fragments, and for this, there is a significant possibility of missing active fragments owing to the protein being locked into a conformation, in the crystals used for the soaking studies, that does not allow the interaction of fragments that require induced fit to bind. Although, no data are available on the false negative rate for fragment screening by X-ray crystallography it may be significant for certain targets. Each fragment screening technique has its advantages; X-ray crystallography provides immediate structural information, NMR provides binding site and affinity information of a very high quality while bioassays provide functionally relevant activity data for larger collections of fragments. The best approach is to combine the methods in order to maximize their value to fragment-based drug discovery. NMR and biochemical screening of fragments are complementary orthogonal methods that can be used individually or in concert to provide the most effective way of addressing each new biological target of interest. The strength of biochemical screening is that its throughput allows large fragment collections to be screened in a short length of time. This ensures that the most ligand efficient diverse starting points are available for medicinal chemists to select for subsequent optimisation. A further advantage is that screening related targets using generic biochemical assay formats enables insights into target selectivity from the outset. The large number of fragment hits that are obtained through use of biochemical screening of

large fragment libraries can be effectively triaged ahead of crystallography by the use of protein NMR. Thus, the most effective way to perform fragment screening is not to rely on a single method but to use orthogonal methods in concert.

5. Conclusion

HTS is the most widely applicable technology delivering chemistry entry points for drug discovery programs, however it is well recognized that even when compounds are identified from HTS they are not always suitable for further medicinal chemistry exploration. It is evident that in the future the overwhelming number of emerging target will dramatically increase the demand put on HTS and that this will call for new hit and lead generation strategies to curb costs and enhance efficiency. The collections of large pharmaceutical companies are approaching approximately one million entities, which represents historical collections (intermediates and precursors from earlier medicinal or agrochemical research programs), natural products and combinatorial chemistry libraries. This about one order of magnitude higher than ten years ago when HTS and combinatorial chemistry first emerged. However today purchasing efforts in many pharmaceutical companies are directed towards constantly improving and diversifying the compounds collections and making them globally available for random HTS campaigns. The combinatorial explosion- meaning the virtually number of compounds that are synthetically tractable-has fascinated and challenged chemists ever since the inception of the concept. Independent of the library designs, the question of which compounds should be made from the huge pool of possibilities always emerges immediately, once the chemistry is established and the relevant building blocks are identified. The original concept of "synthesize and test", without considering the targets being screened, was frequently questioned by the medicinal chemistry community and is nowadays considered of lower interest due to the unsatisfactory hit rates obtained so far. As a consequence there is now a clear trend to move away from huge and diverse "random" combinatorial libraries towards smaller and focused drug-like subsets. Hit and lead generation are key processes involved in the creation of successful new medicinal entities and it is the quality of information content imparted through their exploration and refinement that largely determines their fate in the later stages of clinical development. The combination of virtual screening and parallel and medicinal chemistry, in conjuction with multi-dimensional compound-property optimization, will generate a much-improved basis for proper and timely decisions about which lead series to pursue further.

6. References

Freire, E. (2008). Do enthalpy and entropy distinguish first in class from best in class?. Drug discovery today 13(19-20): 869-874.

Houghten, R.A., Wilson, D. B, & Pinilla, C. (2000). Drug discovery and vaccine development using mixture-based synthetic combinatorial libraries. Drug discovery today 5(7): 276-285.

Jhoti, H., Cleasby, A., Verdonk. M., & Williams, G. (2007). Fragment-based screening using X-ray crystallography and NMR spectroscopy. Current opinion in chemical biology 11(5): 485-493.

Marasco, D., Perretta, G., Sabatella, M., & Ruvo, M. (2008). Past and future perspectives of synthetic peptide libraries. Curr Protein Pept Sci. 9(5):447-67.

Schapira, M., Raaka, B.M., Das, S., Fan, L., Totrov, M., Zhou, Z., Wilson, S.R., Abagyan, R., & Samuels H.H. (2003). Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. Proceedings of the National Academy of Sciences of the United States of America 100(12): 7354-7359.

Schulz-Gasch, T. & Stahl, M.. (2003). Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. Journal of molecular modeling 9(1): 47-57.

Shin, D.S., Kim, D.H., Chung, W.J., & Lee, Y.S. (2005). Combinatorial solid phase peptide synthesis and bioassays. Journal of biochemistry and molecular biology 38(5): 517-525.