Computational Chemistry as an Integral Component of Lead Generation

Olivier Roche* and Wolfgang Guba

F. Hoffmann-La Roche Ltd Pharmaceuticals Division CH-4070 Basel, Switzerland

Abstract: From library shaping to ADME-Tox prediction *via* virtual screening, computational chemistry is an integral component of Lead Generation. It provides a series of tools that help focusing on compounds with a balanced pharmacodynamic and ADME-Tox profile together with a high potential to optimize potency and selectivity.

Keywords: Lead Generation, virtual screening, ADME, library design, lead-likeness, frequent hitters, similarity, clustering, complexity.

INTRODUCTION

Pharmaceutical industry is facing a productivity gap which is reflected by a reduced number of drug applications to the regulatory authorities, an increase in development times and, concomitantly, sharply rising costs for development [1, 2]. To find a solution to this problem, Lead Generation groups have been established to thoroughly assess the optimization potential of hits in the early stage of the drug discovery process. Their mission is to deliver high-quality leads with balanced pharmacodynamics and Absorption Distribution Metabolism Elimination Toxicity (ADME-Tox) profiles. This goal can only be achieved by making knowledge-based decisions about advancing or dropping a candidate based on a multidisciplinary approach [3]. Fig. (**1**) describes the workflow of Lead Generation and focus on the hit identification phase. The paper details each step involved in this phase and emphasizes the role of Computational Chemistry for selecting the most promising compound classes.

Screening of chemical libraries is usually the starting point of the Hit Generation process. The quality of the libraries has a dramatic impact on the hits generated either by high-throughput screening (HTS) [4] or by virtual screening. A major endeavor has been initiated to clean the libraries by *in silico* filtering. These computational filters will be categorized into two classes, i.e. "hard filters" which remove compounds from further assessment and "soft filters" raising alerts for potential liabilities of molecules. The "hard filters" are directly related to the physicochemical properties. They are usually more reliable than "soft" filters, which model more complex phenomena like bioavailability, metabolism or promiscuous binding. Thus, the "hard" filters are used to shape libraries by removing undesired compounds [5], improving the physicochemical properties (permeability, solubility) and reducing the molecular complexity [6]. All together, these filters enhance the lead-likeness nature of the libraries [7]. Concomitantly, library partitioning into target-based subfamilies has been initiated in order to promote more efficient focused screens [8].

Depending on the initial knowledge available about a specific target, many different virtual screening techniques can be considered. Our focus will be on ligand-based design for early hit finding *via* similarity searching [9] and chemical feature-based pharmacophore filtering [10].

Once hits are validated, i.e. compounds have been active in at least two independent biological assays, the first task is to cluster them by chemotypes [11] and assess their chemical tractability [12]. Then, medicinal chemists need to identify the most promising chemotypes for further investigation. At this crucial decision point, "soft filters" are applied to prioritize the different classes. Since the "soft filters" are the results of quantitative structure activity or structure-property relationships (QSAR/QSPR), they are intrinsically less accurate. So, they are designed to raise alerts about possible pitfalls during the hits refinement process. The alerts can be as diverse as cytochrome P450 liability (CYP450) [13], hERG K^+ channel binding [14], drug-likeness [15, 16] or Frequent Hitters (FH) [17]. The challenge of "soft filters" is to provide, before any synthesis, a multi-dimensional overview of the optimization potential of each series so that the most balanced series in terms of pharmacodynamics and ADME-Tox are selected. Consequently, powerful visualization tools are needed to efficiently navigate in this multi-dimensional space [18].

HARD FILTERS

Hard filters are defined by deriving rules for property ranges from the analysis of molecules in databases. These rules are directly connected with the 2-dimensional molecular structure and can be modified to optimize the properties of libraries in the design process. Whereas the pioneering phase of combinatorial chemistry tended to maximize size and diversity, the current focus is on the improvement of the chemical and pharmacological properties of the screened compounds. Hard filters are commonly applied to reduce the number of false positive hits and to favor lead-like molecular properties.

Reactivity

At Roche, more than one hundred functionalities have been defined by a global team of experienced medicinal chemists. This expert knowledge has been encoded as a substructure matching tool using the SMARTS language and can be quickly applied to any new library [19].This list covers reactive and unstable compounds as well as covalent

*Address correspondence to this author at the F. Hoffmann-La Roche Ltd Pharmaceuticals Division CH-4070 Basel, Switzerland; Tel: +41-61 688 74 08; Fax: +41-61 688 90 26; E-mail: olivier.roche@roche.com

1389-5575/05 $50.00+.00

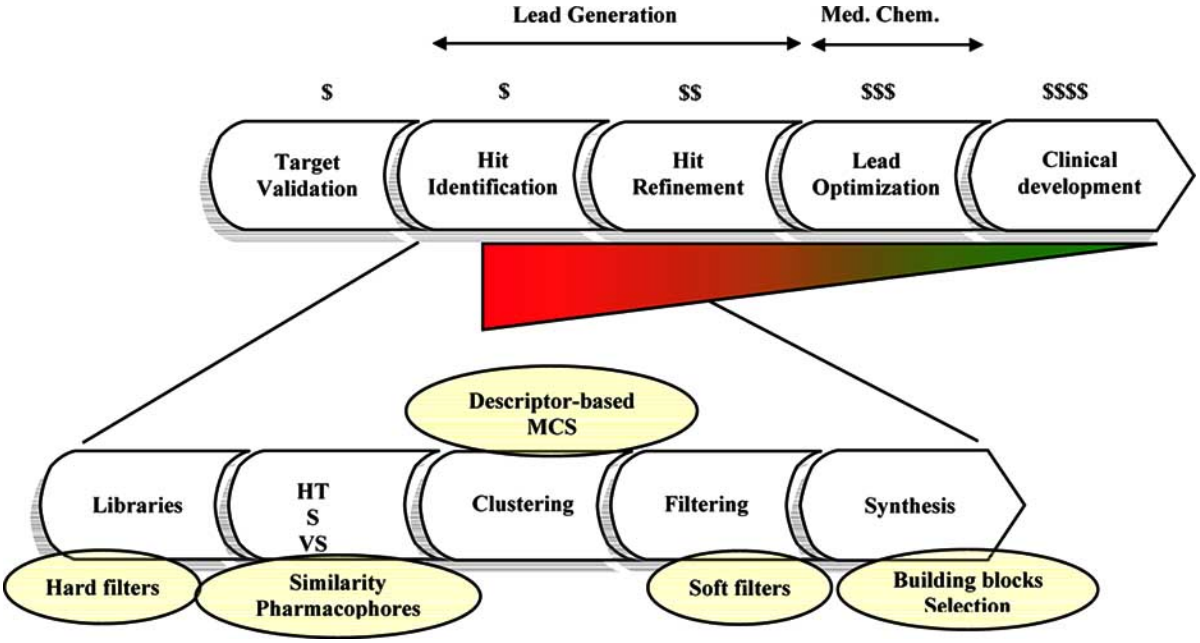

Fig. (1). The lead generation process with a focus on the hit identification phase. The yellow ellipses represent the contributions of computational chemistry. The shaded triangle indicates the compounds attrition rate in the value-adding chain.

binders [5]. It has been extended to unwanted features (e.g. poly-acids, alkyl aldehydes, poly-halogenated phenols...), which are chemically unattractive starting points for a hit to lead optimization and often produce non-optimizable SAR patterns. In addition, some particular chemotypes are also filtered like catechols, detergent or steroids. Nevertheless, depending on the project, the filtered chemotypes can be partially or entirely restored.

Physicochemical Properties

Since the aim of most therapeutics is oral bioavailability, permeability and solubility should be assessed as early as possible. In general, poor solubility is related to high lipophilicity, and hydrophilic compounds show poor permeability. The widely applied "Rule of 5" (Ro5) from Lipinski *et al*. [20, 21] predicts that poor absorption or permeation is more likely if two or more of the following criteria are met (exceptions: antibiotics, antifungals, substrates for transporters): > 5 H-bond donors (OH, NH), molecular weight > 500, clogP > 5, > 10 H-bond acceptors (sum of N and O). The analysis of particular drug data sets like central nervous system (CNS) drugs leads to additional parameters. The polar surface area (PSA) seems to play a pivotal role in passive diffusion through membranes: PSA < 140Å^2 for intestinal absorption and PSA < 80Å^2 for blood-brain barrier penetration [22]. The number of rotatable bonds and the number of rings are reflecting the flexibility of the molecule, which may also have an influence on oral bioavailability [23]. Besides this qualitative approach, QSAR models have also been developed for predicting human *in vivo* intestinal permeability. Independently of the methods, the permeability correlates with combinations of general hydrogen bonding (PSA), hydrogen bond donor (sum of HBD) and lipophilicity descriptors (clogP) [24].

Lipophilicity is a key parameter influencing membrane permeability, drug absorption, distribution and clearance. Several algorithms for logP calculations are available with each approach showing weaknesses for different structural classes. Thus, a multivariate profiling with logP prediction programs is recommended to compensate for errors. The distribution coefficient logD would be physiologically more relevant than logP, because the charge state at a given pH (e.g. blood pH 7.4, intestinal pH 6.5 etc.) is considered. However, the available pKa calculators are not equally well parametrized for all structural fragments and the error may be unacceptably high. The situation is even worse for solubility prediction, the effects of counter-ions and crystal forms are only taken into account indirectly by using the experimental melting point as an additional parameter, which very often is not available. At present there are no methods available to reliably predict the pharmaceutically relevant solubility range of up to 100 μg/ml.

Without accurate pKa calculations, the protonation state of the molecules and their charge cannot be assigned properly, which limits their use in QSAR or QSPR models.

Complexity

Having too many functionalities and too large molecules as starting points for chemistry reduces the opportunity for "hit refinement". This observation is supported by the theoretical work from M. Hann *et al*., who define the complexity of the ligand as the number of possible interaction sites (pharmacophores). They show that an increasing complexity of the ligands is correlated with a reduced chance to match the pharmacophoric pattern of the receptor [6]. The number of bits set in the Daylight fingerprint is proposed as a metric for molecular complexity. An alternative is suggested by Rücker *et al*., who use substructure, subgraph and walk counts for defining the complexity [25].

Lead-Likeness

Complexity and lead-likeness are intimately related. Indeed, the lead-likeness concept aims at valorized smaller and less lipophilic molecules [26], since the lead optimization phase generally increases the molecular weight and the lipophilicity of the initial compound [7]. Whereas the Ro5 has been derived from the analysis of fully optimized molecules, the physicochemical property ranges of a lead-like molecule are more stringent, leaving room for further improvements. An upper limit of 350 Dalton for molecular weight, a clogP range between 1 and 3 and the presence of maximum one charge (preferably a secondary or tertiary amine) are recommended. Moreover, the analysis of the structural difference between drugs and their lead reveals that they are closely related. This emphasizes the importance of having high quality leads [27].

Targeted Libraries

In parallel to the compounds quality improvement, a major effort has been recently initiated to partition libraries into smaller target family-directed sub-libraries. The goal is to prioritize a subset of compounds that are more likely to generate hits. The efficiency of the partitioning is dependent on the accuracy of the underlying models. The large amount of knowledge available about G-protein coupled receptors (GPCR) makes this family very attractive. On the one hand, Balakin *et al*. used a statistical classification method to discriminate between GPCR-ligands and non-GPCR ligands [8]. On the other hand, Müller has analyzed each family based on the privileged structures concept [28], which has been extensively reviewed by Horton *et al.* [29].

The application of "hard filters" based on physicochemical properties will bias libraries towards compounds more suitable for optimization after a screening campaign.

VIRTUAL SCREENING

Whereas HTS campaigns are limited by the development of suitable assays and by the costs, virtual screening campaigns fully depend on the reference compounds that are retrieved from literature or patents. In the context of ligand-based design, the amount of knowledge already available will drive the selection of a particular virtual screening technique. If only few ligands are known and no structure activity relationship (SAR) has been previously described, similarity searches are an interesting starting point.

Similarity

The approach is based on the concept that molecules with closely related structures will have similar biological activities. Since this assumption is not always valid, Martin *et al.* emphasize the need to have a small number of similar compounds in the libraries to confirm a potential hit but not too many as to compromise diversity [30]. In similarity searches, no assumption is made on the chemical features responsible for the binding.

However, there is a high level of noise since low and medium affinity compounds do not have a perfect match with the target, which is due either to missing functionalities or to parts of the molecule not engaged in binding. For performing a similarity analysis, molecules need to be characterized with numerical descriptors, such as calculated molecular properties or values derived from the two or three-dimensional structure of the molecules [31]. Since these descriptors are not fully correlated with the binding event, additional noise is introduced. Each combination of descriptors will provide a different description of the molecules leading to different ranking of compounds [32]. This is true as well for similarity metrics [33]. To overcome this problem, we have implemented a consensus scoring scheme to combine the results of various similarity analyses. In theory, consensus scoring outperforms any single scoring for a statistical reason: the mean value of repeated samplings tends to be closer to the true value [34]. Various methods exist to combine the results. The simplest approach would be to pick up molecules that are identified by at least two different similarity metrics. A more sophisticated one would be to use data fusion rules [35]. Data fusion is a process of combining inputs from various sources. Primarily created for sensors measurements, it has been applied to aircraft,

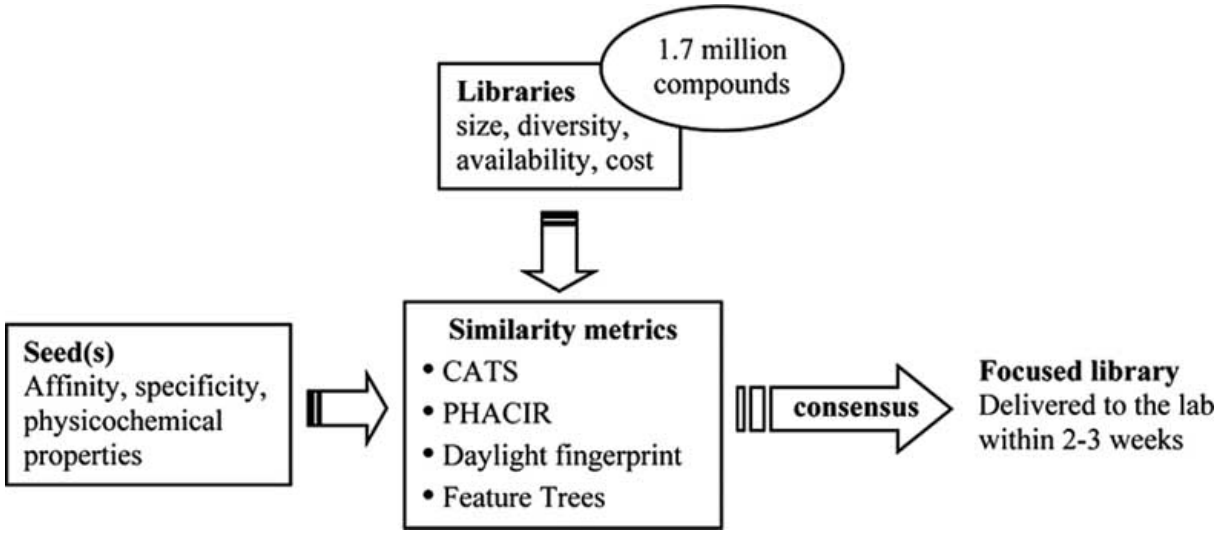

Fig. (2). The components of the consensus similarity screening.

medical imaging, internet searches, etc… Several fusion rules exist in data fusion. The more stable rule is the SUM, where the sum of all the rank positions is assigned to each object. In our case, each molecule will be ranked based on the sum of its rank in each similarity metric.

The consensus approach is very conservative, but in association with filtered proprietary and external vendors' libraries, it allows to perform quick SARs without synthetic effort. So far, at Roche, chemists have access to the in-house plated collection plus eight external vendors' libraries representing one million compounds that are available within 2-3 weeks. The consensus is performed through a web interface on the results of four similarity metrics: Daylight Fingerprint [19], Feature Trees [36], Cats [37] and Phacir [38] - see Fig. (**2**).

Pharmacophore-Based Database Mining

When more information is available about the relevant functionalities involved in the binding, a 3D pharmacophore model can be established for virtual screening. This information can be provided by SARs coming from literature, HTS or similarity searches. The strength of such an approach is its ability to retrieve different chemotypes matching the same pharmacophoric pattern. On the one hand, the pharmacophore model can be applied to screen existing libraries. The Catalyst software [10] has proven to be successful in filtering pre-generated conformers' databases in many projects [39]. On the other hand, the model can be used to generate molecules *de novo* [40]. They are usually used as a template for further chemistry designs.

Clustering

Clustering is an important aspect of the analysis of screening hit lists where the hits have to be partitioned into meaningful chemical classes and prioritized. Several clustering methods are available and there is no clear recipe to achieve this goal. Whether hierarchical methods, such as Ward's clustering or non-hierarchical methods, such as Jarvis-Patrick, k-nearest neighbors, self-organizing maps, are applied the clustering results depend on many parameters such as descriptors selection, data normalization, similarity metrics…[11] Moreover, hierarchical clustering will not reveal the number and the distribution of the structural families before an arbitrary threshold is chosen and it is even worse for partitional clustering where the number of clusters needs to be set a priori.

Whereas the grouping of compounds with respect to properties is an established and fast procedure, the partitioning into chemotypes is not straightforward. Constant effort has been pursued in Peter Willett's team to implement and validate graph fitting algorithms that identify the maximum common subgraph (MCS) between two molecules, leading to improvement of the efficiency of clique-detection algorithm efficiency either by heuristics approaches [41], reduced graphs [42], or association with pharmacophores [43]. The MCS approach seems promising, as it deals directly with chemical structures. However, MCS calculations are quite time-consuming and their validity for chemotype identification within the workflow of hit selection still needs to be established [44].

An alternative method has been proposed by Xu [45]. The scaffold-based classification approach (SCA) starts with the identification of all the non-redundant scaffolds, forming different topological classes. Then, all the structures falling in the same classes are sorted in ascending order of structural complexity. The least complex molecule, i.e the least substituted, will be the center of the class.

As soon as some interesting clusters have been highlighted, an important decision should be taken about the series to be further optimized.

Chemical Tractability

Synthetic tractability is one of the main criteria for selecting hit classes. Retrosynthetic analysis procedures have provided chemistry guidelines for the selection of both scaffolds and building blocks easily accessible by combinatorial chemistry for the exploration of the SAR. A broadly used approach is based on the Retrosynthetic Combinatorial Analysis Procedure (RECAP) rules that fragment molecules based on 11 predefined bond types. It can be used either to automatically shred compounds or, with some modification, to build them *de novo* [12]. Nevertheless, most of the time, chemical tractability depends on the experience of the project chemists and the availability of the starting material. For that purpose and also to assess patentability, databases like SciFinder [46] and Derwent [47] are used.

SOFT FILTERS

Usually, the only experimental data available to select the appropriate clusters is a dose-response curve: IC50, Ki or EC50. The role of the computational chemist is now to provide further information about ADME-Tox liabilities for each selected class. The "soft filters" are used to prioritize the different chemotypes by raising alerts with respect to binding to cytochrome 450 (CYP450), human serum albumin (HSA) or hERG as well as frequent hitters identification.

Frequent Hitters

The frequent hitters (FH) are defined as molecules showing up as hits in many different assays covering a wide range of targets [17]. Since we have already filtered the covalent binders and unstable compounds (see "hard filters"), this definition encompasses the promiscuous inhibitors (non-specific binders) described by McGovern *et al.* [48], and also includes compounds perturbing assay or detection methods. The FH model has been built on a mixed approach. The first step consists of identifying molecules which were more than 8 times among the best 1000 hits across 161 HTS assays. Then, this list has been submitted to the vote of eleven independent medicinal chemistry teams. Based on this data set an *in silico* prediction model (three-layered neural network) was built. As an independent validation set, 31 promiscuous ligands published by McGovern *et al*. [48] have been used to assess if the FH model is stable. 25 out of 31 have been predicted successfully, the false negatives are depicted in Fig. (**3**).

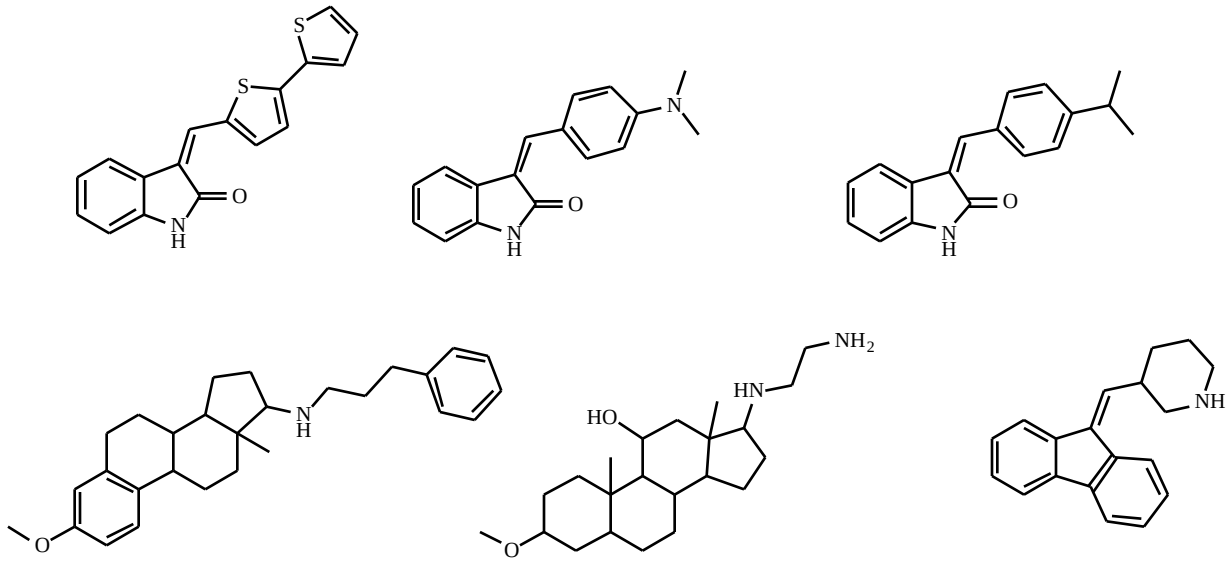

Fig. (3). Experimentally determined promiscuous ligands, which are not identified by the Frequent Hitter *in silico* filter.

Drug-Likeness

Whereas lead-likeness analyzes molecular property ranges, drug-likeness is based on patterns derived from oral drugs. The assumption is that drugs have favorable ADME-Tox properties, including permeability and solubility. Dozen of classification methods have been applied to discriminate between drugs and non-drugs sets. All have been quite successful as reviewed by Walters *et al*. [49]. Nevertheless, these models are "black boxes" since the results cannot be traced back to the molecular structure. Moreover, the selection of a relevant non-drugs data set is far from trivial because non-drug-likeness is inferred from the lack of a drug history. To avoid this problem, various groups have focused on the description of the drug properties and extracted simple rules [50] or pharmacophores [51].

ADME-Tox

The final step for a full profiling of compounds is the prediction and measurement of ADME-Tox properties. Even if recent advances in *in vitro* ADME-Tox technologies have enhanced the throughput of some assays (absorption, solubility, CYP450 inhibition…) to an unprecedented level, the improvement of others is still difficult due to the underlying mechanisms (metabolic stability, hERG liability, cytotoxicity…) Moreover, some properties can only be measured *in vivo*: bioavailability, volume of distribution, half-life.... Therefore, only a complementary use of experimental and *in silico* ADME-Tox methods will allow to achieve the throughput imposed by chemistry [52]. The *in silico* prediction of ADME-Tox properties from the chemical structure has been reviewed previously [53, 54]. The main limitation is the predictability, often coming from the availability of suitable data sets. Indeed, a sufficient quantity and diversity of the molecules as well as controlled quality and homogeneity of the experimental results is the only way to achieve reliable models. Mechanistic or expert models are usually more suited for generalization than the statistical models, which are limited to the chemical space of the training set. According to Clark, the blood-brain barrier (BBB) permeation models have reached their optimum given the available data sets [22]. Hence, new improvements can only come from new data. For that reason, major pharmaceutical companies have a clear advantage since they can use their proprietary data sets to develop the models and use the public heterogeneous data sets for validation. This is the protocol we have followed to build our own HSA, CYP450 3A4 and hERG filters [13,14,55]. As "hit identification" is dealing with compounds coming from HTS or virtual screening, the ADME-Tox models have to be "global" and need to be dynamically refined with newly generated experimental data.

If ADME properties are difficult to predict, toxicology models are even more complicated due to the few data points available, the intrinsic variability of the end-points and also the large diversity of underlying mechanisms. Commercial softwares are based on QSAR models [56], expert models DEREK [57] or both MCASE [58].

All these ADME-Tox prediction tools are regarded as qualitative to prioritize the series that will be further optimized by chemistry. The combination of these soft filters with the physicochemical properties and biological results provides a multi-dimensional matrix, which needs to be analyzed.

Visualization

It is extremely important for the decision making process that the chemists can visualize all the properties and extract patterns and trends. In addition, these values should be linked to the chemical structures of interest. Spotfire [18] and Miner3D [59] are powerful tools providing different ways to quickly select specific property ranges. There are also more dedicated packages like SARNavigator [60] or ClassPharmer [61] that perform the calculations of properties, clustering and visualization. Finally, major pharmaceutical companies have implemented their own in-house solution like the Hits Analysis Database (HAD) described by Shen [62].

Up to the end of the "hit identification" process, no resources need to be committed to the synthesis of new compounds. Chemistry efforts will start as soon as the

chemical classes will be selected. Then, the computational chemistry support will focus on series specific models ("local" models) during the optimization process. This knowledge-based decision is now driven by information from heterogeneous sources such as biological data, physicochemical properties, tractability and patentability, which has been compiled to assess as early as possible the optimization potential of the screening hits.

CONCLUSION

From library shaping to the first synthesized molecules, computational chemistry is supporting chemists in the selection of compounds for further optimization. From millions of compounds to a particular scaffold, the task is to identify the molecules that will have the right balance between potency and ADME-Tox properties. For that purpose, many *in silico* filters have been established to transform the current knowledge into computational models, which are applied for multivariate profiling of novel molecules. However, most of these filters are only well suited for a large number of compounds since they are based on statistical models and none of them should be applied on a single molecule. Therefore, they fit nicely with the Lead Generation phase where the goal is to reduce the number of compounds and to focus on the most promising ones. When the chemistry effort has started on particular chemotypes (Lead Optimization), QSAR and QSPR approaches should be valorized.

Nowadays, large number of computational tools are available to support medicinal chemists' decisions in all phases of Lead Generation. Specifically, the "hit identification" process benefits from the constant improvements of computational tools since they cover a wide range of techniques such as library design, virtual screening, clustering and property modeling. Thus, computational filters complement the experience and creativity of the medicinal chemists [63].

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues in Molecular Modeling/Cheminformatics for scientifically stimulating discussions and valuable contributions. The fruitful collaboration with colleagues from Medicinal Chemists, Roche Compound Depository, Assay Development & HTS and the Molecular Properties Group at F. Hoffmann-La Roche Ltd in Basel is greatly acknowledged.

ABBREVIATIONS

HTS = High-Throughput Screening

ADME-Tox = Absorption Distribution Metabolism Elimination Toxicity

QSAR = Quantitative Structure-Activity Relationship

QSPR = Quantitative Structure-Property Relationship

CYP450 = Cytochrome

FH = Frequent Hitters

Ro5 = Rule of 5

CNS = Central nervous system

PSA = Polar surface area

GPCR = G-protein coupled receptor

SAR = Structure Activity Relationship

MCS = Maximum common subgraph

SCA = Scaffold-based Classification Approach

RECAP = Retrosynthetic Combinatorial Analysis Procedure

HSA = Human Serum Albumin

BBB = Brain Blood Barrier

REFERENCES

[1] Frantz, S.; Smith, A. *Nature Rev. Drug Discov.*, **2003**, *2*, 95.
[2] Drews, J. *Drug Discov. Today*, **2003**, *8*, 411.
[3] Alanine, A.; Nettekoven, M.; Roberts, E.; Thomas, A.W. *Comb. Chem. High Throughput Screen.*, **2003**, *6*, 51.
[4] Fox, S.; Wang, H.; Sopchak, L.; Farr-Jones, S. *Journal of Biomolecular Screening*, **2002**, *7*, 313.
[5] Rishton, G.M. *Drug Discov. Today*, **1997**, *2*, 382.
[6] Hann, M.M.; Leach, A.R.; Harper, G. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 856.
[7] Oprea, T.I.; Davis, A.M.; Teague, S.J.; Leeson, P.D. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1308.
[8] Balakin, K.V.; Tkachenko, S.E.; Lang, S.A.; Okun, I.; Ivashchenko, A.A.; Savchuk, N.P. *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 1332.
[9] Willett, P.; Barnard, J.M.; Downs, G.M. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 983.
[10] Kurogi, Y.; Guner, O.F. *Curr. Med. Chem.*, **2001**, *8*, 1035.
[11] Downs, G.M.; Barnard, J.M. *Rev. Comp. Chem.*, **2002**, *18*, 1.
[12] Lewell, X.Q.; Judd, D.B.; Watson, S.P.; Hann, M.M. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 511.
[13] Zuegge, J.; Fechner, U.; Roche, O.; Parrott, N.J.; Engkvist, O.; Schneider, G. *Quant. Struct.-Act. Relat.*, **2002**, *21*, 249.
[14] Roche, O.; Trube, G.; Zuegge, J.; Pflimlin, P.; Alanine, A.; Schneider, G. *ChemBioChem.*, **2002**, *3*, 455.
[15] Sadowski, J.; Kubinyi, H. *J. Med. Chem.*, **1998**, *41*, 3325.
[16] Ajay; Walters, W.P.; Murcko, M.A. *J. Med. Chem.*, **1998**, *41*, 3314.
[17] Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E.M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G. *J. Med. Chem.*, **2002**, *45*, 137.
[18] Spotfire; Spotfire, Inc.: http://www.spotfire.com, 2003.
[19] Daylight; Daylight Chemical Information Systems, Inc.: http://www.daylight.com, 2002.
[20] Lipinski, C.A. *J. Pharm. Tox. Methods*, **2000**, *44*, 235.
[21] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug Deliv. Rev.*, **1997**, *23*, 3.
[22] Clark, D.E. *Drug Discov. Today*, **2003**, *8*, 927.
[23] Veber, D.F.; Johnson, S.R.; Cheng, H.-Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. *J. Med. Chem.*, **2002**, *45*, 2615.
[24] Winiwarter, S.; Ax, F.; Lennernas, H.; Hallberg, A.; Pettersson, C.; Karlen, A. *J. Mol. Graphics Modell.*, **2003**, *21*, 273.
[25] Rucker, G.; Rucker, C. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1457.
[26] Rishton, G.M. *Drug Discov. Today*, **2003**, *8*, 86.
[27] Proudfoot, J.R.J.R. *Bioorg. Med. Chem. Lett.*, **2002**, *12*, 1647.
[28] Muller, G. *Drug Discov. Today*, **2003**, *8*, 681.
[29] Horton, D.A.; Bourne, G.T.; Smythe, M.L. *Chem. Rev.*, **2003**, *103*, 893.
[30] Martin, Y.C.; Kofron, J.L.; Traphagen, L.M. *J. Med. Chem.*, **2002**, *45*, 4350.
[31] Livingstone, D.J. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 195.
[32] Sheridan, R.P.; Kearsley, S.K. *Drug Discov. Today*, **2002**, *7*, 903.
[33] Holliday, J.D.; Salim, N.; Whittle, M.; Willett, P. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 819.

[34] Wang, R.; Wang, S. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1422.
[35] Ginn, C.M.R.; Willett, P.; Bradshaw, J. *Perspect. Drug Discov. Des.*, **2000**, *20*, 1.
[36] Rarey, M.; Dixon, J.S. *J. Comput.-Aid. Mol. Des.*, **1998**, *12*, 471.
[37] Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed.*, **1999**, *38*, 2894.
[38] PHACIR; CallistoGen, AG, **2002**.
[39] Langer, T.; Krovat, E.M. *Curr. Opin. Drug Discovery Dev.*, **2003**, *6*, 370.
[40] Stahl, M.; Todorov, N.P.; James, T.; Mauser, H.; Boehm, H.J.; Dean, P.M. *J. Comput.-Aid. Mol. Des.*, **2002**, *16*, 459.
[41] Raymond, J.W.; Gardiner, E.J.; Willett, P. *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 305.
[42] Gillet, V.J.; Willett, P.; Bradshaw, J. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 338.
[43] Rhodes, N.; Willett, P.; Calvet, A.; Dunbar, J.B.; Humblet, C. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 443.
[44] Raymond, J.W.; Blankley, C.J.; Willett, P. *J. Mol. Graphics Modell.*, **2003**, *21*, 421.
[45] Xu, J. *J. Med. Chem.*, **2002**, *45*, 5311.
[46] SciFinder; American Chemical Society: http://www.cas.org/-SCIFINDER/, **2002**.
[47] Derwent; Thomson Scientific: http://www.derwent.com, **2002**.
[48] McGovern, S.L.; Caselli, E.; Grigorieff, N.; Shoichet, B.K. *J. Med. Chem.*, **2002**, *45*, 1712.
[49] Walters, W.P.; Murcko, M.A. *Adv. Drug Deliv. Rev.*, **2002**, *54*, 255.
[50] Egan, W.J.; Walters, W.P.; Murcko, M.A. *Curr. Opin. Drug Discovery Dev.*, **2002**, *5*, 540.
[51] Muegge, I.; Heald, S.L.; Brittelli, D. *J. Med. Chem.*, **2001**, *44*, 1841.
[52] Yu, H.; Adedoyin, A. *Drug Discov. Today*, **2003**, *8*, 852.
[53] Clark, D.E.; Grootenhuis, P.D.J. *Curr. Opin. Drug Discovery Dev.*, **2002**, *5*, 382.
[54] van de Waterbeemd, H.; Gifford, E. *Nature Rev. Drug Discov.*, **2003**, *2*, 192.
[55] Kratochwil, N.A.; Huber, W.; Muller, F.; Kansy, M.; Gerber, P.R. *Biochem. Pharmacol.*, **2002**, *64*, 1355.
[56] TOPKAT; Accelerys Inc., 9685 Scranton Road. San Diego, CA 92121-3752 USA, **2002**.
[57] Greene, N.; Judson, P.N.; Langowski, J.J.; Marchant, C.A. *SAR QSAR Environ. Res.*, **1999**, *10*, 299.
[58] MCASE; MultiCASE, Inc. 23811 Chagrin Blvd. Ste 305, Beachwood, OH 44122 USA., **2002**.
[59] miner3d; Dimension 5: http://www.miner3d.com, **2002**.
[60] SARNavigator; Tripos: http://www.tripos.com/sciTech/inSilico-Disc/dataAnalysis/sarnavigator.html, **2002**.
[61] ClassPharmer; BioReason: http://www.bioreason.com/-classpharmer.php, **2002**.
[62] Shen, J. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 1668.
[63] Kubinyi, H. *Nature Rev. Drug Discov.*, **2003**, *2*, 665.