Virtual Darwinian Drug Design: QSAR Inverse Problem, Virtual Combinatorial Chemistry, and Computational Screening

J. V. de Julián-Ortiz*

Unidad de Investigación de Diseño de Fármacosy Conectividad Molecular. Facultad de Farmacia, Universitat de València. Spain.

Abstract: The generation of diversity and its further selection by an external system is a common mechanism for the evolution of the living species and for the current drug design methods. This assumption allows us to label the methods based on generation and selection of molecular diversity as "Darwinian" ones, and to distinguish them from the structure-based, structure-modulation approaches. An example of a Darwinian method is the inverse QSAR. It consists of the computational generation of candidate chemical structures and their selection according to a previously established QSAR model. New trends in the field of combinatorial chemical syntheses comprise the concepts of virtual combinatorial synthesis and virtual or computational screening. Virtual combinatorial synthesis, closely related to inverse QSAR, can be defined as the computational simulation of the generation of new chemical structures by using a combinatorial strategy to generate a virtual library. Virtual screening is the selection of chemical structures having potential desirable properties from a database or virtual library in order to be synthesized and assayed. This review is mainly focused on graph theoretical drug design approaches, but a survey with key references is provided that covers other simulation methods.

INTRODUCTION

As Balaban pointed out in 1985 [1], the essence of chemistry is the combinatorics of atoms according to definite rules. Making a simple and mechanistic paraphrase of this statement we might say that the essence of biology is the combinatorics of molecules according to definite, chemical rules. The generation of diversity and its further selection by an external system is a common mechanism to the evolution of the living species and to the current drug discovery methods. This assumption allows us to label such methods, based on generation and selection within a molecular diversity set, as *Darwinian* ones. In the Darwinian Drug Design (DDD) scheme, the keystone is the molecular selection step within the diverse set, whereas the structure-based one is focused on the rational building of the best-candidate molecular structures. Thus, thedifference lies in the scope and philosophy of each approach. If the phenomena that condition the drug design were perfectly known, only the second scheme makes sense. Since this knowledge is beyond the actual science, the DDD approach affords valuable tools to find even unexpected new active compounds.

In view of the definitions above, the following drug search techniques may be considered to be Darwinian:

EXPERIMENTAL TECHNIQUES

· Screening of Natural or Synthetic products / Random Selection / Trial and Error

· Parallel Synthesis

· Combinatorial Synthesis

· High Throughput Screening

*Address correspondence to this author at the Unidad de Investigación de Diseño de Fármacosy Conectividad Molecular. Facultad de Farmacia, Universitat de València. Spain, E-mail: julian@goya.combios.es

1386-2073/01 $28.00+.00

· Combinatorial Biosynthesis

COMPUTATIONAL TECHNIQUES

· Database Search / Virtual Screening

· Systematic Inverse QSAR methods, also referred as: graph reconstruction, generation of molecular graphs, inverse imaging, design of molecules from QSAR, molecule building from QSAR equations, inverse structure generation, etc.

· Random structure generation

· Design of Libraries

· Virtual Combinatorial Synthesis

· Genetic-Algorithm-based Drug Design

These computational techniques suggest new molecular structures, setting up the complete structure or through building blocks. An experimental validation is required in either case, and it will be accomplished by a Darwininan experimental method whether the size of the *generandum* is high enough.

Economic pressures explain the interest of the pharmaceutical industry in combinatorial chemistry related methods in order to increase the effectiveness in lead generation and optimization. In the mid-1990s the pharmaceutical market growth suffered an important decrease while the research investment was considerably amplified with respect to the previous decade [2]. The new developments in DDD are the reaction of the pharmaceutical industry to achieve more efficiency.

Going on with the simile of the Darwinian selection, in this review first we deal with the establishing of the environment: the elaboration of the model. Second, we analyze the proliferation of potential inhabitants: the raw generation of a pool of virtual chemical structures; and third, the natural selection process: the problem of selecting the best-fitted structures. A wider view of graph theoretical approaches is given in this article. For a more deep approach to molecular dynamics and other simulations applied to ligand design see, for example, the review by Apostolakis and Caflish published in 1999 [3].

ACTIVITY MODELING

A given environment can be characterized through the description of the living species found in it. Similarly, molecular structures that show a given activity characterize their *pharmacological niche.* The first problem that a quantitative structure-property relationship (QSPR) model needs to solve is the translation of the chemical structure into quantitative information. A few numerically expressible chemical features were used in QSAR at the beginning. These descriptors were physicochemical parameters related to the chemical forces involved in the interaction of the drug and its surrounding molecular environment. Mainly, additive substituent constants were used, which belonged to three categories [4]: a) Parameters that stands for charge distributions, electronic effects of substituents in aromatic rings, and electrostatic influences, i.e., Hammett σ, in different versions. b) Parameters that give account of hydrophobicity; such as Zahradnik β and Hansch π, and c) Several steric parameters such as Hansch's or Taft's. The Randic index [5], another steric paramenter initially conceived as a quantification of the ramification degree, opened a new way in QSAR, as will be seen.

The physicochemical parameters are selected in a QSAR analysis in the belief that they mirror salient features in the mechanism of biological action, and that it is possible to infer mechanistic models from these equations. As a matter of fact, in many occasions the obtained correlations are purely phenomenological and do not reflect a causal mechanism. This can be due to intercorrelations among the physicochemical parameters, or contributions of several mechanisms to a given value of the activity. The true character of the physicochemical parameters is often complex and they give account of superimposed effects [6]. Declared empirical methods such as the so-called Free-Wilson analysis [7] or the DARC/PELCO approaches [8]

are also based on the additivity concept of group contributions to the biological activity.

Mainly, four fundamental theories support the chemical models: Quantum Theory, Statistical Mechanics, Group Theory and Graph Theory. Other theories that probably will afford new points of view to chemistry in the near future are the Theory of Information [9,10] and the Theory of Chaos. Quantum chemistry supplies a theoretical overall approach alternative to the physicochemical parameters, which are partial aspects of more fundamental molecular properties.

Graph theory has been useful in modeling chemistry-related concepts such as compounds, intermediates, synthons, reactions, physical changes, reaction mechanisms or molecular point clouds. In Chemical Graph Theory, molecular structures are normally represented as hydrogen-depleted graphs, whose vertices and edges act as atoms and covalent bonds, respectively. Chemical structural formulas can be then assimilated to undirected and finite multigraphs with labeled vertices, commonly known as molecular graphs. Stereoisomerism can also be characterized by application of the Group Theory to the molecular graph [1], by embedding the graph on coordinate grids [11,12] or by assigning specific weights to the vertices [13]. The most important applications of molecular graphs comprise the QSPR, the enumeration and unique representation of compounds for nomenclature and documentation purposes, and the computational structure elucidation.

Graph-theoretical Indices, also known as Topological Indices, were developed. They are descriptors that characterize a molecular graph and are able to give account of their structural properties, in spite of the fact that a single number representation of a structure results in a loss of information. The first indices that can be regarded as graph-theoretical are the Wiener index [14] and the Platt index [15], published in 1947. The work of these pioneers was discontinued with some exceptions [16], until the investigations on topological chemistry of Hosoya [17], Randic [5], Rouvray [18], Balaban [19], Bonchev [20], Gutman [9], Trinajstic [21] and other scientists [22] in the 1970s.

Kier and Hall, who presented their Molecular Connectivity method in 1975, applied and developed the ideas of Randic [23]. The basic assumption is that the structural information can be coded in a family of indices, each one carrying a given type of substructural information. This technique was a powerful stimulus in the search of new QSAR approaches and indices based on the chemical graph concept, which were successful in the correlation of different properties [24]. Among the more recent developments in this field we can cite the Electrotopological State (E-state) [25], the Topological Charge Indices [26], and the Topological Sub-Structural Molecular Design (TOSS-MODE) [27].

Kier and Hall have been developing and applying the E-state indices since 1990 as an approach able to characterize the molecular structure combining both electronic and topological properties. In principle, the indices are defined for atoms within the molecules, allowing the characterization of pharmacophores, the evaluation of the influence of substituents in congeneric series, the measure of molecular similarities or diversity, the handling of molecular fragments for database organization or searching, etc.

Topological Charge Indices have been proposed as tools to characterize the intramolecular charge transfers, and have revealed their usefulness in the prediction of several pharmacological properties [28]. TOSS-MODE is based on descriptors that are able to represent the quantitative contribution of each substructural fragment. An important advantage of this method resides in the fact that the identification of fragments responsible for the activity permits the explicit generation of possible pharmacophores. For a survey on new trends in SAR with topological indices see the review by Devillers released in 2000 [29].

Generally, a property P can be linearly related with one or several descriptors X_i by using statistical methods:

$$P = a_0 + \sum_1 a_i X_i$$

Constants a_0, a_i are correlation coefficients for the regression. Non-linear equations such as the quadratic ones have been often used. The application of non-linear methods to the modeling of biological activities of compounds is a growing field in recent years, mainly with the development of the computational intelligence techniques: evolutionary algorithms, fuzzy systems, artificial neural networks and simulated annealing. Other interesting approaches to the activity modeling comprise the learning machine-expert system [30]. These topics are beyond the scope of the present review.

Computationally more complex, 3D QSAR has reached widespread use in structure modulation and *de novo* design. 3D-based models of the molecular structure, geometrical models, give account of the stereoisomerism in a more natural fashion than topological ones, provided a criterion of orientation is previously introduced. Only 3D macromolecular structures give true information on the mechanistic aspects of QSAR. The excellent books [31] and reviews [32] released on this topic make it unnecessary to give extension to this subject here.

The best-known Darwinian technique within this field is the 3-D Database Searching based on the structure of receptor sites [12, 33]. In relation to this, a great effort is devoted to the estimation of drug binding affinities from structure [34]. An interesting approach is the Comparative Binding Energy (COMBINE) Analysis [35]. QSARs are deduced from interaction energies of ligand-macromolecule complexes by using molecular mechanics calculations and partial least squares. Scoring functions have also demonstrated its efficiency in the prediction of binding affinities [36,37]. A good example is VALIDATE [38], a combined strategy that calculates binding enthalpies by molecular mechanics and binding entropies by heuristic estimation of physicochemical properties.

Other computational approaches do not require the determination of the receptor structure. Among these methods we can cite the Quantum Molecular Similarity [39], as well as the CoMFA (Comparative Molecular Field Analysis) approach [40], HASL (Hypothetical Active Site Lattice) [41], SOMFA (Self-Organizing Molecular Field Analysis) [42], MTD (Minimal Topological Difference) [43], and WHIM (Weighted Holistic Invariant Molecular) approaches [44]. Another interesting approach when the receptor structure is unknown has been the construction of receptor models using, for example, genetic algorithms [45].

Topology and geometry afford two alternative, closely related, complementary, and equally valid ways to model molecular structure. These are, respectively, the graph and the spatial representations. The last one, also termed three-dimensional structure, implicitly carries the topological information, but says nothing about the fourth dimension. Molecules are not static entities. They are continuously changing their conformations. This limiting aspect has been determinant for the development of techniques able to give account of the temporal coordinate, such as the molecular dynamics simulation which

Table 1. Concepts Involved in Two Important Models of Molecular Structure

	Molecular graph	**3D model**
Domain	Topology	Geometry
Also implied	Information theory	Physics
Naturally gives	Structural invariants	Stereochemistry
Problems	Stereochemistry	Alignment The fourth dimension
Solutions	Group theory Grids Specific weights	Optimization algorithms Molecular Dynamics 4D-QSAR

solves the problem by means of a heavy mathematical burden [46], or the 4D-QSAR formalism [47]. By contrast, a molecular graph implicitly carries not only three-dimensional information (many computer programs convert graphs into 3-D structures) but information of the true structural invariants non-affected by vibrational nor conformational changes. In addition, the computational effort is always lighter, since no problems of alignment or minimization are involved with graphs. Table I shows the differences and complementary aspects of the two representations.

Since the space of possible conformations depends on the topological structure, following with the genetic simile, the topological structure can be assimilated into a kind of molecular genotype, while the conformation designates the phenotype. This genotype contains the molecular identity features. By contrast, the phenotypical structure actually depends on features external to the molecule itself, such as the molecular environment, temperature, etc.

Another division of the techniques used in the modeling of activity can be established for those that rely on the molecular mechanism, or if the information they manage is purely structural. In the last case the key is similarity. Techniques involving the explicit modeling of the receptor structure are of the first kind, while all the others are of the last type. Interesting achievements within the latter approach are the techniques developed in order to discriminate drug-like from non-drug-like molecular structures [3]. These results point out that mere structural similarity patterns can play a sound role in drug discovery beyond, or previously to, the elucidation of mechanisms of action.

The following part of this review concerns the raw generation of virtual chemical structures, mainly based on the Molecular Graph concept. Virtual DDD methods in which the generation of chemical structures is the key factor are briefly described. It is surprising that to date just a few research groups have published in this field.

THE GENERATION OF CHEMICAL STRUCTURES. COMBINATION-FOCUSED TECHNIQUES

Genetic mutation in its wide sense is, long term, the mechanism responsible for biological diversity. The differential feature of the DDD methods is the central role played by the generation of molecular structures, real or simulated. The history of building chemical diversity began with the problem of quantifying the isomers for a given empirical formula. An algorithm for calculating the number of structural isomers of aliphatic compounds for certain carbon atom contents was already proposed by Cayley in the second half of the 19th century [48]. It was a wholly solved issue by the 1930s [49,50] The magnitude of the combinatorial explosion was made clear from the beginning [51]. For example, the number of disubstitution products of the paraffins containing 20 carbon atoms and two given unlike substituents was calculated to be 88,594,746 [49].

A second step beyond the counting of chemical diversity was achieved with effective algorithms for the generation of isomeric structures. Probably, the best example of this achievement is the project DENDRAL [52], which was one of the first computer programs and expert systems for structural analysis of organic compounds. DENDRAL identified molecular structures using the empirical formulas, the basic knowledge of chemical topology (such as the tetravalence of the carbon atom), data of molecular fragmentation in mass spectrometry, knowledge of the stability (or plausibility) of the different molecular candidates and, finally, NMR data.

Somewhat more recently, Randic successfully constructed a complete list of graphs on a given number of nodes and edges by using a procedure for unique enumeration of atoms in a molecule [53]. Typically, the generation of molecular structures has been mainly devoted to the canonical indexing, the constructive enumeration of non-redundant graphs, the computer generation of isomeric structures [1, 54], and the molecular structure elucidation [55]. The achievements in stereoisomer generation are noteworthy, especially

for computational structural elucidation purposes [56]. However, the methods developed in these areas are not fully adequate for inverse QSAR, if procedures for simulating successive substitutions in the scaffold of a given parent structure are required.

In 1988 the Gálvez group released a doctoral thesis on computer assisted drug design in which was developed a software package for computerized molecular synthesis [57]. This method allowed the molecular construction from scratch or, alternatively, used a scaffold denominated base structure and a set of substituents constituted by hydrocarbon substructures and functional groups. The substructural fragments were non-cyclic having four or less carbon atoms and bond orders of one to three. The fragments and functional groups were computationally assembled to the base structure at the previously defined attachment sites and/or between them. These could be joined up by each one of their available atoms, and the formation of multiple bonds and cyclic structures was allowed in these steps. For each new *synthesized* molecule, the computer program *decided* if it was potentially active or not, according to a previous QSAR linear model based on Randic-Kier-Hall type indices.

The user chose optimal interval values for each prediction equation. The program also allowed the user to select the number of structures to be generated with an adjustable screening level in the construction process, from the relentless exhaustive search of each isomer to wide leaps within the space of possible graphs. This procedure generated an adjustable molecular diversity. The common strategy consisted on initial runs with wide intervals for the acceptable QSAR equation values and high diversity. In successive runs, designed around the new selected structures, the intervals were narrowed and the diversity was decreased. Finally, the best potential active compounds were synthesized and tested. The results of such tests were used to refine the QSAR model.

The method has been useful in the design of molecular structures with analgesic activity. By using benzene as base structure with attachment sites in several ring positions, active compounds were obtained, including among others, 2-(1-propenyl)phenol (**1**) [58], 2,4-dimethyl acetophenone (**2**) [59], and 1-(*p*-chlorophenyl)propanol (**3**) [60]. By using pyridazine as base structure, 3-methyl pyridazine (**4**) was selected [61]. All of these compounds, depicted in Fig. (1), experimentally showed greater analgesia than acetylsalicylic acid [61], and 76% of the molecular structures predicted to be active showed some experimental analgesia.

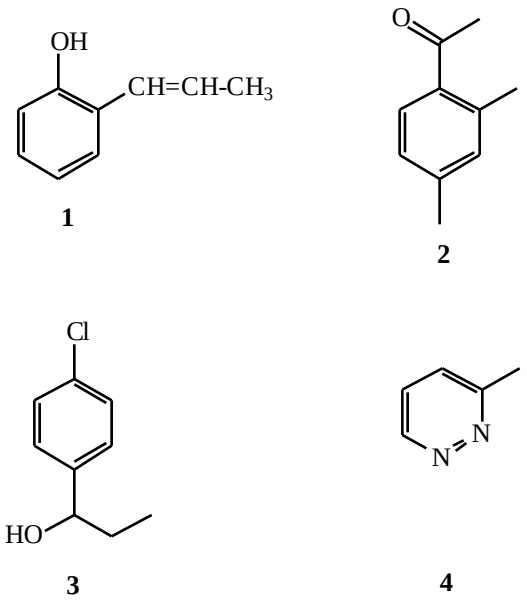

Fig. (1). Analgesic structures designed by inverse QSAR.

In contrast to molecular codes, molecular invariants or single-index QSAR equations do not allow the direct reconstruction of a molecule. The Zefirov group succeeded in reconstruction of graphs through an algorithm focused on edge types [62]. From these, vertex types are encoded by restricting relations, and the graphs are reconstructed via the Faradzhev algorithm [63], narrowing down the possibilities for candidate structures. In 1991, a more developed algorithm for generating non-redundant molecular graphs of substituted derivatives of a given structure was elaborated by the same group for inverse QSAR studies. In a preliminary step, the complete set of substituents with prescribed structural features was constructed from *elementary* fragments, coupled by single bonds [64]. All the fragments could contain heteroatoms and rings and were classified by their number of additional

substituents: terminal, linear or branched, having respectively none, one, and two or more supplementary attachment sites. The construction of each substituent ended when a new terminal fragment was obtained by combination of elementary fragments that fulfill the requirements. No ring systems were built in this step. The same substituted fragment could be obtained by different means (intersection problem), being detected by an additional subroutine. In the end, this algorithm provided non-duplicate target molecular graphs by assembling substituents to all positions of the scaffold. Automatic selection of potentially active structures used QSARs as filters. Later developments have included the use of indices characterizing the molecular shape [65]. Zefirov *et al.* have also developed the SMOG program for generation of chemical structures with required or forbidden structural fragments, invariant substructures, several atomic valence states, specific treatment of aromaticity and other features [66].

The generation of potentially active structures by random combination of database fragments and further selection by topological similarity and statistical techniques has been also studied [67]. The database of fragments including rings, ring systems, acyclic fragments and functional groups was built from a 200,000 compound library. Every fragment was weighted with its frequency of occurrence in the library, and this parameter conditioned the probability of choosing it. Fragments chosen from the database were assembled by randomly assigned free sites. The process could be tuned by changing the weight of each fragment, among other parameters. It continued until a prefixed molecular mass was attained. This method generated many unstable or unusual structures. In a first chemical selection, these generated structures were eliminated in a topological similarity probe test using structurally simple, widely used and very effective drugs. These were: atenolol, captopril, cimetidine, naproxen and ranitidine. The similarity between generated structures and drugs was measured by using the atom pairs descriptors [68] and the topological torsion [69]. In the second selection, the statistical relationship between a given biological activity and descriptors in such a training set is used to calculate a *trend vector* in order to predict the activity of generated structures. This procedure is similar to the chemical similarity selection, but here it is used a training set of compounds with the same pharmacological activity. The random character of the method makes the generation of redundant structures unlikely. In 20,000 generated structures, only 30 were duplicates.

For a brief summary of other approaches used in the inverse problem for non-drug molecules, and the description of a genetic algorithm for virtual polymer generation based on the presence of active substructures, see reference [70]. Genetic algorithms have reached wide use in the design of polymers and other linear structures, which can be parameterized as *chromosomes*. Crossover simulation between chromosomes allows the exploration of a wide space of possible structures [71]. A genetic approach would be welcome that is able to work with generalized building blocks, probably in connection with simulated annealing techniques.

Each new designed compound possessing an arbitrary structure introduces an arbitrary difficulty in its synthesis. This principle is the bottleneck of any kind of molecular design. In order to minimize this factor, it is a good solution to pay attention to an experimentally easy chemical reaction and then to build the new structures using it. This is one of the reasons for the success of the DDD methods known as combinatorial synthesis. Two different approaches, *targeted* and *broad*, have been developed in order to satisfy two different objectives, respectively, lead optimization and lead identification [72]. New computational tools such as the Virtual Library Design were necessary to assist the combinatorial chemical syntheses of large libraries of compounds and render a filtered set of chemical structures to be assayed by further high-throughput screening. During the first stages of this revolution, libraries of peptides and peptide-like molecules were synthesized leading to the identification of novel active compounds [73], but it was soon apparent that there was interest in

using synthons other than aminoacids to extend combinatorial libraries to small organic non-oligomeric structures [74].

The Darwinian character of this methodology is clear since the number of compounds produced as part of a library is usually very large and it is understood that only a tiny fraction of all library compounds will exhibit any bioactivity. Strategies to enhance the probability of finding active compounds in the beginning steps of high-throughput screenings by maximizing the representation and the structural diversity have been developed [75]. Experimental design [76] and clustering techniques [77] has been used for this purpose. Hassan *et al.* have selected diverse subsets of structures by using a single-point-mutation stochastic method for the optimization of diversity functions, defined in terms of distances in multidimensional spaces of descriptors [78]. Approaches designed in order to exclude non-drug-like structures from virtual libraries have been also tackled [3, 79]. Applications of ligand-receptor docking methodology in library design have been reviewed in reference [80].

One step beyond, the computational simulation of combinatorial synthesis soon appeared. Thus, the concept of chemical combinatorial synthesis gave way to virtual combinatorial synthesis, and virtual library design has evolved until the notion of computational or virtual screening [81] in which the number of chemical structures to be assayed is smaller. These developments have been applied to peptoids [82] as well as other non-oligomeric [83] virtual libraries in order to merge two apparently contradictory strategies: combinatorial chemistry and rational drug design.

Lahana *et al.* have developed an approach based on the virtual construction of a combinatorial library, a static computational screening method using molecular graph derivable requirements learned from the known active compounds, and a final selection of candidates using molecular dynamics simulation analysis [84]. The target was to obtain new decapeptides showing significant immunosuppressive action. Seven positions were varied using six possible amino acids, which were selected in order to preserve a hydrophobic distribution similar to the known active compounds. With these restrictions, the number of possible peptides was 279,936. Value ranges defining activity were calculated for thirty structural and graph-theoretical descriptors. Finally, a set of 13 of these descriptors with low inter-correlation was selected. In order to be considered for further selection, a given peptide needed to have its 13 descriptors within the intervals defined by the active molecules. This computational screening selected 26 out of the library of 279,936 peptides. The principal component analysis of molecular dynamics trajectories for the conformations of the 26 peptides, expressed as autocorrelation vectors, allowed the selection of four peptides. Thus only four peptides shared the conformational space occupied by the training set. All of them showed immunosuppressive activity in mice. One of the designed molecules displayed an activity 100 times higher than the known active peptides. For model testing purposes, a fifth peptide of the set of 26 that did not surpass the second selection, was also synthesized and tested, giving no activity *in vivo*. This fact indicates that 4-D dynamic information must not be neglected for this biological activity. This scheme is, up to now, the best example of a paradigm combining the two models of the molecular structure, topological and geometrical. The excellent result obtained must be a stimulus for the use of *topology prior to geometry* in the search of new drugs.

In the next section, we analyze approaches whose key factor is the selecting of structures within the pool of possible ones, limiting the combinatorial generation explosion. Again, the number of articles published following these schemes is relatively modest as yet.

SELECTING CHEMICAL STRUCTURES. SELECTION-FOCUSED TECHNIQUES

The environment selects their living organisms by survival of the fittest. By this way the species improve their adaptation. The quality of the

selection step is determinant for the success of every method of molecular design. The touchstone in the intelligent solving of problems is based on the reduction of random search. More complex models than simple Darwinian postulates are necessary in order to avoid combinatorial explosion in structure generation. The more the virtual combinatorial syntheses are somehow limited to the useful candidates, the more the computational problem is simplified. Pioneers in this field were Zefirov *et al.*, [62] and Klopman *et al.* who defined local distance vertex indices and gave examples for inverse structure generations [85]. The basic ideas can be seen sketched in the works of these researchers.

Kier and Hall have developed a much more elaborated formalism to readily reach target structures. QSAR equations constituted by molecular connectivity indices have enough structural information so that new molecules may be designed from those equations. This process was labeled as an inverse imaging problem [86]. These authors deduced a set of equations to relate low-order path counts to vertex degrees [87], and described ways to construct graphs from a set of vertex degrees [88,89].

A target interval for the investigated property or activity was established, together with QSAR equations based on low-order molecular connectivity indices. Target values for these indices were deduced. Then, a formal scheme for conversion of the connectivity indices into several possible path counts was applied. A path count of a given order or length is the number of times that a linear graph of such length is found in a graph. The conversion is possible because each connectivity index of a given order determines a finite number of possible path counts of this order [86, 87]. Path counts can be exactly converted into degree counts, i.e., the number of graph vertices (atoms) having a given number of edges (bonds) attached to them [87]. The following step was the construction of connection tables from the vertex degrees [89] that give the possible graphs. Alternatively to this matrix, it was possible to calculate edge-type counts [87,88] via the Faradzhev algorithm [63]. The path counts, vertex degrees sets, as well as the edge type counts, are graph *primitives* that are the basis for graph reconstruction into a reduced set of molecular structures. The reconstruction must be both exhaustive and non-redundant [54]. This process generated the candidate set. About one half of the solutions produced had the prescribed value of the property. Finally, target graphs were obtained by filtering through the best QSAR equation having higher order indices [88]. Incorporation of heteroatoms and bonding schemes are added using the information resident in the original database. The probability that the target set possesses property values in the target range is limited by the experimental error in the original data set upon the equation is based.

Kvasnicka and Pospíchal have proposed an inverse imaging method based on simulated annealing [90]. The annealing algorithms are methods capable of locating absolute minima in hyper-surfaces with high probability. A randomly generated molecular graph initiates the process. The application of this method requires the random perturbation of molecular graphs to generate additional ones, starting from a randomly selected atom, and the rest of the code is subsequently changed. The acceptance of the new generated codes is made by a Monte Carlo method, the Metropolis criterion [91]. Following the analogy with the annealing of solids, all particles arrange themselves in the low energy state, assuming that the maximum temperature is high enough and the cooling is carried out sufficiently slowly. After the prescribed number of successive graph-perturbation steps using the Metropolis algorithm, the temperature is decreased. Redundancy is avoided by using adjacency matrices in semicanonical form [92]. The algorithm is terminated for some small value of the temperature at which the state is considered frozen. The minimization of an arbitrary function, the linear combination of the Wiener and the Randic indices, was studied by this scheme for molecular graphs with and without cycles. Another example was the application to the design of hydrocarbon structures with a desired NMR shift for a given atom. Beyond the results that

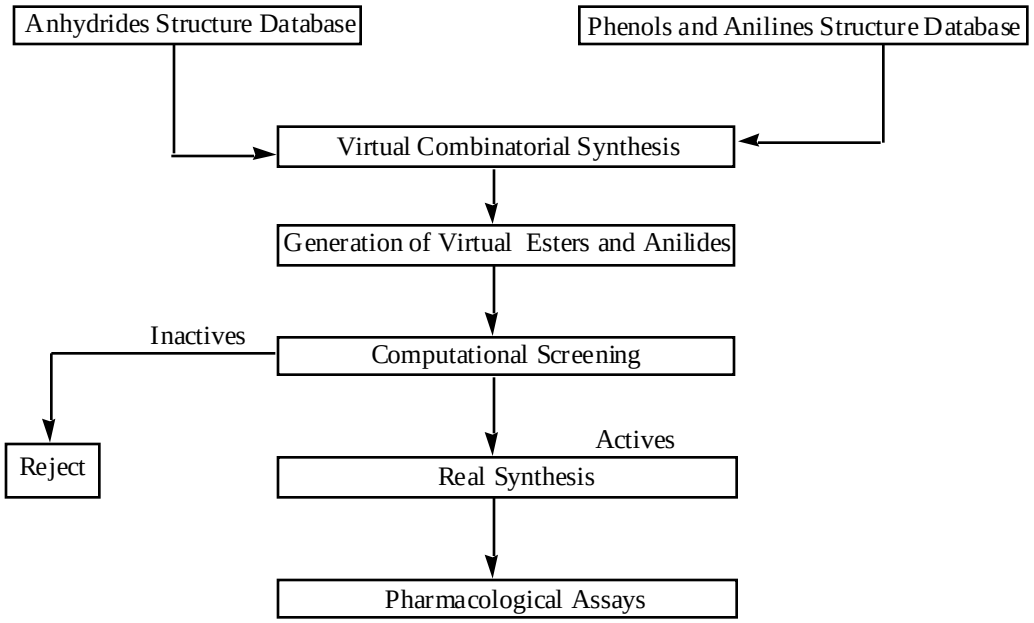

Fig. (2). Steps incorporating virtual combinatorial synthesis and virtual screening involved in obtaining new active compounds.

were obtained, the greatest value of this work was in opening a new approach.

Gálvez *et. al.* have developed a virtual combinatorial technique that minimize the unnecessary combinatorial explosion [93]. A QSAR model was established with different molecular connectivity indices that were correlated with some properties of a group of topical anti-herpetic drugs in order to act as successive filters. The four finally selected functions corresponded to a pharmacokinetic, a pharmacodynamic and a microbiological property, and a classification function obtained from linear discriminant analysis.

In order to find new chemical structures with potential activity and easy synthesis, two fragment databases were designed to generate a virtual library of esters and amides. Figure (2) shows a graphic representation of the concepts involved in the design stage by virtual combinatorial synthesis - computational screening. Previously to the generation of the virtual library, the contributions of each fragment to the values of the connectivity functions were calculated. When the sum of the fragment contributions to a limiting property for a given virtual compound was far enough from the limits of the prefixed intervals, this combination was skipped. This process could be enhanced by incorporating composition rules of graph-theoretical indices to the algorithm [94]. The selected compounds, shown in Fig. (3), were synthesized and tested *in vitro*. All five of these compounds were shown to have anti-herpes virus activity with potencies comparable to the known antiviral agent foscarnet.

Techniques for the identification of new potential ligands for known receptors based on molecular three-dimensional structure have arisen. Problems to be solved in this domain comprise not only the structure generation but also the explosion of conformational possibilities as well as the classic alignment problem. The techniques are based on searching libraries of structures or fragments to find molecules complementary to the binding site or on the construction of molecular structures *de novo* to maximize favorable interactions with the binding site [95]. An example is the the program LEGEND [96] that constructs possible ligand molecules by adding atoms one by one. It randomly assigns attachment, position, and atom and bond type, using the backtracking algorithm and force field. Heteroatoms are exclusively positioned where they can hydrogen-bond the receptor. Final structures are selected following energetic and structural criteria.

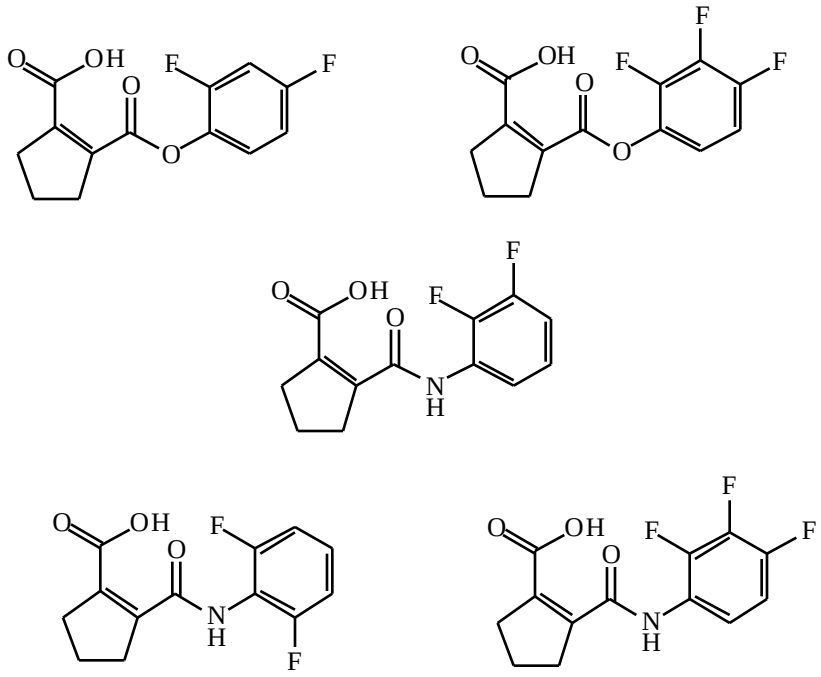

Fig. (3). Antiviral structures designed by virtual combinatorial chemical syntheses and selected by computational screening as described in Figure 2.

The most used *de novo* design program is LUDI [97]. It builds complementary structures to the receptor from a database of fragments, and the interaction energy is estimated by a simple scoring function based on electrostatic charges, hydrogen bonds, hydrophobic contact areas and entropy due to rotatable bonds [36]. For a review on the topic of ligand design see reference 3.

CONCLUSIONS AND FUTURE PROSPECTS

Pharmaceutical industry requires, more than ever, new approaches able to answer the challenge of discovering new lead drugs with the minimum cost. Although some developments towards the end of the 20th century have led to initial excitement, it seems that there are no magic recipes for the design of new bioactive compounds. In order to achieve this objective, no approach should be underestimated. The main part of the research in drug design is currently focused on the mechanism-based approach using computer simulations. With few resources devoted to the topological approaches, much of the effort in the application of more sophisticated schemes now directed elsewhere might be refocused in this area not only for their lower cost but for their proven efficiency.

Inverse QSAR and virtual combinatorial synthesis – computational screening techniques can be considered as Darwinian algorithms since they operate as a simulation of the natural evolution process by natural selection. Deliberate mutations of a parent structure undergo a selection process managed by a QSAR paradigm. Future trends in this field will integrate evolutionary algorithms in these models. Parallel or cluster-distributed computation is also needed in order to improve the yield of the generation process, and this probably would allow the use of mechano-quantical or molecular dynamics descriptors.

Virtual DDD using graph-theoretical models can be more selection-focused methods since composition rules for the descriptors used are available. By contrast, 3-D based Virtual DDD will remain as more irreducible combination-focused methods because the molecular structure can not be exactly inferred from constituent fragments.

The methods described until now operate in a one step Darwinian way. It is possible to devise paradigms in which successive generations of libraries of compounds, in a closer similarity with the evolution mechanism, could be generated and

tested. The compounds resulting in a first generation could be varied in an *orthogonal way* into successive generations or by applying optimization techniques. Genes create enzymes, enzymes create natural products. The possibility of modifying units in the enzyme genetic sequence in order to obtain desired modifications in the final products could introduce a combinatorial genetic engineering.

ACKNOWLEDGMENT

The author sincerely thanks Profs. Ernesto Estrada, Ramon Carbó-Dorca, Lemont B. Kier, Francisco Torrens and Zeno Simon for their help. Also thanks to Profs. Jorge Gálvez and Ramón García-Domenech for their support.

REFERENCES

[1] Balaban, A.T. *J. Chem. Inf. Comput. Sci.*, **1985**, *25*, 334-343.

[2] Brown, D. *Mol. Diversity*, **1996**, *2*, 217-222.

[3] Apostolakis, J., Caflisch, A. *Comb. Chem. High Throughput Screen.*, **1999**, *2*, 91-104.

[4] a) Glasser, A.C. *Methods Find. Exp. Clin. Pharmacol.*, **1984**, *6*, 563-569. b) Hansch, C., Leo, A., Hoekman, D. *Exploring QSAR, Hydrophobic, Electronic, and Steric Constants*, ACS Professional Reference Book, American Chemical Society, Washington, DC, **1995**. c) Braban, M., Pop, I., Willard, X., Horvath, D. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 1119-1127.

[5] Randic, M. *J. Am. Chem. Soc.*, **1975**, *97*, 6609-6615.

[6] Verloop, A. in *Drug Design* E. J. Ariëns, ed Vol III, Chapter 2, pp. 133-187. Academic Press. New York, **1972**.

[7] a) Bruice, T.C., Kharasch, N., Winzler, R.J. *Arch. Biochem. Biophys.*, **1956**, *62*, 305-317. b) Free, S.M., Wilson, J.W. *J. Med. Chem.*, **1964**, *7*, 395-402. c) Bocek, K., Kopecky, J., Krivucová, M. *Experientia*, **1967**, *23*, 1038.

[8] a) Mercier, C., Fabart, V., Sobel, Y., Dubois, J.E. *J. Med. Chem.*, **1991**, *34*, 934-942. b) Mercier, C., Sobel, Y., Dubois, J.E. *Ann. Math. Chem.*, **1992**, *3*, 199-257.

[9] Gutman, I., Ruscsic, B., Trinajstic, N., Wilcox, C.F.Jr. *J. Chem. Phys.*, **1975**, *62*, 3399-3405.

[10] Chen, J., Yang, Y., Jia, H., Lee, T. *J. Math. Chem.*, **1994**, *16*, 125-136.

[11] a) Randic, M., Jerman-Blazich, B., Trinajstic, N. *Computers Chem.*, **1990**, *14*, 237-. b) Bognanov, B., Nikolic, S., Trinajstic, N. *J. Math. Chem.*, **1990**, *5*, 305-306.

[12] Balaban, A.T. *SAR & QSAR Environ. Res.*, **1998**, *8*, 1-21.

[13] a) Pogliani, L. *J. Chem. Inf. Comput. Sci.*,**1994**, *34*, 801-804. b) Estrada, E. *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 708-713. c) Schultz, H.P., Schultz, E.B., Schultz, T.P. *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 864-870. d) Julián-Ortiz, J.V. de' Gregorio Alapont, C. de., Ríos-Santamarina, I., García-Domenech, R., Gálvez, J. *J. Mol. Graphics Model.*, **1998**, *16*, 14-18. e) Golbraikh, A., Bonchev, D., Tropsha, A. *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 147-158.

[14] Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17-20.

[15] Platt, J.R. *J. Chem. Phys.*, **1947**, *15*, 419-420.

[16] Gordon, M., Scantlebury, G.R. *Trans. Faraday Soc.*, **1964**, *60*, 605-.

[17] Hosoya, H. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332-2339.

[18] Rouvray, D.H. *Math. Chem.*, **1975**, *1*, 125-.

[19] Balaban, A.T. *Chem. Phys. Lett.* **1982**, *89*, 399-404.

[20] Bonchev, D., Trinajsti_, N. *J. Chem. Phys.* **1977**, *67*, 4517-4533.

[21] Trinajstic, N. *Chemical Graph Theory. Vol. I* , CRC Press: Boca Raton, FL, **1983**.

[22] Devillers, J., Balaban, A.T. *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers: Amsterdam, Nederlands. **2000**.

[23] a) Kier, L.B., Murray, W.J., Randic, M., Hall, L.H. *J. Pharm. Sci.* **1976**, *65*, 1226-1230. b) Kier, L. B., Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*, Academic Press: New York, **1976**.

[24] a) Gozalbes, R., Julián-Ortiz, J.V. de; Antón-Fos, G.M., Gálvez, J., García-Domenech, R. *Chromatographia*, **2000**, *51*, 331-337. b) Cercós-del-Pozo, R. A., Pérez-Giménez, F., Salabert-Salvador, M. T., García-March, F. J. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 178-184. c) Gregorio Alapont, C. de; Kier, L. B., Hall, L. H. QSAR modeling with the electrotopological state indices: Corticosteroids. *J. Comput.-Aided Mol. Design*, **1998**, *12*, 557-561. d) Ivanciuc, O; Ivanciuc, T., Balaban, A.T. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 395-401. e) Pogliani, L. *J. Phys. Chem.*, **1996**,. *100*, 18065-18077. f) Patra, S.M., Mishra, R.K. *Bull. Chem. Soc. Jpn.*, **1997**, *70*, 997-1001.

[25] a) Kier, L.B., Hall, L.H. *Pharm. Res.*, **1990**, *7*, 801-807. b) Kier, L.B., Hall, L.H. *Molecular Structure Description. The Electrotopological State*, Academic Press: San Diego, **1999**.

[26] Galvez, J., García-Domenech, R., Salabert-Salvador, M.T., Soler, R. *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 520-525.

[27] a) Estrada, E., Peña, A., García-Domenech, R. *J. Comput.-Aided Mol. Design*, **1998**, *12*, 583-595. b) Estrada, E. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 90-95. c) Estrada, E. *SAR & QSAR Environ. Res.*, **2000**, *11*, 55-73. d) Estrada, E., Uriarte, E., Montero, A., Teijeira, M., Santana, L., De Clercq, E. *J. Med. Chem.*, **2000**, *43*, 1975-1985. e) Estrada, E., Peña, A. *Bioorg. Med. Chem.*, **2000**, *8*, 2755-2770.

[28] Galvez, J., García-Domenech, R., Salabert-Salvador, M.T., Soler, R. *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 272-284.

[29] Devillers, J. *Curr. Opin. Drug Discovery Develop.*, **2000**, *3*, 275-279.

[30] Sternberg, M.J.E., Lewis, R.A., King, R.D., Muggleton, S. *Farady Discuss.*, **1992**, *93*, 269-280.

[31] a) Kubinyi, H., ed., *3D-QSAR in Drug Design. Theory. Methods. Applications*, ESCOM: Leyden, **1993**. b) Rappé, A.K., Casewit, C.J. *Molecular Mechanics Across Chemistry*, University Science Books: Susalito, CA. **1997**. c) Veerapandian, P., ed., *Structure-based drug design*, Marcel Dekker: New York, **1997**.

[32] a) Green, S.M., Marshall, G.R. *Trends Pharmacol. Sci.*, **1995**, *16*, 285-291. b) Marrone, T.J., Briggs, J.M., McCammon, J.A. *Annu. Rev. Pharmacol. Toxicol.*, **1997**, *37*, 71-90. c) Kubinyi, H. *Curr. Opin. Drug Discovery Develop.*, **1998**, *1*, 16-27. d) Lemmen, C., Lengauer, T.J. *Comput. Aided Mol. Des.*, **2000**, *14*, 215-232

[33] a) Humlet, C., Dunbar, J.B., Jr. *Annu. Rep. Med. Chem.*, **1993**, *28*, 275-284. b) Doucet, J.P., Panaye, A. *SAR & QSAR Environ. Res.*, **1997**, *7*, 1-24. c) Lessel, U.F., Briem, H. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 246-253. d) Schuffenhauer, A., Gillet, V.J., Willett, P. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 295-307.

[34] a) Gilson, M.K., Given, J.A., Head, M.S. *Chem. Biol.*, **1997**, *4*, 87-92. b) Kubinyi, H., Folkers, G., Martin, Y.C., eds., *3D-QSAR in Drug Design. Volume 2*, Kluwer Academic Publishers: Dordrecht, **1998**.

[35] a) Ortiz, A.R., Pisabarro, M.T., Gago, F., Wade, R.C. *J. Med. Chem.*, **1995**, *38*, 2681-2691. b) Pérez, C., Pastor, M., Ortiz, A.R., Gago, F. *J. Med. Chem.*, **1998**, *41*, 836-852.

[36] Böhm, H.-J. *J. Comput.-Aided Mol. Design*, **1994**, *8*, 243-256.

[37] a) Ajay, Murcko, M.A. *J. Med. Chem.*, **1995**, *38*, 4953-4967. b) Bamborough, P., Cohen, F.E. *Curr. Opin. Drug Struct. Biol.*, **1996**, *6*, 236-241. c) Jain, A.N. *J. Comput.-Aided Mol. Design*, **1996**, *10*, 427-440. d) Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V., Mee, R.P. *J. Comput.-Aided Mol. Design*, **1997**, *11*, 425-445.

[38] Head, R.D., Smythe, M.L., Oprea, T.I., Waller, C.L., Green, S.M., Marshall, G.R. *J. Am. Chem. Soc.*, **1996**, *118*, 3959-3969.

[39] a) Carbó-Dorca, R., Arnau, M., Leyda, L. *Int. J. Quant. Chem.*, **1980**, *17*, 1185-1189. b) Carbó-Dorca, R., Besalú, E., Amat, L., Fradera, X. *J. Math. Chem.*, **1995**, *18*, 237-246. c) Amat, L., Carbó-Dorca, R., Ponec, R. *J. Med. Chem.*, **1999**, *42*, 5169-5180. d) Carbó-Dorca, R., Robert, D., Amat, L., Gironés, X., Besalú, E. *Molecular Quantum Similarity in QSAR and Drug Design. Lecture Notes in Chemistry. Vol. 73.* Springer Verlag: Berlin, **2000**. e) Carbó-Dorca, R., Amat, L., Besalú, E., Gironés, X., Robert, D. *THEOCHEM*, **2000**, *504*, 181-228.

[40] a) Cramer, R.D. III; Patterson, D.E., Bunce, J.D. *J. Am. Chem. Soc.*, **1988**, *110*, 5959-5967. b) Cramer, R.D. III; DePriest, S., Patterson, D.E., Hetch, P. in *3D-QSAR in Drug Design*. H. Kubinyi, ed., pp. 443-485. ESCOM: Leyden, **1993**.

[41] Doweyko, A.M. *J. Med. Chem.*, **1988**, *31*, 1396-1406.

[42] Robinson, D.D., Winn, P.J., Lyne, P.D., Richards, W.G. *J. Med. Chem.*, **1999**, *42*, 573-583.

[43] a) Simon, Z. in *3D-QSAR in Drug Design*. H. Kubinyi, ed., pp. 307-319. ESCOM: Leyden, **1993**. b) Sulea, T., Kurunczi, L., Oprea, T.I., Simon, Z. *J. Comput. Aided Mol. Des.*, **1998**, *12*, 133-146.

[44] Bravi, G., Gancia, E., Mascagni, P., Pegna, M., Todeschini, R., Zaliani, A. *J. Comput. Mol. Des.*, **1997**, *11*, 79-92.

[45] Walters, D.E., Hinds, R.M. *J. Med. Chem.*, **1994**, *37*, 2527-2536.

[46] Sagui, C., Darden, T.A. *Annu. Rev. Biophys. Biomol. Struct.*, **1999**, *28*, 155-179.

[47] a) Hopfinger, A.J., Wang, S., Tokarski, J.S., Jin, B., Albuquerque, M., Madhav, P.J., Duraiswami, C. *J. Am. Chem. Soc.*, **1997**, *119*, 10509-10524. b) Venkatarangan, P., Hopfinger, A.J. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1141-1150.

[48] Cayley, A. *Philos. Mag.*, **1877**, *3*, 34.

[49] Henze, H.R., Blair, C.M. *J. Am. Chem. Soc.*, **1934**, *56*, 157.

[50] Pólya, G. *Acta Math.*, **1937**, *68*, 145.

[51] Otter, R. *Ann. Math.*, **1948**, *49*, 583-599.

[52] a) Lederberg, J. *Proc. Natl. Acad. Sci. U.S.A.*, **1965**, *53*, 134. b) Lederberg, J., Sutherland, G. L., Buchanan, B. G., Feigenbaum, E. A., Robertson, A. V., Duffield, A. M., Djerassi, C. *J. Am. Chem. Soc.*, **1969**, *91*, 2973-2976.

[53] Randic, M. *J. Chem. Phys.*, **1974**, *60*, 3920-3928.

[54] a) Masinter, L.M., Sridharan, N.S., Lederberg, J., Smith, D.H. *J. Am. Chem. Soc.*, **1974**, *96*, 7702-7714. b) Mitrofanov, Ju.P., Raznikov, V.V., Shukurov, V.A. *Zh. Anal. Khim.*, **1982**, *37*, 1477-1483. c) Trinajstic N., Jericevic Z., Knop, J.V., Müller, W.R., Szymanski, K. *Pure Appl. Chem.*, **1983**, *55*, 379-390. d) Bangov, I.P. *MATCH*, **1983**, *14*, 235-246. e) Christie, B.D., Munk, M.E. *J. Chem. Inf. Comput. Sci.*, **1988**, *28*, 87-93. f) Kvasnicka, V., Pospíchal, J. *J. Chem. Inf. Comput. Sci.*, **1990**, *30*, 99-105. g) Bangov, I.P. *J. Chem. Inf. Comput. Sci.*, **1990**, *30*, 277-289. h) Kerber, A., Laue, R., Moser, D. *Anal. Chim. Acta*, **1990**, *235*, 221-228. i) Contreras, M.L., Valdivia, R., Rozas, R. *J. Chem. Inf. Comput. Sci.*, **1992**, *32*, 323-330. j) Balaban, T.S., Filip, P.A., Ivanciuc, O. *J. Math. Chem.*, **1992**, *11*, 79-105. k) Bytautas, L., Klein, D. J. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 1063-1078. l) Ouyang, Z., Yuan, S., Brandt, J., Zheng, C. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 299-303. m) Lukovits, I. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 361-366.

[55] a)Carhart, R.E., Smith, D.H., Brown, H., Djerassi, C. *J. Am. Chem. Soc.*, **1975**, *97*, 5755-5762. b) Carhart, R.E., Smith, D.H., Gray, N.A., Nourse, J.G., Djerassi, C. *J. Org. Chem.*, **1981**, *46*, 1708-1718. c) Dubois, J.-E., Carabedian, M., Dagane, I. *Anal. Chim. Acta*, **1984**, *158*, 213-233. d) Bremser, W., Fachinger, W. *Magn. Reson. Chem.*, **1985**, *23*, 1056-1071. e) Funatsu, K., Miyabayashi, N., Sasaki, S. *J. Chem. Inf. Comput. Sci.*, **1988**, *28*, 18-28. f) Christie, B.D., Munk, M.E. *J. Chem. Inf. Comput. Sci.*, **1988**, *28*, 87-93. g) Christie, B.D., Munk, M.E. *J. Am. Chem. Soc.*, **1991**, *113*, 3750-3757.

[56] a) Nourse, J.G. *J. Am. Chem. Soc.*, **1979**, *101*, 1210-1216. b) Balasubramanian, K. *Theor. Chem. Acta*, **1979**, *51*, 37-54. c) Nourse, J.G., Carhart, R.E., Smith, D.H., Djerassi, C. *J. Am. Chem. Soc.*, **1979**, *101*, 1216-1223. d) Nourse, J.G., Smith, D.H., Carhart, R.E., Djerassi, C. *J. Am. Chem. Soc.*, **1980**, *102*, 6289-6295. e) Abe, H., Hayasaka, H., Miyashita,Y., Sasaki, S. *J. Chem. Inf. Comput. Sci.*, **1984**, *24*, 216-219. f) Zlatina, L.A; Elyashberg, M.E. *MATCH*, **1992**, *27*, 191-207. g) Razinger, M., Balasubramanian, K., Perdih, M., Munk, M.E. *J. Chem. Inf. Comput. Sci.*, **1993**, *33*, 812-825. h) Ivanov, J., Karabunarliev, S., Mekenyan, O. *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 234-243. i) Contreras, M.L., Trevisiol, G.M., Alvarez, J., Arias, G., Rozas, R. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 475-482. j) Bytautas, L., Klein, D. J. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 803-818. k) El-Basil, S. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 572-. l) Almsick, M. van Dolhaine, H., Hönig, H. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 956-966.

[57] a) Bernal, J.M. *Desarrollo de un nuevo método de diseño molecular asistido por ordenador. Su aplicación a fármacos betabloqueantes y benzodiazepinas*. Doctoral thesis. Universitat de València (Estudi General), Spain, **1988**. b) Gálvez, J., García-Domenech, R., Bernal, J. M., García-March, F. J. *An. Real Acad. Farm.* **1991,** *57*, 533-546.

[58] Gálvez, J., García-Domenech, R., Moliner, R., García-March, F. J. Spanish patent P9101034, 1991.

[59] Gálvez, J., García-Domenech, R., Moliner, R. ; García-March, F. J. Spanish patent P9101134, 1991.

[60] García-March, F. J., García-Domenech, R., Gálvez, J., Antón-Fos, G.M., Julián-Ortiz, J.V. de,Giner-Pons, R., Recio-Iglesias, C. *J. Pharm. Pharmacol.*, **1997**, *49*, 10-15.

[61] García-Domenech, R., García-March, F.J., Soler, R.M., Gálvez, J., Antón-Fos, G.M., Julián-Ortiz, J.V. de *Quant. Struct.-Act. Relat.*, **1996**, *15*, 201-207.

[62] Baskin, I.I., Gordeeva, E.V., Devdariani, R.O., Zefirov, N.S., Palyulin, V.A., Stankevich, M.I. *Doklady Akad. Nauk. SSSR*, **1989**, *307*, 613-617.

[63] Faradzhev, I.A. in *Algorithmic Investigations in Combinatorics.* I.A. Faradzhev, ed., pp. 11-19. Nauka: Moskow, **1978**.

[64] a)Tratch, S.S., Lomova, O.A., Sukhachev, D.V., Palyulin, V.A., Zefirov, N.S. *J. Chem. Inf. Comput. Sci.*, **1992**, *32*, 130-139. b) Lomova, O.A., Sukhachev, D.V., Kumskov, M.I., Palyulin, V.A., Tratch, S.S., Zefirov, N.S. *MATCH*, **1992**, *27*, 153-174.

[65] Skvortsova, M.I., Baskin, I.I., Slovokhotova, O.L., Palyulin, V.A., Zefirov, N.S. *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 630-634.

[66] Molchanova, M.S., Shcherbukhin, V.V., Zefirov, N.S. . *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 888-899.

[67] Nilakantan, R., Bauman, N., Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.*, **1991**, *31*, 527-530.

[68] Carhart, R.E., Smith, D.H., Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.*, **1985**, *25*, 64-73.

[69] Nilakantan, R., Bauman, N., Dixon, J.S., Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.*, **1987**, *27*, 82-85.

[70] Venkatasubramanian, V., Chan, K., Caruthers, J.M. *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 188-195.

[71] Milne, G.W.A. *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 639-644.

[72] Zheng, W., Cho, S. J., Waller, C.L., Tropsha, A. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 738-746.

[73] a) Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A., Gordon, E. M. *J. Med. Chem.*, **1994**, *37*, 1233-1251. b) Zuckermann, R.N., Martin, E.J., Spellmeyer, D.C., Stauber, G.B., Shoemaker, K.R., Kerr J.M., Figliozzi, G.M., Goff, D.A., Siani, M.A., Simon, R.J., Banville, S.C., Brown, E.G., Wang, L., Richter, L.S., Moos, W.H. *J. Med. Chem.* **1994**, *37*, 2678-2685. c) Gruber, B. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 580-587.

[74] a) Gordon, E.M., Barrett, R.W., Dower, W.J., Fodor, S.P.A., Gallop, M.A. *J. Med. Chem.* **1994**, *37*, 1385-1401. b) Bunin, B.A., Ellmam, J.A. *J. Am. Chem. Soc.*, **1992**, *114*, 10997-10999. c) DeWitt, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6909-6913. d) Bunin, B.A., Plunkett, M.J., Ellmann, J.A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4708-4712. e) Plunkett, M.J., Ellmann, J.A. *J. Am. Chem. Soc.*, **1995**, *117*, 3306-3307. f) Murphy, M.M., Schullek, J.R., Gordon, E.M., Gallop, M.A. *J. Am. Chem. Soc.*, **1995**, *117*, 7029-7030. g) Thompson, L.A., Ellman, J.A. *Chem. Rev.*, **1996**, *96*, 555-600. h) Kurth, M.J., Randall, L.A.A., Takenouchi, K. *J. Org. Chem.*, **1996**, *61*, 8755-8761. i) Sim, M.M., Lee, Ch.L., Ganesan, A. *J. Org. Chem*, **1997**, *62*, 9358-9360. j) Warr, W.A. *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 134-140.

[75] a) Bures, M.G., Martin, Y.C. *Curr. Opin. Chem. Biol.*, **1998**, *2*, 376-380. b) Bayada, D.M., Hamersma, H., Geerestein, V.J. van *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1-10. c) Leach, A.R., Bradshaw, J., Green, D.V.S., Hann, M.M., Delany, J.J. III *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1161-1172. d) Matter, H., Pötter, T. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1211-1225. e) Darvas, F., Dormán, G., Papp, Á. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 314-322. f) Pickett, S.D., McLay, I.M., Clark, D.E. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 263-272. g) Martin, E., Wong, A. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 215-220. h) Stanton, R.V., Mount, J., Miller, J.L. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 701-705. i) Agrafiotis, D.K., Lobanov, V.S. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 1030-1038. j) Gorse, D., Lahana, R. *Curr. Opin. Chem. Biol.*, **2000**, *4*, 287-294.

[76] Martin, E.J., Blaney, J.M., Siani, M.A., Spellmeyer, D.C., Wong, A.K., Moos, W.H. *J. Med. Chem.*, **1995**, *38*, 1431-1436.

[77] Brown, R.D., Martin, Y.C. *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 572-584.

[78] Hassan, M., Bielawski, J.P., Hempel, J.C., Waldman, M. *Mol. Diversity*, **1996**, *2*, 64-74.

[79] Wang, J., Ramnarayan, K. *J. Comb. Chem.*, **1999**, *1*, 524-533.

[80] Kirkpatrick, D.L., Watson, S., Ulhaq, S. *Comb. Chem. High Throughput Screen.*, **1999**, *2*, 211-221.

[81] Hopfinger, A.J., Reaka, A., Venkatarangan, P., Duca, J.S., Wang, S. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1151-1160. d) Baxter, C.A,; Murray, C.W., Waszkowycz, B., Li, J., Sykes, A., Bone, R.G.A., Perkins, T.D.J., Wylie, W. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 254-262.

[82] a) Sheridan, R.P., Kearsley, S.K. *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 310-320. b) Zheng, W., Cho, S.J., Tropsha, A. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 251-258. c) Cho, S.J., Zheng, W., Tropsha, A. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 259-268.

[83] Li, J., Murray, Ch.W., Waszkowycz, B., Young, S.C. *Drug Discovery Today*, **1998**, *3*, 105-112.

[84] Grassy, G., Calas, B., Yasri, A., Lahana, R., Woo, J., Iyer, S., Kaczorek, M., Floc'h, R., Buelow, R. *Nature Biotechnol.*, **1998**, *16*, 748-752.

[85] Klopman, G., Raychandhury, C. *J. Chem. Inf. Comput. Sci.*, **1990**, *30*, 12-19.

[86] Kier, L.B., Hall, L.H., Frazer, J.W. *J. Chem. Inf. Comput. Sci.*, **1993**, *33*, 143-147.

[87] Kier, L.B., Hall, L.H., Frazer, J.W. *J. Chem. Inf. Comput. Sci.*, **1993**, *33*, 148-152.

[88] Hall, L.H., Dailey, R.S., Kier, L.B. *J. Chem. Inf. Comput. Sci.*, **1993**, *33*, 598-603.

[89] Kier, L.B., Hall, L.H. *Quant. Struct.-Act. Relat.*, **1993**, *12*, 383-388.

[90] Kvasniecka, V., Pospíchal, J. *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 516-526.

[91] Metropolis, M., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.: Teller, E. *J. Chem. Phys.*, **1953**, *21*, 1087-1092.

[92] Kvasnicka, V., Pospíchal, J. *J. Math. Chem.*, **1992**, *9*, 181-196.

[93] Julián-Ortiz, J.V. de,Gálvez, J., Muñoz-Collado, C., García-Domenech, R., Gimeno-Cardona, C. *J. Med. Chem.*, **1999**, *42*, 3308-3314.

[94] a) Mekenyan, O., Bonchev, D., Balaban, A.T. *J. Math. Chem.*, **1988**, *2*, 347-375. b) Diudea, M. V., Kacsó, I. E. *MATCH*, **1991**, *26*, 255-269.

[95] a) Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. *J. Mol. Biol.*, **1982**, *161*, 269-288. b) Moon, J.B., Howe, W.J. *Proteins: Struct. Func. Genet*., **1991**, *11*, 314-328. c) Nishibata, Y., Itai, A. *Tetrahedron*, **1991**, *47*, 8985-8990. d) Lawrence, M.C., Davis, P.C. *Proteins: Struct. Func. Genet.*, **1992**, *12*, 31-41. e) Verlinde, C.L.M.J., Rudenko, G., Hol, W.G.J. *J. Comput.-Aided Mol. Design*, **1992**, *6*, 131-147. g) Rotstein, S.H., Murcko, M.A. *J. Med. Chem.*, **1993**, *36*, 1700-1710.

[96] Nishibata, Y., Itai, A. *J. Med. Chem.*, **1993**, *36*, 2921-1928.

[97] a) Böhm, H.-J. *J. Comput.-Aided Mol. Design*, **1992**, *6*, 61-78. b) Böhm, H.-J. *J. Comput.-Aided Mol. Design*, **1996**, *10*, 265-272.