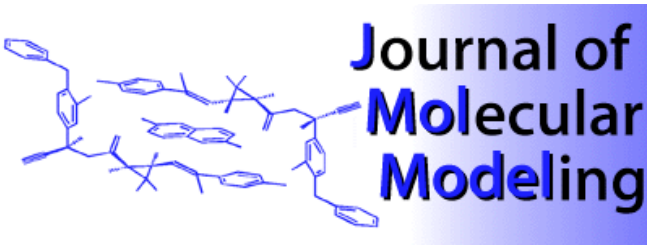
J. Mol. Model. **2000**, *6*, 358 – 378



FULL PAPER

The Search for the Spatial and Electronic Requirements of a Drug

Sandra Handschuh and Johann Gasteiger

Computer-Chemie-Centrum, Institut für Organische Chemie, Universität Erlangen-Nürnberg, Nägelsbachstr. 25, D-91052 Erlangen, Germany. E-mail: gasteiger@ccc.chemie-uni-erlangen.de

Received: 15 November 1999/ Accepted: 25 January 2000/ Published: 29 February 2000

Abstract A new program called GAMMA (genetic algorithm for multiple molecule alignment) has been developed for the superimposition of several three-dimensional chemical structures. Superimposition of molecules and evaluation of structural similarity is an important task in drug design and pharmaceutical research. Similarities of compounds are determined by this program either based on their structural or their physicochemical properties by defining different matching criteria. These matching criteria are atomic properties such as atomic number or partial atomic charges. The program is based on a combination of a genetic algorithm with a numerical optimization process. A major goal of this hybrid procedure is to address the conformational flexibility of ligand molecules adequately. Thus, only one conformation per structure is necessary and the program can work even when only one conformation of a compound is stored in a database. The genetic algorithm optimizes in a nondeterministic process the size and the geometric fit of the overlay. The geometric fit of the conformations is further improved by changing torsional angles combining the genetic algorithm and the directed tweak method. The determination of the fitness of a superimposition is based on the Pareto optimization. As an application the superimposition of a set of Cytochrome P450c17 enzyme inhibitors has been performed.

Keywords Genetic algorithm, Three-dimensional superposition of structures, Cytochrome P450c17 inhibitors, Structural similarities, Alignment of 3D structures

Introduction

A century ago the spatial requirements for a molecule to act as a specific drug were compared to the requirements for a key fitting into a given lock. In recent decades molecular modeling techniques have established and refined this simple picture. Much effort has been devoted to determining the shape of the lock by X-ray structure determination of receptor proteins in order to then be able to select the proper key, a ligand that will fit into the active site of the receptor.

However, the number of receptors where the 3D structure is known is still small compared to the number of known receptors. It is quite clear that many proteins can never be crystallized or their structure will dramatically change when taken out of their natural environment, such as for membrane proteins. In such a situation, other experimental techniques such as NMR spectroscopy can be applied to derive the 3D structure of a protein. However, for a long time we will have to live with a situation where the 3D structure of the receptor is *not* known.

Correspondence to: J. Gasteiger

Dedicated to Professor Paul von Ragué Schleyer on the occasion of his 70th birthday

How can we then learn anything about the spatial requirements of a ligand to fit into such a receptor whose structure is not known? Can we learn something about how a key should look from a series of keys that fit into a certain lock? Can we learn about the spatial and electronic requirements for a ligand by comparing a set of ligands that are known to bind to the receptor of interest?

In this publication we will present a method that attempts to extract the largest three-dimensional substructure that is common to a set of molecules or ligands, much in the same way as the comparison of a set of keys will give us the essential features necessary for a key to fit into a given lock.

The investigation of a series of ligands binding to the same receptor is usually performed by defining the similarities between the ligands through a pharmacophore pattern. A pharmacophore defines the three-dimensional arrangement of substructural units such as hydrogen bonding or hydrogen accepting sites or hydrophobic areas in a molecule. Usually, no more than three or four such sites are identified for the definition of a pharmacophore in order not to make pharmacophore searches too time-consuming.

The pharmacophore pattern of a set of ligands can be derived from the largest 3D substructure that these compounds have in common. The methods initially developed for searching for the three-dimensional maximum common substructure (MCSS) only worked with a single, rigid conformation, not taking into account the conformational flexibility of the ligands.[1–4] The first detailed study of distance-based methods for 3D similarity searching was published by Pepperrell and Willett.[5] More recently Sheridan *et al.*[6] reported on distance based methods using several conformations for each structure. Angle-based and fragment-based methods[7, 8] like those of Fisanick *et al.*[9] have also been used to calculate 3D similarities. Overviews on 3D substructure and pharmacophore searching are contained in several sections of the Encyclopedia of Computational Chemistry.[10-13]

In our group, we have developed an approach for MCSS search that is based on atom mappings. The approach was initially developed to be applied to the constitution of a molecule as given by a connection table.[14] However, it was shown that this method can be extended to the 3D structure of a molecule including conformational flexibility of the ligands.[13, 15]

In order to perform the search for the largest three-dimensional substructure, a universal access to the 3D structure of molecules is necessary. Such an approach can be provided by the automatic 3D structure generator CORINA.[16–18] All 3D structures investigated in this report have been obtained by CORINA which can provide a single low-energy conformation of any organic molecule.

The 3D substructure search starts with one conformation for each structure and investigates the conformational flexibility during the optimization process. A "query-directed" conformational search technique was implemented [14] by combining evolutionary theories with a numerical optimizer. [19] This hybrid technique of a genetic algorithm combined with a directed tweak method based on numerical optimization is a flexible search system that accounts for conformational flexibility by rotation around single bonds during the optimization process.

This hybrid method, called GAMMA (genetic algorithm for multiple molecule alignment) has been extended to the simultaneous superimposition of a set of conformationally flexible molecules. An approach was chosen that is able to include specific knowledge to the problem. Thus, special pharmacophore features have been implemented to explore the physicochemical atom properties, such as electronegativities or atomic charges, can be chosen as matching criterion. It is possible to preselect atoms that have to or should be part of the substructure. Rotatable bonds are automatically recognized or alternatively can be selected by the user and last but not least entire molecules can be chosen as rigid or flexible.

General principles of the genetic algorithm

Genetic algorithms (GAs) represent robust optimization methods that are based on the mechanisms of natural selection and genetics.[20–23] They can solve problems involving large search spaces efficiently, and thus, can even be applied to problems beyond the reach of classical exhaustive search methods.[21,22] A GA imitates nature's methods for adapting to a changing environment. Optimization therefore does not start from a single point, but from a population of starting points that is randomly generated. These starting points

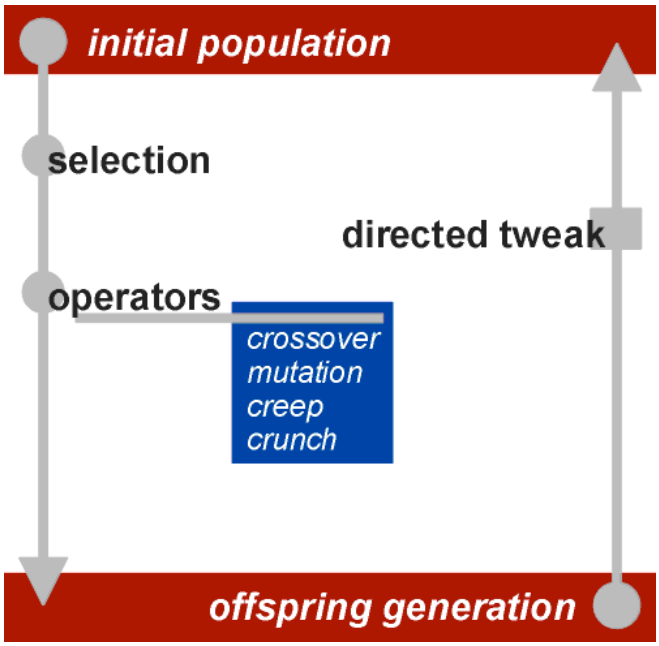


Figure 1 *Outline of a complete GA run showing the application of the genetic operators onto an initial population of superpositions. The new generation is submitted to the directed tweak mechanism, which improves on the geometric fit of the individuals*

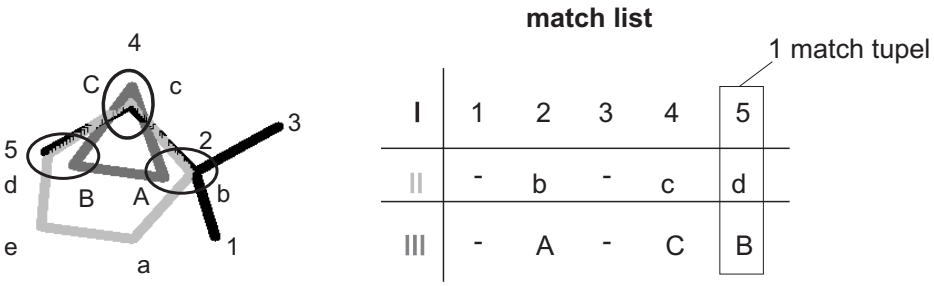


Figure 2 *The first chromosome in a multiple superimposition problem: the match list. The number of molecules,* n, *is here 3, consisting of the molecules I, II, and III. The size of the substructure,* N, *the number of matching atoms is also 3*

correspond to the chromosomes or individuals of a population representing potential solutions to the search problem. The individuals must be represented by a special coding of the parameters of the function to be optimized. Quite often a bit string is chosen for this purpose. In the present case, we have chosen two chromosomes, one representing the superposition of atoms, the other the torsional angles. The genetic operators *selection*, *mutation*, and *crossover* are applied iteratively to the population. In the method presented here, two additional operators called *creep* and *crunch* that are tailored to the specific problem have been implemented. In the search for the maximum 3D common substructure two conflicting criteria must be optimized: the number of matching atoms between two molecules has to be maximized, whereas the deviations in the coordinates of the superimposed atoms must be minimized. These two criteria are monitored separately by a so-called Pareto fitness. The Pareto fitness is not based on one fitness function but on several parameters that are treated independently of each other. After the selection process, the genetic operators are applied to the chromosomes and a new population forms the offspring generation. One complete GA run begins with the initialization of the individuals and ends with obtaining one set of optimized solutions after cycling through all generations. Figure 1 shows the general outline of the genetic algorithm.

Genetic algorithms are not based on a deterministic procedure. Therefore, optimization by a GA does not necessarily arrive at the optimum solution. In order to alleviate this problem, an additional method, the *directed-tweak* [18] procedure was implemented to match the conformations of the molecules to be overlaid. The geometric fitness of the offspring population is assessed by minimizing differences in the conformations during the *directed tweak* procedure. The result is only part of the fitness values, the geometry of the molecules is not changed. This helps to prevent the loss of genetic variety during the optimization process. The actual conformations are changed once by a directed tweak optimization after the last generation. Usually more than one GA run is performed to arrive at an optimum solution.

The chromosomes

A major task in adapting a genetic algorithm to a specific problem is the encoding of the individuals of the population, i.e., the representation of the genetic information by chromosomes.[19] We have chosen an approach that represents an individual by two independent chromosomes. The first chromosome consists of an atom mapping that is coded by integers and represented as a fixed-length linked match list (Figure 2). The match list is defined by the number, *n*, of molecules to be superimposed and the size of the substructure *N* (number of complete match tupels, Figure 2: $N = 3$).

The match list is a fixed-length linked list comprising all individuals. Each atom may appear only once. To initialize the match lists, first, all permutations of atom mappings are generated. The maximum number of possibilities is $N_x N_y N_z$, where N_x is the number of atoms in molecule x.

In this process, different criteria can be chosen for the atoms that are matched. Either one requires matching atoms to have the same atomic number or to have a certain physicochemical property, e.g., partial atomic charge, in a given interval.

A list of matching atoms is built by randomly selecting match tupels from the initialized complete set of match tupels. The number of individuals to be built, i.e., the size, *i*, of the population is set at the beginning. If an atom is doubly referenced, i.e., an atom appears twice in the match list, after

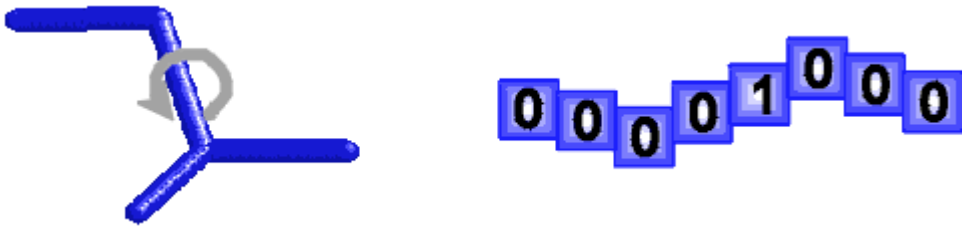
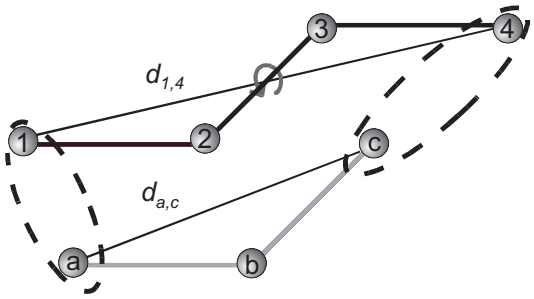


Figure 3 *The chromosomes for coding torsion angles by 8 bits for each torsion angles*



$$D = \frac{1}{4}\frac{n\cdot(n-1)}{2}\sum_{\substack{i,j\\ j\neq i}}^{N}\sum_{\substack{k,l\\ l\neq k}}^{n}(d_k(i,j)-d_l(i,j))^2$$

N	= number of match pairs
n	= number of molecules
i,j	= indices of match tupels to be compared
$d_k(i,j)$, $d_l(i,j)$	= atom distances of match tupel i to match tupel j in molecule k and molecule l

here: $d_k(i,j) = d_{1,4}$

$d_l(i,j) = d_{a,c}$

$N = 2$

$n = 2$

Figure 4 *The two optimization criteria: size of the substructure* N, *and the distance parameter* D

random combination of match tupels, this atom is changed into an atom that is not yet part of the list, or, alternatively, into a zero mapping, (i.e., it is removed from a match list). Zero mappings are marked by a dash in Figure 2.

The second chromosome consists of a bit string representing torsional angles in the flexible molecules (Figure 3). Each single bond that has at either end at least one multi-atom substituent (e.g. a methyl group), but is not a ring bond, is defined as flexible. Each torsional angle is binary coded by 8 bits. Thus, the torsion angles of –180° up to +180° are represented by integer values of 0 to 255. The integer values are then binary Gray-coded.[19] Gray-coding is a specific presentation of an integer value by a bit string. The smallest possible change of the angles is 1.4° (360/256). All torsional angles are concatenated to one bit string. Thus, each bit string of torsional angles has the length of $8n_{tor}$, with n_{tor} being the number of torsional angles in all molecules (in Figure 3, $n_{tor} = 1$).

The two chromosomes, the match list and the bit string of torsion angles, represent together one individual. Several individuals build the population of one generation.

Optimization criteria

The search for the MCSS of a set of molecules takes into account two optimization criteria: the size of the substructure, as given by the number, *N*, of matches (Figure 3), and the geometric fit of the matching atoms as represented by a distance parameter (Figure 4). The distance parameter, *D*, consists of the sum of the squared differences of corresponding atom distances in the molecules.

D is related to the root mean square (*rms*) error of the distances of corresponding atoms in an optimized superimposition. This means that high D-values correlate with a bad geometric fit, as high rms-values do. The D-value has been shown to be sufficient to evaluate the geometric fit during the optimization process, above all because the calculation does not take large computation times. The D-value is a relative fitness value, in contrast to the rms-value that is absolute and can be used for the comparison of superimpositions of different structures. The *rms* value is, however, subject to large changes even if the mapping changes only slightly. Therefore, the distance value, *D*, is better adapted to the spe-

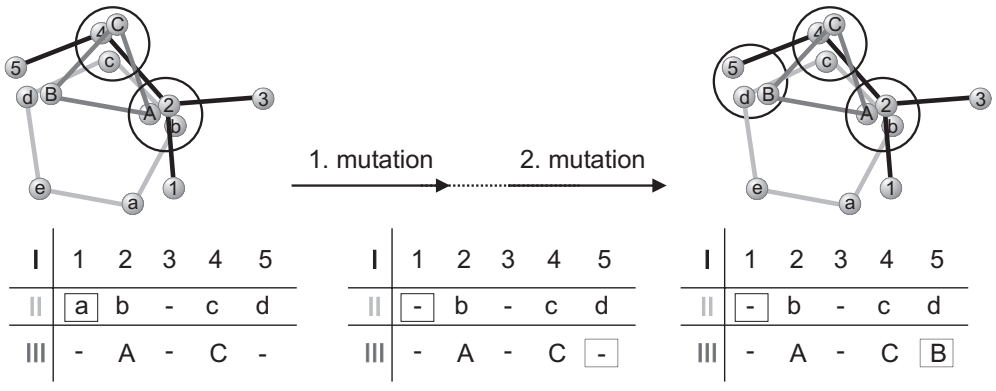
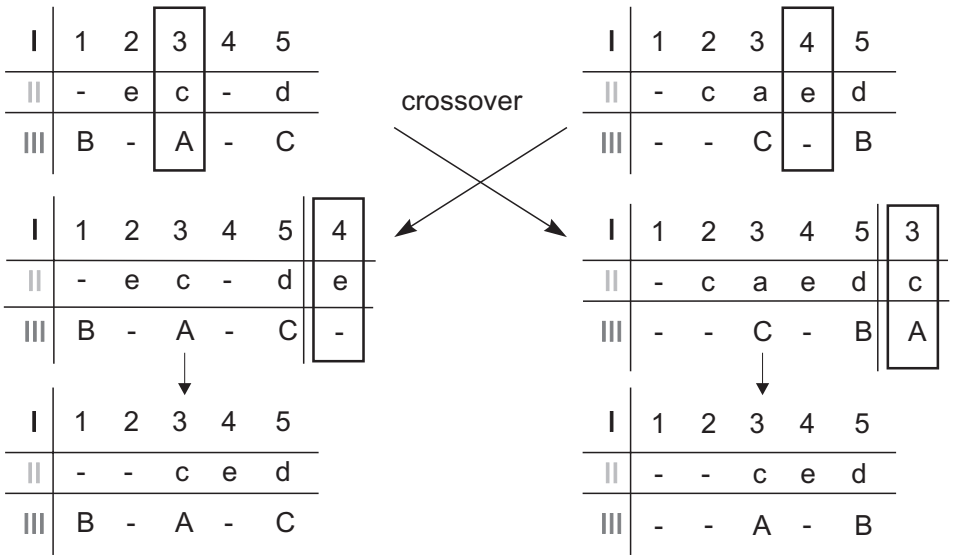


Figure 5 *The mechanism of the multiple mutation operator onto the match lists. If a superimposition contains n (e.g. 3) molecules, n-1 (here 2) mutations are performed. For each mutation, an atom of a match tupel is selected and changed into another atom or a zero mapping (marked here by a rectangle)*

Figure 6 *The mechanism of the multiple crossover operator onto the match lists*



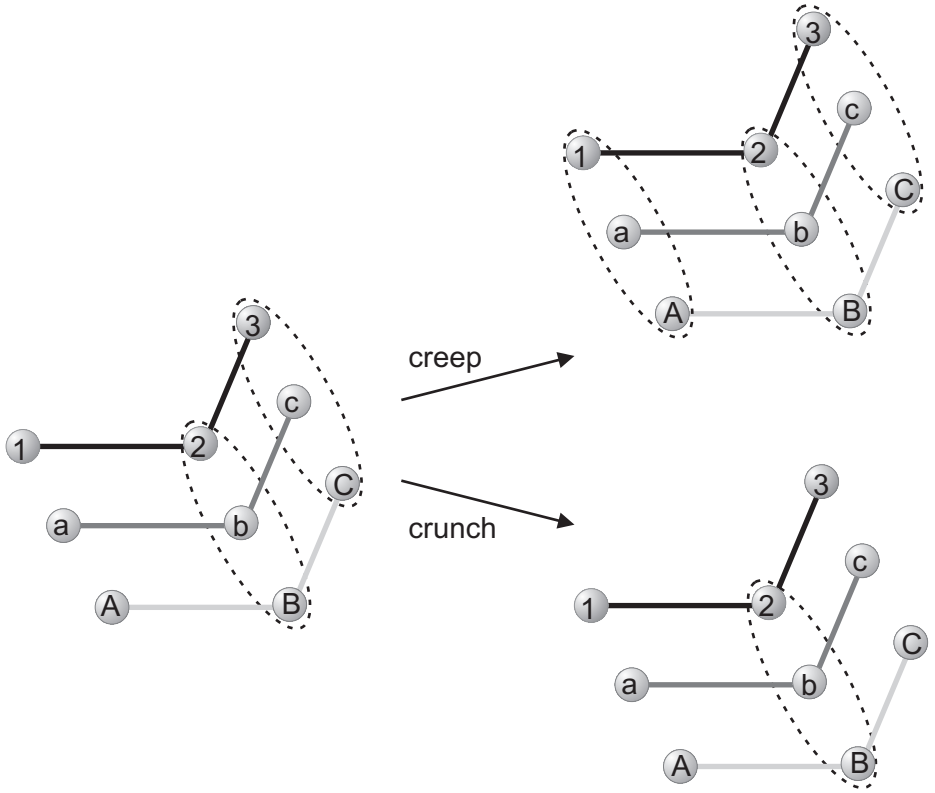
cific use during a GA optimization than the *rms* value. The *rms* value of the obtained superimposition is calculated only once at the end of each GA run in order to present the results.

Genetic and non-genetic operators

Mutation The genetic operators change the two chromosomes of the individuals, the match list and the coding of the torsional angles, in a different manner.

The mutation operator working on the match lists randomly changes atom tupels (Figure 5). To mutate the match list of a superimposition of *n* molecules, *n-1* mutation points are selected at random (Figure 5). The atoms of all molecules except those of the first (largest) molecule in the match list can be mutated. The corresponding atom of a match tupel (for the first mutation, an atom of the second molecule, for the second mutation an atom of the third molecule, etc.) is mutated by obeying the following boundary condition: none of the atoms is allowed to appear more than once in the re-

Figure 7 *The mechanism of the creep and crunch operators. The creep operator leads to a larger substructure, whereas the crunch operator produces a smaller substructure*



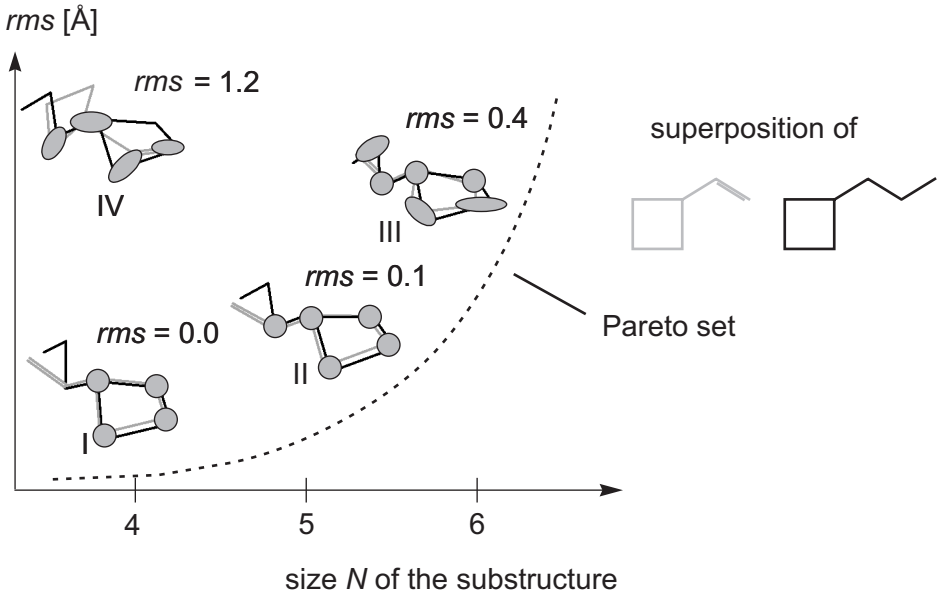


Figure 8 *The Pareto set of a superimposition of vinylcyclobutane and n-propylcyclobutane*

sulting match list and the criterion which atoms are allowed to be matched has to be taken into account. Hence, the atom considered must be changed into one that is not yet in the match list. If all atoms of a molecule are already referenced, the atom is changed into a zero mapping. Zero mappings can also be introduced randomly (see 1. mutation in Figure 5: the match tupel 1,a,- is mutated to 1,-,-).

Mutation in the chromosome representing the torsion angles inverts one bit of a binary coded torsion angle string (1 into 0 or vice versa). As mentioned before, each chromosome of the torsional angles is a bit-string of length $8n_{tor}$ (n_{tor} = number of rotatable bonds in all molecules). One mutation is performed for each angle. The first mutation point is selected randomly and each additional point has an 8 bit dis-

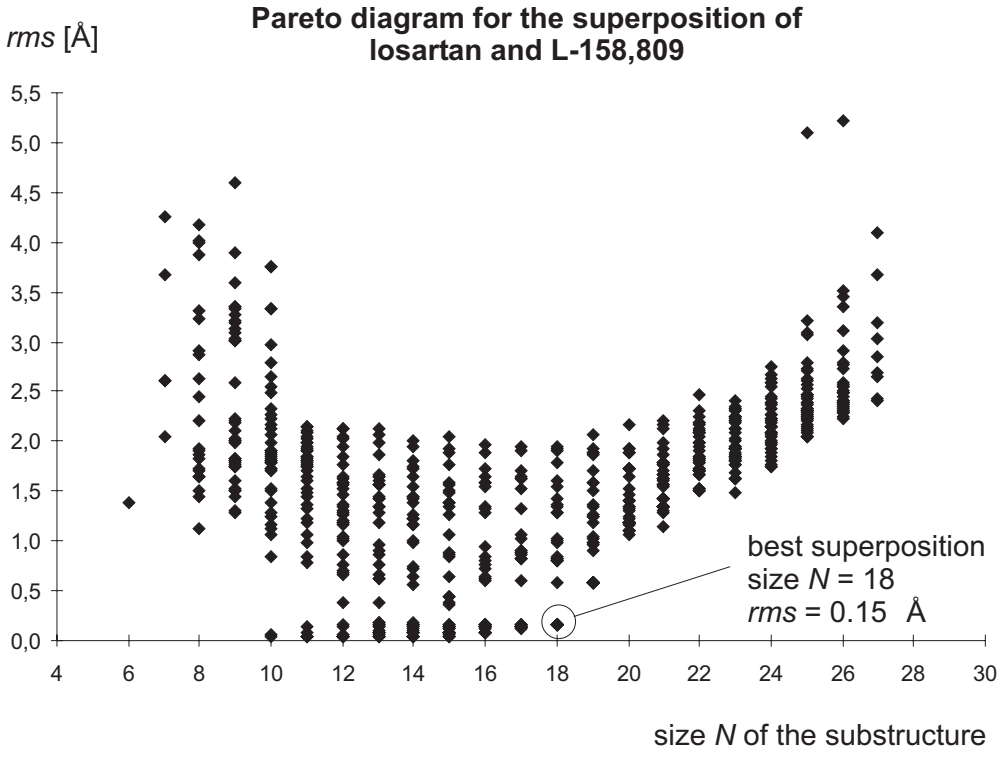


Figure 9 *The Pareto diagram of the superimposition of losartan and L-158,809 obtained in 40 GA runs. Each run results in one Pareto set: one best solution for each substructure size*

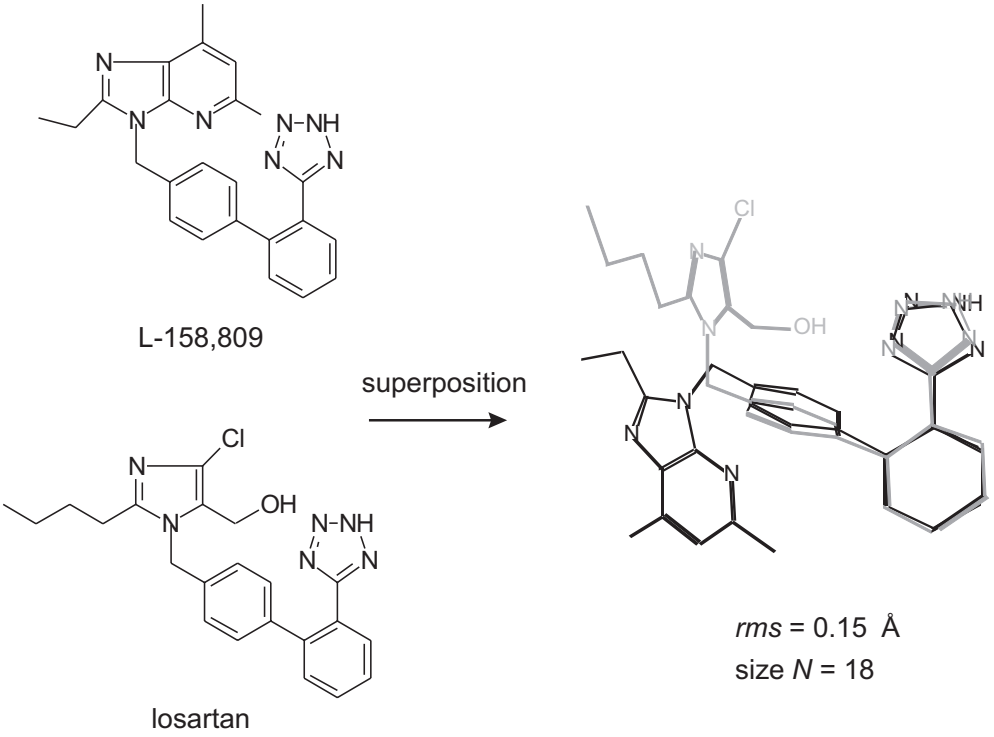


Figure 10 *The best superimposition of losartan and L-158,809 extracted from 40 GA runs*

tance to the first one. Thus, an equal distribution of the mutations on the torsional angles is guaranteed.

Crossover The crossover operator exchanges random parts of two individuals, i.e., partial substructures, and combines partial solutions of the MCSS search problem in a new, and potentially better, way. Two points are chosen randomly in the match lists of two parental individuals (Figure 6).[24] The information string that is to be crossed is contained in between these two points. Each partial list must be of equal length and is copied to the tail of the other parental individual. In this step, double references may be introduced that later have to be deleted: If an atom of molecule I appears twice in the match list, the corresponding original match pair must be replaced by the new one that was copied to the tail (e.g. in Figure 6 match tupel 4,e,- replaces 4,-,- and 3,c,A replaces 3,a,C). Any double references remaining after this process in molecule II and III are replaced by randomly chosen ones conforming to the constraints (the matching criteria). If there are no more atoms that obey these restrictions, a zero mapping must be introduced. This procedure ensures that the match lists always have the same length and that each atom is referenced only once.

The crossover operator working on the representation of the torsion angles is a one-point crossover. One point must be chosen randomly at the same position in the two parental strings. Crossover exchanges two parts of the two chosen parental strings of torsion angles. This leads to new conformations for which the geometric fit has to be assessed.

Two non-genetic operators: creep and crunch Two additional operators were developed to improve the efficiency of the GA: The *creep* and the *crunch* operators. These operators do not act stochastically like the genetic operators crossover and mutation but make use of knowledge specific to the problem to be solved, the MCSS search problem.[25] Hence, they are called 'knowledge-augmented operators'.[19]

The *creep* operator increases the size of the substructure by adding a matching tupel of atoms to the match list while obeying restrictions imposed by the spatial arrangement of the atoms (Figure 7). The new matching atom tupel must not cause a large increase in the *rms* value of the original match. In this way, the creep operator leads to a "hill climbing" mechanism in the GA.

The *crunch* operator (Figure 7) acts as an antagonist to the creep operator in reducing the size of the substructure. The goal of the crunch operator is to eliminate match pairs that are responsible for bad geometric distance parameters. This operation should help to avoid the search becoming trapped in local minima during the optimization process.

The Pareto fitness of individuals

The MCSS search is a multi-criteria optimization problem, where the notion of optimality is difficult to define. Two main principal parameters contribute to the fitness of a superimposition and have to be optimized: the size of the substructure and its geometric fit. The substructure size must be as large as possible, whereas the deviation in the positions of the superimposed atoms should be as low as possible. These criteria are contradictory as a larger substructure may decrease the geometric fit. An optimum must be found that takes both criteria into account. Vilfredo Pareto developed a con-

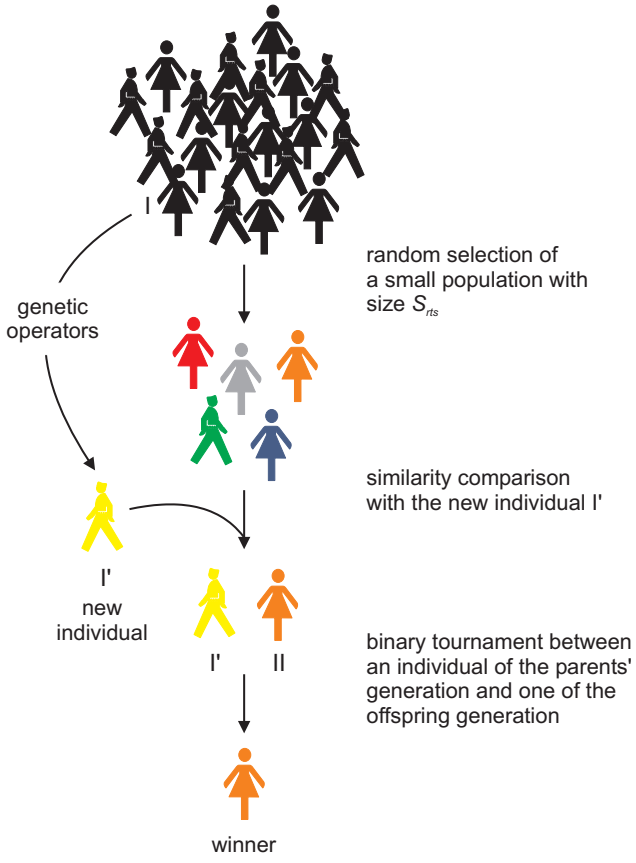


Figure 11 *The restricted tournament selection. Tournaments are held only between similar individuals. This guarantees the preservation of variety in genetic information*

cept for solving multi-criteria optimization.[19, 20] Pareto optimization means that an optimized state is reached if none of the parameters can be improved further without making another one worse.

Pareto optimality applied to the MCSS search problem in a three-dimensional space results in simultaneously maximizing the size of the substructure and optimizing the geometric fit. This does not result in obtaining only one probably perfect substructure but for each possible size of the common substructure an optimal geometric fit is produced.

The application of Pareto optimization to the superimposition of vinylcyclobutane and propylcyclobutane is shown in Figure 8. The atoms marked in gray are those of the common substructure. The result of a Pareto optimization is a set of common substructures for which the geometric deviation cannot be minimized further. Four different superimpositions are shown in Figure 8, three of them corresponding to Pareto optimality. Superimposition I dominates superimposition IV as it has a smaller *rms* value for the same number of matching atoms (Figure 8). In this sense, superimposition I represents a Pareto optimal solution because no other substructure can be found which has a better geometric correspondence. Superimpositions II and III are also members of the Pareto set and no other superimpositions having the same sizes and better geometric fits can be found. Taken together, superimpositions I, II, and III represent the set of equivalent Pareto solutions and none dominates the other.

For each specific superimposition a Pareto diagram can be calculated. It presents the development of the *rms* error with the size, *N*, of the substructure during the GA optimization runs. Figure 9 shows the Pareto diagram for the superimposition of two angiotensin antagonists losartan and L-158,809 (Figure 10).[26] Forty GA optimization runs were performed. The figure shows the set of Pareto solutions consisting of one superposition with lowest *rms* for each number of matching atoms. From among this Pareto optimality set, the superimposition with a substructure size of 18 and an *rms* error of 0.15 Å might be chosen to be the best one. The corresponding point in the Pareto diagram is indicated by a circle.

This superimposition of losartan and L-158,809 extracted from the Pareto diagram is shown in Figure 10.

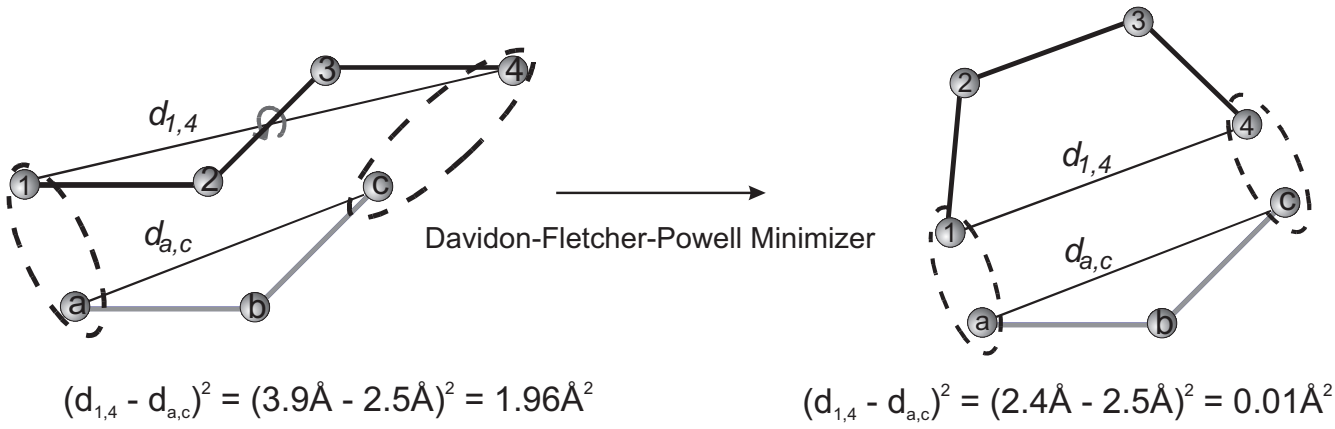


Figure 12 *The directed tweak method is used to minimize differences in the conformations of the two structures to be superimposed*

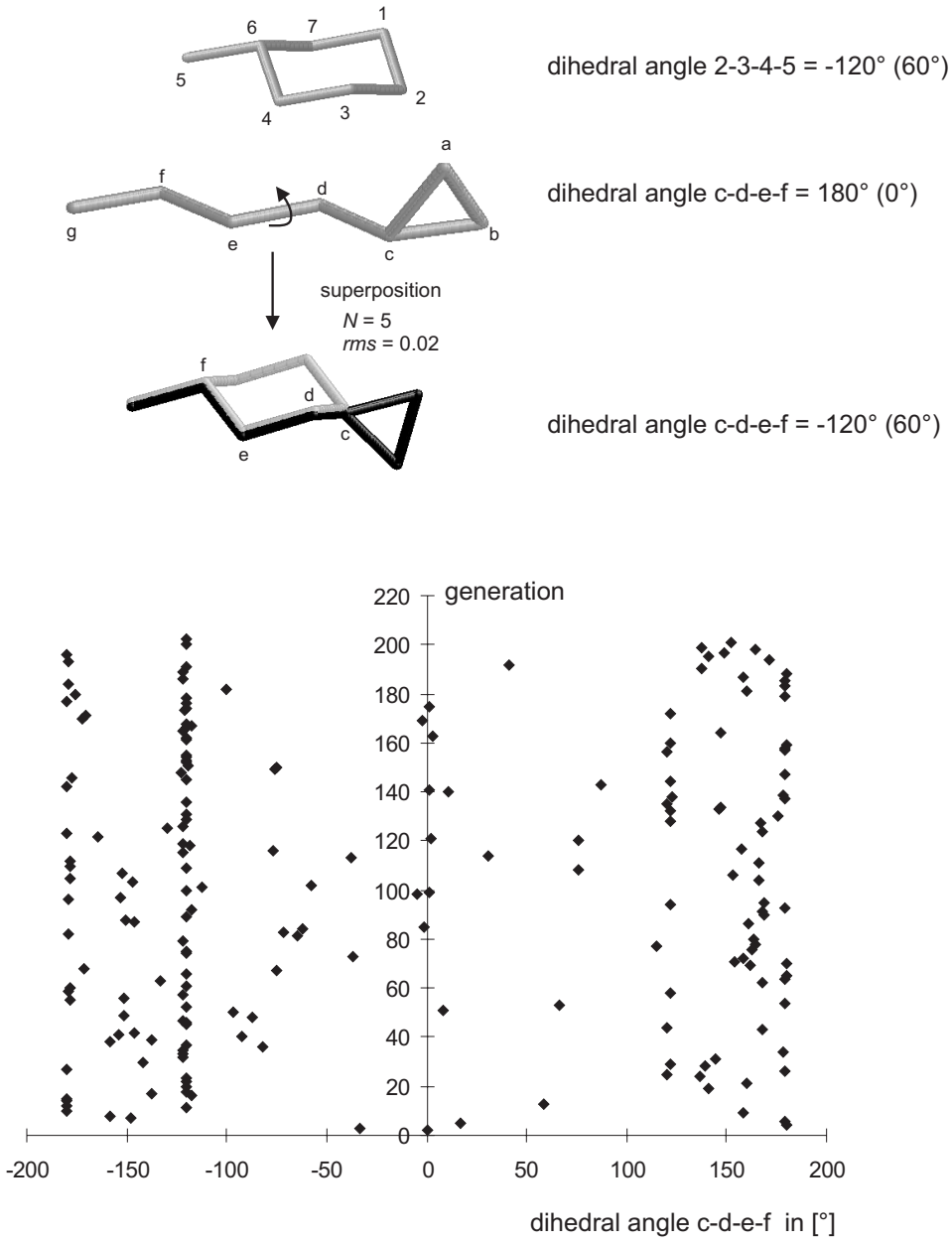


Figure 13 *Adaptation of the conformations of methylcyclohexane and n-butylcyclopropane*

Restricted tournament selection

Selection drives the optimization and causes evolutionary pressure: The selection operator moves individuals from one generation into the next one based on their relative fitness. This corresponds to Darwin's evolution theory of 'survival of the fittest'. Most of the GAs described in literature make use of the procedure of *roulette wheel selection*:[21, 22, 27–29] Each individual is assigned to a sector of a roulette wheel with the sector being proportional to the fitness of the individual: The better the fitness the larger the sector. Hence, the size of the sector corresponds to the probability of an individual being selected as a parent of the next generation. In order to prevent convergence to a suboptimal solution the population must consist of diverse and relevant members and the rapid decrease of genetic variety is to be prevented. We have decided to choose a special selection type to prevent premature loss of genetic information that might occur in a roulette wheel selection procedure. This alternative is called *restricted tournament selection* (RTS) (Figure 11).[27] Restricted tournament selection is found to be useful for solving multimodal problems and is a modification of a binary tournament selection. In a binary tournament selection, tournaments for a place in the new population are held between pairs of individuals chosen at random from the entire population. In this sense "restricted" means that tournaments are

not made between any individuals chosen at random from the entire population but only between similar individuals.

Thus, restricted tournament selection (Figure 11) is based on the concept of local competition. The winners of each tournament are moved into the next generation. An element I is chosen randomly from the basic population and changed by the operators of the GA into a new element I'. For each I' a small population with an optional member size S_{rts} is selected from the basic population. The individual II that is most similar to I' among the chosen individuals is saved. I' must then compete with II for a place in the new population. This form of binary tournament restricts an individual from competing with individuals too different to it. Hence, the variety in the information is maintained. A further advantage of the described mechanism of RTS is the possibility for the so-called continuous selection. A continuous selection allows individuals from different generations (e.g. II and I' in Figure 11) to compete with each other.

Matching the conformations - directed tweak

The directed tweak method reported by T. Hurst [18] was implemented in our procedure. The objective was to combine non-deterministic genetic mechanisms with a numerical optimizer in order to improve potential solutions. After each generation, the geometric fit of each individual or superimposition is improved by mapping torsional angles by the directed tweak method.

The technique makes use of the Davidon-Fletcher-Powell optimizer [30] to minimize differences in the conformations (Figure 12). The squared differences of the distances of corresponding atom pairs (i.e., 1,4 and a,c in Figure 12) are used to minimize the differences in the geometry of the superimposed structures by changing torsion angles. The superimpositions are not limited to low-energy conformations. This allows one to find also conformations of ligands that correspond to those found in the binding of a ligand to a receptor but do not correspond to low-energy conformations in the free state. However, an energy penalty value is added to the distance parameter *D* if a close contact of non-binding atoms is found in a conformation.

The superimposition of methylcyclohexane and n-butylcyclopropane (Figure 13) shows the adaptation of the conformations of two molecules during the optimization process. The dihedral angle c-d-e-f of n-butylcyclopropane fits onto the rigid cyclohexane conformation after rotation around the bond d-e.

The optimization process covers the entire conformational space of bond d-e during the generations. Figure 13 shows the distribution of the dihedral angle c-d-e-f of n-butylcyclopropane during one run of the genetic algorithm. The optimization culminates most often in an angle of -120° (or 180°-120° = 60°), which corresponds to the conformation of cyclohexane (dihedral angle 2-3-4-5).

Special features of the program

Close contact check of van der Waals radii

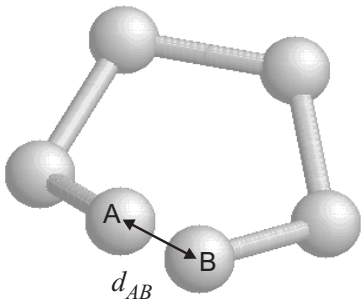
The algorithm treats atoms as points that have no spatial expansion. To prevent conformations from having an overlap of van der Waals radii, the distances of non-bonded atoms are calculated and compared with the sum of the corresponding van der Waals radii (Figure 14).

If a close contact is found, the distance parameter D is multiplied by a penalty factor. Thus, the conformation penalty is part of an optimization criterion. Consequently, individuals representing an unfavorable conformation obtain a high D parameter (bad geometric fitness) and will never dominate in a Pareto tournament.

Matching criteria

Many surface properties, e.g., hydrogen bonding potential, electrostatic potential, or hydrophobicity, are responsible for high receptor binding affinities. These binding properties are mainly based on dipole-dipole interactions and are related to various electronic effects. Thus, we have built into our procedure the option that criteria other than the atomic number, such as physicochemical properties of atoms, ranges of par-

Figure 14 *Close contact check of van der Waals radii*



$$\min_{AB} = 0{,}75 \cdot (vdW(A) + vdW(B))$$

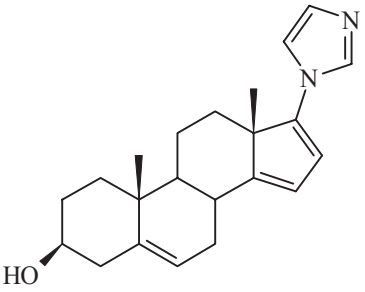
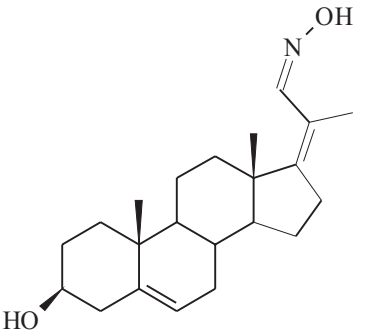
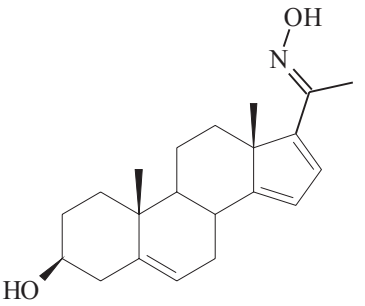
$$fac = \frac{d_{AB}}{\min_{AB}}$$

if

fac < 0.5 : close contact

fac > 0.5 : no close contact

Table 1a *Cytochrome P450c17 inhibitors of the A/B/C/D-ring systems and their $IC_{50\ Human}$-values (progesterone, 25 μM)*

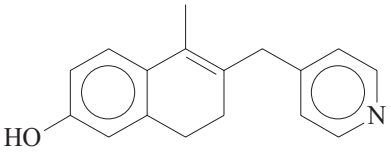
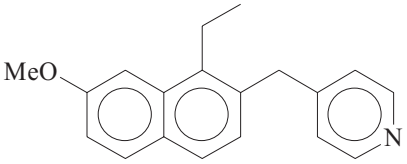
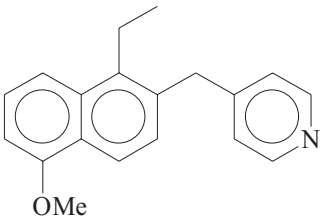
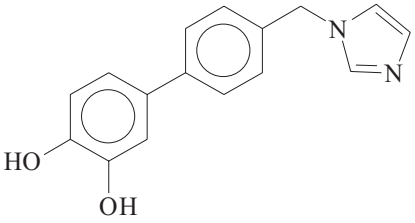
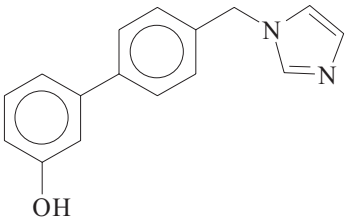
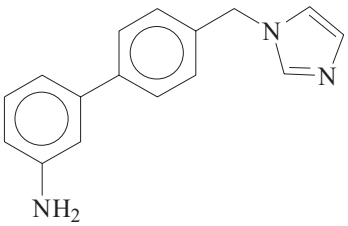
Chemical structure	Name	$IC_{50\ human}$ [μM]
A/B/C/D-ring systems		
	Imidq, **1**	0.04
	MH61	0.077
	MH55	0.17

tial atomic charges (sigma, $q\sigma$, pi, $q\pi$, or total q_{tot}), electronegativity, χ, or polarizability, α, can be chosen as restrictions in the superimposition. Other atom properties, such as descriptors whether atoms are in aromatic and non-aromatic rings, or are ring or non-ring atoms, can also be selected as mapping conditions. These physicochemical parameters are calculated by the program package PETRA.[31, 32] The atoms to be overlaid must conform to the given matching criterion or interval of the physicochemical property. For example, if the matching criterion is chosen to be total atomic charges, q_{tot}, and the interval selected to be $q_{tot} = \pm\ 0.05\ e$, then for an atom of the first molecule with q_{tot} = -0.2 e, only atoms in the interval of q_{tot} = [-0.25, -0.15] are allowed to build match tupels with this first atom.

Cytochrome P450c17-inhibitors

The cytochrome P450c17 enzyme (17-α-hydroxylase/C17-20 lyase) is a key enzyme for the androgen and glucocorticoid biosynthesis.[33] Like most cytochrome P450 isoenzymes, P450c17 also has heme as prosthetic group. Substances conjugate to this enzyme by coordinating to the central iron atom at one end and by a hydrogen bond at the other end of their skeleton. Thus, substances with a high affinity to the enzyme should have a free electron pair (e.g. a nitrogen atom) and at least one hydrogen bond acceptor or donor. Inhibition of the 17-α-hydroxylase/C17-20 lyase is a promising concept for the treatment of prostate carcinoma. How-

Table 1b *Cytochrome P450c17 inhibitors of the A/B- and A/C-ring systems and their $IC_{50\ Human}$-values (progesterone, 25 µM)*

Chemical structure	Name	$IC_{50\ human}$ [µM]
A/B-ring systems		
	BW13, **2**	0.036
	BW61, **4**	0.074
	BW62, **5**	0.085
A/C-ring systems		
	BW112, **3**	0.087
	BW95	0.13
	BW99	0.21

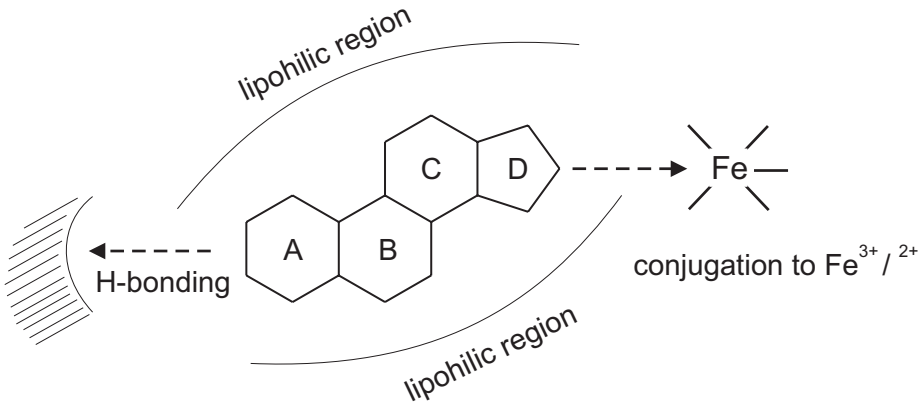


Figure 15 *The steroidal skeleton and the coordinations of its structural parts inside the P450c17 binding pocket*

ever, up to now no inhibitors have been found with an acceptable selectivity for the P450c17 enzyme.[34, 35]

P450c17 inhibitors can be separated into three structural subclasses corresponding to their steroidal skeleton: A/B/C/D-ring systems (steroids), A/B-ring systems (naphthalene compounds) and A/C-ring systems (biphenyl compounds). In Table 1 the chemical structures, their names, and their $IC50_{human}$-values (progesterone, 25 μM)[33, 34] are given. Imidq, **1**, is the most active steroidal ligand with a rather rigid skeleton. BW13, **2**, and BW112, **3**, are conformationally flexible ligands, with **2** the most active 3,4-dihydronaphthalene ligand, and **3** the most active biphenyl ligand.

We will illustrate with this example different options of the program that can be used to gain more insight into the structural requirements for a ligand to fit and bind into the pocket of the receptor. It will be shown how information already known can be used to impose restrictions on a program run in order to try to confirm these hypotheses.

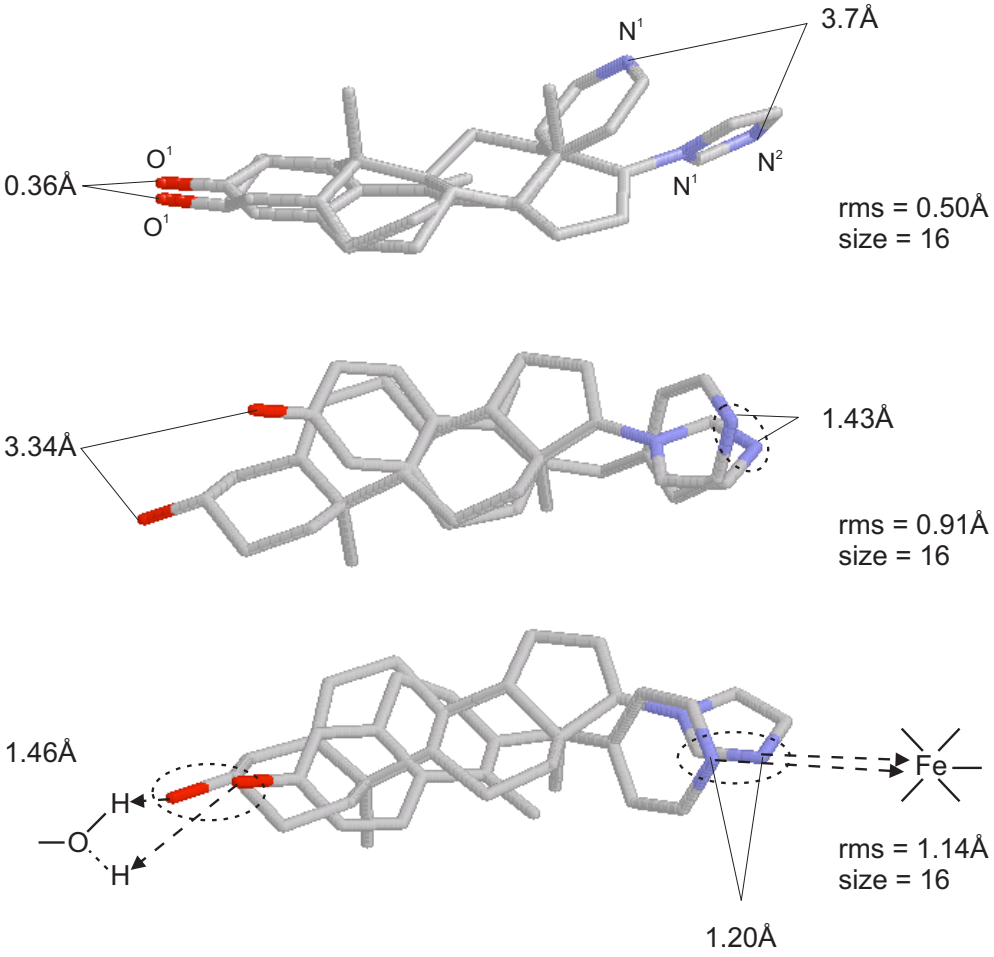


Figure 16 *The superimposition of **1** (rigid) and **2** (flexible). The required matching atom pairs are marked by a dashed circle. No restrictions are required for the upper superimposition, the matching atom pair **1**: N^2/**2**:N^1 is required for the middle superimposition, and **1**: N^2/**2**:N^1 and **1**: O^1/**2**:O^1 is required for the lower superimposition*

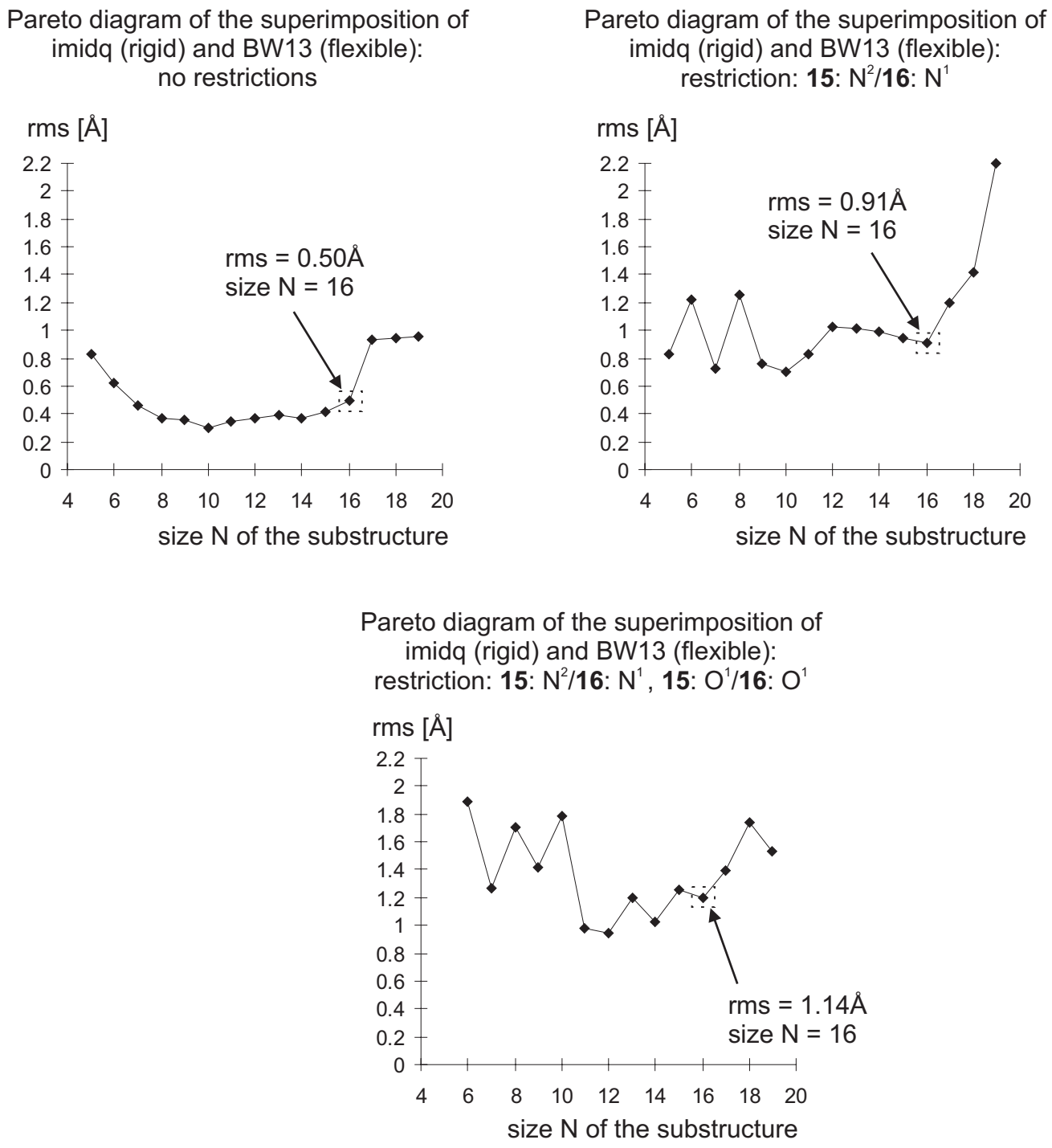


Figure 17 *The best Pareto diagrams of 40 GA experiments for the superimposition of imidq,* ***1*** *and BW13,* ***2*** *with different restriction (see Figure 16)*

The compounds shown coordinate through their basic nitrogen atom to the central iron atom inside the P450c17 enzyme and bind through a hydrogen bond acceptor (OH-, OMe-, NR_2-group) to a corresponding donor atom of the enzyme binding pocket (Figure 15).

Knowledge-based superimpositions

First we will explore the overall geometric fit and try to find out whether atoms necessary for binding can be located.

Structural homologies must be evaluated for the most active compounds of each structural subclass (**1**, **2** and **3**). Imidq,

1, can provide a template for the superimposition with BW13, **2**, and BW112, **3**, because **1** has a comparatively rigid steroidal skeleton. The conformation of **1** was set to be rigid during the optimization process, whereas the conformations of all other molecules were treated as flexible. During superimpositions not involving the steroid **1** both molecules were treated as flexible.

The compounds of the three structural subclasses have quite different numbers of atoms. Thus, an unbiased rigid superimposition leads to results that do not consider any requirements imposed by the three-dimensional structure of the enzyme binding pocket. In Figure 16 three different superimpositions of compound **1**, an A/B/C/D-ring system and **2**, an A/B-ring system are shown. In all of the three superimpositions, it was required that the atoms to be superimposed have the same atomic number. Always the best results of the 40 GA experiments are presented. The differences in the superimpositions result from the restrictions imposed on the substructures to be matched. The first superimposition was calculated without any restrictions (Figure 16, top). The probability for each atom to be part of the substructure was the same. The second superimposition requires the alignment of the atom pair **1**: N^2/**2**:N^1 (Figure 16, center). The third superimposition, in addition, requires the atom pair **1**: O^1/**2**:O^1 to match (Figure 16, bottom). Thus, knowledge about a specific binding mode of the ligands inside the enzyme is taken into account by forcing important atom pairs to match.

The restrictions lead to a decrease in the overall geometric fit. The *rms* values are 0.50 Å without any restrictions and 0.91 Å and 1.14 Å with restrictions. A potential reason for the increase in the *rms* values are the different sizes of the structures to be aligned. Without any restrictions, the smaller molecule will fit onto any part of the larger molecule (Figure 16, top). In this case, BW13, **2**, aligns exactly with the A/B/C-ring of imidq, **1**. This leads to a lower *rms* value, but ignores the fact that similarities must be investigated at both ends of the skeleton.

The Pareto diagrams of the best GA experiments out of 40 runs of all three superimpositions are shown in Figure 17. Usually, the *rms*-value increases for increasing substructure sizes. However, the Pareto diagrams of Figure 17 show that, using random based genetic algorithms, it is possible to receive better geometric fits for larger substructures than for smaller substructures.

A substructure size of 16 atoms was extracted for the comparison of the superimpositions. This number of atoms presents the minimum with the largest number of matching atoms of the Pareto diagram before a further increase in the *rms* values.

The Pareto diagrams show that the *rms* values are lower for all substructure sizes in the case where no restrictions were imposed. Imposing restrictions onto the atoms to be matched allows the investigation of the fit of the structures by considering knowledge about pharmacophore parts of the compounds.

Each nitrogen atom or oxygen atom of **1** and **2** coordinates to corresponding parts inside the receptor binding pocket. Thus, no large distances between the nitrogen atoms or the oxygen atoms in a superimposition can be accepted. In the lower superimposition of Figure 16 the distances are 1.20 Å (N-N) and 1.46 Å (O-O). Thus, a coordination to similar positions inside the receptor binding pocket would be possible. Although, the superimpositions show for the entire structures a worse atomic fit, the hydrophobic parts are aligned well onto each other. The importance of short distances of the nitrogen atoms is higher than those of the oxygen atoms. The nitrogen atoms coordinate to a fixed iron atom inside the

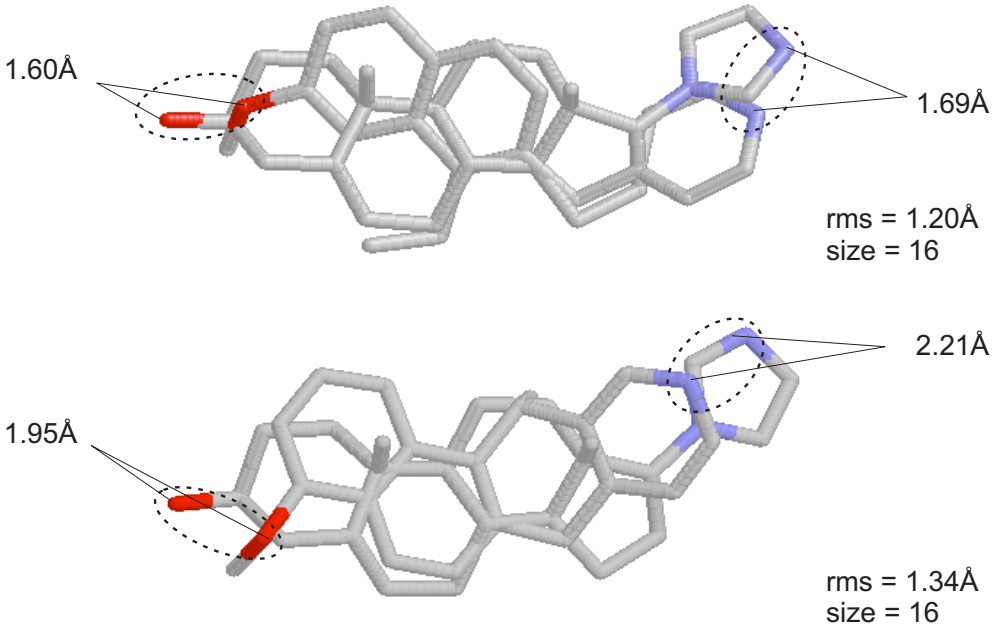


Figure 18 *The superimposition of **1** (rigid) and **4** (flexible, top) or **5** (flexible, bottom). The restricted matching atom pairs are marked by a dashed circle. Top: **1**: N^2/**4**: N^2 and **1**: O^1/**4**: O^1; bottom: **1**: N^1/**5**: N^2, **1**: O^1/**5**: O^1*

heme molecule. In contrast, the oxygen atoms bind to conformationally flexible parts of the enzyme through hydrogen bonds (Figure 16, bottom) which have also degrees of freedom through rotation around the carbon-oxygen bond.

The three A/B-ring system compounds **2**, **4**, and **5** (naphthalene derivatives) each have an oxygen atom on the A-ring. However, the site of substitution is different in all three cases. The next investigations should clarify whether, nevertheless, an alignment of ligands can be found that allows superposition of the oxygen atoms. For this purpose, again, the rather rigid steroid skeleton of **1** was kept rigid. In compound **4**, the oxygen atom of the 7-OMe group was required to be superimposed onto the oxygen atom of the OH-group of **1**. In compound **5**, the oxygen atom of the 5-OMe group was superimposed onto the OH group of **1** (Figure 18).

Thus, although the oxygen atoms in the three naphthalene derivatives **2**, **4** and, **5** are positioned at three different substitution sites (positions 6, 7, and 5, respectively), nevertheless in all three cases an alignment was found that was able to superimpose the oxygen and nitrogen atoms.

However, the superimpositions of both **4** and **5** with **1** (*rms* = 1.2 Å and 1.34 Å) lead to a lower geometric fit than the superimposition of **1** with the 6-OH substituted dihydronaphthalene, **2** (Figure 16, *rms* = 1.14 Å). In addition, the distances of the nitrogen pairs (**4**: 1.69 Å, **5**: 2.21 Å) and oxygen pairs (**4**: 1.60 Å, **5**: 1.95 Å) are relatively high. This decreased fit of **4** and **5** onto **1** compared to the fit of **2** onto **1** is reflected by a decrease in biological activity.

The ligand **3** has two hydroxyl groups. The next investigations explore whether GAMMA can provide information as to which one of the two hydroxyl groups is better suited for binding. The superimposition of imidq, **1**, (A/B/C/D-ring system) and BW112, **3**, (A/C-ring system) is shown in Figure 19. **3** and **1** do not differ much in their numbers of atoms. Compound **3** has two hydroxyl groups (3'-OH and 4'-OH) at the A-ring, in contrast to the A/B-ring compound **2**, that has only one hydroxyl group in position 6. To compare the superimposition of **3** and **1** with the superimposition of **2** and **1**, first only the 3'-OH group of **3** was taken into account (Figure 19, top). Then, a superimposition was performed by considering the 4'-OH group of **3** (Figure 19, bottom).

By requiring the 3'-OH group of **3** to be matched onto the oxygen atom of **1**, a better geometric fit of **3** onto **1** is obtained than the geometric fit of **2** and **1** (Figure 19, top, and Figure 16, bottom). The distance of the corresponding oxygen atoms d(O-O) is 0.47 Å and of the nitrogen atoms d(N-N) it is 0.95 Å. Thus, important pharmacophore points are closer in the superimposition of **3** and **1** than in the superimposition of **2** and **1** (Figure 16, d(O-O): 1.46 Å, and d(N- N): 1.20 Å).

The superimposition of **3** and **1** obtained by requiring the matching atom pairs to be **1**: N^2/**3**: N^2 and **1**: O^1/**3**: 4'O^1 leads to a higher *rms* value and, therefore, to a worse geometric fit than in the previous case when the 3'-O^1 atom of **3** is required to match (*rms* = 1.13 Å *vs* 0.72 Å). The distances of the oxygen atoms (d(O-O): 0.74 Å and 1.03 Å) and of the nitrogen atoms (d(N-N): 0.95 Å and 1.86 Å) are also lower than for the atom pairing **1**: O^1/**3**: 3'O^1. In addition, the plane of the A-biphenyl ring of **3** is more or less perpendicular to the A-ring plane of the rigid steroidal skeleton of **1** in the case of requiring the 4'-O^1 of **3**. This allows the conclusion to be made that in compound **3** it is the 3'-OH group that is involved in binding.

In order to explore this hypothesis further, the compounds BW112, **3**, (A/C-ring system) and BW13, **2** (A/B-ring system) were superimposed, in one case the oxygen atom of the

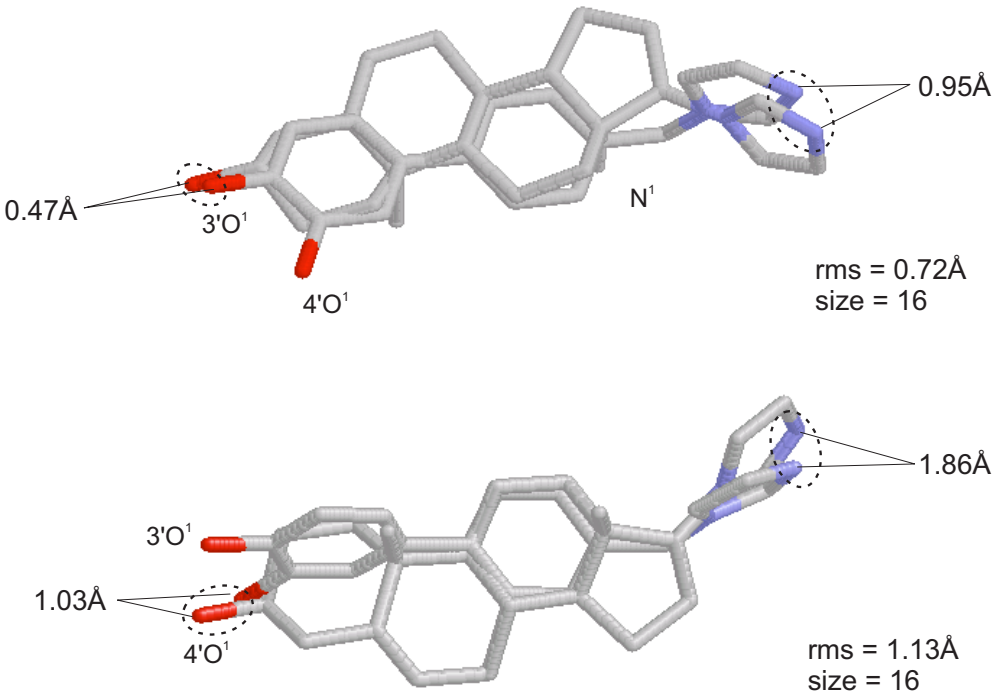


Figure 19 *The superimposition of **1** (rigid) and **3** (flexible). The required matching atom pairs are marked by a dashed circle. Top: **1**: N^2/**3**: N^2 and **1**: O^1/**3**: 3'O^1; bottom: **1**: N^2/**3**: N^2 and **1**: O^1/**3**: 4'O^1*

Table 2 *The* rms*-values of the calculated superimpositions of some P450c17-inhibitors and the corresponding distances of the N-N and O-O atom pairs.*

superimposition	required matching atom pair	rms [Å]	d(N-N) [Å]	d(O-O) [Å]
1 with **2**	N^2-N^1, O^1-O^1	1.14	1.20	1.46
1 with **4**	N^2-N^1, O^1-O^1	1.20	1.60	1.69
1 with **5**	N^2-N^1, O^1-O^1	1.34	2.21	1.95
1 with **3**	N^2-N^2, O^1-3'O^1	0.72	0.95	0.74
1 with **3**	N^2-N^2, O^1-4'O^1	1.13	1.86	1.03
2 with **3**	N^1-N^2, O^1-3'O^1	0.57	1.41	0.10
2 with **3**	N^1-N^2, O^1-4'O^1	0.82	0.21	0.90
4 with **3**	N^1-N^2, O^1-3'O^1	1.16	1.14	2.00
4 with **3**	N^1-N^2, O^1-4'O^1	0.52	0.88	0.24
5 with **3**	N^1-N^2, O^1-3'O^1	1.03	1.54	2.38
5 with **3**	N^1-N^2, O^1-4'O^1	0.68	0.97	0.49

3'-OH group of **3** required to match onto the oxygen atom of **2**, in the other case, the oxygen atom of the 4'-OH group of **3** requires to be matched onto the oxygen atom of **2**. In both investigations, both molecules **2** and **3** were allowed to be flexible. These two compounds differ again in their numbers of atoms.

The first superimposition (Figure 20, top) requires the atom pairs **2**: N^1/**3**: N^2 and **2**: O^1/**3**: 3'O^1, while in the second superimposition (Figure 20, bottom) the atom pairs **2**: N^1/**3**: N^2 and **2**: O^1/**3**: 4'O^1 are enforced.

The first superimposition (Figure 20, top) shows a partial alignment of the biphenyl-C-ring of BW112, **3**, onto the 1-methyl and 2-methylene group of the 3,4-dihydronaphthalene, **2**. This presents a rather good geometric fit (*rms* = 0.57 Å) of the skeletons and of the oxygen atoms (d(O-O): 0,10 Å), but also leads to quite a large distance of the nitrogen atoms (d(N-N): 1.41 Å). A shorter distance of the nitrogen atoms (d(N-N): 0.21 Å, d(O-O): 0.9 Å) in the second superimposition (Figure 20, bottom) corresponds to a larger shift (*rms* = 0.82 Å) in the orientation of the skeletons of both structures. This fact does not exclude a similar binding mode because, nevertheless, the hydrophobic parts of both structures are aligned quite well. Again, on the basis of the lower *rms* we come to the conclusion that binding occurs through the 3'-OH group of **3**.

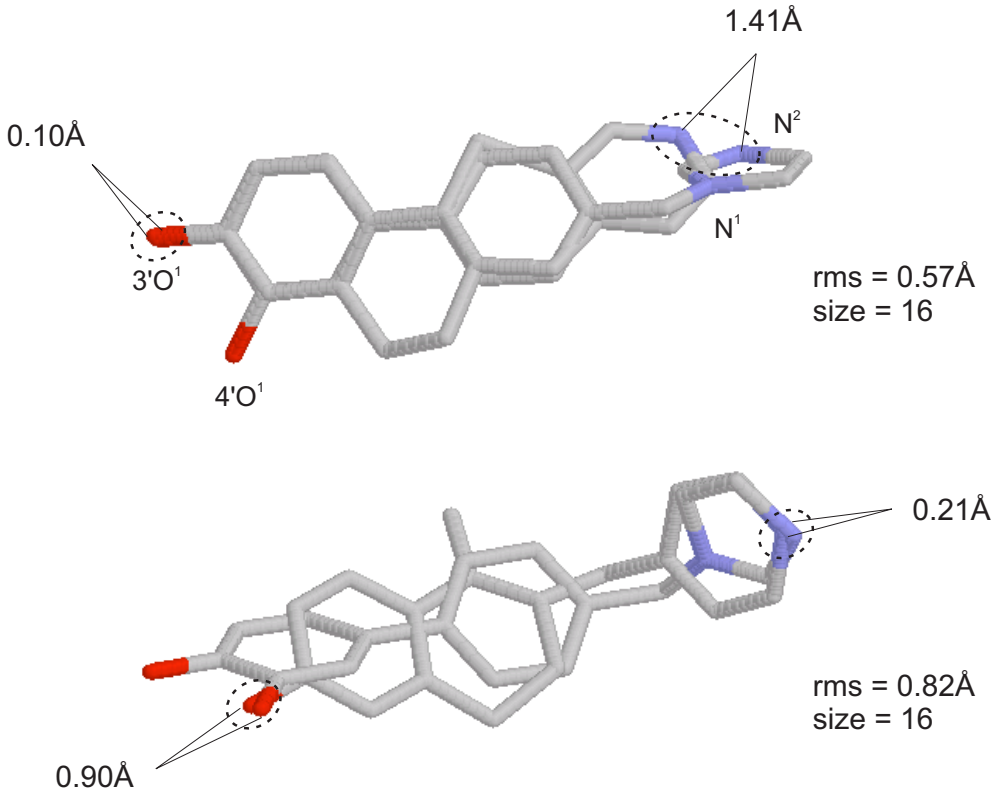


Figure 20 *The superimposition of* ***2*** *(flexible) and* ***3*** *(flexible). The restricted matching atom pairs are marked by a dashed circle. Top:* ***2****:* N^1*/***3****:* N^2 *and* ***2****:* O^1*/***3****:* 3'O^1*, bottom:* ***2****:* N^1*/* ***3****:* N^2 *and* ***2****:* O^1*/***3****:* 4'O^1

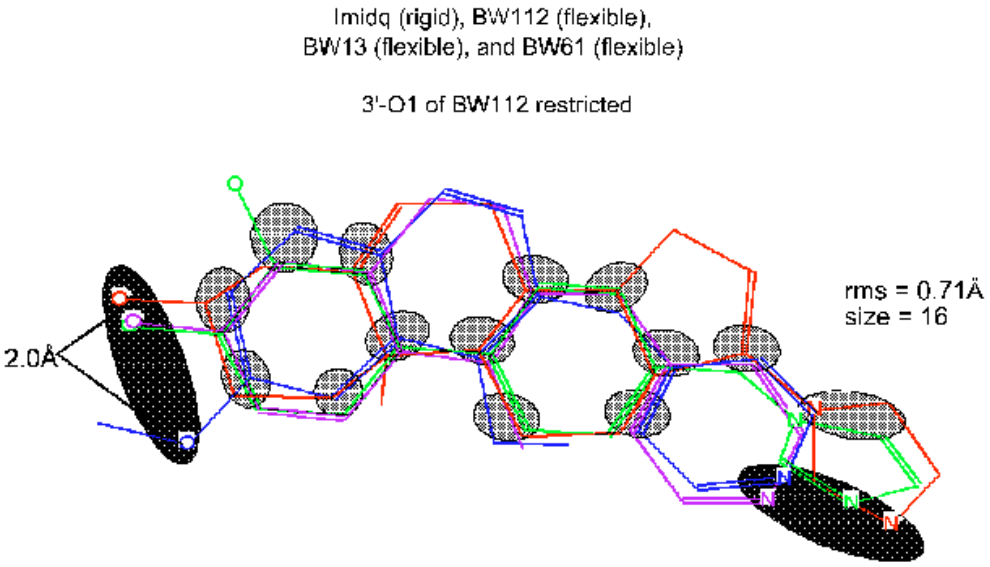


Figure 21 *The result of the simultaneously superimposition of imidq,* ***1*** *(red), BW112,* ***3*** *(green), BW13,* ***2*** *(magenta), and BW61,* ***4*** *(blue). The substructure is marked by gray circles. The matching atom pairs* ***1****:* O^1*/****2****:* O^1*/****3****: 3'*O^1*/****4****:* O^1*, and* ***1****:* N^2*/****3****:* N^2*/****2****:* N^1*/****4****:* N^1 *are restricted and marked by dark gray circles*

Table 2 shows the results of each binary superimposition of compounds **1**, **2**, **3**, **4**, and **5**,. In addition, Table 2 includes the superimpositions of **3** with **4** and **5**, which have not yet been shown before. To evaluate the goodness of the alignment, the *rms* values for a constant substructure size (16 atoms) is given. Furthermore, the distances of the corresponding nitrogen and oxygen atom pairs are considered.

The following conclusions are based on the *rms* values of the geometric fit and in addition on the distances of the restricted matching atom pairs.

The superimposition of the steroidal template **1** with the 6-substituted 3,4-dihydronaphthalene compound, **2**, (*rms* = 0.72 Å) is worse than the superimposition of **1** with the biphenyl system **3**, by restricting the 3'-oxygen atom of **3** (*rms* = 0.57 Å). However the alignment of **1** with **2** (*rms* = 0.72 Å) is better than the alignment of **1** and **3** by restricting the 4'-oxygen atom of **3** (*rms* = 1.13 Å). The alignment of **1** with the 7- and 5-substituted 3,4-dihydronaphthalene compounds **4** and **5** leads to worse geometric fits than the alignment of the **1** with the 6-subsituted 3,4-dihydronaphthalene, **2**. In contrast to the superimposition of **3** and **1**, the geometric fit of the superimposition of **3** with both 7- and 5-substituted 3,4-dihydronaphthalene compounds (A/B-ring systems) **4** and **5** is better by restricting the 4'-oxygen atom of **3** (*rms* = 0.52 Å and 0.68 Å).

In summary, the superimposition of the most active compounds **1** (A/B/C/D-ring system), **2** (A/B-ring system), and **3** (A/C-ring system) show that the biphenyl compounds (A/C-ring systems) have a binding mode similar to that of the steroidal compounds (A/B/C/D-ring systems) by coordina-

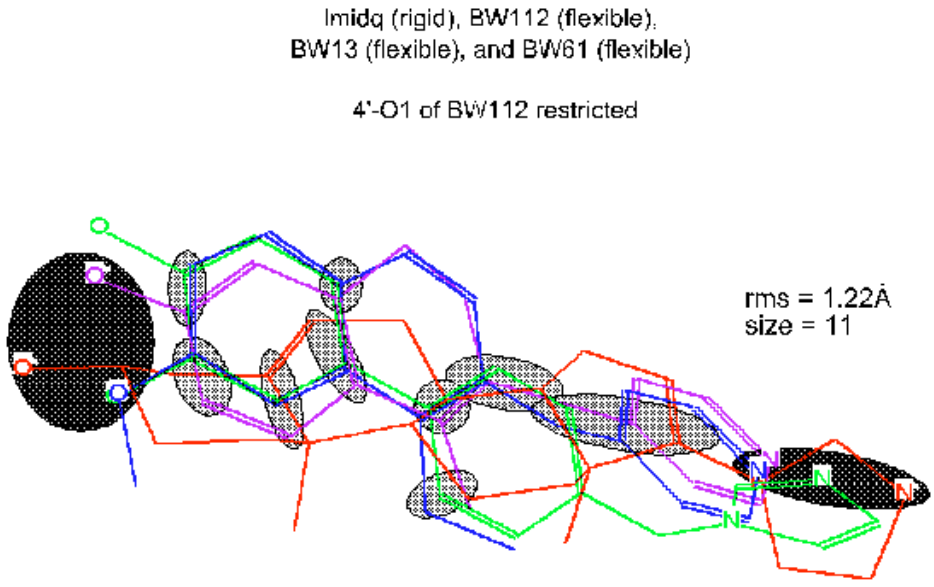


Figure 22 *The result of the simultaneously superimposition of imidq,* ***1*** *(red), BW112,* ***3*** *(green), BW13,* ***2*** *(magenta), and BW61,* ***4*** *(blue). The substructure is marked by gray circles. The matching atom pairs* ***1****:* O^1*/****2****:* O^1*/****3****: 4'*O^1*/****4****:* O^1*, and* ***1****:* N^2*/****3****:* N^2*/****2****:* N^1*/****4****:* N^1 *are restricted and marked by dark gray circles*

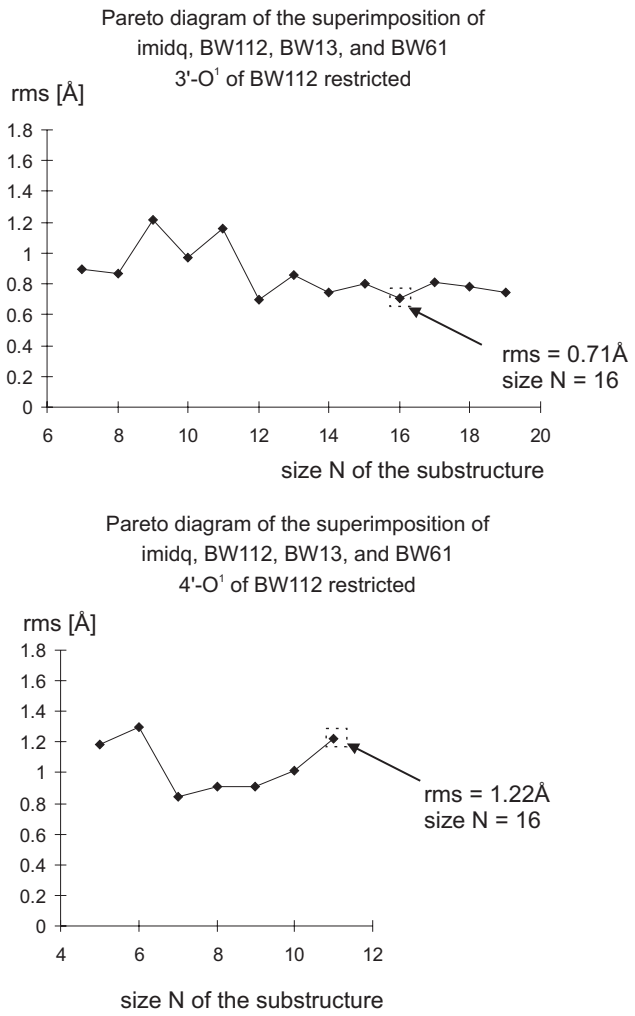


Figure 23 *The Pareto diagrams of two superimpositions of imidq,* ***1****, BW112,* ***3****, BW13,* ***2****, and BW61,* ***4****. Top: restriction of the 3'O of* ***3****, bottom: restriction of the 3'O of* ***3***

tion of the 3'-hydrogen bond acceptor to the receptor binding pocket. The superimposition of the steroid **1** with the biphenyl compound **3** leads to a better geometric fit than the superimposition with the three dihydronaphthalene compounds **2**, **4** and **5**. Whereas the geometric fit of the less active compounds **4** and **5** with the rigid skeleton of **1** is worse than in the geometric fit of the more active compound **2**. In the case of the lowest active compound **5**, the geometric fit is again worse than for compound **4**. Thus, a higher deviation in the fit corresponds to lower activity.

The superimpositions of **3** with **1** and **2** show that the coordination of the 3'-hydrogen bond acceptor of **3** seems to be the preferred one. On the other hand, the 5- and 7-methoxy substituted naphthalene compounds **4** and **5**, which have a lower affinity to the enzyme than the 6-substituted naphthalene compound **2**, show a better alignment with **3** by enforcing an overlap of the 4'-OH group of **3** than when the 3'-OH group was used. Thus, no clear decision can be drawn by only comparing sets of two molecules. We will therefore investigate the simultaneous superimposition of several molecules in the next section.

The superimpositions calculated with GAMMA by considering specific knowledge about the receptor binding pocket show that binding modes can be predicted even for flexible compounds. Affinity potentials of compounds without known biological activity to the P450c17 enzyme can be estimated by superimposing them onto a compound with a known biological profile to the P450c17 enzyme. Three different values are to be evaluated, the *rms* value of the geometric fit, the distances of the atoms that coordinate to the central iron atom of the heme molecule, and the distances of potential hydrogen bond acceptors.

As of now, only the superimpositions of two molecules have been investigated. This might not give the full picture of the pharmacophore as accidental similarities might be found. A clearer picture of the of the spatial and electronic requirements of a drug can only be obtained when more ligands are compared much in the same way as a better idea of requirements for a key to fit into a given lock can be obtained by comparing more keys. This is explored in the next section.

Multiple superimpositions

The simultaneously superimposition of the four highly active P450c17 inhibitors **1**, **2**, **3**, and **4** is shown in Figures 21 and 22. The two superimpositions differ in the restriction imposed on the oxygen atom of compound **3**: In Figure 21 the 3'-OH group of **3** was chosen whereas in Figure 22 the 4'-OH group of **3** was selected. The conformation of **1** was kept rigid, whereas the molecules **2**, **3**, and **4** were allowed to be flexible during the calculation of the superimposition.

In the binary superimposition of **3** and **4** the oxygen atom of the OMe-group of **4** is better aligned to the 4'-oxygen atom of **3**. Thus, in the simultaneous superimposition, compound **4** fits better than the other compounds when the 4'-oxygen atom of **3** is used. A comparison of the superimpositions in Figures 21 and 22 clearly shows that a better geometric fit can be obtained by requiring the 3'-oxygen atom of **3.** The superimposition that aligns the oxygen of the 3'-OH group (Figure 21) has an *rms* value of 0.71 Å, whereas the superimposition that aligns the oxygen of the 4'-OH shows an *rms* value of 1.22 Å. The distance of the oxygen atoms of the OMe-group of **4** is nearly 2.0 Å to the cluster of the aligned oxygen atoms of the compounds **1**, **2**, and **3** in the case of restricting the 3'-oxygen atom of **3** (Figure 21).

Figure 23 shows the Pareto diagrams of the two superimpositions of **1**, **2**, **3**, and **4**. Again a substructure size of 16 atoms has been extracted of the GAMMA run. In the first Pareto diagram (Figure 23, top, 3'-O of **3** is restricted) the *rms* value is rather constant when increasing the size of the substructure. The second Pareto diagram (Figure 23, bottom, 4'-O of **3** is restricted) shows that this superimposition has a lower geometric fit for all substructure sizes related

with a maximum substructure size of only 11 atoms. Substructure sizes higher than 11 atoms could not be found during the optimization. This again proves, that the 3'-OH group of **3** is more favorable for binding to the P450c17 enzyme. Based on the simultaneous superimposition of four ligands, we can now clearly come to the conclusion that binding occurs through the 3'-OH group of **3**.

Summary

The program GAMMA (genetic algorithm for multiple molecule alignment) described here can be used to superimpose and align several structures independently of the conformation chosen initially. An unlimited number of structures can be treated. Only one conformation per structure is necessary and, thus, the program can work even when only one conformation of a compound is stored in a database. The automatic elucidation of structural similarities has its particular efficiency in a hybrid method, the combination of a genetic algorithm and a numerical optimization method, the directed-tweak method. The genetic algorithm process leads to an optimization of the assignment of the atoms in the form of match lists. An optimization of the geometric fit by adapting the conformations of molecules to each other is obtained by the combination of the genetic algorithm with the directed-tweak technique. The genetic algorithm is further improved by two additional operators which are tailored to the superposition problem: the creep and crunch operators. The problems studied here must optimize several conflicting criteria. Therefore, always a set of so-called Pareto solutions is obtained at the end of each GA run. During optimization of the superposition, the conformations of the structures are adapted to each other.

The program allows one to choose a template structure as rigid and then adapts the conformation of the other molecules onto this template. Rotatable bonds are found automatically or, alternatively, can be selected by the user. Special features like selecting an atom list that has or should be part of the match list or defining different matching criteria allow to include specific knowledge to the pharmacophore search problem. The calculated maximum common substructure (MCSS) of a set of three-dimensional structures having the same biological activity can indicate certain pharmacophore regions and points.

Acknowledgements We thank the BMBF and the Merck KGaA for financial support of our research and the "Fonds der Chemischen Industrie" for a scholarship to Sandra Handschuh. We appreciate the contributions of Dr. Markus Wagener to the approach outlined here.

Supplementary material available 3D atomic coordinates for the superimpositions shown in Figures 16, 18, 19 and 20 are available in PDB format.

References

1. Sheridan, R. P.; Nilakantan, R.; Rusinko III, A.; Baumann, N.; Haraki, K. S.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 255-260.
2. Greene, J.; Kahn, S. D.; Savoi, H.; Sprague, P.; Teig, S. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297-1308.
3. Martin, Y. C.; Danaher, E. B.; May, C. S.; Weininger, D. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 15-29.
4. Van Drie, J. H.; Weininger, D.; Martin, Y. C. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 225-251.
5. Pepperrell, C. A.; Willett, P.. *J. Comput.-Aided Mol. Des.* **199**, *5*, 455-474.
6. Sheridan, R. P.; Miller, M. D.; Underwood, D. J.; Kearsley, S. K. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 128-136.
7. Bath, P. A.; Poirrette, A. R.; Willett, P.; Allen, F. H. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 141-147.
8. Lauri, G.; Bartlett, P. A. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 51-66.
9. Fisanick, W.; Cross, K. P.; Rusinko III, A. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 664-674.
10. Güner, O. F.; Henry, D. R. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; Wiley & Sons: Chichester, UK, 1998; pp2988-3003.
11. Henry, D. R.; Özkabak, A. G. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; Wiley & Sons: Chichester, UK, 1998; pp 543-552.
12. Willett, P. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; Wiley & Sons: Chichester, UK, 1998; pp 2748-2756.
13. Milne, G. W. A. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; Wiley & Sons: Chichester, UK, 1998; pp 2046-2056.
14. Wagener, M.; Gasteiger, J. *Angew. Chem.* **1994**, *106*, 1245-1248; *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1189-1192.
15. Handschuh, S.; Wagener, M.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220-232.
16. Sadwoski, J.; Gasteiger, J. *Chem. Rev.* **1993**, *93*, 2567-2581.
17. Sadowski, J.; Gasteiger, J.; Klebe, G. *J. Chem. Inf. Sci.* **1994**, *34*, 1000-1008.
18. CORINA, version 2.4; Molecular Networks GmbH, Erlangen, Germany, 1999; http://www.mol-net.de
19. Hurst, T. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 190-196.
20. Goldberg, D. E. *Genetic Algorithms in Search Optimization and Machine Learning*; Addison-Wesley: New York, 1989.
21. Fonseca, C. M.; Fleming, P. J. In *Proceedings of the 5th International Conference on Genetic Algorithms*; Forrest,

S., Ed.; Morgan Kaufmann Publishers: San Mateo, CA, 1993; pp 416-423.
22. Jones, G. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; Wiley & Sons: Chichester, UK, 1998; pp 1127-1136.
23. Venkatasubramanian, V.; Sundaram, A. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds.; Wiley & Sons: Chichester, UK, 1998; pp 1115-1127.
24. Oliver, I. M.; Smith, D. J.; Holland, J. R. C. In *Proceedings of the 2nd International Conference on Genetic Algorithms*; Grefenstette, J. J. Ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1987; pp 224-230.
25. Davis, L. In *Proceedings of the 3rd International Conference on Genetic Algorithms*; Schaffer, D., Ed.; Morgan Kaufmann Publishers: San Mateo, CA, 1989; 61-69.
26. Wexler, R. R.; Greenlee, W. J.; Irvin, J. D.; Goldberg, M. R.; Prendergast, K.; Smith R. D.; Timmermans, P. B. M. W. M. *J. Med. Chem.* **1996**, *39*, 625-656.
27. Jones, G.; Willett, P.; Glen, R. C. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532-549.
28. Harik, G. R. In *Proceedings of the 6th International Conference on Genetic Algorithms*; Eshelman, L. J., Ed.; Morgan Kaufmann Publishers: San Francisco, CA, 1995, pp 24-31.
29. Brown, R. D.; Jones, G.; Willett, P. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 63-70.
30. Press, H. W., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing;* Cambridge University Press: Cambridge, 1997.
31. Gasteiger, J. In *Physical Property Prediction in Organic Chemistry*; Jochum, C.; Hicks, M. G.; Sunkel, J., Eds.; Springer Verlag: Heidelberg, 1988, 1pp 19-138.
32. PETRA, version 2.6; Molecular Networks GmbH, Erlangen, Germany, 1999; http://www.mol-net.de
33. Bosche, H v. d. *J. Steroid Biochem. Mol. Biol.* **1992**, *43*, 1003-1021.
34. Wachall, B. G.; Hector, M.; Hartman, R. W. *Bioorg. Med. Chem.* **1999**, in press.
35. Wachall, B. thesis, Universität Saarbrücken, 1999.