Fast Continuum Electrostatics Methods for Structure-Based Ligand Docking

Catherine Tenette-Souaille, Nicolas Budin, Nicolas Majeux
and Amedeo Caflisch*

Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich, Switzerland
Tel: (41 1) 635 55 21
FAX: (41 1) 635 57 12

*Corresponding author. Email: caflisch@bioc.unizh.ch

This is a **chapter for the book**
Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications
A.K. Ghose and V.N. Viswanadhan editors

February 25, 2000

1 Introduction

To discover drugs against human diseases a plethora of methodologies have been developed since ever. These involve many research fields and activities ranging from purely theoretical to experimental. Soon after computers were sold at reasonable prices, many investigators in both pharmaceutical companies and nonprofit research institutions realized the almost unlimited opportunities that the computational and data handling power of computers can offer for the very difficult task of drug discovery. After the information technology revolution of the last three decades, the coming years will be characterized by major successes in genomics and proteomics. The cloning and sequencing of the human genome as well as progresses in high throughput approaches to solve protein structures will generate very valuable information on an ever increasing number of potential drug targets. Drug design is significantly facilitated if the three-dimensional conformation of the protein target is known at atomic level of detail. The large efforts invested in the determination of protein structures by pharmaceutical and biotechnology companies are the practical proof that the knowledge of the three-dimensional conformation of a protein target is of paramount importance.

In this chapter we first review a number of approaches for structure based computer-aided design. A detailed description of the continuum electrostatic approach developed in our research group for docking library of small to medium-size fragments is then presented. Finally, an application to the p38 mitogen-activated protein (MAP) kinase is discussed and a brief outlook ends this chapter.

2 Structure-based ligand design approaches

Computational tools that exploit the knowledge of the three-dimensional structure of a protein target are used for *de novo* design [1, 2], improvements of lead compounds, and to help in the selection of monomers to focus combinatorial libraries [3]. Prioritization is done by empirical and knowledge-based scoring functions or force field energy functions [4]. Ligands are built by connecting small molecular fragments or functional groups, often rigid, or even atoms. In the latter case, the

methods have shown significant flexibility with respect to the structures that can be obtained [5, 6, 7]. The main disadvantage of compounds generated by atom-based approaches is that they often have complicated structures and are in most cases very difficult to synthesize. Hence, methods that build new compounds by combining predefined fragments are more popular. The number of newly created bonds is small and therefore it is easier to control the chemistry, i.e., the synthesizability and the chemical stability of the designed molecules. Furthermore, fragments are easily modeled, since the model parameters, like partial charges, periodicity and force constant of torsion angles, are assumed to depend mainly on the fragment and only to a lesser extent on the rest of the structure.

Fragment-based ligand design may be achieved in two ways. In the first one, small fragments are docked in the active site. The best positions of each fragment type are retained and connected to generate candidate ligands. Alternatively, an anchor fragment is docked in the binding site and the ligand is grown starting from it. Both approaches should not be considered as mutually exclusive, but rather as complementary since they are useful to generate candidate ligands with different physico-chemical characteristics and structural properties.

2.1 Methods based on the connection of docked fragments

This approach has the advantage that the fragments occupy optimal positions and are oriented such that their interaction with the protein is favorable. On the other hand, the geometry of the bonds connecting the fragments to each other or to a central template is not optimal and has to be accepted with a certain tolerance initially. The mapping of a binding site and fragment assembly into complete ligands can be performed by separated programs [8, 9] or integrated in a single computational tool [10]. For site mapping, two main approaches have emerged: the first is based on binding site shape descriptors and the second on multiple-copy techniques.

With the program GRID [11], Goodford pioneered the use of molecular probes to explore the surface of a protein and search for energetically favorable positions. The interaction energies are then mapped onto a grid which describes the regions

of attraction between a probe and the protein. The surface descriptors thus obtained can be used to screen a 3D database of small compounds. This task can be performed by the program CLIX [12]. For each molecule, CLIX attempts to make a pair of substituents spatially coincide with a pair of favorable interaction sites proposed by GRID.

Several docking programs can map a protein binding site using small to medium size molecular fragments, either rigid or partly flexible. The program DOCK was first based on rigid docking and the use of geometrical criteria to judge the complementarity between receptor and ligand, and was therefore fast enough to screen databases for leads [13]. DOCK uses spheres complementarity to the receptor molecular surface to create a space filling negative image of the receptor site. Several atoms of the ligand are matched with spheres which define the binding site. Flexibility [14] and a force-field like energy function for scoring [15] were included in later development of the program. DOCK has been used to find novel micromolar inhibitors of enzymes [16, 17]. FlexX [18] is a program for the fast docking of medium sized flexible ligands. It first positions a fragment of the ligand by mapping three interaction centers of the fragment onto three interaction points of the receptor. The ligand is then constructed in an incremental way and Böhm's empirical function is used for scoring [19]. It is fast enough to allow screening of small databases of ligands. The docking of hydrophobic fragments has been slightly improved [20] and the algorithm has been extended to predict the location of water molecules in the binding site [21].

A number of genetic algorithms have been suggested for docking [22, 23, 24]. They combine speed with simplicity of concept. For example, GOLD [22] is based on a genetic algorithm that encodes in the chromosomes the values of the dihedral angles around rotatable bonds and positions the ligand in the binding site by a simple least squares fitting that maximizes the number of intermolecular hydrogen bonds. It also allows flexibility around bonds to hydrogen bonds donors and acceptors in the receptor. GOLD uses a force field with a simple approximation of solvation consisting of precalculated atom type-based hydrogen bond energies. The method has been tested on 100 complexes leading to a success rate, defined by the

authors in a rather subjective way, of about 70% for redocking into the complexed conformation [25]. Docking into the unbound conformation was tested on only three examples and gave mixed results [25]. An accurate treatment of solvation is essential for docking into a flexible binding site [26].

Multiple-copy techniques use numerous fragment replicas, each transparent to the others but subject to the full force field of the receptor, to determine energetically favorable positions and orientations (functionality maps) of small molecules or functional groups in the binding site of a protein [27, 28, 29] or RNA [30]. Although the multiple copy simultaneous-search (MCSS) method was originally proposed in the context of a rigid receptor [28], it was extended to allow for ligand and receptor flexibility [29, 31, 32].

Recently, we have developed a new continuum solvation approach that can be used for efficiently docking fragments into a rigid receptor (SEED [8]). It combines the advantages of shape descriptors and multiple-copy methods. Polar and apolar vectors are distributed on the surfaces of the receptor binding site and the fragments, and matched with each other, allowing an exhaustive docking on a discrete space. The main advantage of SEED over other docking programs is the comprehensive treatment of electrostatic solvation effects in an efficient and accurate manner. SEED will be described in further details in section 3.

The large amount of structural information in the functionality maps can be exploited by other programs. The docked fragments can be linked together with smaller [9, 33] or larger [34, 35, 36] linkers. The program CAVEAT [34, 35] was designed to search, in an interactive manner, 3D databases for molecular frameworks that can position functional groups in specific relative orientations. CAVEAT focuses on relationships between bonds; methods are implemented to identify and classify structural frameworks. The HOOK algorithm [36] connects a set of functional groups previously docked in a binding site using "skeletons" from a database, on which "hooks" are defined. The linkage is accomplished by fusing the hooks with two or more methyl groups from the functional groups. Computational combinatorial ligand design (CCLD) [9] is also based on docking of functional groups with MCSS or SEED [8]. The fragments are ranked according to an approximated

binding free energy. After classifying positions into overlapping (i.e., mutually excluding) and bonding (i.e., possibly bound by small linkers) pairs, CCLD creates ligands by linking the docked fragments with the most favorable of small linkers. To avoid combinatorial explosion, growing is discontinued when the average binding free energy of the fragments in the new ligand exceeds a user-specified threshold.

Some programs integrate both site mapping and fragment assembly. LUDI [10, 37] makes extensive use of empirical information derived from structural databases. Interaction sites that indicate possible positions for functionalities complementary to the receptor are defined and used to dock fragments from a library. Alternatively, the output of GRID can be used for the definition of interaction sites. The fragment are fitted on the interaction sites with the algorithm published by Kabsch [38] and are connected with small linkers. Interaction geometries were derived from structural data on small organic molecules [39, 40]. The scoring function used in LUDI is empirical [19]. The program was extended to take into account the synthetic accessibility of the constructed molecules [41]. The program SPROUT [42] can deal either with a three-dimensional experimental structure of the receptor or with a pharmacophore model derived from known inhibitors. Target sites in the binding pocket are identified and labeled by type. Fragments, from a library presorted according to atomic and molecular properties, are selected and overlaid on a target site. Once fragments have been docked into all of the target sites, the linking procedure is performed taking into account the identity of the fragments. In the second phase, atom types are exchanged with others of the same hybridization state in order to find a combination with optimal interactions with the binding site.

In an effort to remain close to the progresses of modern chemistry, a number of computational tools were further developed to facilitate the design of combinatorial libraries. CombiDOCK is a modifed version of DOCK to efficiently screen a large combinatorial library for a receptor [43]. CombiDOCK first positions the scaffolds in the binding site and, for each scaffold orientation, all potential fragments are attached. The interactions between substituents and receptor are individually scored and factorial combinations of fragments are suggested. In LUDI, a new procedure

has been implemented to focus the design on a chemical reaction, amenable to parallel chemistry [41, 44].

2.2 Methods based on the progressive build-up of ligands

Ligand build-up is a powerful stepwise strategy for *de novo* ligand design. It starts with a seed fragment placed in an appropriate region of the binding site. New ligands are then grown by sequentially appending building blocks (fragments or atoms). To avoid combinatorial explosion, a large fraction of all building blocks is discarded at every step according to some heuristic scoring. This method has the advantage that the newly formed chemical bonds have a correct geometry and that the intraligand interactions can be taken into account during the design. On the other hand, build-up approaches have difficulties to generate ligands that bind to different pockets if these are separated by gap regions that do not allow specific interactions. Moreover, the success of the growing procedure and therefore the quality of the designed molecules depends dramatically on the position of the seed, since the latter is usually kept fixed. The seed position(s) can be determined from X-ray or NMR structures of ligand-protein complexes. If no structure is available, seeds must be obtained by manual or computer-aided docking. Many programs that implement the build-up strategy have been described in the literature and the following list is not exhaustive.

GenStar [45] and LEGEND [46, 47] use single atoms as building blocks. GenStar grows sequentially structures which are entirely composed of sp3 carbons. It allows branching and rings formation. For each new atom generated, several hundred candidate positions with acceptable bond geometries are generated. Each position is scored based on a simple binding site contact model and the selected position is chosen at random among the highest scoring cases. LEGEND works in a similar way but uses the MM2 force field [48]. The choice of the atom type is driven by the protein electrostatic potential value at the atom position.

GroupBuild [49] and SMoG [50, 51] use a library of organic compounds to design ligands. In both programs, each candidate fragment is attached to the growing structure and rotated around the new bond in fixed increments. In GroupBuild a

standard molecular mechanics potential function is used to rank the candidates. The chosen fragment rotamer is randomly selected among the top 25% fragment positions. SMoG uses a knowledge based potential for the ranking. The lowest energy rotamer's acceptance is determined by a Metropolis Monte Carlo criterion which compares the new energy per atom with and without the candidate fragment.

The build-up strategy has also been implemented in the programs GROW [52], LUDI [41], and PRO_LIGAND [53]. Their library of fragments is however restricted to amino acids and amino acid derivatives. This has the advantage that the designed ligands are synthetically accessible, but the explored chemical space is relatively small. Moreover the energetics of the ligand can be studied by well parameterized force fields. On the other hand, peptides, besides their poor pharmacological properties, represent special problems due to their great conformational flexibility. This latter is taken into account by using multiple conformers for each amino acid. The main differences between these three programs lie in the scoring functions used to rank the ligands, and in the way the conformation libraries for the amino acids are generated. GROW's scoring function is based on the AMBER force field [54] supplemented by a solvent accessible surface approximation of solvation [55]. LUDI and PRO_LIGAND use empirical scoring functions combined with a rule based interaction site strategy [19, 56]. The GROW and PRO_LIGAND libraries contain low energy conformations whereas LUDI uses conformations extracted from high-resolution protein structures.

2.3 Binding energy

Ligand design involves the extension of the docking problem into chemical space. The degrees of freedom to be optimized are not only the positional and conformational variables of a particular compound, but, additionally, its chemical identity. This point of view makes one important problem in the field of ligand design particularly clear: the quality of the scoring or energy function used to evaluate the different solutions. When the search space is very limited as for example in the first programs that performed rigid docking [13] a very simple energy function based on geometrical criteria was sufficient to recognize the correctly docked structures.

When flexibility in the ligand (and the protein) is allowed, the effect of solvation has to be taken into account [26] to avoid sampling irrelevant parts of the conformational space. A simple example shows the higher quality requirements on the scoring function for design purposes: assuming the charge on an atom in a designed ligand is a (continuous or discrete) variable of the optimization, any simple force field based energy function would tend to maximize the total charge [57]. This is however in disagreement with empirical data. Although sometimes high affinity may be due to ionic interactions, often the desolvation penalty of full charges on the ligand and the protein is stronger than the direct interaction. Desolvation is the change in the solvation energy of the ligand and the receptor due to the displacement of high dielectric solvent by low dielectric solute upon complex formation. This further indicates that the scoring function should correspond to a difference between the free and the complexed states. The calculation of such differences is not necessary in docking because the term corresponding to the free state identically cancels. Accurate and reliable prediction of the absolute binding free energy for a medium-sized flexible ligand is currently beyond the limits of routine calculations, since it also includes finding the most probable conformations in water and averaging with the correct thermodynamic weights. Furthermore, in ligand design free energies are assumed to be additive, although it is clear that this is only a crude approximation [58]. The main task for a scoring function in a ligand design program is to find the conformations with the lowest energies for every chemical species (be that an atom, fragment or complete ligand) and in the case of different chemical entities (for example a benzene and a guanidinium docking in the same binding pocket) to decide which yields the lowest binding free energy. Both tasks and especially the latter are not straightforward and will most probably have to be addressed at different levels of accuracy during different stages of the design process. Recently, methods based on the combination of several models (multi-layer scoring system, consensus scoring) have been shown to increase the predictivity [59] and to reduce the number of false positives suggested by individual scoring functions [60].

3 The SEED approach

Figure 1 shows a flow-chart of the library docking program SEED. A brief explanation is given here while further details of the method, e.g., the clustering procedure and evaluation of the van der Waals interactions, can be found in the original paper [8]. Different types of fragments are docked in the order specified by the user. After each fragment placement the binding energy is estimated. The binding energy is the sum of the van der Waals interaction and electrostatic energy with continuum solvation. Each fragment type is docked after all placement–energy evaluations of the preceding fragment type have been made. The fragment docking procedure and electrostatic energy evaluation are outlined in this section. For the docking of a library of 100 fragments into a rigid binding site of about 25 residues, the latest version of SEED requires about 5 hours of CPU time on a single processor (195 MHz R10000 or PentiumIII 550 MHz). For more than one processor the speed-up is linear so that the docking of a library of 1000 fragments would require about 6 hours on a cluster of eight commodity-price processors.

3.1 Fragment docking

The binding site of the receptor is defined by a list of residues which are selected by the user. Fragments are docked using the polar vectors if they have at least one H-bond donor or acceptor. Due to this definition some "polar" fragments can have considerable hydrophobic character (e.g., diphenylether). Therefore they are also docked by the procedure for nonpolar fragments unless otherwise specified by the user.

Docking of polar fragments. These are docked so that one or more hydrogen bonds with the receptor is formed. The fragment is then rotated around the H-bond axis to increase sampling. Fig. 2 shows the sampling of docked positions for pyrrole and acetone around a tyrosine side chain. Ideal and close-to-ideal hydrogen bond geometries are sampled in a discrete but exhaustive way.

Docking of nonpolar fragments. The hydrophobicity maps [61] are used to dock nonpolar fragments. The points on the receptor SAS are ranked according to the sum of van der Waals interaction and receptor desolvation [61, 62], and the n most hydrophobic points (where n is an input parameter) are selected for docking. As an illustrative example, Fig. 3 shows the most hydrophobic points on the ATP binding site of the p38 MAP kinase. For both the fragment and the receptor, apolar vectors are defined by joining each point on the SAS with the corresponding atom. Finally, nonpolar fragments are docked by matching an apolar vector of the fragment with an apolar vector of the receptor at the optimal van der Waals distance. To improve sampling, additional rotations of the fragment are performed around the axis joining the receptor atom and fragment atom (Fig. 2). To increase efficiency nonpolar fragments are discarded without calculation of the electrostatic energy if the van der Waals interaction is less favorable than a threshold value.

For both polar and nonpolar fragments, the docking is exhaustive on a discrete space. The discretization originates from the limited number of preferred directions and rotations around them. Fragment symmetries are checked only once for every fragment type and are exploited to increase the efficiency in docking.

3.2 Electrostatic energy with continuum solvation

The main assumption underlying the evaluation of the electrostatic energy of a fragment-receptor complex is the description of the solvent effects by continuum electrostatics [62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72]. The system is partitioned into solvent and solute regions and appropriate values of the dielectric constant are assigned to each region. In this approximation only the intra-solute electrostatic interactions need to be evaluated. This strongly reduces the number of interactions with respect to an explicit treatment of the solvent. Moreover it makes feasible the inclusion of solvent effects in docking studies where the equilibration of explicit water molecules would be a major difficulty. In docking and even more so in ligand design, the electrostatic effects of the solvent have to be modelled accurately; it has been shown that the continuum dielectric model provides an efficient and useful approximation of molecules and molecular complexes

in solution [62, 63, 73].

The difference in electrostatic energy in solution upon binding of a fragment to a receptor can be calculated as the sum of the following three terms [9, 65]:

- Partial desolvation of the receptor: electrostatic energy difference due to the displacement of high dielectric solvent by the fragment volume.
- Screened receptor-fragment interaction: intermolecular electrostatic energy between the fragment and the receptor in the solvent.
- Partial desolvation of the fragment: electrostatic energy difference due to the displacement of high dielectric solvent by the receptor volume.

The definition of the solute volume, i.e., the low dielectric volume, is central in the evaluation of these energy terms with a continuum model. The solute-solvent dielectric boundary is described by the molecular surface (MS) of the solute [74]. A grid covering the receptor is utilized to identify the low dielectric volume. In a first step the volume occupied by the isolated receptor is defined on the grid. Subsequently for every position of a docked fragment the volume enclosed by the MS of the fragment-receptor complex is identified.

The screened fragment-receptor interaction and the fragment desolvation are evaluated with a grid-based implementation [62, 63] of the generalized Born (GB) approximation [68, 69, 70, 71, 72]. The GB approach would be too time consuming for the evaluation of the desolvation of the receptor which is calculated as described below.

Receptor desolvation. The electrostatic desolvation of the receptor accounts for the loss of receptor-solvent favorable electrostatic interactions due to the removal of part of the highly polarizable solvent to accommodate the volume of the fragment. This contribution always disfavors association and can be calculated within the assumption of continuum electrostatics [62, 63, 64, 68, 69, 70, 71, 72]. The electrostatic energy E of the receptor in solution can be expressed in terms of the electric displacement vector $\vec{D}(\vec{x})$ and of a location dependent dielectric

constant $\epsilon(\vec{x})$ as an integral over the three-dimensional (3D) space R^3 [75]:

$$E = \frac{1}{8\pi}\int_{R^3} \frac{\vec{D}^2(\vec{x})}{\epsilon(\vec{x})}\, d^3x \tag{1}$$

Since $\vec{D}(\vec{x})$ is additive, for point charges it can be rewritten as a sum over all charges i of the receptor:

$$\vec{D}(\vec{x}) = \sum_i \vec{D}_i(\vec{x}) \tag{2}$$

For what concerns the electrostatics, the displacement of solvent by the fragment volume at the surface of the receptor has the only effect of modifying the dielectric properties in the space occupied by the fragment. Over this volume the dielectric constant changes from the solvent value (ϵ_w) to the solute value (ϵ_p). Usually, ϵ_w is set to 78.5 which is the value of water at room temperature, while the value of ϵ_p can range from 1 to 4. In the limit in which $\vec{D}(\vec{x})$ does not change significantly upon docking of the ligand, the variation of the electrostatic energy of the receptor (i.e., the desolvation) can be written according to equation 1 as an integral over the volume occupied by the fragment (V_{fragment}):

$$\Delta E^{desolv} = \frac{\tau}{8\pi}\int_{V_{\text{fragment}}} \vec{D}^2(\vec{x})\, d^3x \tag{3}$$

where $\tau = \frac{1}{\epsilon_p} - \frac{1}{\epsilon_w}$. A 3D grid is built around the receptor and equation 3 becomes:

$$\Delta E^{desolv} = \frac{\tau}{8\pi} \sum_{k \in V_{\text{fragment}}} \vec{D}^2(\vec{x_k}) \Delta V_k \tag{4}$$

where the index k runs over the grid points occupied by the fragment. The grid spacing is usually 0.5 Å. The electric displacement of every charge of the receptor can be approximated by the Coulomb field [62, 71, 76]:

$$\vec{D}(\vec{x}) = \sum_i q_i \frac{(\vec{x} - \vec{x_i})}{|\vec{x} - \vec{x_i}|^3} \tag{5}$$

where $\vec{x_i}$ is the position of the receptor atom i and q_i its partial charge. Equation 5 is an analytical approximation of the total electric displacement and fulfills the condition of validity of equations 3 and 4, i.e., $\vec{D}(\vec{x})$ is independent of the dielectric environment. The receptor desolvation in the Coulomb field approximation results from equation 4 together with equation 5:

$$\Delta E^{desolv} = \frac{\tau}{8\pi} \sum_{k \in V_{\text{fragment}}} \left(\sum_i q_i \frac{(\vec{x}_k - \vec{x_i})}{|\vec{x}_k - \vec{x_i}|^3} \right)^2 \Delta V_k \tag{6}$$

The volume occupied by a docked fragment is the part of the volume enclosed by the MS of the complex that was not occupied by the isolated receptor. It consists of the actual volume of the fragment and the interstitial volume enclosed by the reentrant surface between fragment and receptor.

It is important to note that the desolvation of a charged ion by a small nonpolar sphere at a distance r from the ion varies approximately as $\frac{1}{r^4}$ (equation 6). This is a very short range effect if compared with the ion electrostatic potential which varies as $\frac{1}{r}$.

Screened fragment-receptor interaction. The fragment-receptor interaction in solution is calculated via the GB approximation [68]. In a solvent of dielectric constant ϵ_w, the interaction energy between two charges embedded in a solute of dielectric constant ϵ_p is

$$E_{ij}^{int} = \frac{q_i q_j}{\epsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \tag{7}$$

where $\tau = \frac{1}{\epsilon_p} - \frac{1}{\epsilon_w}$,

$$R_{ij}^{GB} = \sqrt{r_{ij}^2 + R_i^{eff} R_j^{eff} exp\left(\frac{-r_{ij}^2}{4R_i^{eff} R_j^{eff}}\right)} \tag{8}$$

and q_i is the value of the partial charge i, while r_{ij} is the distance between charges i and j. R_i^{eff} is the effective radius of charge i and it is evaluated numerically on a 3D grid covering the solute as described in [62]. It is a quantity depending only on the solute geometry and represents an estimate of the average distance of a charge from the solvent.

The intermolecular interaction energy is calculated as:

$$E^{int} = \sum_{\substack{i \in fragment \\ j \in list_i}} E_{ij}^{int} \tag{9}$$

where $list_i$ contains the receptor atoms belonging to the neighbor list of atom i. The electrostatic neighbor list includes all the receptor atoms of the van der Waals neighbor list and one atom for every charged residue whose charge center is within a distance of 13 Å from the closest binding site residue. Supplementing the van der Waals neighbor list with a monopole approximation of distant charged

residues dramatically reduces the error originating from the long range electrostatic interactions.

Fragment desolvation. The fragment intramolecular energy in solution is calculated with the GB formula as described in [62]:

$$E = \sum_{i \in fragment} E_i^{self} + \sum_{\substack{i>j \\ i,j \in fragment}} \left(\frac{q_i q_j}{\epsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \right) \tag{10}$$

where the two sums run over the partial charges of the fragment. Equation 10 differs from equation 9 due to the presence of the *self-energy* term $\sum_i E_i^{self}$. This term is not zero only in the case of intramolecular energies. E_i^{self} is the *self-energy* of charge i and represents the interaction between the charge itself and the solvent. It is calculated as [62, 71]:

$$E_i^{self} = \frac{q_i^2}{2R_i^{vdW}\epsilon_p} - \frac{q_i^2 \tau}{2R_i^{eff}} \tag{11}$$

where R_i^{vdW} is the van der Waals radius of charge i.

The difference in the intramolecular fragment energy upon binding to an uncharged receptor in solution is:

$$\Delta E = E^{docked} - E^{free} \tag{12}$$

where E^{docked} and E^{free} are the energies in solution of the fragment bound and unbound to the receptor, respectively. They are evaluated according to equation 10. For the unbound fragment (E^{free}) the effective radii are calculated considering as solute the volume enclosed by the molecular surface of the fragment. For the bound fragment (E^{docked}) the solute is the volume enclosed by the molecular surface of the receptor-fragment complex. E^{free} is evaluated only once per fragment type, while E^{docked} is recalculated for every fragment position in the binding site.

3.3 Validation

The approximations inherent to our continuum electrostatic approach were validated by comparison with finite difference solutions of the Poisson equation [8]. For this purpose, the three electrostatic energy terms were calculated with SEED

and UHBD [67, 77] for a set of small molecules and ions distributed over the binding site of thrombin and at the dimerization interface of the HIV-1 aspartic protease monomer. The molecule set included acetate ion, benzoate ion, methylsulfonate ion, methylammonium ion, methylguanidinium ion, 2,5-diketopiperazine, and benzene. The total number of fragment-receptor complexes analyzed were 1025 for thrombin and 1490 for the HIV-1 protease monomer. The agreement between the two methods is very good, and better for solute dielectric constant of 4.0 than 1.0 (see Table II of [8]). It was also shown that systematic errors (slope $\neq 1$) are independent of the receptor and the solute dielectric constant and consequently can be corrected by the use of appropriate scaling factors for the different energy terms [62].

4 Illustrative application

4.1 The MAP kinase family

Mitogen-activated protein (MAP) kinases are essential enzymes for intracellular signalling cascades because they phosphorylate several regulatory proteins. They are responsive to hormones, cytokines, environmental stresses, and other extracellular stimuli, and are activated by a dual phosphorylation of a threonine and a tyrosine residues in a TXY motif in the so-called phosphorylation lip. A MAP kinase is characterized by its downstream substrates and by the kinases by which it is preferentially activated. The best characterized MAP kinases are the extra-cellular-signal regulated kinases ERKs (TEY activation motif) [78, 79], the c-Jun N-terminal kinases JNKs (TPY motif) [80, 81, 82], and p38 (TGY motif) [83, 84, 85, 86]. p38 MAP kinase (also called CSBP2) plays a role in processes as diverse as transcriptional regulation, production of interleukins, and apoptosis of neuronal cells. Inhibitors of p38 activity could therefore be useful as a treatment strategy for inflammatory and neurodegenerative diseases.

All unphosphorylated forms of MAP kinases have a similar topology, which consists of two domains separated by a substrate-binding cleft [87]. The N-terminal domain incorporates the glycine-rich loop, which contains the ATP binding motif

GXGXXG, while the C-terminal domain contains the magnesium sites, the catalytic base, and the phosphorylation "lip" with the TXY motif (Fig. 4). Yet, some noticeable differences affect the geometry of the ATP-binding site: *i*) the phosphorylation lips vary in sequence and structure, *ii*) the relative orientation of the two domains is different leading to different domain interface shapes, and *iii*) some residues in the ATP-binding site are different, e.g., at position 106 (sequence number according to p38) there is a Thr in p38, a Gln in the ERKs, and a Met in the JNKs.

4.2 Specific inhibitors of p38

The CSAID^{TM} (cytokine suppressive anti-inflammatory drugs) class of anti-inflammatory compounds inhibits the synthesis of cytokines, such as interleukin-1 and tumor necrosis factor, by specific inhibition of the p38 MAP kinase [83, 88, 89]. They have a common chemical pattern: a central five-membered ring, either imidazole or pyrrole, substituted by a pyridine or a pyrimidine ring at C-5, a fluorinated or iodinated phenyl ring at C-4, and a third substituent at position N-1 or C-2 (Fig. 5). These low–molecular weight inhibitors and their analogs bind to the ATP-binding cleft of the unactivated form of p38 and are competitive with respect to ATP. They are potent inhibitors, with IC_{50} in the nanomolar range, and highly selective for p38 compared to the other MAP kinases.

Both biochemical and structural data suggest that the specificity towards p38 kinases is determined by differences in nonconserved regions within or near the ATP binding site. In particular, a site-directed mutagenesis study demonstrated the crucial role of Thr106 [90]. The determination of the three-dimensional structures of the apo, unphosphorylated human and murine p38 MAP kinases, free and bound to various small-sized inhibitors, allowed to gain more insights into the structural affinity and specificity determinants of the p38 binding site [91, 92, 93, 94]. The binding modes of the inhibitors share common properties: the pyridine or pyrimidine nitrogen acts as hydrogen bond acceptor for the Met109 backbone NH and the phenyl ring is inserted into a mainly hydrophobic pocket, delimited by the Lys53 and Thr106 side chains. The former feature is analogous to what is observed

in complexes with ATP and with other protein kinase inhibitors [95, 96, 97, 98]. Conversely, the latter is unique to p38. The third substituent of the central scaffold may also be involved in inhibitor selectivity by interacting with various residues of the glycine-rich loop and occasionally with the Asp168 side chain. The imidazole scaffold is hydrogen-bonded to the Lys53 side chain. Despite these common properties of inhibitor binding, the detailed organization of the binding cleft differs in the structures of the complexes between p38 and two series of inhibitors [93, 94]. The loop containing the ATP binding motif adopts different orientations, leading to a more or less open site, and the side chain orientation of some critical residues, e.g., Lys53, Thr106, and Asp168, depends on the presence or absence of the inhibitor in the site, as well as on the type of inhibitor [93, 94]. Overall, the ATP binding site of p38 displays a remarkable flexibility.

4.3 Docking of molecular fragments with SEED

Fragment library. A library of 70 mainly rigid fragments ranging in size from 7 to 31 atoms was used in this study. It contains 17 apolar fragments (no hydrogen bond donors or acceptors), 39 polar and neutral compounds, and 14 fragments with one or two formal charges (Table 1). Many of the molecular frameworks found frequently in known drugs [99] are included (e.g., benzene, pyridine, naphthalene, 5-phenyl-1,4-benzodiazepine, etc.) and some of them can be used for the synthesis of combinatorial libraries in the solid phase [100] or by portioning and mixing [101]. Fragment structures were generated with the molecular modelling program WITNOTP (A. Widmer, Novartis Pharma Basel, unpublished). For each fragment type, all of the low-energy conformations are included in the library (e.g., cis and trans for 2-butene). Partial charges were assigned with an iterative method based on the partial equalization of orbital electronegativities (MPEOE, [102, 103]). The MPEOE implementation in WITNOTP reproduces the all-hydrogen CHARMm22 parameter set (Molecular Simulations Inc.) for proteins and proteinaceous fragments. Fragment coordinates were minimized with the program CHARMM [104] and the CHARMm22 parameter set to an average value of the energy gradient of 0.01 kcal/mol Å using a linear distance-dependent dielectric function.

SEED input parameters. The same input parameters as in Table I of the original SEED paper [8] were used except for the following ones. The interior dielectric constant was set to 4 to model the electronic polarizability and dipolar reorientation effects of the solute [105]. The number of apolar vectors was increased from 100 to 150 because of the very large binding site (see below). Finally, the radius of the probe sphere for the definition of the SAS (used for the selection of the apolar vectors) was set to 1.4 Å (instead of 1.8 Å) to better define microcavities and small crevices. A preselection of the positions of the fragments was made by discarding those whose geometric center was outside a sphere of 10 Å radius centered on the center of the binding site. All polar fragments were docked using both polar and apolar surface points.

System set-up. The structure of the human p38 with the inhibitor SB203580 inside the ATP binding site [94] was downloaded from the PDB (code 1a9u [106]). The water molecules and the inhibitor were removed. Hydrogen atoms were added with the program WITNOTP. Partial charges were assigned to p38 with the MPEOE method [102, 103] implemented in WITNOTP which, as mentioned above, reproduces the all-hydrogen MSI CHARMm22 parameter set. The following thirty-two residues of the ATP binding site were used by SEED to position the apolar and polar surface points for docking: Val30, Gly31, Ser32, Gly33, Ala34, Tyr35, Gly36, Ser37, Val38, Cys39, Val50, Ala51, Val52, Lys53, Glu71, Leu75, Ile84, Gly85, Leu86, Leu104, Val105, Thr106, His107, Leu108, Met109, Gly110, Ala111, Asp112, Asn155, Leu167, Asp168, and Arg173.

Description of the binding site. On Fig. 3 the most hydrophobic regions of the ATP binding site are displayed together with the most hydrophobic points. The hydrophobic regions were determined using the method developed by Scarsi *et al.* [61]. Five hydrophobic regions of concave shape were found. They will be referenced as pockets 1 to 5 henceforth. Region 5 is almost flat and pockets 2 and 4 are less concave than pockets 1 and 3. Pocket 1, located between the Thr106 and Lys53 side chains, is occupied by the phenyl group of the known inhibitors, while

pocket 2, lined by the Thr106 and Met109 side chains, is occupied by the pyridine or pyrimidine cycle. The N-substituent of the central imidazol or pyrrole groups is in contact with pocket 5, close to the Val30 and Val38 side chains. Surprisingly, pockets 3 and 4 are empty in the available crystal structures of the MAP kinase p38/inhibitor complexes.

The backbone NH of Met109 and the side chains of Lys53, Tyr35, Arg67, and Arg173 are hydrogen bond donors in the protein binding site, that can interact with candidate ligands. Interesting hydrogen bond acceptors are in the side chains of Asp168, Glu71, and Tyr35.

4.4 p38 functionality maps

The library used in this study contains 4-(4-fluorophenyl)-1-methyl-5-(4-pyridyl)-imidazole (Compound **1**, Fig. 5) that is a close analog of a class of potent CSAID^{TM} inhibitors of p38 [83, 88, 89]. Compound **1** has three of the four rings of the known inhibitors but lacks the C-2 substituent of the imidazole. Figure 6 shows that SEED docks compound **1** in the right orientation (heavy atom root mean square deviation of 0.9 Å from the position of the inhibitor SB203580, PDB code 1a9u [94]) and ranks it as best, among the 70 fragments of the library, with a very favorable van der Waals energy (Table 2). Furthermore, the aromatic rings of the second and third best fragments, diphenylether and dibenzocyclohexane, overlap the pyridine and phenyl rings of the inhibitor SB203580. To better describe the SEED results, the functionality map discussion is divided into three subsections according to the polar character of the fragments.

4.4.1 Fragments docked using the apolar vectors

Common trends are observed for most of the apolar fragments containing a phenyl ring, on one hand, and for most of the alkyl groups, on the other hand. Generally, all hydrophobic pockets accommodate apolar fragments, but some preference is observed for pockets 1 to 4, while pocket 5 is less favorable because it is almost flat. As expected, the electrostatic intermolecular term and the desolvation of the apolar fragment are negligible. We describe the results obtained with benzene and

cyclohexane, as representative functional groups. Diphenylether, the fragment with the second best binding energy in the library after compound **1**, is also analyzed to describe the behavior of slightly polar fragments docked as apolar fragments. In general, the functionality maps of nonpolar fragments are consistent with the available structural data. Moreover, they indicate that the binding affinity might be improved by supplementing the known ligands with substituents that fill the pockets 3 and/or 4.

Benzene. A large energy gap (about 4 kcal/mol) is observed between the representative of the first cluster and the representatives of the other clusters. Moreover, the nine remaining members of the first cluster also have a more favorable energy than the positions in the following clusters. This difference is mainly due to the large favorable van der Waals term (Table 2). The first cluster of benzene is embedded in the hydrophobic pocket 1, where the phenyl group of the known inhibitors is placed. The orientation of its members is also similar to that observed in the crystallographic structure (Fig. 7). The representatives of the remaining clusters display close energy values. The other apolar fragments containing a phenyl ring (e.g., naphthalene, tetraline, N-methyl indole, and dibenzocyclohexane) show the same behavior: there is a large energy gap (from 2.5 to 4.0 kcal/mol) between the position of the first cluster and the other cluster representatives. The phenyl ring of the best position is located in the hydrophobic pocket 1, except for N-methyl indole whose substituted five-membered ring is placed in hydrophobic pocket 1.

Cyclohexane. Very close binding energies are obtained for the first four cluster representatives (Table 2), although they are in four different regions of the binding site. However the repartition between the different energy terms is not similar: the fragments found in the hydrophobic pockets 1 and 3 display a more favorable van der Waals term than those in the pockets 2 and 4. Conversely, the electrostatic desolvation term of the receptor is more favorable in the latter regions than in the former ones. The same trend is observed for benzene and for all apolar fragments, which are small enough to fill only one pocket. This means that better van der

Waals contacts can be achieved in the pockets 1 and 3 to the detriment of electrostatic interactions of the receptor polar groups found in these regions. Except for adamantane and propane, no energy gap is observed between the representative positions of the first clusters of the aliphatic fragments. For the other alkyl groups the position ranking is different for each fragment type, but the total binding energies are close to one another. The large, almost spherical, adamantane fragment has a large energy gap (2.5 kcal/mol) between the second and the third cluster representatives. The first two clusters are located in the pockets 2 and 4 which are large enough to contain this fragment while cluster 3 is in between. A 2.0 kcal/mol energy gap is observed between the representative of the first cluster of propane located in the hydrophobic pocket 1 and the remaining ones.

Diphenylether. This fragment contains two phenyl rings linked by an oxygen atom. Here, it was docked using both the apolar and polar vectors but the best binding modes were obtained with the former. Due to the size-dependence of the van der Waals interaction, the intermolecular energy is more favorable than for benzene and cyclohexane and diphenylether ranks as nr. 2 among the 70 compounds of the library. Energy gaps larger than 2.0 kcal/mol exist between the representatives of the first and the second clusters and between the second and the remaining ones. In the best binding mode diphenylether matches the pyridine and phenyl ring of the SB203580 inhibitor. Some positions allow a weak hydrogen bond between the diphenylether oxygen and the side chains of Tyr35 and Lys53.

Fragments containing a phenyl or pyridyl ring usually have the aromatic ring in pocket 1. The position of the first cluster of 5-phenyl-1,4-benzodiazepine is very interesting, as the 5-phenyl substituent is located in pocket 1, while the benzene ring is in pocket 2. For fragments with mainly apolar character, the binding energies obtained by apolar docking are more favorable than those obtained by polar seeding (data not shown), mainly because of a more favorable van der Waals term. The regions occupied by the clusters are also different. For example, large fragments, as diphenylether or 5-phenyl-1,4-benzodiazepine, do not penetrate deeply into the binding site when they are docked using polar vectors. They are placed at the

entry of the binding site, being hydrogen-bonded to the side chain of Tyr35 or to the backbone of the neighboring residues.

4-(4-fluorophenyl)-1-methyl-5-(4-pyridyl)-imidazole. Another interesting example of apolar docking concerns compound **1**, a close analog of a series of potent p38 inhibitors [83, 88, 89]. Compound **1** has the best binding energy of the 70 fragments of the library and is very close to the positions of the inhibitors in the crystallized complexes with p38 [94] (the distances between corresponding heavy atoms range between 0.8 and 1.0 Å). Interestingly, although it was placed using the apolar vectors, the best binding mode has a good electrostatic interaction (-1.9 kcal/mol) and makes two hydrogen-bonds, with the backbone NH of Met109 and the Lys53 side chain (Fig. 6), that are identical to those observed for the inhibitors in the crystal structures. Similarly to what was observed for other fragments docked using the apolar vectors, the representative of the first cluster displays a large gap with the other cluster representatives (3 kcal/mol).

4.4.2 Polar groups

Pyridine and pyrrole are analyzed as they represent hydrogen bond acceptors and donors, respectively. Their electrostatic interaction energy is somewhat more favorable than the one of the fragments without donors and acceptors. However, the dominant term is still the van der Waals energy because the receptor desolvation penalty often compensates the favorable electrostatic interaction.

Pyridine. Pyridine, as well as other fragments containing a similar ring, are preferentially located in the hydrophobic pocket 2, accepting a hydrogen bond from the backbone NH of Met109, as the pyridyl group of the known inhibitors [94]. The main chain NH of Met109 is indeed the privileged partner of fragments with a hydrogen bond acceptor. The orientation of the members of the first cluster is very close to that of the corresponding ring in the crystallographic structure (Fig. 7). The energy gap between the first cluster representative and the others is almost 2 kcal/mol, although other cluster representatives of pyridine interact with the Met109 NH. The difference is distributed among the two intermolecular terms

and the electrostatic desolvation, but the latter seems to play a significant role comparing the energy profiles of the first position and of the fourth and the fifth ones, given in Table 3. The other protein partners for hydrogen bond acceptors are mainly Lys53 and Tyr35 side chains. The latter is preferred by large fragments, like diphenylether or 5-phenyl-1,4-benzodiazepine, for steric reasons, because it is at the entry of the ATP binding site.

Pyrrole. Pyrrole and fragments with hydrogen bonds donors also interact preferentially with the region around Met109. The main hydrogen bonds acceptors are the main chain CO group of Met109, Gly110 and Val30. More seldom are the side chain atoms of Tyr35, Glu71 and Asp168. Energy differences between cluster representatives are rather small. Many fragments in the library contain both donors and acceptors and occupy most of the pockets which are favorable for either pyridine or pyrrole. For example, the representative position of the first cluster of α-carboline is involved in two hydrogen bonds with the protein, with the backbone donor and acceptor of Met109.

4.4.3 Charged groups

The binding energy of functional groups with formal charge(s) is much less favorable than the one of neutral fragments (Table 4). This is consistent with the lack of formal charges in the known inhibitors and indicates that the electrostatic desolvation penalty cannot be neglected. Small or negligible energy differences are observed between cluster representatives of charged groups, which suggests that there is no really favorable region for the charged fragments of the SEED library inside the ATP binding site of the p38 MAP kinase. The results obtained for acetate and methylammonium are described as exemplar negatively and positively charged fragments, respectively (Table 4).

Acetate. The preferred protein partners for acetate are the Lys53 and Arg67 side chains, which are located above the hydrophobic pocket 3. This is also the case for most of the negatively charged fragments. The methyl group of acetate of the fourth cluster members is positioned in pocket 1. Arg173 and Tyr35 are the other donors

interacting with these groups. Interestingly, the best methylphosphate (charge of -2) is at the entry of the binding cleft and interacts with the loop containing the consensus Gly-X-Gly-X-X-Gly sequence which is known to bind the phoshpate group of ATP. However it has a binding energy of about 0 kcal/mol, which is much less favorable than the one of the neutral fragments discussed above. This suggests that the binding affinity of ATP is not due to the triphosphate moiety.

Methylammonium. Conversely to the negatively charged fragments, no positively charged compound is deeply buried into the binding site. They are mainly found in region 4, interacting with the backbone oxygen atom of Gly110 and Met109, as well as with the Asp112 side chain. Other partners are the Val30 carbonyl group and Asp168 side chain.

5 Future directions

The Internet is providing an excellent opportunity for computer-aided drug design; user-friendly, interactive and platform-independent WWW-based tools for molecular modelling exist since several years [107, 108, 109] and more will emerge. In the near future, it is expected that multicenter ligand design will become a reality with several researchers working on the same target by library docking and calculation of binding affinities distributed on several computers connected by very fast networks.

It is clear from this and other chapters that drug design is a really multidisciplinary research field; we hope that this book will spur interest and enthusiasm for computer-aided drug design in chemists, physicists, biologists, and computer scientists. Although the field is not in its infancy anymore, new ideas and multidisciplinary approaches are required to meet the two main challenges: the accurate estimation of binding affinity [4, 110], and the large amount of data emerging from the genomics and proteomics endeavors [111, 112, 113].

Acknowledgements

We thank Dr. Chantal Carrez and Marco Scarsi for helpful discussions. This work was supported by Aventis Pharma (Vitry-sur-Seine, France) and the Swiss National Science Foundation (grant no. 31-53604.98 to AC).

References

[1] A. Caflisch, R. Wälchli, and C. Ehrhardt. Computer–aided design of thrombin inhibitors. *News Physiol. Sci.*, 13:182–189, 1998.

[2] H. Kubinyi. Structure-based design of enzyme inhibitors and receptor ligands. *Curr. Opin. Drug Design Discov.*, 1:4–15, 1998.

[3] L. Weber, S. Wallbaum, C. Broger, and K. Gubernator. Optimization of the biological activity of combinatorial libraries by a genetic algorithm. *Angew. Chem. Int. Ed.*, 34:2280–2282, 1995.

[4] J. Apostolakis and A. Caflisch. Computational ligand design. *Combinatorial Chemistry and High Throughput Screening*, 2:91–104, 1999.

[5] D. Pearlman and M. Murcko. CONCEPTS: new dynamic algorithm for *de novo* drug suggestion. *J. Comput. Chem.*, 14:1184–1193, 1993.

[6] D. K. Gehlhaar, K. E. Moerder, D. Zichi, C. J. Sherman, R. C. Ogden, and S. T. freer. *De novo* design of enzymes inhibitors by Monte Carlo ligand generation. *J. Med. Chem.*, 38:466–472, 1995.

[7] N. P. Todorov and P. M. Dean. A branch-and-bound method for optimal atom-type assignment in *de novo* ligand design. *J. Comput.-Aided Mol. Design*, 12:335–349, 1998.

[8] N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhard, and A. Caflisch. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. *Proteins: Structure, Function and Genetics*, 37:88–105, 1999.

[9] A. Caflisch. Computational combinatorial ligand design: Application to human α-thrombin. *J. Comput.-Aided Mol. Design*, 10:372–396, 1996.

[10] H. J. Böhm. The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J. Comput.-Aided Mol. Design*, 6:61–78, 1992.

[11] P. J. Goodford. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, 28:849–857, 1985.

[12] M. C. Lawrence and P. C. Davis. CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins*, 12:31–41, 1992.

[13] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, 161:269–288, 1982.

[14] C. M. Oshiro, I. D. Kuntz, and J. S. Dixon. Flexible ligand docking using a genetic algorithm. *J. Comput.-Aided Mol. Design*, 9:113–130, 1995.

[15] E. C. Meng, B. K. Shoichet, and I. D. Kuntz. Automated docking with grid-based energy evaluation. *J. Comp. Chem.*, 13:505–524, 1992.

[16] R.L. Desjarlais, G. L. Seibel, I. D. Kuntz nad P. S. Furth, J. C. Alvarez, P. R. Ortiz de Montellano, D. L. DeCamp, L. M. Babe, and C. S. Craik. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. U.S.A.*, 87:6644–6648, 1990.

[17] B. K. Shoichet, R. M. Stroud, D. V. Santi, I. D. Kuntz, and K. M. Perry. Structure-based discovery of inhibitors of thymidylate synthase. *Science*, 259:1445–1450, 1993.

[18] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, 261:470–489, 1996.

[19] H. J. Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Design*, 8:243–256, 1994.

[20] M. Rarey, B. Kramer, and T. Lengauer. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics*, 15:243–250, 1999.

[21] M. Rarey, B. Kramer, and T. Lengauer. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins*, 34:17–28, 1999.

[22] G. Jones, P. Willet, and R. C. Willet. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput.-Aided Mol. Design*, 9:532–549, 1995.

[23] G. Jones, P. Willet, and R. C. Willet. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 245:43–53, 1995.

[24] K. P. Clark and Ajay. Flexible ligand docking without parameter adjustment across four ligand-receptor complexes. *J. Comp. Chem.*, 16:1210–1226, 1995.

[25] G. Jones, P. Willet, R. C. Glen, A. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267:727–748, 1997.

[26] J. Apostolakis, A. Plückthun, and A. Caflisch. Docking small ligands in flexible binding sites. *J. Comput. Chem.*, 19:21–37, 1998.

[27] R. Elber and M. Karplus. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.*, 112:9161–9175, 1990.

[28] A. Miranker and M. Karplus. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins*, 11:29–34, 1991.

[29] R. Rosenfeld, Q. Zheng, S. Vajda, and C. DeLisi. Computing the structure of bound peptides: application to antigen recognition by class I MHCs. *J. Mol. Biol.*, 234:515–521, 1993.

[30] F. Leclerc and M. Karplus. MCSS-based predictions of RNA binding sites. *Theor. Chem. Acc.*, 101:131–137, 1999.

[31] A. Caflisch, H. J. Schramm, and M. Karplus. Design of dimerization inhibitors of HIV-1 aspartic proteinase: a computer-based combinatorial approach. *J. Comput.-Aided Mol. Design*, 14:161–179, 2000.

[32] C. Stultz and M. Karplus. MCSS functionality maps for a flexible protein. *Proteins: Structure, Function and Genetics*, 37:512–529, 1999.

[33] D. Joseph-McCarthy, J. M. Hogle, and M. Karplus. Use of multiple copy simultaneous search (MCSS) method to design a new class of Picornavirus Capsid binding drugs. *Proteins*, 29:32–58, 1997.

[34] P. A. Bartlett, G. T. Shea, S. J. Telfer, and S. Waterman. *Molecular recognition: chemical and biological problems*, pages 182–193. Royal Society of Chemistry, London, 1989.

[35] G. Lauri and P. A. Bartlett. CAVEAT: a program to facilitate the design of organic molecules. *J. Comput.-Aided Mol. Design*, 8:51–66, 1994.

[36] M. B. Eisen, D. C. Wiley, M. Karplus, and R. E. Hubbard. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins*, 19:199–221, 1994.

[37] H. J. Böhm. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Design*, 6:593–606, 1992.

[38] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A32:922–923, 1976.

[39] G. Klebe and T. Mietzner. A fast and efficient method to generate biologically relevant conformations. *J. Comput.-Aided Mol. Design*, 8:583–606, 1994.

[40] G. Klebe, T. Mietzner, and F. Weber. Different approaches toward an automatic structural alignment of drug molecules: Applications to sterol mimics, thrombin and thermolysin inhibitors. *J. Comput.-Aided Mol. Design*, 8:751–778, 1994.

[41] H. J. Böhm. Towards the automatic design of synthetically accessible protein ligands: peptides, amides and peptidomimetics. *J. Comput.-Aided Mol. Design*, 10:265–272, 1996.

[42] V. Gillet, A. P. Johnson, P. Mata, S. Sike, and P. Williams. SPROUT: a program for structure generation. *J. Comput.-Aided Mol. Design*, 7:127–153, 1993.

[43] Y. Sun, T. J. A. Ewing, A. G. Skillman, and I. D. Kuntz. CombiDOCK: structure-based combinatorial docking and library design. *J. Comput.-Aided Mol. Design*, 12:597–604, 1998.

[44] H. J. Böhm, D. W. Banner, and L. Weber. Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J. Comput.-Aided Mol. Design*, 13:51–56, 1999.

[45] S. H. Rotstein and M. A. Murcko. GenStar: a method for de novo drug design. *J. Comput.-Aided Mol. Design*, 7:23–43, 1993.

[46] Y. Nishibata and A. Itai. Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron*, 47:8985–8990, 1991.

[47] Y. Nishibata and A. Itai. Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation. *J. Med. Chem.*, 36:2921–8, Oct 1 1993.

[48] N.L. Allinger. MM2. A hydrocarbon force field utilizing *V1* and *V2* torsional terms. *J. Am. Chem. Soc.*, 99:8127–8134, 1977.

[49] S. H. Rotstein and M. A. Murcko. GroupBuild: a fragment-based method for de novo drug design. *J. Med. Chem.*, 36:1700–10, 1993.

[50] R.S. DeWitte and E.I. Shakhnovich. SMoG: *de novo* design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.*, 118:11733–11744, 1996.

[51] R.S. DeWitte and E.I. Shakhnovich. SMoG: *de novo* design method based on simple, fast, and accurate free energy estimates. 2. Case studies in molecular design. *J. Am. Chem. Soc.*, 119:4608–4617, 1997.

[52] J. B. Moon and W. J. Howe. Computer design of bioactive molecules: a method for receptor-based de novo ligand design. *Proteins*, 11:314–28, 1991.

[53] D. Frenkel, D. E. Clark, J. Li, C. W. Murray, B. Robson, B. Waszkowycz, and D. R. Westhead. PRO_LIGAND: an approach to de novo molecular design. 4. Application to the design of peptides. *J. Comput.-Aided Mol. Design*, 9:213–25, 1995.

[54] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Jr Profeta, and P. A. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106:765–784, 1984.

[55] T. Ooi, M. Oobatake, M. Némethy, and H. A. Scheraga. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. USA.*, 84:3086–3090, 1987.

[56] D. E. Clark, D. Frenkel, S. A. Levy, J. Li, C. W. Murray, B. Robson, B. Waszkowycz, and D. R. Westhead. PRO_LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules. *J Comput Aided Mol Des*, 9:13–32, 1995.

[57] B. K. Shoichet, A. R. Leach, and I. D. Kuntz. Ligand solvation in molecular docking. *Proteins: Structure, Function and Genetics*, 34:4–16, 1999.

[58] A.E. Mark and W.F. van Gunsteren. Decomposition of the free energy of a system in terms of specific interactions. implications for theoretical and experimental studies. *J. Mol. Biol.*, 240:167–176, 1994.

[59] S. So and M. Karplus. A comparative study of ligand-receptor complex binding affinity prediction methods based on glycogen phosphorylase inhibitors. *J. Comput.-Aided Mol. Design*, 13:243–258, 1999.

[60] P. S. Charifson, J. J. Corkery, M. A. Murcko, and W. P. Walters. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.*, 42:5100–5109, 1999.

[61] M. Scarsi, N. Majeux, and A. Caflisch. Hydrophobicity at the surface of proteins. *Proteins: Structure, Function and Genetics*, 37:565–575, 1999.

[62] M. Scarsi, J. Apostolakis, and A. Caflisch. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A*, 101:8098–8106, 1997.

[63] M. Scarsi, J. Apostolakis, and A. Caflisch. Comparison of a GB solvation model with explicit solvent simulations: potentials of mean force and conformational preferences of alanine dipeptide and 1,2-dichloroethane. *J. Phys. Chem. B*, 102:3637–3641, 1998.

[64] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.*, 157:671–679, 1982.

[65] M. K. Gilson and B. H. Honig. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Structure, Function and Genetics*, 4:7–18, 1988.

[66] D. Bashford and M. Karplus. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, 29:10219–10225, 1990.

[67] M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comp. Phys. Comm.*, 62:187–197, 1991.

[68] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[69] G. D. Hawkins, C. J. Cramer, and D. G. Trulhar. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.*, 246:122–129, 1995.

[70] G. D. Hawkins, C. J. Cramer, and D. G. Trulhar. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, 100:19824–19839, 1996.

[71] M. Schaefer and M. Karplus. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.*, 100:1578–1599, 1996.

[72] Di Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A*, 101:3005–3014, 1997.

[73] T. J. Marrone, M. K. Gilson, and J. A. McCammon. Comparison of continuum and explicit models of solvation: Potential of mean force for alanine dipeptide. *J. Phys. Chem.*, 100:1439–1441, 1996.

[74] F. M. Richards. Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng.*, 6:151–176, 1977.

[75] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons: New York, 1975.

[76] R. Luo, J. Moult, and M.K. Gilson. Dielectric screening treatment of electrostatic solvation. *J. Phys. Chem. B*, 101:11226–11236, 1997.

[77] M. E. Davis and J. A. McCammon. Solving the finite difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. *J. Comput. Chem.*, 10:386–391, 1989.

[78] T. G. Boulton, G. D. Yancopoulos, J. S. Gregory, C. Slaughter, C. Moomaw, J. Hsu, and M. H. Cobb. An insulin-stimulated protein kinase similar to yeast kinases involved in cell cycle control. *Science*, 249:64–67, 1990.

[79] T. G. Boulton, J. S. Gregory, and M. H. Cobb. Purification and properties of extracellular signal-regulated kinase 1, an insulin-stimulated microtubule-associated protein 2 kinase. *Biochemistry*, 30:278–286, 1991.

[80] B. Dérijard, M. Hibi, I. H. Wu, T. Barrett, B. Su, T. Deng, M. Karin, and R. J. Davis. JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. *Cell*, 76:1025–1037, 1994.

[81] J. M. Kyriakis, P. Banerjee, E. Nikolakaki, T. Dai, E. A. Rubie, M. F. Ahmad, J. Avruch, and J. R. Woodgett. The stress-activated protein kinase subfamily of c-Jun kinases. *Nature*, 369:156–160, 1994.

[82] T. Kallunki, B. Su, I. Tsigelny, H. K. Sluss, B. Dérijard, G. Moore, R. Davis, and M. Karin. JNK2 contains a specificity-determining region responsible for efficient c-Jun binding and phosphorylation. *Genes Dev.*, 8:2996–3007, 1994.

[83] J. C. Lee, J. T. Laydon, P. C. McDonnell, T. F. Gallagher, S. Kumar, D. Green, D. McNulty, M. J. Blumenthal, J. R. Heys, S. W. Landvatter, and P. R. Young. A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature*, 372:739–746, 1994.

[84] J. Han, J. D. Lee, L. Bibbs, and R. J. Ulevitch. A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells. *Science*, 265:808–811, 1994.

[85] J. Rouse, P. Cohen, S. Trigon, M. Morange, A. Alonso-Llamazares, D. Zamanillo, T. Hunt, and A. R. Nebreda. A novel kinase cascade triggered by stress and heat shock that stimulates MAPKAP kinase-2 and phosphorylation of the small heat shock proteins. *Cell*, 78:1027–1037, 1994.

[86] N. W. Freshney, L. Rawlinson, F. Guesdon, E. Jones, S. Cowley, J. Hsuan, and J. Saklatvala. Interleukin-1 activates a novel protein kinase cascade that results in the phosphorylation of Hsp27. *Cell*, 78:1039–1049, 1994.

[87] F. Zhang, A. Strand, D. Robbins, M. H. Cobb, and E. J. Goldsmith. Atomic structure of the MAP kinase ERK2 at 2.3 A resolution. *Nature*, 367:704–711, 1994.

[88] A. Cuenda, J. Rouse, Y. N. Doza, R. Meier, P. Cohen, T. F. Gallagher, P. R. Young, and J. C. Lee. SB 203580 is a specific inhibitor of a MAP kinase homologue which is stimulated by cellular stresses and interleukin-1. *FEBS Lett.*, 364:229–233, 1995.

[89] J. C. Lee and J. L. Adams. Inhibitors of serine/threonine kinases. *Curr. Opin. Biotechnol.*, 6:657–661, 1995.

[90] J.-M. Lisnock, A. Tebben, B. Franz, E. A. O'Neill, G. Croft, S. J. O'Keefe, B. Li, C. Hacker abd S. de Laszlo, A. Smith, B. Libby abd N. Liverton, J. Hermes, and P. LoGrasso. Molecular basis for p38 protein kinase inhibitor specificity . *Biochemistry*, 37:16573–16581, 1998.

[91] Z. Wang, P. C. Harkins, R. J. Ulevitch, J. Han, M. H. Cobb, and E. J. Goldsmith. The structure of the mitogen-activated protein kinase p38 at 2.1 Åresolution. *Proc. Natl. Acad. Sci. U.S.A.*, 94:2327–2332, 1997.

[92] K. P. Wilson, M. J. Fitzgibbon, P. R. Caron, J. P. Griffith, W. Chen, P. G. McCaffrey, S. P. Chambers, and M. S. S. Su. Crystal structure of p38 mitogen-activated protein kinase. *J. Biol. Chem.*, 271:27696–27700, 1996.

[93] L. Tong, S. Pav, D. M. White, S. Rogers, K. M. Crane C. L. Cywin, M. L. Brown, and C. A. Pargellis. A highly specific inhibitor of human p38 MAP kinase binds in the ATP pocket. *Nature Struct. Biol.*, 4:311–316, 1997.

[94] Z. Wang, B. J. Canagarajah, J. C. Boehm, S. Kassisa, M. H. Cobb, P. R. Young, S. Abdel-Meguid, J. L. Adams, and E. J. Goldsmith. Structural basis of inhibitor selectivity in MAP kinases. *Structure*, 6:1117–1128, 1998.

[95] U. Schulze-Gahmen, J. Brandsen, H. D. Jones, D. O. Morgan, L. Meijer, J. Vesely, and S. H. Kim. Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins*, 22:378–391, 1995.

[96] M. Mohammadi, G. McMahon, L. Sun, C. Tang, P. Hirth, B. K. Yeh, S. R. Hubbard, and J. Schlessinger. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science*, 276:955–960, 1997.

[97] J. Singh, E. M. Dobrusin, D. W. Fry, T. Haske, A. Whitty, and D. J. McNamara. Structure-based design of a potent, selective, and irreversible inhibitor of the catalytic domain of the erbB receptor subfamily of protein tyrosine kinases. *J. Med. Chem.*, 40:1130–1135, 1997.

[98] W. F. De Avezado, H. J. Mueller-Dieckman, U. Schulze-Gahmen, P. J. Worland, E. Sausville, and S. H. Kim. Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase. *Proc. Natl. Acad. Sci. U.S.A.*, 93:2735–2740, 1996.

[99] G.W. Bemis and M. A. Murcko. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.*, 39:2887–2893, 1996.

[100] C.J. Andres, D.J. Denhart, M.S. Deshpande, and K.W. Gillman. Recent advances in the solid phase synthesis of drug-like heterocyclic small molecules. *Combinat. Chem. and High Throughput Screening*, 2:191–210, 1999.

[101] A. Furka and W.D. Bennett. Combinatorial libraries by portioning and mixing. *Combinat. Chem. and High Throughput Screening*, 2:105–122, 1999.

[102] J. Gasteiger and M. Marsili. Iterative partial equalization of orbital electronegativity. A rapid access to atomic charges. *Tetrahedron*, 36:3219–3288, 1980.

[103] K.T. No, J.A. Grant, and H.A. Scheraga. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J. Phys. Chem.*, 94:4732–4739, 1990.

[104] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

[105] M. K. Gilson and B. H. Honig. Energetics of charge-charge interactions in proteins. *Proteins: Structure, Function and Genetics*, 3:32–52, 1988.

[106] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Jr Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542, 1977.

[107] A. R. Leach. Structure-based selection of building blocks for array synthesis via the World-Wide Web. *J. Mol. Graphics Modell.*, 15:158–160, 1997.

[108] P. Ertl. World Wide Web-based system for the calculation of substituent parameters and substituent similarity searches. *J. Mol. Graphics Modell.*, 16:11–13, 1998.

[109] D. J. Wild and C. J. Blankley. VisualiSAR: a Web-based application for clustering, structure browsing, and structure-activity relationship study. *J. Mol. Graphics Modell.*, 17:85–89, 1999.

[110] Ajay and M. A. Murcko. Computational methods to predict binding free energy in ligand–receptor complexes. *J. Med. Chem.*, 38(26):4953–4967, 1995.

[111] R. Sánchez and A. Sali. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. *Proc. Natl. Acad. Sci. USA.*, 95:13597–13602, 1998.

[112] E. M. Marcotte, M. Pellegrini, M. J. Thompson, T. O. Yeates, and D. Eisenberg. A combined algorithm for genome-wide prediction of protein function. *Nature*, 402:83–86, 1999.

[113] A. J. Enright, I. Iliopoulos, N. C. Kyrpides, and C. A. Ouzounis. Protein interaction maps for complete genomes based on gene fusion events. *Nature*, 402:86–89, 1999.

[114] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran.* Cambridge University Press, 1992.

[115] A. T. Hagler, E. Huler, and S. Lifson. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.*, 96:5319–5327, 1974.

[116] N. Pattabiraman, M. Levitt, T. E. Ferrin, and R. Langridge. Computer graphics in real-time docking with energy calculation and minimization. *J. Comput. Chem.*, 6:432–436, 1985.

[117] E. C. Meng, B. K. Shoichet, and I. D. Kuntz. Automated docking with grid–based energy evaluation. *J. Comput. Chem.*, 13:505–524, 1992.

[118] B.A. Luty, Z.R. Wasserman, P.F.W. Stouten, C.N. Hodge, M. Zacharias, and J.A. McCammon. A molecular mechanics/grid method for evaluation of ligand-receptor interactions. *J. Comput. Chem.*, 16:454–464, 1995.

[119] A. Nicholls, K. A. Sharp, and B. Honig. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11:281–296, 1991.

[120] M. L. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221:709–713, 1983. Description of the method to calculate the "connolly" surfaces.

[121] P. Kraulis. Molscript, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, 24:946–950, 1991.

Table 1: Fragment library

Fragment	Nr. of atoms heavy	Nr. of atoms total	Nr. of confs.	H-bond Acc.	H-bond Don.	Formal charges
Apolar fragments						
ethane	2	8	1	-	-	0
propane	3	11	1	-	-	0
cyclopropane	3	9	1	-	-	0
2-methylpropane	4	14	1	-	-	0
1-butene	4	12	4	-	-	0
2-butene	4	12	2	-	-	0
2-methyl-2-butene	5	15	1	-	-	0
2,2-dimethylpropane	5	17	1	-	-	0
cyclopentane	5	15	1	-	-	0
benzene	6	12	1	-	-	0
cyclohexane	6	18	1	-	-	0
adamantane	10	26	1	-	-	0
dekaline	10	28	1	-	-	0
naphthalene	10	18	1	-	-	0
N-methylindole	10	19	1	-	-	0
tetraline	10	22	1	-	-	0
dibenzocyclohexane	14	26	1	-	-	0
Neutral fragments with one or more hydrogen bond groups						
dimethylsulfoxyde	4	10	1	1	-	0
isopropanol	4	12	2	1	1	0
imidazole	5	9	1	1	1	0
N-methylacetamide	5	12	1	1	1	0
pyrrole	5	10	1	-	1	0
N-methyl-methylsulfonamide	6	13	1	2	1	0
oxazolidinone	6	11	1	2	1	0
pyridine	6	11	1	1	-	0
pyrimidine	6	10	1	2	-	0
2-pyrrolidinone	6	13	1	1	1	0
4-thiazolidinone	6	11	1	1	1	0
delta-valero-lactam	7	16	2	1	1	0
3,4-dihydroxy-tetrahydrofurane	7	15	4	3	2	0
phenol	7	13	1	1	1	0
tetrahydro-2-pyrimidinone	7	15	1	1	2	0
cytosine	8	13	1	2	2	0
1,2-dihydroxy-benzene	8	14	1	2	2	0
1,2-dihydroxy-cyclohexane	8	20	4	2	2	0
2,5-diketo-1,4-piperazine	8	14	2	2	2	0
uracil	8	12	1	2	2	0
indole	9	16	1	-	1	0
2-methyl-3-amino-N-methylbutanamide	9	23	18	1	2	0

Table 1: continued

Fragment	Nr. of atoms heavy	total	Nr. of confs.	H-bond Acc.	Don.	Formal charges
adenine	10	15	1	3	2	0
3,6-dimethyl-2,5-diketo-1,4-piperazine	10	20	3	2	2	0
isoquinoline	10	17	1	1	-	0
N-formyl-L-proline	10	19	8	3	1	0
quinazoline	10	16	1	2	-	0
quinoline	10	17	1	1	-	0
tetrahydro-quinoline	10	21	1	-	1	0
guanine	11	16	1	3	3	0
meso-inositol	12	24	1	6	6	0
alpha-carboline	13	21	1	1	1	0
beta-carboline	13	21	1	1	1	0
diphenylether	13	23	1	1	-	0
5-methyl-3-methylsulfoxyde-acetophenone	13	25	8	2	-	0
2,3,4-furantricarboxylic-acid	14	18	8	7	3	0
5-phenyl-1,4-benzodiazepine	17	29	2	2	-	0
5-phenyl-1,4-benzodiazepine-2-one	18	30	2	2	1	0
4-(4-fluorophenyl)-1-methyl-5-(4-pyridyl)--imidazole **(1)**	19	31	2	2	-	0
Charged fragments						
methylammonium	2	8	1	-	1	+
methylamidine	4	11	1	-	2	+
methylguanidine	5	13	1	-	3	+
tetrahydropyrrole	5	15	1	-	1	+
piperidine	6	18	1	-	1	+
benzamidine	9	18	2	-	2	+
5-amidine-indole	12	22	2	-	3	+
acetate	4	7	1	2	-	−
methylsulfonate	5	8	1	3	-	−
benzoic acid	9	14	1	2	-	−
L-proline	8	17	2	2	1	+ −
D-proline	8	17	2	2	1	+ −
piperazine	6	18	1	-	2	++
methylphosphate	6	9	1	4	-	− −

Table 2: Cluster representatives of fragments docked using the apolar vectors*

Rank	Intermolecular		Elect. desolvation		ΔG_{bind}†	Site‡
	vdW	Elec	Receptor	Fragment		
benzene						
1	-14.7	-0.6	3.4	0.3	-11.6	1
2	-10.1	0.0	2.0	0.3	-7.8	2
3	-11.3	0.3	3.1	0.3	-7.7	3
4	-10.7	0.5	2.8	0.3	-7.2	3
5	-11.0	0.1	3.6	0.3	-7.0	1
cyclohexane						
1	-11.3	0.2	2.4	0.0	-8.6	4
2	-13.1	0.1	4.6	0.0	-8.5	3
3	-11.1	0.1	2.7	0.0	-8.4	2
4	-13.0	-0.0	4.7	0.0	-8.3	1
5	-9.7	0.2	3.6	0.0	-7.8	2,5
diphenylether						
1	-21.3	-0.9	5.5	0.5	-16.2	1,2
2	-19.9	-0.6	6.6	0.5	-13.4	1,2
3	-18.5	0.1	6.5	0.5	-11.4	3
4	-14.4	-0.5	3.8	0.4	-10.7	2,4
5	-17.0	-0.6	6.4	0.5	-10.7	3
4-(4-fluorophenyl)-1-methyl-5-(4-pyridyl)-imidazole						
1	-26.0	-1.9	9.7	1.1	-17.1	1,2
2	-20.1	0.5	4.6	0.9	-14.1	2,4
3	-18.8	-1.0	5.9	1.0	-12.9	2,4
4	-18.7	0.2	5.6	1.0	-11.8	2,4
5	-21.1	-1.9	10.9	1.0	-11.2	1,2

*All energies are in kcal/mol.

†Sum of the values in the four preceding columns, i.e., intermolecular interaction and electrostatics desolvation energies.

‡Numbering of the hydrophobic pockets as defined in Fig. 3.

Table 3: Cluster representatives of fragments docked using polar vectors*

Rank	Intermolecular		Elect. desolvation		ΔG_{bind}†	HB partners
	vdW	Elec	Receptor	Fragment		
pyridine						
1	-9.1	-0.6	1.4	0.8	-7.4	Met109-NH
2	-8.5	-2.0	4.1	0.8	-5.6	Lys53
3	-7.6	0.1	1.3	0.8	-5.4	Met109-NH
4	-9.2	-1.6	4.9	0.8	-5.1	Lys53
5	-9.1	-2.2	6.1	0.8	-4.4	Lys53
pyrrole						
1	-6.6	-0.3	1.1	0.3	-5.5	Met109-CO
2	-5.0	-0.5	0.8	0.3	-4.4	Val30-CO
3	-5.0	-0.4	1.0	0.3	-4.0	Val30-CO
4	-4.8	-0.5	1.3	0.3	-3.7	Met109-CO
5	-4.7	-0.3	1.1	0.3	-3.6	Gly110-CO

*All energies are in kcal/mol.

†Sum of the values in the four preceding columns, i.e., intermolecular interaction and electrostatics desolvation energies.

Table 4: Cluster representatives of charged fragments*

Rank	Intermolecular		Elect. desolvation		ΔG_{bind}†	HB partners
	vdW	Elec	Receptor	Fragment		
acetate						
1	-6.1	-7.5	3.6	7.6	-2.4	Arg67-$H^{\eta 21}$ Lys53-$H^{\zeta 3}$
2	-5.9	-7.7	3.5	7.7	-2.4	Arg67-$H^{\eta 21}$ Lys53-$H^{\zeta 3}$
3	-5.4	-8.1	3.4	7.9	-2.2	Arg67-$H^{\eta 21}$ Lys53-$H^{\zeta 3}$
4	-8.5	-4.2	3.0	7.6	-2.1	Lys53-$H^{\zeta 1}$
5	-4.9	-8.2	3.3	8.2	-1.6	Lys53-$H^{\zeta 3}$ Arg67-$H^{\eta 21}$
methylammonium						
1	-1.9	-6.1	1.2	3.9	-2.9	Gly110-CO
2	-3.5	-3.7	1.2	3.2	-2.8	Asp112-$O^{\delta 2}$
3	-3.8	-3.6	0.9	4.3	-2.3	Gly110-CO
4	-5.4	-5.0	1.0	7.3	-2.0	Gly110-CO
5	-1.3	-6.9	1.6	4.7	-1.9	Met109-CO

*All energies are in kcal/mol.

†Sum of the values in the four preceding columns, i.e., intermolecular interaction and electrostatics desolvation energies.

Figure legends

Figure 1. Flow-chart of the program SEED. The fast van der Waals energy is evaluated by trilinear interpolation [114] from a grid (look-up table) using the geometric mean approximation [115, 116, 117, 118]. The fast total energy includes solvation and will be presented elsewhere (Majeux et al., submitted).

Figure 2. Relaxed-eyes stereoview of benzene, acetone, and pyrrole docked by SEED around a tyrosine side chain. The benzene groups are distributed above and below the plane of the tyrosine phenolic ring, while pyrrole and acetone fragments are involved in hydrogen bonds with the tyrosine hydroxyl as donors and acceptors, respectively. Carbon atoms are black, oxygen and nitrogen atoms dark gray, and hydrogen atoms light gray. Hydrogen bonds are drawn with dashed lines.

Figure 3. ATP binding site in the p38 MAP kinase (PDB file 1a9u, [94]). (Top) The hydrophobic pockets are colored in green using the hydrophobicity map approach [61]. They are referenced by numbers. The figure was made with GRASP [119]. (Bottom) The yellow spheres represent the most hydrophobic points [61] in the binding site. The residues lining the binding site are displayed with cylinders. The molecular surface, calculated with the Connolly algorithm [120], is displayed by white dots.

Figure 4. Molscript picture [121] of the p38 MAP kinase (PDB file 1a9u, [94]).

Figure 5. Chemical structure of a series of potent and selective p38 inhibitors [94]. X = N, CH; Y = N, CH; R_1 = H, methylcyclopropyl, piperidinyl; R_2 = H, 4-methylsulfinylphenyl, piperidinyl; R_3 = H, F; R_4 = H, I, CF_3. Compound **1** of the SEED fragment library is 4-(4-fluorophenyl)-1-methyl-5-(4-pyridyl)-imidazole, i.e., $R_1 = CH_3$, R_2 = H, R_3 = F, R_4 = H, X = CH, Y = N.

Figure 6. Relaxed-eyes stereoview of the p38 MAP kinase binding site (thin

lines) with the SB203580 inhibitor [94] (medium lines, green carbon atoms), and a close analog, i.e., compound **1** (thick lines), docked by SEED. Oxygen atoms are colored in red, nitrogen in blue, sulfur in yellow, and hydrogen in cyan. Hydrogen bonds are shown by a red dashed line.

Figure 7. Relaxed-eyes stereoview of the ten best benzenes and pyridines docked by SEED into the p38 MAP kinase. The bound conformation of the SB203580 inhibitor [94] is displayed to show that the ten best benzenes and pyridines match the corresponding groups of SB203580.