Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

A. MAY, S. EISENHARDT, J. SCHMIDT-EHRENBERG,
F. CORDES

Rigid Body Docking for Virtual Screening

ZIB-Report 03-47 (December 2003)

Rigid Body Docking for Virtual Screening

**A. May, S. Eisenhardt, J. Schmidt-Ehrenberg,
F. Cordes**

Abstract

A recently developed algorithm allows Rigid Body Docking of ligands to proteins, regardless of the accessibility and location of the binding site. The Docking procedure is divided into three subsequent optimization phases, two of which utilize rigid body dynamics. The last one is applied with the ligand already positioned inside the binding pocket and accounts for full flexibility. Initially, a combination of geometrical and force-field based methods is used as a Point-Matching strategy, considering only Lennard-Jones interactions between the target and pharmaceutically relevant atoms or functional groups. The protein is subjected to a Hot Spot Analysis, which reveals points of high affinity in the protein environment towards these groups. The hot spots are distributed into different subsets according to their group affiliation. The ligand is described as a complementary point set, consisting of the same subsets. Both sets are matched in $\mathbb{R}^3$, by superimposing members of the same subsets. In the first instance, steric inhibition is nearly neglected, preventing the system's trajectory from trapping in local minima and thus from finding false positive solutions. Hence the exact location of the binding site can be determined fast and reliably without any additional information. Subsequently, errors resulting from approximations are minimized via Fine-Tuning, this time considering both Lennard-Jones and Coulomb forces. Finally, the potential energy of the whole complex is minimized. In a first evaluation, results are rated by a reduced scoring function considering only non-covalent interaction energies. Exemplary Screening results will be given for specific ligands.

Key words. Pharmacophore, Pharma Site, Point-Matching, identification of active sites.

Mathematics subject classification. 92C40, 62P10.

Introduction

Structure based Virtual Screening (VS) is a theoretical method for the prediction of binding affinities between small ligands, e.g. potential drugs, and pharmaceutically interesting targets, e.g. proteins, on the basis of structural information of the participating molecules. The process, which brings the two structures together is called Docking. The Docking problem can be subdivided into two steps. Exploring the conformational space of ligands which bind to target molecules and scoring this set, i.e. ranking it in accordance to the binding affinity. The binding affinity is estimated rather than calculated, accounting for the term 'scoring' function. The different strategies of scoring [28] like *shape and chemical complementary scores*, *force field scoring*, *empirical scoring functions*, and *knowledge based scoring functions* are more or less combinations of ensemble-averaged terms and constitute a compromise between exactness and computational effort. They are not the topic of the current article, which concentrates on the docking step.

A variety of docking methods exists, allowing for flexibility of the ligand to different extent. To name a few there are conformational ensemble methods [16], genetic algorithms and evolutionary programming [13, 14, 31, 11, 27], fragment based methods with a place and join approach [8, 15, 26, 21] or incremental construction [36, 18, 35]. Usage of internal degrees of freedom as in distance geometry methods [23] allow fast searches, whereas simulation or force-field based methods [3, 25, 22, 33, 38] , and tabu searches [7] emphasize the physical modelling of the problem.

Only a few methods exist regarding protein flexibility. Principally, MD is a convenient simulation technique when implying full-flexibility, however the crucial point in this scope is runtime. Wasserman et al. developed a method for MD simulations with flexible proteins by dividing the protein into rigid and flexible regions [39, 22]. Knegtel et al. use energy-weighted and geometry-weighted averages of protein structure ensembles [17]. Claussen et al. use a set of protein structures, as implemented in the program FlexE, an expansion of FlexX. A so-called united protein representation is derived from a superposition of several experimentally determined structures. Only regions comprising structural variations are handled as alternatives, similar backbone regions or side-chains are combined in one region [8].

With respect to computational costs, full Rigid Docking is invincible. The target can be represented via a pre-computed grid, while the ligand is regarded as a rigid body, i.e. the ligand has only 6 degrees of freedom. Nevertheless, sampling all orientations of the ligand versus the target with exhaustive search algorithms demands considerable computational effort. Thus, geometrical or point complementary methods are employed after performing a discretization. The basic idea is to match points, characterized by certain features which represent molecular properties of ligand and protein, in 3D space. Those points are matched if a ligand feature is superimposed with a complementary protein feature. Complementarity can be achieved sterically, i.e. via surface complementarity, or physico-chemically, e.g. a hydrogen bond donor and its corresponding acceptor. A transformation is desired, which maximizes the number of simultaneous matches. As the number of protein features usually exceeds the number of ligand features by far, the problem sometimes is referred to as "needle in the haystack". Three major strategies are being applied.

Clique searches use distance compatibility graphs. With respect to docking, the nodes mirror all matches between protein features and ligand features, edges are links between two compatible (mostly distance compatible) nodes. Thus, a clique is a set of distance compatible (usually within a tolerance ϵ) features. Maximizing the number of simultaneous matches means searching for maximum cliques. This is a common problem in graph theory and it is NP-complete. Nevertheless there are fast algorithms dealing with this problem in practice.

The pioneer software in Molecular Docking called DOCK developed by Kuntz et al. [20, 19, 24] generates a set of spheres inside the binding pocket of the protein using molecular surfaces. The greater the diameter of the spheres, the greater is the volume possibly occupied by the ligand. The ligand is similarly modelled by a set of spheres, either in the same fashion, by spheres inside of it, or by taking atom positions as the centers of the spheres. Maximum clique detection yields an initial orientation of the ligand inside the binding pocket. Being a standard computer science method, hashing facilitates fast access to data. Characteristic features of a data object are transformed to a hash key (usually a short fixed-length value) via a hashing function. Those hash keys can be easily compared to a template via the hashing function, since only the hash values have to be checked. Particularly *Geometric hashing* can be easily used to match features in $\mathbb{R}^{3n}$. It has been developed in the context of computer vision and is capable of handling partial matches. This is important, since not necessarily all points of the ligand have to or merely can be matched with protein points in Molecular Docking. Some functional group or organic residue of the ligand may not take part in the binding process.

Fischer et al. [9] were the first to address geometric hashing to the molecular docking problem, using the sphere representation discussed above. *Pose clustering* is a pattern recognition technique. Two point sets of features can be matched in the following way. Both point sets are divided into all possible triplets, each of the triplets of both sets is matched to each of the triplets of the other set. A transformation can be calculated for every pair of triangles accordingly. Cluster analysis of the sets of transformations yields orientations with high numbers of matching features.

The Docking procedure is divided into three subsequent optimization phases, two of which utilize rigid body dynamics. The last one is applied with the ligand already positioned inside the binding pocket and accounts for full flexibility.

In order to use conformational ensembles for Virtual Screening, a Rigid Body Docking modul has to be developed first. This might seem as an obsolete approach, but Rigid Docking is fast, and it can be applied to small or rigid molecules or molecular fragments, or conformational ensembles [28]. The overall strategy can be expanded to semi-flexible Docking, which involves rigid targets but flexible ligands. Full-flexible docking on the basis of a conformation ensemble for proteins can be used either by employing the conformation analysis to macromolecules or by using experimental data.
We will introduce the docking algorithm without discussing the open question which representatives should be drawn out of an ensemble of conformations. The bioactive conformation is not necessarily identical with the conformation featuring the lowest potential energy in the gas phase (induced fit). Fragment Based Docking can circumvent this problem, but neglects cooperative effects in the course of binding.

Actually, a self-evident Redocking experiment [8] is presented to demonstrate the ability of the method to find the active site of proteins and orientate and reproduce known 3D-structures of protein-ligand complexes, obtained from NMR Spectroscopy or X-ray crystallography.

Theory

Free Energy of Binding

In Molecular Mechanics the spatial position of a molecule with n atoms is represented and uniquely determined by a set P of $3n$ coordinates $q \in \Omega \subset \mathbb{R}^{3n}$, and will further be referred to as positioning. We are looking for a specific positioning $P_L^* \{q_1, ..., q_3 n\} \in \Omega_L$ of the ligand, relatively to a positioning $P_P \{q_1, ..., q_3 m\} \in \Omega_P$ of the protein, so that the binding affinity $A(P_L) : \Omega_L \to \mathbb{R}$ between both sets reaches its maximum. Therefore molecular docking can be described as an optimization problem:

$$A(P_L^*) = max!$$

The existence of a solution P_L^* can be assumed for non-convex steady functions A, if the set Ω_L is compact. Uniqueness is not a necessary condition since there could be several binding sites having identical affinity values. Since the change in free energy $\Delta_R G < 0$ holds for all spontaneous chemical reactions, the optimization problem can be written as:

$$\Delta_R G(P_L^*) = min!$$

The positioning P_L^* is also called the correct binding mode. For conformational dynamics the canonical NVT-ensemble is generally employed. The corresponding thermodynamic potential to an NVT-ensemble is the free energy $F(T, V, N)$, rather than the Gibbs free energy $G(T, p, N)$.

$$F(T, V, N) = U - TS$$

$$dF = dU - TdS - SdT$$

For an NVT-ensemble the number of particles N, the volume V, and the temperature T are constant. Thus, the change in free energy can be described as follows:

$$dF = dU - TdS$$

At first glance this approximation is crude, as it means substituting the Gibbs free energy $G(T, p, N)$ for the free energy $F(T, V, N)$. However, at constant atmospheric pressure the change in volume of the system target-ligand-water during the binding reaction will be considerably small. Of course binding is not only enthalpy-driven, but often also entropy-driven. Thermodynamically two different types of reactions can be observed. Mostly molecular recognition is enthalpy-driven. Albeit, especially big lipophilic ligands sometimes bind to their targets despite a positive enthalpic contribution because of a higher absolute change in entropy [6]. Since calculation of entropy changes lies far beyond a screening experiment it will be neglected within all explicit energy calculations. With this most drastic approximation we yield:

$$dF \approx dU$$

Molecular Mechanics provides us with the required models to calculate the interaction energy between target and ligand, i.e. the change in intrinsic energy. If we further assume, that the average kinetic energy per degree of freedom is invariant, then:

$$dU \approx dV$$

where dV is the change in potential energy. Thus, our optimization problem in its final approximation is:

$$\Delta V(P_L^*) = min!$$

Note that not only solvent entropic effects, but also hydrational contributions to the enthalpy H, arising from the presence of solvent molecules, are ignored.

Gradient methods in rigid body dynamics

Minimization techniques using gradient methods include steepest descent, conjugate gradient, Newton, and Quasi-Newton methods [34]. All these strategies are renowned to minimize a given function from a starting coordinate set, while the internal distances between points in $\mathbb{R}^3$ may vary. Minimizing a function under the rigid body constraint implies keeping internal distances fixed. Moving a point mass in a conservative force field along the negative gradient will doubtlessly minimize the field function. The same holds for a rigid body's center of mass. A rigid body does not only have translational degrees of freedom, but also rotational ones. If a rigid body experiences a force $\vec{F_r}$ instantiated by a gradient of potential energy, it also experiences a torque $\vec{T}$, unless the force is applied to the center of mass $\vec{\gamma}$ only. The single forces $\vec{f_i}$ and torques $\vec{t_i}$, corresponding to the point masses[1] belonging to the rigid body, are:

$$\vec{f_i} = -gradV(\vec{x_i}) \tag{1}$$

$$\vec{t_i} = (\vec{x_i} - \vec{\gamma}) \times \vec{f_i} \tag{2}$$

The resultant force $\vec{F_r}$ and torque $\vec{T}$ acting on the rigid body can be obtained via superposition.

$$\vec{F_r} = \sum \vec{f_i}$$

$$\vec{T} = \sum \vec{t_i}$$

The question arises, if a rotation according to this torque will minimize the field function in the same way as a translation along the downhill gradient does? Mathematical physics [5] gives the following definitions and theorems from which the laws of rigid body motion can be derived:

Definition 1 *A rigid body is a system of point masses, constrained by holonomic relations expressed by the fact that the distance between all the points is constant:*

$$|\vec{x_i} - \vec{x_j}| = r_{ij} = const. \tag{3}$$

and

Theorem 2 *The configuration manifold of a rigid body is a six-dimensional manifold, namely,* $\mathbb{R}^3 \times SO(3) = SE(3)$*(the direct product of a three-dimensional space* $\mathbb{R}^3$ *and the group SO(3) of its rotations).*

In other words, a rigid body has 6 degrees of freedom 3 translational ones and 3 rotational ones. The rigid body motion can be split into a translation applied to the center of mass and a rotation, according to the following theorems [5]:

Theorem 3 *Under the free motion of a rigid body, its center of mass moves uniformly and linearly.*

[1] In the following all summation symbols go over all point masses i

Theorem 4 *A free rigid body rotates about its center of mass as if the center of mass was fixed at a stationary point 0.*

We consider an inertial coordinate system K (reference frame) in $\mathbb{R}^{3N}$ and a moving coordinate system k fixed to a rigid body (body-fixed frame) also in $\mathbb{R}^{3N}$. The coordinates of the mass points i in k are given by $\vec{r_i} = (\vec{x_i} - \vec{\gamma})$. According to theorem 4 only rotations are allowed in coordinate system k. Those rotations are elements of $SO(3)$. As $SO(3)$ is isomorphic to $\mathbb{R}^3$ we can define a rotation vector $\vec{\varphi}$:

$$\vec{\varphi} \in \mathbb{R}^3 \simeq SO(3)$$

An infinitesimally small rotation is an element of $SO(3)$, as well.

$$\vec{d\varphi} \in SO(3)$$

We define the projection of an infinitesimally small movement in $SO(3)$ on $\mathbb{R}^3$ by:

$$\vec{dr_i} = \vec{d\varphi} \times \vec{r_i} \tag{4}$$

In fact, $\vec{d\varphi}$ represents a rotation around the axis $\vec{d\varphi}$ by the angle $||\vec{d\varphi}||$. If we consider a system comprised of a rigid body in a conservative field, an infinitesimally small change in potential energy V can be described as the sum of all forces $\vec{f_i}$ acting on all the single point masses of the rigid body multiplied by the distance $\vec{dr_i}$ by which each point mass is moved in $\mathbb{R}^3$.

$$dV = -\sum \langle \vec{f_i}, \vec{dr_i} \rangle$$

where $\langle\rangle$ denotes the scalar product. Disregarding any translation we can apply the projection as defined in 4 and we yield.

$$dV = -\sum \langle \vec{f_i}, \vec{d\varphi} \times \vec{r_i} \rangle$$

Since $\langle \vec{f_i}, \vec{d\varphi} \times \vec{r_i} \rangle$ is a spat product, we can make use of its cyclic shifting properties.

$$dV = -\sum \langle \vec{d\varphi}, \vec{r_i} \times \vec{f_i} \rangle$$

Application of the constraints of rigid body motion (see 3), yields the fact that the angle $\vec{d\varphi}$ is identical for all the point masses. Thus, it can be pre-drawn.

$$\begin{aligned} dV &= -\langle \vec{d\varphi}, \sum \vec{r_i} \times \vec{f_i} \rangle && (5) \\ &= -\langle \vec{d\varphi}, \vec{T} \rangle && (6) \\ & && (7) \end{aligned}$$

$\sum \vec{r_i} \times \vec{f_i}$ is identical with the definition of the torque $\vec{T}$. Since the vectors $\vec{d\varphi}$ and $\vec{T}$ are composed as follows,

$$\vec{d\varphi} = \begin{pmatrix} d\varphi_1 \\ d\varphi_2 \\ d\varphi_3 \end{pmatrix}, \; and \; \vec{T} = \begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix}$$

$$dV = -T_1 d\varphi_1 - T_2 d\varphi_2 - T_3 d\varphi_3$$

A comparison of coefficients yields:

$$\vec{T} = \begin{pmatrix} -\frac{\partial V}{\partial \varphi_1} \\ -\frac{\partial V}{\partial \varphi_2} \\ -\frac{\partial V}{\partial \varphi_3} \end{pmatrix}$$

Thus, in $SO(3)$ the following relation holds:

$$\vec{T} = -gradV(\vec{\varphi})$$

Whereas, in $\mathbb{R}^{3N}$,

$$\vec{F} = -gradV(\vec{r})$$

A gradient method, which minimizes the energy of a rigid body is set up by a translation of the center of mass and a rotation around the rotation axis defined by the torque.

Modelling

Extended Pharmacophore Model (EPM)

As mentioned in the introduction the rigid docking problem can be modeled as a Point-Matching problem in $\mathbb{R}^3$ between two complementary subsets of points belonging to the ligand and target. Therefore, pharmacologically predominant interactions, accounting for specificity and binding mode predefinition in molecular recognition have to be explored. These are for example hydrogen bonds or Lennard-Jones interactions of aromatic functional groups. Now we can classify these interactions, and all the atoms of the ligand respectively:

1. H-Bond Donor
2. Non-aromatic H-Bond Acceptor
3. Atom belonging to an aromatic group
4. Aromatic H-Bond Acceptor
5. Atom belonging to none of the above groups

We will call subsets 1-4 pharma groups and a representation of a molecule based on these pharma groups an extended pharmacophore model (EPM). The classification into pharma groups is accomplished according to the typing rules in MMFF94 [12].

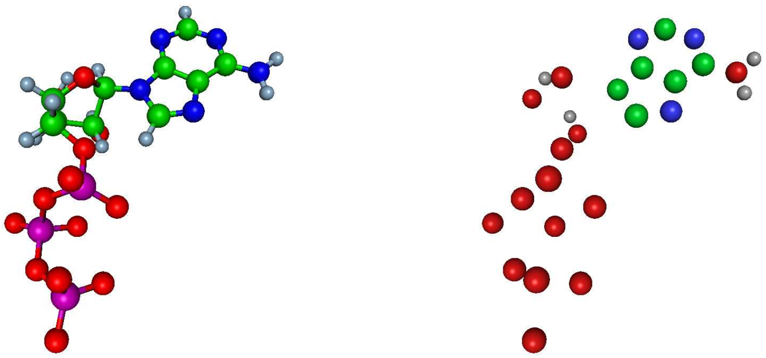

Figure 1: Left: ATP in ball and stick representation, bond types are not depicted. Carbons are displayed in green, hydrogens in grey, oxygens in red, nitrogens in blue, and phosphorus atoms in purple Right: Extended Pharmacophore Model (EPM) of ATP. The heterocyclic rings show aromatic behavior (green) and some aromatic atoms are additionally acceptors (blue). Hydrogens are regarded as H-bond donors (grey) in MMFF94, donors and acceptors can possess both acceptor properties (red).

Pharma Site

Neglecting dynamic information results in a very powerful computational simplification with respect to rigid body docking. By this constraint, the order of the costly non-bonded interactions between the macromolecule with N and a ligand with n atoms can be reduced to $O(Nn)$. Now, the interactions between a macromolecule and a small molecule can be pre-computed, by shifting a probe atom with distinct properties in $\mathbb{R}^3$ in the expected range of the considered interaction. The appropriate function, e.g. Lennard-Jones energy, is evaluated in distinct intervals for the three directions in space. Thus, a 3D lattice is acquired containing the appropriate normalized interaction energy at a given resolution, usually between 0.25 and 1 Å. During runtime, the potential energy function for interatomic interaction can then be evaluated very fast with $O(n)$ by multiplying physical attributes, e.g. charges, of the small molecule by the function value at the current location. This value can be obtained by interpolation on the grid values.

The grid representation can be used, for creating point groups of the target molecule, which fulfill criteria of complementarity against the EPM. Attractive points or hot spots in 3D space can be selected by extracting a sufficiently large number n of points per grid, having the lowest potential energy values. Performing this hot spot analysis for each grid which contains the potential field of a given pharma group, we yield a point set consisting of four subsets of hot spots, which we call pharma groups again. The classification into pharma groups are according to the respective potential value. The whole set will be referred to as a pharma site.

We can expect, that for most cases of good binding the pharma site and the pharmacophore share a common but unknown subset. Figure 2 shows HIV-1-protease together with its pharma site. Note, that the active site is the spatial location featuring the highest 'hot spot'-density. This involves also the highest number of possible simultaneous interactions between EPM and pharma site and thus the highest number of possible positionings, that may lead to a binding position.

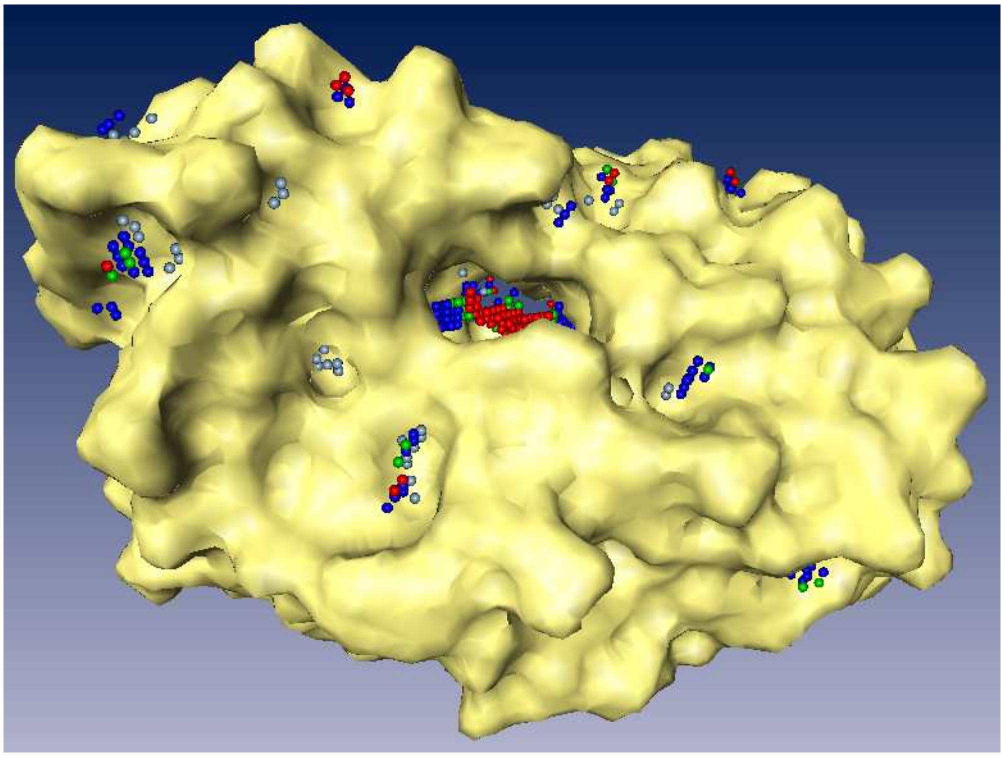

Figure 2: HIV-1-protease in molecular surface view together with its pharma site. Donors in grey, acceptors in red, aromatics in green and aromatic acceptors in blue.

Docking Algorithm

The designated docking algorithm is divided into three subsequent optimization phases, two of which utilize rigid body minimization. The last one is applied with the ligand already positioned inside the binding pocket and accounts for full flexibility.

1. **Point-Matching**
2. **Fine-Tuning**
3. **Relaxation**

Point-Matching

An attractive force between pharma site and the extended pharmacophore model (EPM) is introduced, which allows the solution of the Point-Matching by energy minimization of the rigid body coordinates defined by the EPM. Given two different point sets x_n, y_m with $n \in \{1, ..., N\}$ denotes the point set of the ligand and $m \in \{1, ..., M\}$ denotes the point set of the target. A so called pharma force field is defined by the force $\vec{F}_{nm}$ between the sets:

$$\vec{F}_{nm} = w_m \delta_{p_n p_m} \frac{\vec{x_n} - \vec{y_m}}{\|\vec{x_n} - \vec{y_m}\|^2} \tag{8}$$

with the corresponding potential

$$E_{nm} = -w_m \delta_{p_n p_m} ln(\|\vec{x_n} - \vec{y_m}\|) \tag{9}$$

and the torque

$$\vec{T}_{nm} = (\vec{x_n} - \vec{\gamma}_E) \times \vec{F}_{nm} \tag{10}$$

w_m is the weight, which results from the hot spot analysis, i.e. the Lennard Jones potential at the corresponding grid point belonging to a pharma group. $\delta_{p_n p_m}$ is 1 if points belong to the same pharma group and 0 else. $\vec{\gamma}_E$ is the geometrical center of gravity of the EPM. The single forces and torques are summed up yielding a resultant force $\vec{F}$ and a resultant torque $\vec{T}$, which will be used in the energy minimization algorithm.

For each minimization step the EPM is translated and rotated as a rigid body. The translation direction is governed by the direction of the model force, acting on the geometrical center of gravity $\vec{\gamma}_E$ of the EPM. The rotation axis is determined by the torque. The **Basic Rigid Body Docking Scheme** for Point-Matching so far reads as follows:

1. Choose a starting position for the rigid body.
2. Compute a resultant force $\vec{F}$ and torque $\vec{T}$ acting on the rigid body as obtained from the employed force field

3. Calculate a translational step-size $x(\|\vec{F}\|)$ and a rotational step-size $\alpha(\|\vec{T}\|)$ and apply a minimization step to the rigid body by translating and rotating it.

4. Continue steps 2-3 until the convergence criterion, i.e. appropriate minima of the force and torque are reached.

5. Rank the acquired solutions according to the chosen scoring function.

This basic scheme can be used for any minimization of a rigid body within a given force field. Especially for the Point-Matching problem following peculiarities and modifications have been considered:

ad 1: Multi-start initialization For Point-Matching the whole minimization scheme is repeated with a distinct number $500 < s < 2000$ of starting positions, which are chosen randomly on a spherical surface around the pharma site's center of gravity. The diameter of the sphere depends on the extension of target and ligand.

ad 2: surface term and cut off The binding cleft is often very steep and narrow. Moreover it is commonly observed that proteins form small cavities, which are in vivo not accessible for ligands. Nevertheless some 'hot spots' can be located beneath the protein surface. The pharma force field modelled so far allows a penetration of target and ligand. An additional constraint must be implied, which guarantees that the global minimum is located at the protein surface. Therefore, another grid is constructed, which contains for every grid point a neighbor list of target atoms. A 3D vector field is derived from the scalar field of neighbors, which is directed towards the protein surface and converges to zero outside or deep inside of the target. Adding this gradient vector to the single forces in every step, the overall solution will be shifted towards the protein surface.

With increasing number of iteractions a decreasing cut-off for the model force is applied, which guarantees that initially the extended pharmacophore model is attracted by all points, according to complementarity. Subsequently, only points in the spatial vicinity of the extended pharmacophore model are regarded to allow local exploration of the pharma site.

ad 3: translation and rotation The translational stepsize x is calculated with a factor k_x proportional to $\|\vec{F}\|$ and for the Point-Matching also to the distance d between the geometrical centers of gravity of the two point sets. If the scalar product of the current iteration's force and the previous iteration's force is negative/positive and the factor k_x exceeds/falls below a distinct value, k_x will be halved/doubled In close vicinity of the minimum,

$d < 2\mathring{A}$, k_x is set to 1, to allow full influence of local environment.
The rotational stepsize α is calculated with a factor k_α proportional to $\|\vec{T}\|$ and for the Point-Matching it also depends on d. If the scalar product of the current iteration's torque and the previous iteration's torque is negative/positive and the factor k_α exceeds/falls below a distinct value, k_α will be also halved/doubled.

ad 5: scoring As can be obtained from Fig. 3 the scoring energy, which results from the pharma force field does not perform reliably enough. Only a weak correlation exists between this scoring energy and the RMSD from the crystal coordinates for Point-Matching solutions of CDK2. The 1000 time-steps corresponding to the lowest RMSDs have nearly identical energy values. For the estrogen receptor the situation is even worse. No correlation between RMSD and scoring energy can be observed (Fig. 3). Therefore, the MMFF94 energy term will be used as a scoring energy in both docking steps, the Point-Matching and the Fine-Tuning.

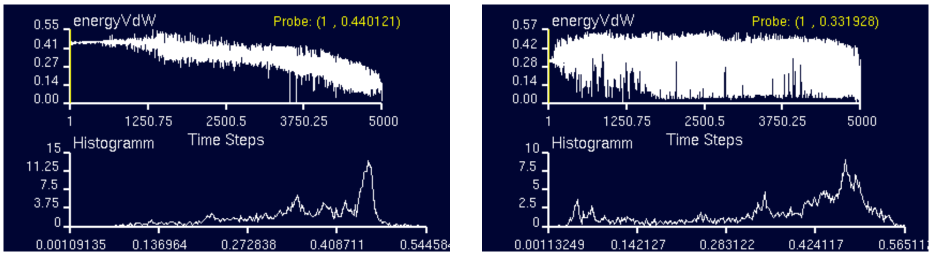

Figure 3: Left: Trajectory containing 5000 Point-Matching solutions for CDK 2 sorted by RMSD from crystal coordinates starting from the smallest value. In the upper graph the individual time-steps are plotted against the scoring energy. Right: Same for the estrogen receptor

Fine-Tuning

The proposed minimization scheme can be used for Point-Matching the EPM onto the pharma site as well as for Fine-Tuning the rigid body orientation of the ligand within the active site of the target. The first uses the pharma force field, whereas the second needs a full atom representation of the molecules e.g. based on the MMFF94 force field [12]. The following modifications of the basic scheme are proven to be reasonable for the Fine-Tuning step:

ad 1: Multi-start initialization The Fine-Tuning starts with the best solutions of the Point-Matching.

ad 2: surface term and cut off The surface term is almost needless for the Fine-Tuning, because steric clashes are implicit considered within the Lennard-Jones interactions. Nonetheless, if the interatomic distance of an atom pair drops below 70% of their van der Waals contact distance, the surface term is switched on and the MMFF contribution is switched off, to prevent effectively too much penetration of the molecules.

ad 3: translation and rotation The modification of translational and rotational stepsize is handled in the same way as in the Point-Matching step, besides the fact, that the distance d between the centers of masses has no influence.

Relaxation

A vital reason for a fully-flexible energy minimization is the fact that coordinates for hydrogens cannot be resolved by X-ray crystallography. When obtaining target structures from the PDB [1], for instance, all hydrogen coordinates have to be computed by geometrical optimization. For this task a Quasi-Newton method proposed by Broyden, Fletcher, Goldfarb, and Shanno (BFGS), as can be found in [34], was employed. Quasi-Newton methods are powerful minimization strategies, since they are quadratically convergent and more exact than steepest descent or conjugate gradient methods[34].
Alternatively, a short Relaxation of the docked complex conformation is performed by Molecular Dynamics [10]. For the target-ligand complexes 500 trajectories comprising 10 time-steps and a step-size of 1 fs are computed. The starting values for momenta were chosen from a velocity distribution according to Maxwell-Boltzmann at a low temperature of 30 Kelvin. In order to conserve the complex coordinates as much as possible and also to reduce computational costs, only those atoms located in the range of 10 Å from the ligand were regarded as flexible. The remaining atoms were considered as a rigid body. The dynamic simulation is performed using a cut off criterion of 10 Å. The complex conformation featuring the lowest *intra*molecular potential energy (according to MMFF94) is selected. For this particular state the *inter*molecular potential energy between ligand and protein is monitored.

Results

The generation of pharma sites for the Point-Matching is based on the computation of grids for every atom type in MMFF94 which belongs to a particular pharma group. MMFF94 includes more than 100 atom types. Generally more than one atom type will belong to every pharma group. A virtual probe atom is shifted through 3D space and the Lennard Jones potential energy is assigned to the appropriate lattice point. The single grids for every pharma group are summed up and a mean grid is calculated for every pharma group, in order to obtain a simplified but specific average potential energy for each pharma group.

Several Redocking experiments were conducted. A first self-evident test of the docking scheme is the reproduction of experimentally detected binding modes. A second validation method is docking a natural ligand of an arbitrary protein into its uncomplexed conformation. All the subsequently studied protein structures were obtained from the PDB [1], as a selection of pharmaceutically promising targets. All hetero-atoms, i.e. water molecules, ions, and complexed ligands, were erased from the files. Ligands were stored in extra files. Bond types were assigned, where required. Subsequently hydrogen coordinates for protein and ligand were determined via geometrical optimization. In a preprocessing phase molecules were parameterized according to MMFF94. Grids were pre-computed using zibMol®, a software developed at ZIB. Grid resolution ranged from 0.5 to 1.0 Å, depending on the target size.
Each Docking was repeated 5 times for each of the test targets, in order to proof the reproductional properties of the algorithm. In all cases, identical solutions were obtained. All results presented here neglect an electrostatic contribution for the Point-Matching step (see Discussion). More details of the used parameter set for all docking steps are listed in the Appendix.

The following targets were docked with their native ligands, as found in the PDB. The experimentally determined binding mode was predicted with an RMSD of less than 1 Å.

Target	PDB-entry
HIV-Reverse Transcriptase[2]	1RT1
β-Trypsin	1BJU
Elastase[2]	1ELA
Arabinose Binding Protein	1ABE
pp60(c-src) SH2 domain[3]	1A07

In the next paragraphs results are presented for five target molecules, which have been studied in more detail.

Cyclin dependent kinase 2

Cyclin dependent kinases (cdk) control the cell division cycle, making them interesting targets in cancer therapy. They consist of a catalytic subunit (cdk1-cdk8) and a regulatory subunit (cyclin A-cyclin H). They are

[2]For this target the active site has been manually determined, since the Point-Matching did not work in this particular case.

[3]Membrane-associated tyrosine-specific kinase encoded by the c-src genes

regulated in different ways, by subunit production or complex formation to name a few. Most importantly they are controlled by reversible phosphorylation of hydroxyl groups, which is a common property of all protein kinases. This broadens the spectrum of application in drug design to diseases like arthritis and diabetes. The majority of known protein kinase inhibitors are competitors of ATP. Development of purin-like inhibitors is an obvious aim [37, 32].

Cyclin dependent kinase 2		
PDB Source:	1HCK	
Ligand:	ATP	
Point-Matching		
RMSD(Å):	2.57	
Non-covalent energy (kJ/mol):	-3762.64	
Fine-Tuning		
RMSD(Å):	1.95	
Non-covalent energy (kJ/mol):	-5318.39	
Lennard-Jones energy (kJ/mol):	281.45	
Coulomb energy (kJ/mol):	-5599.84	
Relaxation	BFGS	MD
Non-covalent energy (kJ/mol):	-5721.34	-7030.98

HIV-1-Protease

HIV-1-protease is the main protease of human immunodeficiency virus, and thus also represents its 'Achilles Heel'. The RNA of the virus is polycistronic, several genes are encoded in one RNA molecule, accounting for a high processing speed. After translation a precursor polypeptide chain has to be cleaved into several polypeptides, allowing for correct folding and thus activity of the different proteins. If the main protease is prevented from dissecting, the precursor polypeptide remains in its unmature state and is digested by the proteasome.

HIV protease is unique in the way that, in contrast to human enzymes it can cleave between phenylalanine and tyrosine or proline.

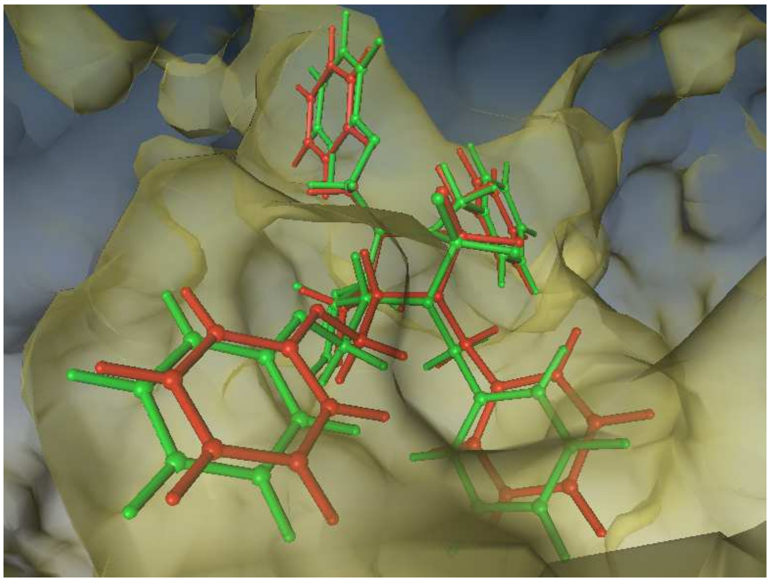

Figure 4: HIV-1-protease depicted in molecular surface representation, HIV-1-protease Inhibitors in ball and stick representation. Experimentally resolved binding mode mapped in green, binding mode according to Point-Matching in red (RMSD:0.77Å).

HIV-1-protease		
PDB Source:	1AJV	
Ligand:	Aha006[4]	
Point-Matching		
RMSD(Å):	0.77	
Non-covalent energy (kJ/mol):	159.90	
Fine-Tuning		
RMSD(Å):	0.22	
Non-covalent energy (kJ/mol):	-283.15	
Lennard-Jones energy (kJ/mol):	-110.83	
Coulomb energy (kJ/mol):	-172.32	
Relaxation	BFGS	MD
Non-covalent energy (kJ/mol):	-292.57	-442.03

α-Thrombin

Thrombin is a serine protease and plays a major role in the initiation of blood coagulation. Inhibitors of thrombin may act as powerful anticoagulant therapeutics.

[4]Cyclic Sulfamide Inhibitor

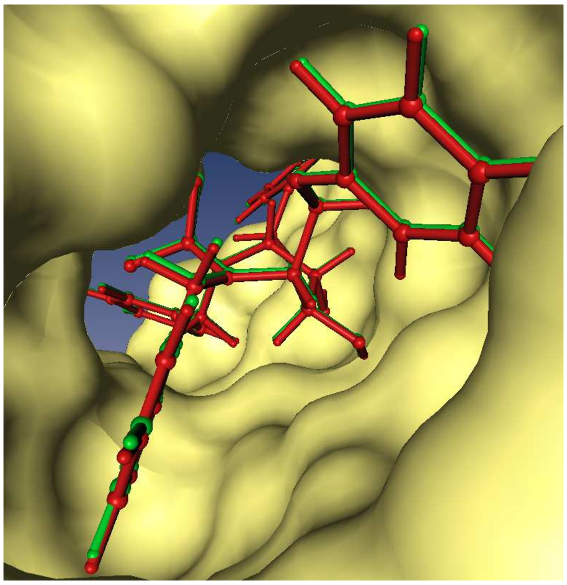

Figure 5: HIV1-protease depicted in molecular surface representation, inhibitors in ball and stick representation. Experimentally resolved binding mode mapped in green, docked binding mode after Fine-Tuning in red (RMSD:0.22Å).

α-Thrombin		
PDB Source:	1DWD	
Ligand:	NAPAP	
Point-Matching		
RMSD(Å):	1.45	
Non-covalent energy (kJ/mol):	410.50	
Fine-Tuning		
RMSD(Å):	0.46	
Non-covalent energy (kJ/mol):	-94.11	
Lennard-Jones energy (kJ/mol):	-92.66	
Coulomb energy (kJ/mol):	-1.46	
Relaxation	BFGS	MD
Non-covalent energy (kJ/mol):	-192.20	-290.71

As a test on robustness the bioactive conformation of NAPAP, obtained from PDB-entry 1DWD, was docked against the unligated conformation as obtained from 1MKX. Although thrombin as complexed in 1DWD is from a human source and thrombin as complexed in 1MKX is from bovine, thrombin is highly conserved and thus NAPAP exhibits binding affinity towards both thrombin molecules as can be seen in PDB-entry 1ETR. The two conformations of thrombin differ only slightly in the vicinity of the binding site. Nearly the same binding mode as observed in 1DWD was predicted. Due to the slight difference of the two conformational states, a hydrogen bond between NAPAP and Gly 216, which can be observed in 1DWD, cannot be formed. As already mentioned in the introduction, small modifications in protein structures can lead to large deviations in complex formation. Hence, only this single example of docking a ligand into a 'foreign' structure is given here.

Ligand binding domain of Human Estrogen-Receptor

The estrogen receptor serves as nuclear transcription factor. Binding of an agonist to the estrogen receptor can trigger the activation or repression of genes. Estrogens are steroids and can induce growth, development and homeostasis of several tissues. The estrogen receptor is an oligomeric protein, consisting of several subunits. For Docking only one of the two ligand binding polypeptide chains (as contained in the PDB-entry 1ERR) was selected as target, since an algorithm capable of distinguishing between several binding sites has not been implemented, yet.

Estrogen-Receptor		
PDB Source:	1ERR	
Ligand:	Raloxifene (Antagonist)	
Point-Matching		
RMSD(Å):	1.14	
Non-covalent energy (kJ/mol):	234.13	
Fine-Tuning		
RMSD(Å):	0.93	
Non-covalent energy (kJ/mol):	103.41	
Lennard-Jones energy (kJ/mol):	220.22	
Coulomb energy (kJ/mol):	-116.75	
Relaxation	BFGS	MD
Non-covalent energy (kJ/mol):	5.64	-306.93

A comparison of the location of the hydroxy-hydrogen atom on the right side of Fig. 6 with the Point-Matching result exhibits that the Fine-Tuning routine draws the Point-Matching solution from a sterically hindered (the oxygen-hydrogen bond penetrates the surface of the binding pocket) to an unhindered positioning.

Fab' Fragment of the Db3 Anti-Steroid Monoclonal Antibody

Steroids are the basic compound class of a variety of hormones.

Only two of the 53 atoms of Progesterone belong to a pharma group, namely two carbonyl oxygen atoms (see Fig. 7). Although, the pharma site outlines the binding pocket very nicely, the Point-Matching fails to

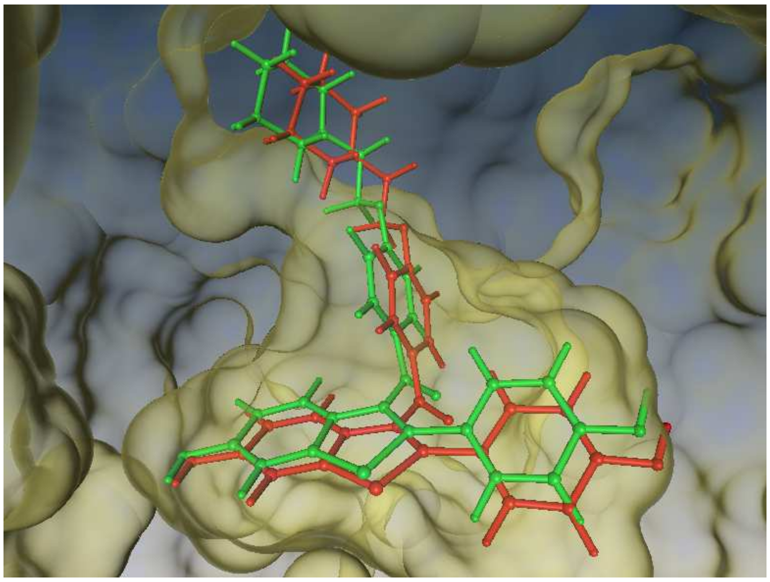

Figure 6: Estrogen receptor depicted in molecular surface representation (transparent), antagonist as contained in 1err in ball and stick representation. Experimentally resolved binding mode mapped in green, docked binding mode after Fine-Tuning in red.

produce a solution, which is close enough to the local minimum of the energy surface for the non-covalent interaction. Therefore, the Fine-Tuning and relaxation step are omitted here. Nevertheless, the binding pocket is found and progesterone is positioned inside of it. The interaction energy of -90.95 kJ/mol suggests, that progesterone is able to bind to the target.

Fab' Fragment	
PDB Source:	1DBB
Ligand:	Progesterone
Point-Matching	
RMSD(Å):	3.79
Non-covalent energy (kJ/mol):	-90.95
Lennard-Jones energy (kJ/mol):	-66.11
Coulomb energy (kJ/mol):	-24.84

Screening

The energy-minimized ligand molecules provided by the Cambridge Crystallographic Data Center (CCDC) the Cambridge Structure Database (CSD) [4] in the CCDC/Astex test set [2] were screened against the Sh2 domain of pp60 (PDB-entry:1A07).

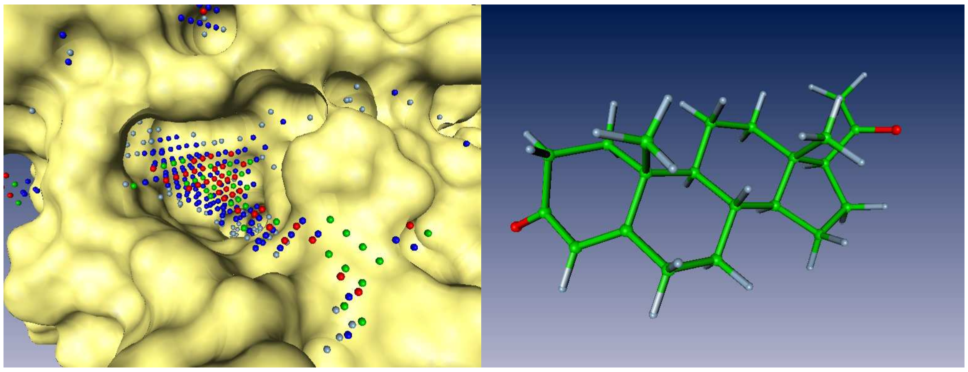

Figure 7: Left: Pharma site of the anti-steroid antibody together with molecular surface. Attentive observation reveals the grid structure of the pharma site. Right: Progesterone featuring two carbonyl oxygen atoms.

The interaction energy has been employed as a default scoring function for the screening. Surprisingly more than 60% of the approximately 300 ligands contained in the CCDC can build a complex with the target featuring a negative non-covalent interaction energy. The bioactive conformation of the 'native ligand' (Ace-Malonyl Tyr-Glu- (N, N-Dipentyl Amine)) is ranked on position 117. Meaning it is rated among the first 40% of the total library, and among the first 60% of the hits.

Small molecules, i.e. molecules having a molecular weight, which is below the weight of the native ligand found in the PDB file, can exert considerably strong interactions in the narrow side pocket of the binding cleft. Therefore, a correlation between interaction energy and molecular weight can be observed.

Discussion

Redocking

The Redocking results are encouraging, 9 out of 10 complexes could be reproduced with an $RMSD$ of less than 2 Å, only for the anti-steroid antibody the $RMSD$ is about 3.8 Å (see above). 8 complexes could even be reproduced with an $RMSD$ below 1 Å. The conclusion arises that if the solution space is properly sampled, the Fine-Tuning routine manages to predict the correct binding mode. Although a random multi-start strategy is applied, the solutions are highly reproducible. This is due to the uniform sampling of the starting sphere, which accounts for most of the computational effort in the Point-Matching phase (see computational effort). The fact that the Docking routine finds the binding site as well as the binding mode for the bioactive conformation of NAPAP, as obtained from the PDB-entry 1DWD, docked against an unligated conformation of thrombin as obtained from 1MKX is a further indication of the robustness of the designated method.

Point-Matching

Estrogen receptor agonist and ATP as ligand of CDK2 are the only ligands comprising atoms from all 4 pharma groups. All the others lack the fourth pharma group, which is made up of aromatic acceptors. In the future an investigation should be conducted, using a larger test set, in order to fathom if the fourth pharma group can be discarded or replaced by a different one, e.g. metal binding atoms. Additionally, the native ligand of the arabinose binding protein (PDB-entry:1ABE) only has donors and acceptors, which does not result in diminished reproductional exactness. Whereas, the fact that the EPM of the native ligand of the anti-steroid antibody only comprises 2 acceptors, leads to an inadequate sampling of the solution space. This can be improved by including an additional pharma group. The Point-Matching algorithm manages to get sufficiently close to the local minimum in all other cases discussed here. Although steric effects were not directly regarded during Point-Matching, they are, to a certain extent, implicitly included in the 'hot spot' distribution. Moreover, all the solutions are shifted towards the protein surface. This strategy works extremely well in both Point-Matching and Fine-Tuning. The percentage of undesired, i.e. sterically hindered, Point-Matching solutions could be diminished to a very low value.

Fine-Tuning

The Fine-Tuning routine manages to diminish the $RMSD$ by 20-60% and as expected also the non-covalent interaction energy in all observed cases, even for the one case where the Point-Matching did not work very well (FAB' fragment of the anti-steroid antibody). The reasons for choosing only a low number of iteration steps, which does not guarantee convergence for every minimization run, are the following:

Hydrogen atoms were added to the pdb-entries without any subsequent energy minimization. Thus one would certainly expect the local minimum to be shifted slightly compared to the minimum yielded after X-ray structure determination. Additionally, refinement of X-ray structures is always combined with energy minimization using force fields like AMBER or CHARMM but not MMFF94.

When Fine-Tuning a set of Point-Matching solutions, most probably there will be some solutions that will be too far from the minimum to converge to it in a reasonable time. These solutions will rather converge to a local minimum, wasting a great amount of computational time leading to no gain. Additionally, finding a distinct value for a gradient convergence criterion is not a trivial task, since the potential energy surfaces, and thus absolute values differ widely between targets. Provided that at least one of the Point-Matching solutions comes as close to the minimum to converge in a feasible number of steps, a strategy using a low maximal number of iteration steps seems to be justifiable. Anyway, subsequently an energy-minimization with a profound convergence criterion should be applied with the best candidates.

The fact that the native binding mode of progesterone could be reproduced with an $RMSD$ of approximately 3.8 Å accounts also for the robustness of the designated procedure. In this case the potential energy surface is not that steep, since there are not many highly specific non-covalent interactions (nearly all atoms of progesterone belong to simple aliphatic residues), that is why the interaction energy is fairly negative. Thus, although the $RMSD$ is above 3 Å, in a virtual screening experiment progesterone would have been selected as a potentially binding ligand.

Relaxation

In a Redocking experiment Relaxation is not necessary, whereas in Screening flexibility is essential, since both the ligand's and the target's bioactive conformation are unknown. Moreover, the interaction energy between target and ligand can serve as a first approximation towards the binding energy. Naturally, when conducting a Screening experiment one would perform an energy minimization for ligand and target independently in the preprocessing phase, to ensure that degrees of freedom that are not influenced by the binding process are energy-minimized only once. Results are given here in order to illustrate the fact that, determination of a minimum energy coordinate set for the target or the target-ligand complex is not a trivial task with a unique solution. Using two different Relaxation strategies we obtain different energies and coordinates for identical starting complexes. This is due to the fact that different strategies, local versus semi-global, as well as different convergence criteria for MD and BFGS were employed. Another crucial point stems from the fact that calculations are performed in the absence of solvent molecules. Thus, partial charges saturated in vivo could attract each other. As a consequence, minimizing over many iteration cycles until the gradient vanishes can yield a conformation that is far from any conformational state occupied in vivo. When determining the coordinate set for Scoring and adding the non-covalent energy contribution to the Scoring value, this observation becomes a highly intricate problem.

Electrostatic contribution

Electrostatic interaction is, though strong in absolute value and in many cases specific (salt bridges), not as highly specific as Lennard-Jones interaction, since it does not decay that fast in dependence of the range.

Additionally, different atom pairs possess different dispersion terms, whereas electrostatic interaction only depends on range and charge. Binding specificity arises out of the three-dimensional structure of the interacting molecules. A single Lennard-Jones interaction may be weak, but specific. If the number of interacting atoms increases, dispersion interactions will become more and more important. The larger the intersection of the surface areas of two interacting molecules, the more dispersion interactions can occur simultaneously and the stronger is the overall attractive or repulsive force. Hence molecular recognition requires steric complementarity, meaning ligand and target must fit together like key and lock. One can assume that electrostatics are often highly specific only in conjunction with hydrogen bonds. Additionally salt bridges can be saturated by water molecules. Presumably, electrostatic forces are less important for a coarse spatial orientation of the ligand inside the binding pocket. Since the relation of the two terms within the pharma force field remains unclear and a positive effect cannot be proven within some test experiments, an eletrostatic term has been neglected in the proposed pharma force field used in the Point-Matching.

Apart from that, Coulomb forces are surely essential for guiding the ligand towards the binding site, but this effect is not subject of this study. For the positioning within the binding site (Fine-Tuning). Otyepka et al., found that Coulomb energy contributes strongly to the overall interaction energy of all observed ligands, but not very specifically, when regarding binding of ligands to CDK2 [32]. Calculated electrostatic energies are usually largely overestimated. This leads to systematic errors in ranking of screening results. Scaling down electrostatic interactions by a uniform dielectric constant is not a decent approach, because in the absence of solvent, dipole interactions have to be scaled down less than interaction between ion pairs [29, 30].

Screening

Due to the overestimation of electrostatics, molecules of low molecular weight get higher scores, because the probability of sterical clashes is considerably low and the number of possible positions in the active site is high. Screening with respect to interaction energy tends to choose molecules with maximal attractive electrostatic interaction. Moreover, the non-covalent interaction energy is correlated with the number of ligand atoms.
A reduced scoring function which is based only on the intermolecular interaction energy derived from a force field, is not valid for Screening a ligand library, at most it can be used as a pre-filter. Taking into account all the approximations connected to the interaction energy between target and ligand, the fact that the native ligand is ranked on position 117 is encouraging

for a physically motivated model of a scoring function. Even for more sophisticated scoring functions the native ligand is not always ranked among the top 10%. A high false positive ratio seems quite inevitable, when scoring is applied only with regard to non-covalent interaction.

Computational effort

Despite several simplifying assumptions, the overall run-time of approximately 5 minutes per processor for docking one ligand conformation is too long. Whereas, runtime decreases considerably with the size of the ligand. A Redocking run for the arabinose binding protein (PDB-entry:1ABE) with arabinose (20 atoms) takes approximately 1 minute. When aiming at semi-flexible Docking, this is a first argument for a strategy with a single fragment placed inside the binding pocket, rather than employing a place and join algorithm.

Acknowledgement The work of A. May and S. Eisenhardt has been supported by AnalytiCon Discovery (Potsdam, Germany). The work of F. Cordes and J. Schmidt-Ehrenberg has been supported by the German Federal Ministry of education and research (grant no. 031U109A/031U209A, Berlin Center for Genome Based Bioinformatics). All visualizations in this work have been rendered using amiraMol® a visualization software package being developed at the Zuse Institut Berlin (ZIB). The authors would like to thank M. Weber, D. Baum and T. Baumeister for helpful discussions.

Appendix

Coarse Docking:	
'Hot spots' per pharma group:	150
Runs for finding the binding site:	100
Runs for Coarse Docking:	1000
Convergence tolerance ϵ:	10^{-8}
Minimum value for dynamic cut-off:	2
Fine-Tuning:	
Input steps for Fine-Tuning:	20
Steps in Fine-Tuning:	100
Initial translational step-size (Å):	0.1
Initial k_x:	0.5
Initial k_α:	0.5
Interval for k_x:	0.2-2.0
Interval for k_α:	0.2-2.0
Cut-off for non-covalent interactions (Å):	14
Cut-off for H-bonds only (Å):	4
Update number for cut off list:	4

Relaxation with BFGS [34]:

gtol:	0.1
tolX:	10^{-15}

References

[1] URL http://www.rcsb.org/pdb.

[2] Url http://www.ccdc.cam.ac.uk/products/life_sciences/validate/astex.php.

[3] R. Abagyan and M. Totrov. High-throughput docking for lead generation. *Current Opinion in Chemical Biology*, 5:375–382, 2001.

[4] Allen, F. H., Bellard, S., Brice, M.D., Cartwright, B. A., Doubleday, A., Higgs, and H. et al. The cambridge crystallographic data centre: computerbased search, retrieval, analysis, and display of information. *Acta Crystallog. sect. B*, 35:2331–2339, 1979.

[5] V. I. Arnold. *Mathematical Methods of Classical Mechanics.* Springer Verlag, New York and Berlin and Heidelberg, 2nd edition, 1989.

[6] H.-J. Böhm, G. Klebe, and H. Kubinyi. *Wirkstoffdesign.* Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford, 1996.

[7] Baxter C.A., Murray C.W., Clark D.E., Westhead D.R., and Eldridge M.D. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins*, 33(3)::367–382, 1998.

[8] H. Claußen. *Effizientes Protein-Ligand-Docking mit flexiblen Proteinstrukturen.* PhD thesis, GMD, 2001.

[9] D. Fischer, R. Norel, H. Wolfson, and R. Nussinov. Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition. *Proteins*, 16(3):278–292, 1993.

[10] D. Frenkel and B. Smit. *Understanding Molecular Simulation. From Algorithms to Applications.* Academic Press, San Diego, 2nd edition, 2002.

[11] D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, D. B. Fogel, L. J. Fogel, and S. T. Freer. Docking conformationally flexible small molecules into a protein binding site through evolutionary programming. In J. R. McDonell, R. G. Reynolds, and D. B. Fogel, editors, *Proceedings of the 4th Annual Conference on Evolutionary Programming.* MIT Press, 1995.

[12] T. A. Halgren. Merck molecular force field. i. basis, form, scope, parametrization, and performance of mmff94. *J. Comp. Chem.*, 17:490–519, 1996.

[13] G. Jones, P. Willett, and R. C. Glen. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, 245:43–53, 1995.

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

[15] A. Kaapro and J. Ojanen. Protein docking. *URL http://www.lce.hut.fi/teaching/S-114.500/k2002/Protdock.pdf.*, Nov. 2002.

[16] S. Kearsley, D. Underwood, R. Sheridan, and M. Miller. Flexibases: A way to enhance the use of molecular docking methods. *J. of Computer-Aided Molecular Design*, 8:565–582, 1994.

[17] R. M. Knegtel, I. D. Kuntz, and C. M. Oshiro. Molecular docking to ensembles of protein structures. *J. Mol. Biol.*, 266:424–440, 1997.

[18] B. Kramer, M. Rarey, and T. Lengauer. Evaluation of the flexx incremental construction algorithm for protein-ligand docking. *Proteins Struct Funct Genet*, 37:228–241, 1999.

[19] I. D. Kuntz. Structure-based strategies for drug design and discovery. *Science*, 257, 1992.

[20] 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.

[21] A. R. Leach and I. D. Kuntz. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comp. Chem.*, 13:pp 730, 1992.

[22] B. A. Luty, Z. R. Wassermann, 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. Comp. Chem.*, 16(4):454–464, 1995.

[23] R. Abagyan M. Totrov. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins*, 1:215–220, 1997.

[24] S. Makino and I. D. Kuntz. Automated flexible ligand docking method and its application for database search. *J. Comp. Chem.*, 18(14):1812–1825, 1997.

[25] M. Mangoni, D. Roccatano, and A. Di Nola. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins*, 35:153–162, 1999.

[26] J.B. Moon and W. J. Howe. Computer design of bioactive molecules: A method for receptor-based de novo ligand design. *PROTEINS: Struct., Func., and Gen.*, 11:pp 314, 1991.

[27] G. M. Morris, D. S. Goodsell, R.S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.*, 19(14):1639–1662, 1998.

[28] I. Muegge and M. Rarey. Small molecule docking and scoring. In D. B. Boyd K. B. Lipkowitz, editor, *Reviews in Computational Chemistry*, volume 17. Wiley-VCH, New York, 2001.

[29] I. Muegge, T. Schweins, R. Langen, and A. Warshel. Electrostatic control of gtp and gdp binding in the oncoprotein p21 (ras). *Structure*, 4:475, 1996.

[30] I. Muegge, H. Tao, and A. Warshel. A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding. *Protein Eng.*, 10:1363, 1997.

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

[32] M. Otyepka, V. Kryštof, L. Havlíček V. Siglerová, M. Strnad, and J. Koča. Docking-based development of purine-like inhibitors of cyclin-dependent kinase-2. *J. Med. Chem.*, 43:2506–2513, 2000.

[33] C. Pérez and A.R. Ortiz. Evaluation of docking functions for protein-ligand docking. *J. Med. Chem.*, 44:3768–3785, 2001.

[34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C - The art of scientific computing.* Cambridge University Press, USA, 2nd edition edition, 1999.

[35] M. Rarey, B. Kramer, and T. Lengauer. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics*, 15(3):243–250, 1999.

[36] 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.

[37] L. Shewchuck, A. Hassell, B. Wisely, W. Rocque, W. Holmes, J. Veal, and L. F. Kuyper. Binding mode of the 4-anilinoquinazoline class of protein kinase inhibitors: X-ray crystallographic studies of 4-anilinoquinazolines bound to cyclin-dependent kinase 2 and p38 kinase. *J. Med. Chem.*, 43:133–138, 2000.

[38] J. Wang, P. A. Kollman, and I. D. Kuntz. Flexible ligand docking: A multistep strategy approach. *PROTEINS: Struct., Func., and Gen.*, 36:pp 1, 1999.

[39] Z. R. Wassermann and C. N. Hodge. Fitting an inhibitor into the active site of thermolysin: A molecular dynamics case study. *Proteins: Struct., Funct., Genet.*, 227(24), 1996.