

A New Method for the Gradient-Based Optimization of Molecular Complexes

JAN FUHRMANN, ALEXANDER RURAINSKI, HANS-PETER LENHOF, DIRK NEUMANN

Center for Bioinformatics, Saarland University, 66123 Saarbrücken, Germany

Received 22 August 2008; Revised 6 October 2008; Accepted 7 October 2008

DOI 10.1002/jcc.21159

Published online 20 November 2008 in Wiley InterScience (www.interscience.wiley.com).

Abstract: We present a novel method for the local optimization of molecular complexes. This new approach is especially suited for usage in molecular docking. In molecular modeling, molecules are often described employing a compact representation to reduce the number of degrees of freedom. This compact representation is realized by fixing bond lengths and angles while permitting changes in translation, orientation, and selected dihedral angles. Gradient-based energy minimization of molecular complexes using this representation suffers from well-known singularities arising during the optimization process. We suggest an approach new in the field of structure optimization that allows to employ gradient-based optimization algorithms for such a compact representation. We propose to use exponential mapping to define the molecular orientation which facilitates calculating the orientational gradient. To avoid singularities of this parametrization, the local minimization algorithm is modified to change efficiently the orientational parameters while preserving the molecular orientation, i.e. we perform well-defined jumps on the objective function. Our approach is applicable to continuous, but not necessarily differentiable objective functions. We evaluated our new method by optimizing several ligands with an increasing number of internal degrees of freedom in the presence of large receptors. In comparison to the method of Solis and Wets in the challenging case of a non-differentiable scoring function, our proposed method leads to substantially improved results in all test cases, i.e. we obtain better scores in fewer steps for all complexes.

© 2008 Wiley Periodicals, Inc. J Comput Chem 30: 1371–1378, 2009

Key words: local optimization; molecular orientation; exponential mapping; docking; unconstrained minimization; parametrization for minimization

Introduction

Computer-aided drug design is a key methodology in the drug development process. For statistical ligand-based approaches, it is often sufficient to reduce a molecule to its chemical formula or a two-dimensional representation. From such simplified representations numerous descriptors are readily calculated. More complex descriptors (e.g., obtained from CoMFA or CoMSIA) may require the positions of each atom in three dimensions. The descriptors are then used to train and validate a predictive model for the interaction of putative new drug molecules with the target at hand by applying statistical learning methods. In contrast to such ligand-based approaches, receptor-based design assesses the interaction between ligand and receptor by searching the conformational space for energetically favorable complexes and estimating the binding free energy.

The conformational sampling necessitates the knowledge of the ligand's and receptor's atoms and some representation thereof. The obvious way to represent the spatial arrangement of a ligand molecule while sampling the conformational space is by defining the Cartesian coordinates of all its atoms as well as a set of bonds that

defines their interconnectivity. The drawback of this method is the high number of degrees of freedom (DOFs) which frustrates many methods in molecular modeling when applied to larger structures (e.g., proteins).

Alternatively, a molecule may be represented by its translation, orientation, and a set of torsional angles, bond angles, and bond lengths resulting in a similar number of DOFs. However, this representation has the advantage that nonrelevant degrees of freedom can be frozen. For a number of applications in molecular modeling, like docking, it is reasonable to constrain molecular flexibility by fixing bond angles and bond lengths and restricting torsional flexibility to rotation around single bonds that connect rigid entities, like ring structures.^{1–3} By using such a compact representation, the number of DOFs is significantly reduced while it is possible to rapidly convert from the compact to the Cartesian representation. This is mandatory for many tasks in molecular modeling, such as ligand receptor docking, which are usually addressed by optimizing an energy or scoring function.

Correspondence to: D. Neumann; e-mail: d.neumann@bioinf.uni-sb.de

A number of meta-heuristics used for ligand receptor docking, like the Lamarckian Genetic Algorithm,⁴ try to improve their results by performing local optimization. These methods can be classified into two distinct categories: approaches which need only function values and methods utilizing the function's derivatives. The first class can be subdivided into deterministic algorithms (e.g., Powell algorithm,⁵ Simplex algorithm⁶), and stochastic methods like the algorithm of Solis and Wets.⁷ The approaches of the second class benefit from employing derivatives of the objective function⁸ and are expected to find better results faster, e.g. “deeper minima” requiring shorter time. Therefore, these approaches are preferable whenever useful derivative information is available.

Nonetheless, the methods of the first class, especially the approach of Solis and Wets,⁷ are widely used in docking applications. There are two main reasons for using these methods: (1) In practice, many scoring functions, especially non force field-based functions, are continuous but not differentiable. For these functions, nongradient-based techniques of the first class seem favorable. (2) Stochastic methods like the Solis and Wets⁷ approach are easily adapted to specific optimization tasks.

On the other hand, the gradient-based methods of the second class are restricted to differentiable objective functions. Furthermore, they are sensitive to singularities, like a loss of DOFs, or to nonminimal parametrizations. Unfortunately, there is no orientational parametrization that is at the same time minimal^{9,10} and free of singularities. Therefore, the main challenge with respect to the compact molecular representation is to find an orientational parametrization that is well suited for gradient optimization.

In Cartesian space, the orientation of a body can be represented in different ways. Using the well-known Euler angle representation resulting in a straightforward calculation of the gradient has the disadvantage of the so called gimbal lock phenomenon,¹¹ that is the loss of DOFs. Generally speaking, this means that in the course of an optimization, a molecule can reach positions in which not all possible further alterations of its orientation are achievable by variations of the three Euler angles. Thus, the unit quaternion, which avoids gimbal lock singularities, has become a quasi standard for orientations.¹² However, because the unit quaternion space is only a subset of \mathbb{R}^4 with three DOFs, direct optimization of the four interdependent unit quaternion values is awkward.⁹

In this work we present a novel method for energy minimization of molecular complexes with special attention to gradient-based optimization of the molecules' orientation. We use the exponential mapping,¹⁰ which transforms a vector in \mathbb{R}^3 to the unit quaternion space. Although this representation is not free of singularities, they can be avoided with only negligible additional computational cost allowing for the application of most gradient-based local optimization algorithms. We demonstrate that an L-BFGS approach^{13,14} is, in our modified version, optimally suited and highly efficient for the given task. Furthermore, we show that negligible straightforward modifications of a continuous but not differentiable function are sufficient to apply our optimization algorithm successfully. To compare the gradient-based optimization of the exponential map representation to the method of Solis and Wets⁷ that operates directly on the four quaternion values, we minimized the energy of seven ligands with increasing complexity in the presence of their corresponding receptors. As an example for energy functions without a continuous derivative, we used the piecewise linear Gehlhaar

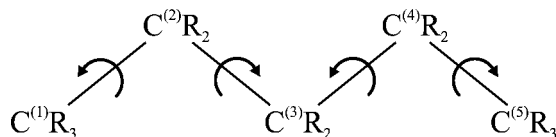


Figure 1. Example for rotatable bonds and molecular centroid with R being arbitrary heavy atoms. If bond $C^{(1)} - C^{(2)}$ is rotated, only atoms R connected to $C^{(1)}$ are moved. If bond $C^{(2)} - C^{(3)}$ is rotated, $C^{(1)}$ and atoms R connected to $C^{(1)}$ and $C^{(2)}$ are moved. Due to symmetry the same holds true for the other two bonds with other indices. This means $C^{(2)}$, $C^{(3)}$, and $C^{(4)}$ are never moved and hence define the molecular centroid.

scoring function¹⁵ which was repeatedly employed for molecular docking.^{16–18} Our novel method shows better performance in all test cases. Especially the energetically more favorable conformations—found already after a few iterations—render the new approach a valuable tool for molecular optimization.

Materials and methods

Molecular Representation

The conformation and position of a molecule in space is uniquely defined by the Cartesian coordinates of its atoms. Often the complete molecular flexibility is abandoned in favor of a reduced set of parameters that is required for representing a molecule. Like many other applications^{3,19,20} we use translation, orientation, and a set of flexible torsional angles that connect rigid compounds. Thus, we need three real values for the translation (t_x, t_y, t_z), one real value for each flexible torsional angle (ϕ_1, \dots, ϕ_n), and a unit quaternion composed of four real values (q_1, q_2, q_3, q_4) for the molecule's orientation. A parameter vector $\mathbf{x} = (t_x, t_y, t_z, q_1, q_2, q_3, q_4, \phi_1, \dots, \phi_n)$ is converted into a molecular conformation by a series of transformations. In the first step, all flexible torsional angles are processed. Because in our case such a rotatable bond is guaranteed not to be part of a ring, it divides the molecule in two substructures. The part containing fewer atoms is rotated while the other one remains stationary (Fig. 1). This procedure is applied to all flexible torsional angles. In the next step, the whole molecule is rotated. To this end, the origin is defined by the average position of all atoms that were not rotated in the first step thus defining a form of molecular centroid. In other implementations the rotation origin is intuitively placed onto the geometric center of the ligand, but this method complicates the computation of the gradient. In the last step, the molecule is moved according to the three translational parameters.

Gehlhaar Scoring Function

A scoring function in ligand-receptor docking is expected to meet multiple requirements. In the first place the scoring function should allow to differentiate native binding poses from decoy structures. Secondly, the score should approximate the binding free energy. Furthermore, it ought to be efficiently computable. In this work, we chose the Gehlhaar function¹⁵ mainly for the following reasons: ease of implementation, sufficient correlation of the rmsd, less frustrated energy landscape compared to other scoring functions, and finally as

Table 1. Atom Types for Nonbonded Interactions.

Atom type	Donor	Acceptor	Both	Nonpolar
Donor	Steric	HB	HB	Steric
Acceptor	HB	Steric	HB	Steric
Both	HB	HB	HB	Steric
Nonpolar	Steric	Steric	Steric	Steric

a test case for an inherently not continuously differentiable function. It must be noted that the Gehlhaar score cannot be used to estimate the binding free energy.

The formula for the score E is composed of one bonded term for the torsional potential E_{tor} and one nonbonded term E_{pair} for a kind of van der Waals interaction

$$E = E_{\text{tor}} + E_{\text{pair}}.$$

For the computation of E_{pair} , the Gehlhaar scoring function distinguishes only four atom types: nonpolar, hydrogen-bond-donor, hydrogen-bond-acceptor, and both-acceptor-and-donor. The interaction between any of these atom types results in two types of nonbonded interaction, namely steric and hydrogen bond contributions (Table 1).

Both interaction types are calculated by an interval piecewise linear function f of the pairwise atom distance d_{ij} of atoms i and j , with each type having different function parameters (Table 2, Fig. 2)

$$E_{\text{pair}} = \sum_{i \neq j} f(d_{ij}).$$

This function is obviously not continuously differentiable so we added a quadratic transition function in an interval of 0.02 \AA length at each junction of the original linear segments (Fig. 2). These functions are uniquely defined by their interpolation conditions.

The term for the torsional energy E_{tor} is similar to that of other scoring functions, but restricted to $\text{sp}^3\text{-sp}^3$ and $\text{sp}^2\text{-sp}^3$ bonds:

$$E_{\text{tor}} = A \cdot (1 + \cos(n \cdot \phi - \phi_0))$$

with $A = 3.0$, $n = 3$, $\phi_0 = \pi$ for $\text{sp}^3\text{-sp}^3$ bonds, and $A = 1.5$, $n = 6$, $\phi_0 = 0$ for $\text{sp}^2\text{-sp}^3$ bonds. The original Gehlhaar function provides a separate energy term for the internal nonbonded interaction of the ligand by assigning a penalty of 10^4 if two ligand atoms that do not share a bond come closer than 2.35 \AA . This kind of energy

Table 2. Parameter Set for Nonbonded Steric and Hydrogen-Bonding Potentials.

	A	B	C	D	E	F
Steric	3.4 \AA	3.6 \AA	4.5 \AA	5.5 \AA	-0.4	20.0
HB	2.3 \AA	2.6 \AA	3.1 \AA	3.4 \AA	-2.0	20.0

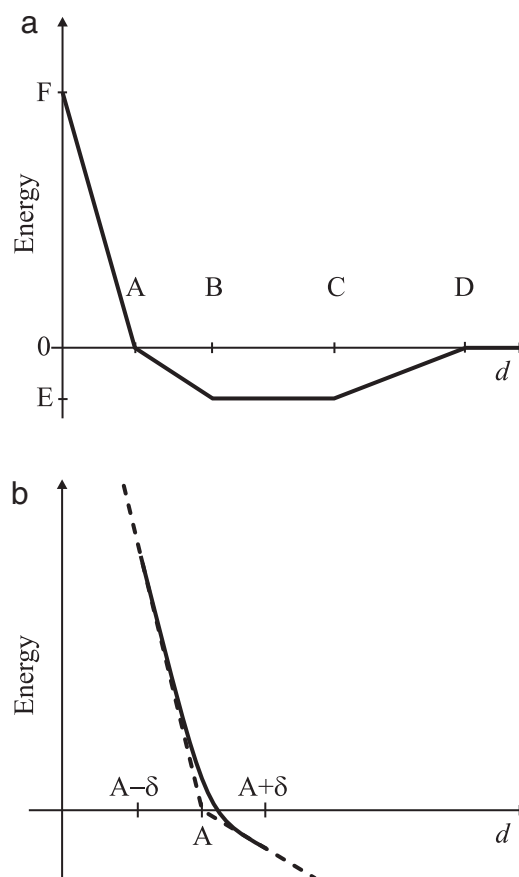


Figure 2. (a) shows the original piecewise linear pairwise potential function used for nonbonded interactions. (b) illustrates the modifications (solid line) applied to the original function (dashed line) to produce a continuously differentiable function.

calculation is entirely unsuited for the computation of a gradient for it is highly noncontinuous. To circumvent this problem we use the same term for internal ligand–ligand interactions as for ligand–receptor interactions.

Gradient Computation and Exponential Mapping

The application of a gradient-based optimizer requires the derivatives of the underlying energy or scoring function E with respect to the parameter vector \mathbf{x} . The Gehlhaar function¹⁵ consists of a pairwise term E_{pair} and a torsional term E_{tor} . Thus, the gradient \mathbf{g} is given by

$$\mathbf{g} := \frac{\partial E}{\partial \mathbf{x}} = \frac{\partial (E_{\text{pair}} + E_{\text{tor}})}{\partial \mathbf{x}}.$$

The gradient of E_{tor} can be easily computed and affects only torsional parameters ϕ_1, \dots, ϕ_n

$$\begin{aligned}\frac{\partial E_{\text{tor}}}{\partial \phi} &= \frac{A \cdot (1 + \cos(n \cdot \phi - \phi_0))}{\partial \phi} \\ &= -n \cdot A \cdot \sin(n \cdot \phi - \phi_0).\end{aligned}$$

To calculate the derivatives for the pairwise interactions ∂E_{pair} we first compute the gradient \mathbf{g}_i for each atom i . This is the sum of all derivatives of pairwise interactions that an atom participates in with \mathbf{v}_i being the position vector of atom i and \mathbf{v}_j being the position vector of the interacting atom j

$$\mathbf{g}_i = \sum_{j \neq i} f'(d_{ij}) \frac{\mathbf{v}_i - \mathbf{v}_j}{\|\mathbf{v}_i - \mathbf{v}_j\|}.$$

Mapping the gradient \mathbf{g}_i of an atom i with position \mathbf{v}_i to an arbitrary parameter r requires the derivative of \mathbf{v}_i with respect to r . $\partial \mathbf{v}_i$ represents the tangential movement of atom i when r varies by an infinitesimal amount and can now be used to calculate the derivative of E_{pair} with respect to r

$$\frac{\partial E_{\text{pair}}}{\partial r} = \sum_i \left(\frac{\partial \mathbf{v}_i}{\partial r} \right)^T \mathbf{g}_i. \quad (1)$$

In the following, we will use eq. (1) to calculate the derivatives of E with respect to specific parameters.

Translational Gradient

Calculating $\partial \mathbf{v}_i$ with respect to a translational parameter t is straightforward because any change in t translates \mathbf{v}_i linearly. Thus, for any translational parameter t , eq. (1) can be reduced to

$$\begin{aligned}\frac{\partial E_{\text{pair}}}{\partial t_x} &= (1, 0, 0) \cdot \sum_i \mathbf{g}_i, \\ \frac{\partial E_{\text{pair}}}{\partial t_y} &= (0, 1, 0) \cdot \sum_i \mathbf{g}_i, \\ \frac{\partial E_{\text{pair}}}{\partial t_z} &= (0, 0, 1) \cdot \sum_i \mathbf{g}_i.\end{aligned}$$

Torsional Gradient

The rapid computation of the torsional gradient has been the subject of numerous scientific studies.²¹ If atoms k and j are connected by a rotatable bond and atom i is moved by rotating this bond (Fig. 3a), the derivative of \mathbf{v}_i with respect to a torsional parameter ϕ_{jk} can be calculated by

$$\frac{\partial \mathbf{v}_i}{\partial \phi_{jk}} = (\mathbf{v}_k - \mathbf{v}_j) \times (\mathbf{v}_i - \mathbf{v}_j). \quad (2)$$

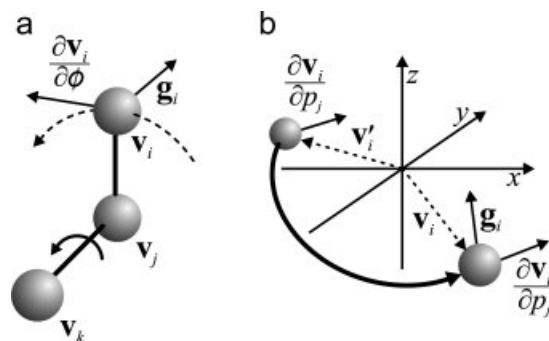


Figure 3. Mapping of nonbonded gradient to (a) torsional and (b) orientational parameter. Please note that in figure (b), $\frac{\partial \mathbf{v}_i}{\partial p_j}$ is first calculated using \mathbf{v}_i' and then translated to \mathbf{v}_i .

Inserting (2) in (1) yields

$$\frac{\partial E_{\text{pair}}}{\partial \phi_{jk}} = \sum_i ((\mathbf{v}_k - \mathbf{v}_j) \times (\mathbf{v}_i - \mathbf{v}_j))^T \mathbf{g}_i.$$

Oriental Gradient

The most challenging part is the computation of the orientational gradient, because, up to now, there is no minimal representation that does not inherit some kind of singularity, e.g. loss of DOFs. Representing the orientation (three DOF) by a unit quaternion does not include such a singularity, but the independent optimization of its four values is awkward.⁹ This is caused by the unit quaternions representing only a subset of the entire four-dimensional quaternion space. To alleviate this problem, we use exponential mapping¹⁰ to map a point $\mathbf{p} = (p_1, p_2, p_3)$ from parameter space \mathbb{R}^3 to q in the unit quaternion space S^3 :

$$q = (q_1, q_2, q_3, q_4) = \begin{cases} (0, 0, 0, 1) & \text{if } \mathbf{p} = (0, 0, 0) \\ (\sin(0.5\|\mathbf{p}\|) \frac{\mathbf{p}}{\|\mathbf{p}\|}, \cos(0.5\|\mathbf{p}\|)) & \text{otherwise} \end{cases}.$$

This enables us to compute the derivative \mathbf{T}_j of the corresponding rotation matrix \mathbf{R} ,

$$\mathbf{T}_j = \frac{\partial \mathbf{R}}{\partial p_j},$$

for each of the three orientation parameters p_j ¹⁰ that can now be used to calculate the gradient $\frac{\partial E_{\text{pair}}}{\partial p_j}$. For a detailed description of the computation of \mathbf{T}_j please refer to the publication by Grassia.¹⁰ For each evaluation of the objective function, the orientational parameter \mathbf{p} is mapped to a unit quaternion q . q is then converted to a rotation matrix \mathbf{R} that defines the molecular orientation

$$\mathbf{R} = \begin{pmatrix} q_1^2 + q_2^2 - q_3^2 - q_4^2 & 2q_2q_3 - 2q_1q_4 & 2q_1q_3 + 2q_2q_4 \\ 2q_1q_4 + 2q_2q_3 & q_1^2 - q_2^2 + q_3^2 - q_4^2 & 2q_3q_4 - 2q_1q_2 \\ 2q_2q_4 - 2q_1q_3 & 2q_1q_2 + 2q_3q_4 & q_1^2 - q_2^2 - q_3^2 + q_4^2 \end{pmatrix}.$$

Let \mathbf{v}'_i be the position of an arbitrary atom i and \mathbf{v}_i the position of the atom after the rotation by \mathbf{R} (Fig. 3(b))

$$\mathbf{v}_i = \mathbf{R}\mathbf{v}'_i.$$

Then

$$\frac{\partial \mathbf{v}_i}{\partial p_j} = \mathbf{T}_j \mathbf{v}'_i.$$

Again inserting in eq. (1) yields

$$\frac{\partial E_{\text{pair}}}{\partial p_j} = \sum_i (\mathbf{T}_j \mathbf{v}'_i)^T \mathbf{g}_i, \quad j = 1, 2, 3.$$

As mentioned earlier, no method for a minimal parametrization of the orientation is free of singularities. This also holds for exponential mapping, where singularities arise if the length of the orientational parameter vector \mathbf{p} approaches 2π . All parameter vectors \mathbf{p} with $\|\mathbf{p}\| = n \cdot 2\pi, n \in \mathbb{Z}_{>0}$ are mapped to the quaternion $q = (0, 0, 0, -1)$. For these parameter vectors, all gradients $\frac{\partial \mathbf{v}_i}{\partial p_j}$ point into the same direction, reducing the number of DOFs to one. Fortunately, all possible orientations can be denoted by parameters within a shell of π around the origin in \mathbb{R}^3 . Thus we only need to take care that the optimization algorithm stays within this shell.

Gradient-Based Local Minimization

Numerical gradient-based local optimization techniques proceed in an iterative fashion by walking downhill on the objective function. In iteration k the best known methods, the so called Newton methods, build up a quadratic model

$$M_k(\mathbf{x}) := E(\mathbf{x}_k) + \mathbf{g}(\mathbf{x}_k)^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{H}(\mathbf{x}_k) \mathbf{x}$$

of the objective function E , where $\mathbf{H}(\mathbf{x}_k)$ and $\mathbf{g}(\mathbf{x}_k)$ are the Hessian and the gradient, respectively, of E at the current position \mathbf{x}_k . If this model is accurate the minimizer* of M_k is a good candidate for the next position \mathbf{x}_{k+1} on E . This approximation requires E to be twice continuously differentiable and needs evaluations of the gradient and the Hessian. Furthermore, if $\mathbf{H}(\mathbf{x}_k)$ is not at least positive semi-definite M_k does not have any minimizer.⁸ The so called quasi-Newton methods replace $\mathbf{H}(\mathbf{x}_k)$ by a matrix \mathbf{B}_k approximating an (assumed) Hessian via collected data (gradients) during the optimization process. This approach only prerequisites a once continuously differentiable objective function and avoids additional evaluations of second derivatives. We use the BFGS[†]

*The position where the minimum is located is the so called minimizer.

†The update scheme for \mathbf{B}_k was discovered independently by Broyden, Fletcher, Goldfarb, and Shanno. It has become known as the BFGS update or BFGS method, named after the initial letters of its inventors.

method,²² which ensures \mathbf{B}_k to be positive definite leading to a unique minimizer $\mathbf{x}_k^{\text{min}}$ of the model

$$\hat{M}_k(\mathbf{x}) := E(\mathbf{x}_k) + \mathbf{g}(\mathbf{x}_k)^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{B}_k \mathbf{x}.$$

To allow for applying our method to large molecules we used a limited memory variant of the BFGS quasi-Newton method (L-BFGS) based on the Strang recurrences^{13,14} which employs only a small amount of collected data to form the model Hessian, say the gradients and positions of the recent m steps. Using this method $\mathbf{x}_k^{\text{min}}$ can be calculated very efficiently with essentially only $n(4m+1)$ multiplications without forming \mathbf{B}_k explicitly, where n denotes the number of parameters. We used $m = 5$ and the adaptive scaling described by Liu and Nocedal.²³

If $\mathbf{x}_k^{\text{min}}$ is acceptable the iteration finishes and $\mathbf{x}_k^{\text{min}}$ becomes immediately \mathbf{x}_{k+1} . In the other case, e.g. $E(\mathbf{x}_k^{\text{min}}) > E(\mathbf{x}_k)$,[‡] either a line search in the direction $\mathbf{d}_k := \mathbf{x}_k^{\text{min}} - \mathbf{x}_k$ is applied[§] or a trust region approach tries to find an acceptable position within a region where \hat{M}_k seems to approximate E very well. However, it is difficult to incorporate our rotational condition in a trust region based approach without significantly increasing computational costs for determining the next trial step. On the other hand, a line search based approach can be easily extended to fulfill this criterion. Let $\mathbf{x}_k^{\text{ori}}$ be the orientational parameters of \mathbf{x}_k and $\mathbf{d}_k^{\text{ori}}$ the corresponding orientational parameters of \mathbf{d}_k , then the general line search problem (or any relaxed version²⁴)

$$\lambda_k := \arg \min_{\lambda > 0} E(\mathbf{x}_k + \lambda \mathbf{d}_k)$$

is restricted to

$$\|\mathbf{x}_k^{\text{ori}} + \lambda_k \mathbf{d}_k^{\text{ori}}\| \leq \frac{3}{2} \pi.$$

This ensures that we always stay in a region far away from the singularity 2π . We used the line search method of Moré and Thuente,²⁴ which can be easily modified to incorporate an upper bound on λ_k . After the line search has finished, in a usual quasi-Newton method the current iteration would terminate yielding $\hat{\mathbf{x}}_k := \mathbf{x}_k + \lambda_k \mathbf{d}_k$ as the next iterate \mathbf{x}_{k+1} . In our case, however, we have to avoid the orientational singularity. Our modified line search ensures that we are in a region where the orientational derivatives are usable but possibly tending to the singularity. Thus, we reparameterize the orientational part $\hat{\mathbf{x}}_k^{\text{ori}}$ of $\hat{\mathbf{x}}_k$ if

$$\|\hat{\mathbf{x}}_k^{\text{ori}}\| \geq \pi$$

‡See for example Dennis and Schnabel⁸ for detailed acceptance rules.

§Note that \mathbf{d}_k is always a descent direction because \mathbf{B}_k is positive definite.

Table 3. Comparison of Our Method to Solis and Wets in Terms of Average Initial and Final Score and Average Number of Function Evaluations.

PDB ID	Ref.	Rotatable bonds/ heavy atoms	Initial score	Our method		Solis and Wets	
				Score	Number of evaluations	Score	Number of evaluations
1FDS	28	0/20	232.5	-50.4	9.80	-12.35	39.41
1FMO	29	2/19	295.7	-56.62	21.89	0.46	46.63
2MCP	30	3/11	199.2	-30.8	16.77	-11.76	29.06
1DWD	31	8/37	714.13	-68.86	34.06	88.48	48.40
1HPV	32	9/35	627.3	-75.13	35.97	107.57	69.86
2R04	33	10/25	770.66	-19.8	62.5	230.7	51.37
1HTF	34	12/41	693.27	-65.93	38.89	117.36	58.17

via replacing $\hat{\mathbf{x}}_k^{\text{ori}}$ by[¶]

$$\left(1 - \frac{2\pi}{\|\hat{\mathbf{x}}_k^{\text{ori}}\|}\right) \hat{\mathbf{x}}_k^{\text{ori}}.$$

Because this is an equivalent orientation¹⁰ (but with better derivatives), this replacement is a reparameterization from a geometric point of view. From a mathematical point of view, this is a well-defined jump on E . In theory, the Euler angle representation could also be dynamically reparameterized to avoid the gimbal lock phenomenon. However, in our case the reparameterization means only to scale the orientational parameters whereas the similar operation on Euler angles implies a sequence of inverse trigonometric functions²⁵ to determine the new parameters. Finally, the current iteration finishes by returning $\hat{\mathbf{x}}_k$ as the next iterate \mathbf{x}_{k+1} .

Solis and Wets Optimization Method

The local search method of Solis and Wets⁷ is a stochastic heuristic for continuous parameter spaces. Its primal purpose is the optimization of functions that do not provide gradient information, e.g. the AUTODOCK scoring function.³ For our comparison we closely followed the version of AUTODOCK 3.1 with the only alterations being due to adjustments to the BALL²⁶ environment. The basic algorithm starts with a random search step and generally follows this direction with random movements as long as the objective function keeps improving. Continued improvements lead to an expansion of the random search steps, whereas continued failing narrows the search. The algorithm iterates until either a maximum number of function evaluations is reached or convergence is established by the random step width falling below a certain threshold value.

BALL

The entire code that was used in this work for the modeling of molecules, scoring functions, etc. was generated using the BALL library.²⁶ BALL is an application framework written in C++ that provides a set of data structures as well a large number of methods in

the field of computational biology. It was designed to be an efficient and robust tool for rapid software prototyping.

Comparison of the Optimization Methods

To test our method on different molecules with varying complexity, we selected seven different ligand-receptor complexes from the Protein Data Base (PDB).²⁷ The ligands span from simple to more complex molecules (Table 3).

For the actual comparison we minimized each ligand 500 times starting from random positions. We recorded for each minimization the best score and the number of function evaluations required to attain a score of 1.0 worse than the best score. This threshold value was selected to account for the inherently different stopping criteria of both methods. For the method of Solis and Wets we partially retained the stopping criterion of the AUTODOCK implementation, that is the falling below a defined step width while we dismissed the maximum number of iterations. This means that, other than in AUTODOCK, the approach of Solis and Wets is always allowed to explore the scoring function to the local minimum. Our method used the following standard stopping criterion, i.e. convergence is assumed if in iteration k

$$\|\mathbf{g}(\mathbf{x}_k)\| < \varepsilon_g(1 + |E(\mathbf{x}_k)|)$$

holds for $\varepsilon_g = 10^{-6}$. However, the compact representation of the molecules has the side effect, especially with increasing flexibility, that small changes of some parameters cause large alterations of the energy value. This effect is still intensified for the gradient and might lead to many iterations around the position of interest without substantially improving the molecular structure. Hence, we used a second stopping criterion, derived from the idea behind the criterion of the method of Solis and Wets combined with an energy criterion that cannot be satisfied if we are too far away from the minimum: Convergence is also assumed if the steps fall below a relative step size and the energy values does not significantly alter any more, i.e. if

$$\|\mathbf{x}_k - \mathbf{x}_{k-1}\| < \sqrt{\varepsilon_E} \frac{(1 + \|\mathbf{x}_k\|)}{100}$$

[¶]We choose π because the derivatives in this region are excellent, while all possible orientations are still representable.

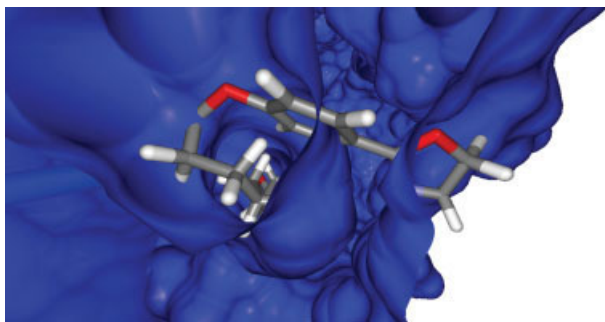


Figure 4. Example for a high energy local minimum. The larger part of the docked ligand 2R04 is situated in the binding pocket on the left while the smaller part penetrates the surface.

and

$$|E(\mathbf{x}_k) - E(\mathbf{x}_{k-1})| < \varepsilon_E(1 + |E(\mathbf{x}_k)|)$$

hold for $\varepsilon_E = 10^{-8}$.

To produce random starting positions we uniformly randomized a ligand's orientation and conformation as well as its translation within a cube of edge length 6 Å focused on the geometric center of the ligands reported binding site.

Results

To compare our approach to the method of Solis and Wets, we used both methods to optimize the positions and conformations of the prepared ligands (start conformations). Table 3 shows the average Gehlhaar-score of 500 minimizations together with the average number of evaluations required to reach a function value at most 1.0 worse than the final score. The number of rotatable bonds corresponds roughly to the complexity of the optimization problem while the average energy before optimization indicates that generally the ligand has multiple van der Waals clashes at the random initial position. The results show that, on average, the score of our method is well below 0 for all ligands. This means that it generally resolves all van der Waals clashes and moves the molecule in a way that it is able to form multiple interactions. Even for more complex ligands, representing more difficult optimization problems, the average score does not deteriorate and seems to be roughly corresponding to the number of heavy atoms. As expected, more complex ligands require more function evaluations to reach the local minimum. In contrast to that, the method of Solis and Wets is able to resolve van der Waals clashes only for simple ligands with both average score and average number of function evaluations being considerably worse compared to our method. As ligands get more complex, the approach of Solis and Wets fails to resolve van der Waals clashes and the scores decline considerably.

There seems to be one outlier, 2R04, for which the results are worse than expected. This is caused by the particular morphology of the binding pocket which forms a longish tube inside the receptor and is located relatively near to the receptor surface. Thus, the likewise elongated ligand can be trapped with one part being situated

in the binding pocket and the other outside the receptor while the center is penetrating the protein producing multiple van der Waals clashes (Fig. 4).

Figure 5 illustrates the performance difference of both methods and the non-deterministic character of the approach of Solis and Wets for 1DWD. In this case, all minimizations started from the same initial position. Our method always converged to a score of -108.49 (solid line) while the best result out of 100 Solis and Wets minimizations was -69.50 (dashed line). On average the approach of Solis and Wets reached a value of 76.74 (the dotted line shows a typical minimization). The method of Solis and Wets required 149 function evaluations to produce its best results, a value that was reached by our method with only 17 function evaluations.

Discussion

When performing docking calculations, the atom positions of the ligand are often described using a compact representation which allows for reducing the DOFs, for example by fixing bond lengths. In some cases, the energy or scoring function is continuous and differentiable which is a prerequisite for the usage of gradient-based algorithms for structure optimization. Despite the advantages of these algorithms, stochastic methods like the approach by Solis and Wets are employed, because the computation of the orientational gradient is difficult. Here, we demonstrated how to solve this issue by using the exponential map to transform a vector in \mathbb{R}^3 to the unit quaternion space and by avoiding the arising singularities. Thus, gradient-based optimization of molecules represented by translation, orientation, and torsional angles is possible employing any continuously differentiable scoring function.

However, our approach is not confined to continuously differentiable objective functions. For energy or scoring functions, which are continuous and not differentiable at a finite number of points, the gradient may be computed by introducing a patch function at

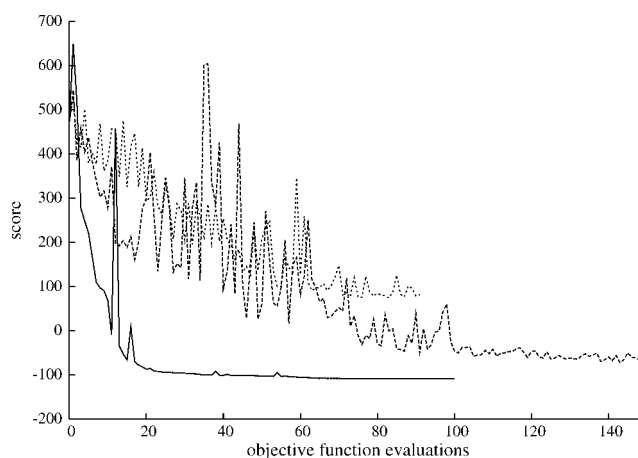


Figure 5. Comparison of one deterministic minimization of our method (solid line) to two different minimizations of Solis and Wets from the same initial position (PDB ID 1DWD). The dashed line is the best result of the approach of Solis and Wets out of 100 minimizations.

the junction of two adjacent continuously differentiable segments of the original objective function. In general, a cubic patch function should meet the requirements to adequately interpolate between the two segments. An example for such a piecewise linear potential function is the Gehlhaar scoring function which was repeatedly employed in protein-ligand docking studies.^{15–18} For this function, a quadratic patch function was sufficient for a smooth transition between adjacent linear portions.

Our approach outperformed the widely employed stochastic Solis and Wets algorithm even for relatively simple optimization problems with few DOFs. The difference in performance became even more substantial with increasing complexity of the optimization problem. For molecules with many rotatable bonds, the approach of Solis and Wets is not able to determine search directions, which are as promising as the directions calculated by the gradient-based approach. The repeated failing to improve the score leads to permanently decreasing stepsizes and finally to the abortion of the optimization possibly without having located a minimum in the energy hypersurface.

In contrast, the internal search direction calculation of our approach is very efficient due to the use of the exponential map and the adoption of the L-BFGS¹⁴ method. The only noticeable additional costs concern the calculation of the gradient of the scoring function. However, in the case of our tested function, the evaluation of the energy value can be extended to compute the gradient efficiently. These additional operations appear to be negligible in comparison to the high number of exploratory energy evaluations of the approach of Solis and Wets. Finally, as a result of the use of derivative information the proposed method reaches usually significantly deeper energy values in fewer steps than the Solis and Wets algorithm.

Conclusion

Our results suggest that the effort to make a scoring function differentiable is worthwhile. When it comes to minimization of molecules that are represented by translation, orientation, and torsional angles, the approach of Solis and Wets⁷ has become a quasi standard procedure. We think that every global optimization method like e.g. the Lamarckian Genetic Algorithm³⁵ that utilizes the algorithm of Solis and Wets for local optimization will benefit when our method is used instead.

References

1. Baxter, C.; Murray, C.; Clark, D.; Westhead, D.; Eldridge, M. *Proteins* 1998, 33, 367.
2. Ewing, T.; Kuntz, I. D. *J Comput Chem* 1997, 18, 1175.
3. Morris, G.; Goodsell, D.; Halliday, R.; Huey, R.; Hart, W.; Belew, R.; Olson, A. *J Comput Chem* 1998, 19, 1639.
4. Hart, W.; Belew, R. *Adaptive Individuals in Evolving Populations: Models and Algorithms*; Belew, R. K.; Mitchell, M.; Eds.; Addison-Wesley Longman Publishing Co.: Boston, MA., 1996; pp. 483–496.
5. Powell, M. *Comput J* 1964, 7, 155.
6. Nelder, J.; Mead, R. *Comput J* 1965, 7, 308.
7. Solis, F.; Wets, R.-B. *Math Opt Res* 1981, 2, 19.
8. J.E. Dennis, J.; Schnabel, R. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations (Classics in Applied Mathematics, 16)*; Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996.
9. Schmidt, J.; Niemann, H. *Using Quaternions for Parametrizing 3–D Rotations in Unconstrained Nonlinear Optimization. In: Vision, Modeling, and Visualization 2001*; Aka GmbH, Stuttgart, Germany, 2001; pp. 399–406.
10. Grassia, F. *J Graph Tools* 1998, 3, 29.
11. Watt, A.; Watt, M. *Advanced Animation and Rendering Techniques—Theory and Practice*; ACM Press: New York, NY, 1992.
12. Karney, C. *J Mol Graph Model* 2007, 25, 595.
13. Nocedal, J. *Math Comp* 1980, 35, 773.
14. Matthies, H.; Strang, G. *Int J Numer Methods Eng* 1979, 14, 1613.
15. Gehlhaar, D.; Verkhivker, G.; Rejto, P.; Sherman, C.; Fogel, D.; Fogel, L.; Freer, S. *Chem Biol* 1995, 2, 317.
16. Verkhivker, G.; Bouzida, D.; Gehlhaar, D.; Reijto, P.; Arthurs, S.; Colson, A.; Freer, S. T.; Larson, V.; Luty, B.; Marrone, T.; Rose, P. *J Comput-Aided Mol Des* 2000, 14, 731.
17. Hou, T.; Wang, J.; Xu, X. *Chin Chem Lett* 1999, 10, 615.
18. Thomsen, R.; Christensen, M. H. *J Med Chem* 2006, 49, 3315.
19. Jones, G.; Willet, P.; Glen, R.; Leach, A. *J Mol Biol* 1997, 267, 727.
20. Trosset, J.-Y.; Scheraga, H. *J Comput Chem* 1999, 20, 412.
21. Abe, H.; Braun, W.; Noguti, T.; Go, N. *Comput Chem* 1984, 8, 239.
22. Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 1987.
23. Liu, D.; Nocedal, J. *Math Program* 1989, 45, 503.
24. Moré, J.; Thuente, D. *ACM Trans Math Softw* 1994, 20, 286.
25. Shoemake, K. In *mphGraphics Gems IV*; Academic Press Professional: San Diego, CA, USA, 1994; pp. 222–229.
26. Kohlbacher, O.; Lenhof, H.-P. *Bioinformatics* 2000, 16, 815.
27. Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res* 2000, 28, 235.
28. Breton, R.; Housset, D.; Mazza, C.; Fontecilla-Camps, J. *Structure* 1996, 4, 905.
29. Narayana, N.; Cox, S.; Shaltiel, S.; Taylor, S.; Xuong, N. *Biochemistry* 1997, 36, 4438.
30. Padalan, E.; Cohen, G.; Davies, D. *Ann Inst Pasteur Immunol* 1985, 136C, 271.
31. Banner, D.; Hadvary, P. *J Biol Chem* 1991, 266, 20085.
32. Kim, E.; Baker, C.; Dwver, M.; Murcko, M.; Rao, B.; Tung, R. D.; Navia, M. *J Am Chem Soc* 1995, 117, 1181.
33. Badger, I.; Minor, I.; Oliveira, M.; Smith, T.; Rossmann, M. *Proteins* 1989, 6, 1.
34. Jhoti, H.; Singh, O. M. P.; Weir, M. P.; Cooke, R.; Murray-Rust, P.; Wonacott, A. *Biochemistry* 1994, 33, 8417.
35. Hart, W. Ph.D. thesis. Department of Computer Science and Engineering, University of California, San Diego, 1994.