JCTC Journal of Chemical Theory and Computation

A Simple Method for Improving Torsion Optimization of Ligand Molecules in Receptor Binding Sites

Jianwei Che*

Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, California 92121

Received September 7, 2004

Abstract: A simple but effective method is introduced for optimizing ligand molecules in torsion space within receptor binding sites. The algorithm makes use of geometric constraints of ligand molecules to search for energetically favorable conformations. It is applied to a conjugate gradient (CG) method as an example. During conformational energy optimization, new line search directions are modified according to the spatial span of rotational groups in ligand molecules. Significant improvements were observed in terms of the abilities both to recover global optimal structures and to obtain lower energy ensembles. This simple algorithm allows rapid implementation and can be incorporated into other conformational energy optimization techniques.

I. Introduction

Energy minimization has been a common and powerful tool in molecular modeling. While global optimization is extremely difficult and rare to achieve except for simple systems, local optimization has been quite useful for studying the properties of a molecular system. For example, it is very important to predict the correct binding configuration of ligand−receptor complexes for rational drug design and lead optimization in the pharmaceutical industry, and one important technique is through multiple conformational energy minimizations of ligand molecules.[1−7] Because of the relative stiffness of bond stretching and angle bending modes, the conformational variation of a ligand molecule mainly comes from the changes of dihedral angles of rotatable bonds (e.g. simple σ bond in a chain). A typical small molecule ligand has several to more than a dozen such rotatable bonds, so its conformational space can be very large and consist of many local energy minimum states. Furthermore, as the number of rotatable bonds increases, the complexity of the system increases exponentially and results in a rough energy landscape, and the energy surface becomes even more complex and rough in a protein binding site.

In classical molecular modeling, energy optimization is normally treated as an applied mathematical problem.[8−11] For example, conjugate gradient (CG) and Monte Carlo (MC) methods are two popular techniques. Other more advanced techniques have been developed, such as truncated Newton conjugate gradient method[12,13] and convex global underestimation (CGU) method.[14] Within the CG framework, different versions have also been developed based on different approaches of updating the line search directions.[15−20] Optimizing a molecular structure is generally done in Cartesian coordinates (CC) (i.e. x, y, and z) or internal coordinates (IC) (e.g. bond length, bond angle, and dihedral angle). They have been implemented in various molecular mechanics and dynamics software.[8−11] Upon noncovalent binding with receptors, it is often convenient to constrain ligand bond lengths and angles at equilibrium positions. This can be done easily in IC but more complicated in CC.

Torsion (i.e. dihedral angle) optimization of ligand molecules is usually efficient because of soft modes of torsional motions. In addition, it has less number of degrees of freedom (DOF) than all atom CC optimization does. Its effectiveness, however, is reduced when applied to ligand molecules in receptor binding sites, because many rotations are restricted by the presence of receptor atoms. The rotation about some ligand rotatable bonds, in particular bonds close to the central core structure, is likely to cause steric clashes. Consequently, only very small steps of rotation can be taken during optimization. To improve torsional optimization under these circumstances, we present here a modified version of this method that considers the geometric structure of biomolecular

* Corresponding author e-mail: jche@gnf.org.

10.1021/ct0499433 CCC: $30.25
Published on Web 05/26/2005

complexes. The modified algorithm was able to obtain global optimal configuration much easier than conventional torsional optimization. In addition to the example application of improving conventional CG, this method can be easily integrated with other existing optimization algorithms such as steepest descent, Powell's method,[17] etc. to increase their performance, particularly for flexible ligand docking. The paper is organized into following sections. In section II, we describe the algorithm in detail. In section III, numerical examples and discussions are presented, and we summarize in section IV.

II. Algorithm

In torsion optimization only dihedral angles are free variables. The system potential energy can be written as

$$U = U(\{\tau_i\}|\{r_j,\theta_k\},i,j,k = 1,2,3...) \quad (1)$$

where τ_i is the ith torsion angle, and r_j and θ_k are jth bond length and kth bond angle, respectively. Note that all rs and θs are fixed parameters in the potential function. Equation 1 can also be written in CC with constraints of fixed bond lengths and angles. It is known that the potential surface in receptor binding sites is very rough, and energy barriers between local minima are often high compared to room temperature. Conventional energy minimization easily gets trapped in local minimum close to initial condition. To overcome the difficulty, soft core potentials (e.g. reduced van der Waals radii) are often used in docking program to ease the steric clashes. Although it introduces softness to the rotation modes restricted by receptor atoms, a drawback is that more false positives can squeeze into the sites, and it is difficult to discriminate between them. Other methods to level off potential barriers[21,22] have also been used.

To further improve the effectiveness of energy minimization with or without modifying potential, we introduce a simple scaling technique that can be combined with many optimization methods. As a demonstration, we will limit our discussion to the CG method, since CG requires a bit more attention in implementation than other simpler methods such as steepest descent. The general principle of torsional scaling, however, is applicable to other algorithms with little modification. In the CG algorithm, a new coordinate that minimizes a given function is generated from an iterative sequence[17]

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{d}_i \quad (2)$$

where $\mathbf{x}_i$ is the coordinate vector at ith iteration, $\mathbf{d}_i$ is the direction vector, and α_i is a scalar of the step length. After each iteration, the new coordinate (e.g. $\mathbf{x}_{i+1}$) locally minimizes the given function along the given search direction (e.g. $\mathbf{d}_i$). The important quantity in CG method is the generation of the direction vector sequence. A simple form is given by[20]

$$\mathbf{d}_{i+1} = -\mathbf{K}\mathbf{g}_{i+1} + \beta_i \mathbf{d}_i \quad (3)$$

where matrix $\mathbf{K}$ is normally chosen to be identity matrix, $\beta_i = \mathbf{g}_{i+1}{}^T\mathbf{K}(\mathbf{g}_{i+1} - \mathbf{g}_i)/\mathbf{d}_i{}^T(\mathbf{g}_{i+1} - \mathbf{g}_i)$, and $\mathbf{g}_i$ is the gradient of the potential function at position $\mathbf{x}_i$. The direction vector sequence must satisfy the conjugacy relation

$$\mathbf{d}^T{}_i\mathbf{A}\mathbf{d}_j = 0, \textit{ for } i \neq j \quad (4)$$

$\mathbf{A}$ is the Hessian matrix of the given potential function, though in practice, most CG implementations avoid explicit or direct calculation of the Hessian matrix.[17,23] In CG, a central step is to calculate the successive conjugate directions $\mathbf{d}_i$, which is determined by β_i. Many variations of CG (e.g. Fletcher-Reeves,[15] Polak-Ribiere,[18] and Shanno's[19,20] CG) differ only by the method of updating line search directions. In the case of exact quadratic functions, they yield identical results. However, because of the nonquadratic nature of molecular potential functions, computer round-off error, and inexact line minimization, they can and normally do give different convergence performance and final results.

Besides many improvements in optimization procedures from mathematic perspectives, less attention has been given to the molecular nature of the problem. When only dihedral angles are free variables, a molecule closely resembles a kinematical system. In fact, this analogy has been used in molecular dynamics and mechanics in torsion space.[24−26] Because torsion movements can introduce nonlocal structural changes for ligand−receptor systems, it limits the optimization step length in line minimization along certain directions. To reduce the limitation, we need to consider the molecular geometric property. It is known from basic mechanics that the corresponding general forces for dihedral angles are torques

$$\mathbf{T}_i = -\nabla_{\tau_i} U = \sum_{j\in D_i} \mathbf{r}_j \times \mathbf{f}_j \quad (5)$$

where τ_i is the rotation vector for dihedral angle i that has the direction of rotation in right-handed convention and magnitude of rotation angle. $\mathbf{T}_i$ is the corresponding torque, which can be calculated from forces and coordinates in CC. D_i denotes the set of all atoms involved in the rotation of torsion angle CC. $\mathbf{f}_j$ is the force acting on atom j, and $\mathbf{r}_j$ is the vector of atom j in the local coordinate frame where ith rotation center vector $\mathbf{O}_i$ is the origin. $\mathbf{r}_j$ can then be calculated from space fixed frame coordinates $\mathbf{R}_j$

$$\mathbf{r}_j = \mathbf{R}_j - \mathbf{O}_i \quad (6)$$

Substitute eq 6 into eq 5, we have

$$\mathbf{T}_i = \sum_{j\in D_i} \mathbf{R}_j \times \mathbf{f}_j - \mathbf{O}_i \times \sum_{j\in D_i} \mathbf{f}_j \quad (7)$$

For a branched molecule, D_i can be partitioned into subgroups if it has additional rotatable bonds. The rotation about the main rotatable bond changes all the atomic positions in the group, and the rotation about a subrotatable bond modifies only the atomic positions within the subgroup. The hierarchical relation forms a tree topology and each branch forms a "cluster". At each level, a cluster can have sublevel clusters if it has rotatable bonds. Otherwise, the cluster is called a leaf. In our present treatment, a rotatable bond is defined as a nonterminal σ bond that does not participate in a ring structure. In principle, the definition can be extended to include double bond and ring bonds. For

computational convenience, we chose the atom that connected to its parent cluster as the rotation center $\mathbf{O}_i$. The overall molecule is treated as a grand cluster, and its center of mass (COM) is the rotation center $\mathbf{O}_{\text{grand}}$. For a given cluster, its torque in the local coordination frame is calculated using eq 5. Therefore, we can rewrite the total torque of a cluster in its local frame in terms of the local torques of its subclusters

$$\mathbf{T}_i = \sum_{D_j \in D_i} [\mathbf{T}_j + (\mathbf{O}_j - \mathbf{O}_i) \times \mathbf{F}_j] \tag{8}$$

where D_j runs through all subclusters of D_i, and $\mathbf{F}_j$ and $\mathbf{O}_j$ are the total force and position of subcluster j. For leaf clusters, only the second terms in the summation survive, and eq 8 recovers eq 5. Similarly, $\mathbf{T}_j$ is also zero for single atoms in cluster i, where $\mathbf{O}_j$ are atomic coordinates and $\mathbf{F}_j$ is its force. Equation 8 clearly illustrates a recursive nature of the torque calculation for the tree topology clusters. In actual implementation, a sweep from tip to base is carried out to calculate $\mathbf{T}_i$ for each cluster using eq 8. Each cluster stored its torque $\mathbf{T}_i$, total force $\mathbf{F}_i$, and rotation center $\mathbf{O}_i$ as its C++ class attributes. As total torque can be calculated recursively from the sweep, the actual gradient along each rotatable bond is the projection of torque. The projection step is trivial and can be carried out after the sweep. We also want to mention that Jain et al.[24] developed a very efficient recursive method for solving equation of motion in IC without explicit inversion of mass matrix. The difference here is that there is no mass matrix in the optimization procedure. Energy minimization in IC is fortunately simpler than the dynamics in this respect. Therefore, we can avoid many matrix operation steps in Jain's[24,25] algorithm. As we can see from eq 8, large clusters tend to have large torques because of their large radius of gyration even if they experience a similar amount of force as smaller ones, and rotation of large clusters will likely create more serious steric clashes with receptor atoms than smaller ones. Because large clusters extend further away from the rotation axis, atomic displacements are amplified for atoms at the tip. It renders line minimization for dihedral angles much less effective if the search direction is dictated by local gradients (e.g. starting direction, and restarting direction), because only a small step length can be taken in each iteration. Therefore, direct optimization of torsion angles (re)started with torques given by eq 5 more likely gets trapped in a local optimal structure close to the initial starting point.

Here, we introduce a novel way of defining initial and restart line search direction in the CG procedure according to the size of each cluster to reduce the restraining effects induced by large clusters. The quantity to characterize the size of a cluster rotating about a given axis is the radius of gyration, defined as l_i for cluster i at a given conformation. When a molecule changes its conformation, the radii of gyration for some or all clusters will also change their values. For a given rotation axis, the radius of gyration of the ith cluster is calculated through the projection of its inertia tensor $\mathbf{I}_i$ with respect to the rotation center

$$\mu_i l_i^2 = \mathbf{i}^T \mathbf{I}_i \mathbf{i}, i = 1,2, \cdots \tag{9}$$

where $\mathbf{i}$ is the unit vector of rotation axis for cluster i. The inertial tensor at any given rotation center is evaluated from the inertia tensor in its COM frame

$$\mathbf{I}_i = \mathbf{I}_i^{\text{com}} + \mu_i(a_i^2 \mathbf{1} - \mathbf{a}_i \mathbf{a}_i^T), i = 1,2, \cdots \tag{10}$$

where $\mathbf{1}$ is the unit tensor, $\mathbf{a}_i$ is the displacement vector from the rotation center to the COM, $a_i^2 = \mathbf{a}_i^T \mathbf{a}_i$, and μ_i is the total mass of cluster i. Similar to eq 8, the inertia tensor of parent cluster is also obtained from their subclusters

$$\mathbf{I}_i^{\text{COM}} = \sum_{D_j \in D_i} \lfloor \mathbf{I}_j^{\text{COM}} + \mu_j(a_{ij}^2 \mathbf{1} - \mathbf{a}_{ij} \mathbf{a}_{ij}^T) \rfloor, i = 1,2, \cdots \tag{11}$$

where $\mathbf{a}_{ij} = \mathbf{R}_j^{\text{COM}} - \mathbf{R}_i^{\text{COM}}$. This is done during the same sweep as torque collection. Inertia tensor $\mathbf{I}_i^{\text{COM}}$ and $\mathbf{R}_i^{\text{COM}}$ are also attributes of cluster class. It is also worth pointing out that normal radius of gyration is not the only choice for characterizing the size of a cluster. For instance, substituting all atomic masses by corresponding van der Waals radii in above equations or simply choosing the largest atomic distance away from the rotation axis in each cluster can all be used to define a cluster size.

Instead of using the negative gradient as initial and restarting direction vectors in conventional CG, we define the initial and restarting direction vectors according to scaled gradients. First, we scale the torque as the following

$$\mathbf{T}_i^s = \mathbf{T}_i / l_i, i = 1,2, \cdots \tag{12}$$

where the superscript s indicates the scaled torque. Define a diagonal scaling matrix $\mathbf{S}$ as

$$\mathbf{S} = \begin{bmatrix} 1/l_1 & & & \\ & 1/l_2 & & \\ & & 1/l_3 & \\ & & & \vdots \end{bmatrix} \tag{13}$$

The scaled initial and restarting direction vectors $\mathbf{g}_t^s$ are obtained from original gradients

$$\mathbf{g}_t^s = \mathbf{S}\mathbf{g}_t \tag{14}$$

Because the scaling factor l_i can be very different for different clusters, the newly constructed line search direction is normally different from the direction used in conventional CG. Note that for a given sequence of CG minimization steps, the subsequent line search directions are determined by conjugacy condition governed by eq 4, applying eq 14 to each iteration will normally break the conjugacy. Therefore, in CG implementation, eq 14 is applied to initial and restarting directions only. Other methods, such as steepest descent, can use eq 14 at each step of minimization. As the result, the scaling transformation chooses a direction that limits the amplification of torques for a large cluster, improving CG optimization to a local/global optimum. Also as expected, the scaling becomes less important when the conformation gets closer to the converged structure.

The first step for torsion optimization is to divide a ligand molecule into clusters. Figure 1 illustrates how a ligand molecule is partitioned. The partition process starts by choosing a base atom, and all atoms that connect to the base

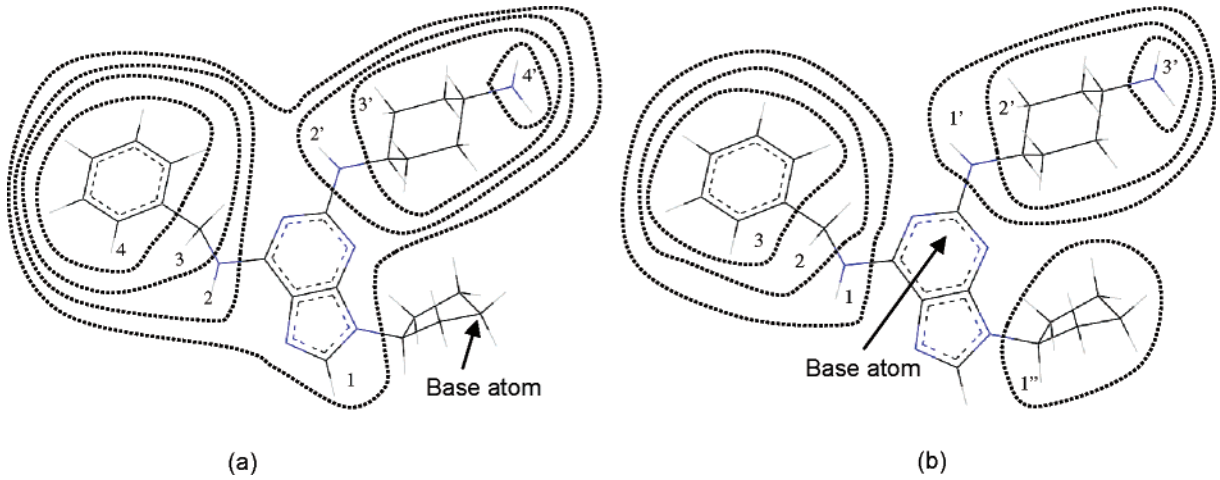

Figure 1. Schematic clustering of molecule, 2-[trans-(4-aminocyclohexyl)amino]-6-(benzyl-amino)-9-cyclopentylpurine. (a) Clusters obtained by choosing a tip atom as the starting root atom. (b) Clusters obtained by choosing the atom closest to COM as the starting root atom.

atom through nonrotatable bonds belong to the base group. When an atom connects to the base group by a rotatable bond, this atom is placed into a subcluster. The base group is always in the grand cluster (i.e. overall molecule). This process is carried out recursively until all atoms are put into proper clusters. In principle, any atom can be a base atom. For instance in Figure 1 (a), a tip atom is the base atom, while in (b) the atom closest to the COM is the base atom. In practice, we found that it was more efficient when the tree structure was balanced in terms of radii of gyration of clusters, for a reason similar to what was previously discussed for large clusters. When a molecule is properly clusterized, the sizes of clusters are kept as small and balanced as possible to reduce the large swinging of tip atoms. Although not rigorously proved, simply choosing the atom closest to COM as the base atom normally gives a satisfactory partition of clusters. Once a ligand molecule is partitioned, energy minimization can be carried out in a straightforward manner. In the next section, we demonstrate that this simple torsion scaling modification can improve the effectiveness of optimizing ligand molecules in binding sites.

III. Numerical Results and Discussion

To study the performance of the modified algorithm, we implemented it in an existing structure optimization module.[27,28] No modification was necessary to the existing module except the initial and restarting directions were recalculated according to eq 14. In our comparitive studies, the exact same multidimensional CG optimization routine was used to carry out three variations of energy minimization: scaled torsion, direct torsion, and all atom minimizations in CC. In all atom CC optimization, line search directions were calculated from atomic forces directly. The inter- and intramolecular interactions were modeled by Merck Molecular Force Field (MMFF94).[29−35] Interactions between ligand and receptor molecules were precalculated on a rectangular grid box that encompassed the whole binding site. The grid spacing is 0.5 Å, and three-dimensional cubic spline interpolation was used to calculate the energy between grid points. In order for the interpolation to give accurate values, we limited the extremely large repulsive pair wise energy by the following equation[36]

$$V_i' = \begin{cases} V_i, & \text{if } V_i < U_s \\ U_s + \dfrac{U_m(V_i - U_s)}{U_m + V_i - U_s}, & \text{if } V_i \geq U_s \end{cases} \quad (15)$$

where V_i is the interaction energy between the ith ligand atom and the receptor molecule, and U_s and U_m are user defined parameters. In our calculations, we use large values $U_s =$ 120 kcal/mol and $U_m =$ 240 kcal/mol in our calculations to maintain the "original" form the potential surface. Although whether this potential accurately represented the molecular system is another question and beyond the scope of this study, our intention was to test the optimization procedure itself even in the very extreme condition. The main purpose for eq 15 in our calculation is for the accuracy of interpolation rather than optimization itself. Equation 15 leveled off only the very top portion of the energy surface (>120 kcal/mol) without any modifications to the lower energy surface. In contrast, conventional potential softening (e.g. reducing van der Waals radii) alters the overall energy surface including the shapes and perhaps the positions of energy minima. To keep the ligand molecule from moving out of binding sites, a harmonic restraint potential is imposed outside the binding box.

Figure 2 depicts the 3D structures of four ligand molecules among those that we used as examples. Our examples included additional 31 ligand−receptor complexes randomly chosen from PDB database, so that they represented a variety of protein−ligand complex structures. These four structures shown here represented very typical ligands and systems from very simple (i.e. few rotatable bonds in 1E1X) to more complicated ones (i.e. more than a dozen rotatable bonds in 1KV2). Their cocrystal PDB codes are labeled above each molecule.

As shown in Figure 1, clusters can have very different sizes. Consequently, the scaling factor can be quite different for different clusters. More importantly, if a cluster contains several levels of subclusters, its radius of gyration l_i (i.e. the

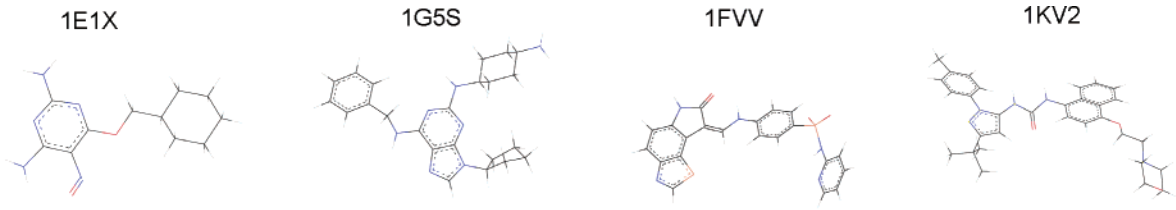

Figure 2. 3D structures of four inhibitor molecules for CDK2. Cocrystal PDB codes are labeled above each structure. Color coding for elements: C: black, H: gray, N: blue, F: green, S: orange, O: red.

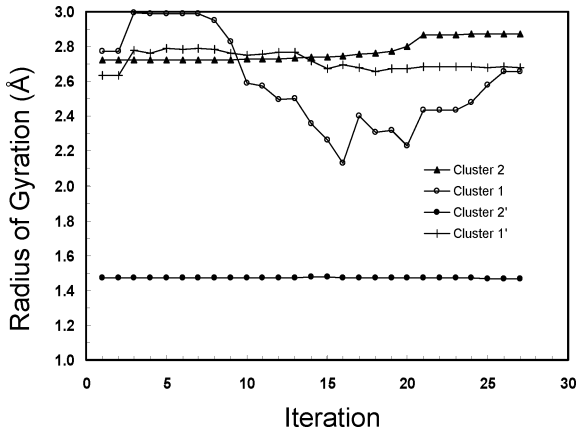

Figure 3. Variation of radius of gyration of all clusters during an optimization for the ligand molecule in Figure 1. The cluster hierarchical structure in Figure 1 (b) is used.

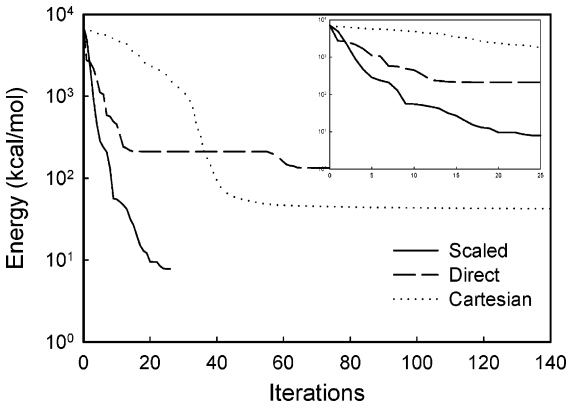

Figure 4. The optimization results of three energy minimization algorithm. The solid line is for scaled torsion optimization, the dashed line is for directed torsion optimization, and the dotted line is for all atom optimization in CC. The inset shows the zoom-in view of the first 25 iterations. The vertical axis for the main graph is in logarithm scale, and it is in linear scale in the inset graph. 100 kcal/mol had been added to the energy values in order to show logarithm scale.

scaling factor) may change significantly during the minimization process. It requires consistent updates of the radius of gyration of all clusters whenever a restarting vector is constructed in CG. The "dynamic" behavior of the radius of gyration during optimization is shown in Figure 3 for a minimization run.

The radii of gyration of some clusters did not change appreciably while others varied significantly during the minimization process. Obviously, the ones that belong to the leaf clusters (omitted in Figure 3) did not change at all (i.e. cluster 1″, 3, and 3′ in Figure 1(b)), and the ones whose subclusters had relatively high symmetry also did not change much (i.e. cluster 2 and 2′ in Figure 1(b)). If l_i of a cluster changed significantly (e.g. cluster 1 had a range of 1 Å variation), it indicated that the cluster conformation has been altered dramatically. The energy changes for this particular minimization process are depicted in Figure 4 for the three optimization algorithms. It shows that the scaled torsion optimization led to a much lower energy state. The inset shows the zoom-in view of the first 25 steps.

Interestingly, while the direct torsion optimization reduced the energy most rapidly in the first few steps, the scaled torsion optimization converged on a lower energy state. The reason is that the direct optimization chose the negative gradients as initial line search direction and therefore was able to go downhill more rapidly than scaled version. However, the initial rapid downhill movement trapped the conformation in a local minimum, so it ended up with a higher energy state. In contrast, the scaled torsion optimization chose the direction that allowed the molecule to have larger configuration change without being pulled into nearby local minima so quickly. This property is particularly appropriate when minimizing ligand molecules in receptors. The variation of l of cluster 1 and the total energy indicated that the final structure was quickly approached in about 25 iterations. To assess the quality of the final converged structures, we compared them to the global optimized crystal structure, which was obtained from all atom minimization of the experimental X-ray structure. Its energy value was −94 kcal/mol and had an RMSD of 0.55 Å with respect to the experimental X-ray structure. More importantly, it had the lowest energy among all configurations studied in our calculations. We believed that this structure could be used as a standard to gauge the performance of the 3 different optimization procedures. As mentioned earlier, direct torsion optimization is more likely to get trapped in a nearby local minimum because of the larger motion of tip atoms. This behavior is quite evident in Figure 4, where all three different methods arrived at different energy states. Each minimization was carried out until gradient RMS was below 0.1 kJ/mol or it reached 500 iterations. Both direct torsion and all atom optimization had difficulties escaping from the local minimum near the initial structure. When the CG (re)started with scaled directions, the molecule was able to find a path to reach the global optimal. The final structure had an energy of −93 kcal/mol and an RMSD of 0.6 Å with respect to the global optimal, and the final energies from direct torsion and all atom CC optimizations were 32 kcal/mol and −60 kcal/mol, respectively. Since each CG minimization iteration finds the local minimum in a particular search direction, it usually

requires multiple (e.g. 3−4) energy updates (i.e. energy evaluations) to bracket and locate the coordinate that minimizes the energy function. In our calculations, normal Amijo and Goldstein criteria[37] were used to terminate each iteration. After an iteration was finished, a new search direction was generated, and the minimization was resumed along the new direction. Here, we gave both iteration and energy update numbers for detailed comparisons. In Figure 4, the scaled method reached the final structure in 26 iterations with energy updates of 109 times, while Cartesian optimization used 406 iterations with energy updates of 1304 times and direct torsion optimization used 76 iterations with energy updates of 285 times to reach their final structures. Although the scaled torsion optimization converged more quickly than direction torsion algorithm in Figure 4, it cannot be generalized. Although the behavior in Figure 4 is not rare, it is not a common phenomenon for a scaled optimization to be significantly faster than direct torsion optimization and achieving much lower energy than CC at the same time. On average, the speed of the scaled method is comparable to direct torsion optimization. Usually, it is very slow for all atom CC optimizations to converge, while the direct torsion optimization often ends with high energy final structures. The whole purpose of the scaled method is to have the advantages from both all atom and direct torsion optimizations while avoiding their major drawbacks. In other words, it is a method able to find low energy structures like all atom CC optimization and as efficiently as the conventional torsion method.

It is usually less meaningful to judge an optimization method for rugged potential functions based on one or a few calculations. A good energy minimization algorithm should consistently produce a lower energy ensemble for a set of random initial conditions, provided that the set is made without biases. To get a global picture of how the scaled torsion optimization can improve an existing direct optimization algorithm for ligand configuration energy minimization, 300 initial configurations for each of the 35 ligand molecules were generated by randomly rotating the rotatable bonds and an overall random rotation translation of the optimized cocrystal structure. The resulting initial structures all had very large RMSD with respect to the global optimal (4 Å−8 Å) and high initial energy (above several thousands kcal/mol). Each configuration was then optimized directly by the three different procedures. Although the initial configurations were far away from global optimal structures and inappropriate for direct molecular dynamics simulations, they were quite relevant in studies that searched for global optima such as docking studies. Unlike molecular dynamics studies of a particular binding configuration, where most conformations in an initial ensemble are generated around the configuration, docking finds the correct one without prior knowledge how a ligand binds. In other words, a much larger configuration space has to be searched to find the best possible binding mode. For any local optimization method, this also means that many calculations are done in a physically irrelevant configuration space. However, whether a given initial configuration leads to a meaningful final state is generally not known a priori. This is particularly true for rugged energy surfaces such as the ones in protein binding sites. Therefore, the initial configurations generated here are very relevant to many practical applications.

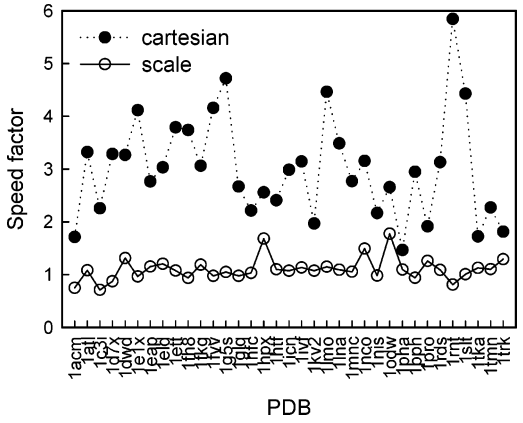

Figure 5. The computational speed factors all-atom Cartesian and scale torsion optimizations scaled by direct torsion method. The dotted line is for Cartesian coordinate optimization, and the solid line is for scaled torsion optimization.

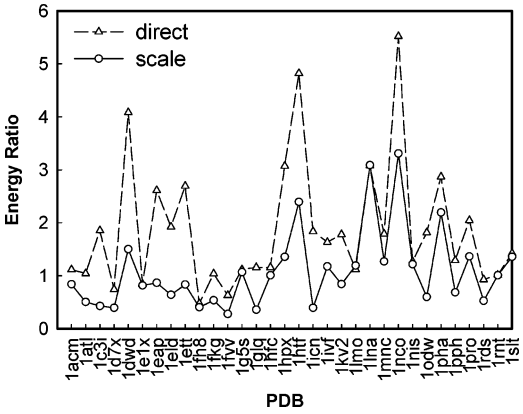

Figure 6. The average energy ratio of direct and scale torsion optimizations measured by all atom CC minimization method. The dashed line is for direct torsion, and the solid line is for scaled torsion optimization.

As we briefly mentioned previously, the scaled method is expected to be as fast as the conventional direct torsion optimization due to its intrinsic similarity. In Figure 5, the computational speed performances of all atom CC and scaled torsion optimizations were illustrated by the ratio between their average iterations and the ones from direct torsion minimization for each system. The larger the factors are, the slower the method is. As we expected, the efficiency of direct torsion optimization was well maintained in the scaled method, and both optimizations were significantly faster than all atom CC optimization.

It is known that all atom optimizations in many cases can achieve energetically more favorable structures than direct torsion optimization, even though it may take significantly longer computing time. The results in Figures 5 and 6 also indicated this behavior. In Figure 6, we plotted the ratios of average energies of all minimized structures from scaled and direct torsion optimizations to the all atom CC method. The ratio was calculated by $(\bar{E} - E_{\min}^{\text{global}})/(\bar{E}_{CC} - E_{\min}^{\text{global}})$, where $\bar{E}$ is the average energy given by a particular method,

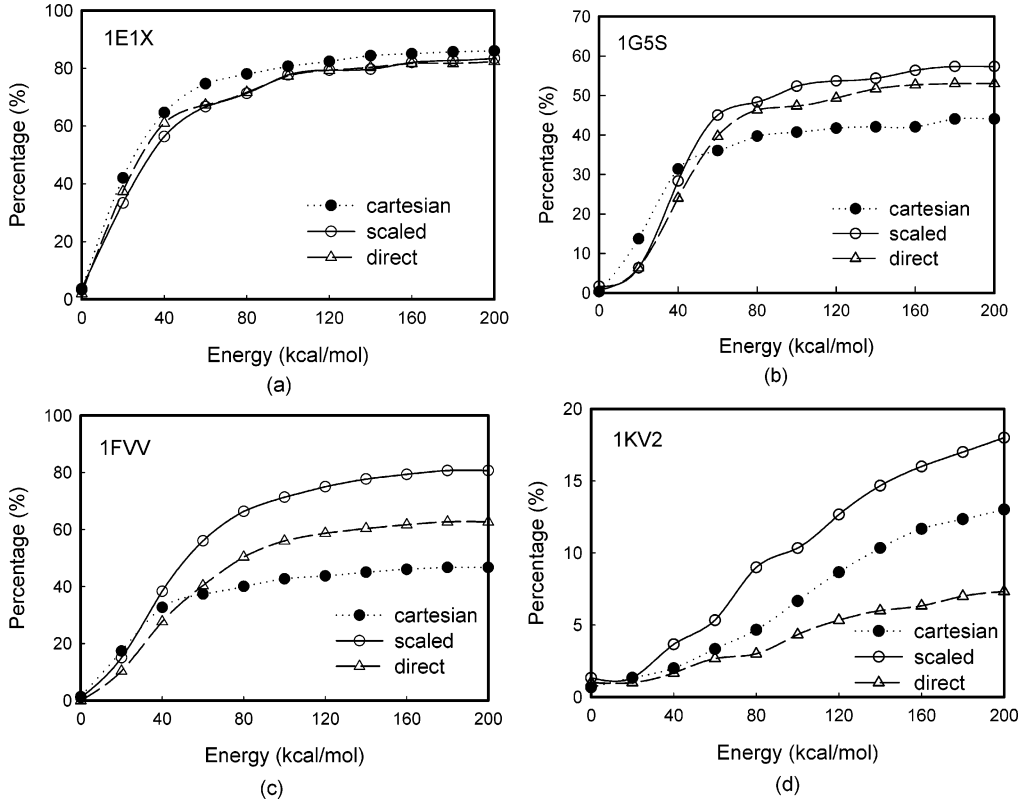

Figure 7. Energy distribution of conformational ensembles obtained from three different optimization procedures for (a) 1E1X, (b) 1G5S, (c) 1FVV, and (d) 1KV2. The initial energy distributions are not shown because they have extremely high energies far beyond the scale of the figure. The first 200 kcal/mol energy window is shown against the accumulate percentage of conformers in each energy bin of 20 kcal/mol. The solid line is for scaled torsion optimization, the dashed line is for direction torsion optimization, and the dotted line is for all atom CC optimization.

$\bar{E}_{CC}$ is the average energy by all atom CC minimization, and E_{min}^{global} is the global minimum energy. Obviously, the scaled optimization consistently outperformed the direct torsion method and gave comparable energies as all atom CC optimizations on average.

Since the scaling is a simple step on top of a conventional optimization algorithm, its overall performance inherits the intrinsic property of the specific optimizer with which it is combined. CG is a local optimization algorithm, and so is the scaled CG optimizer. The added benefit is that it can find better local energy minima more effectively, but it does not guarantee to locate the global optimal. It is also common that different versions of the optimizer result in different final structures from an identical initial structure. It should be noted that effective rotation barriers in torsion optimization are systematically higher than those in all atom optimization because of less flexible molecular structures.[38] We expect the success rate will be further improved if the effective rotational barriers were made consistent with all atom optimization. Although it is not exactly equivalent, a way of estimating it is to further minimize the final structures after torsion optimizations with all atom CC minimization. Our calculations showed that the average energies from both direct and scaled torsion optimization were further reduced significantly and became much lower than the ones given by an all atom CC method alone. The final energy difference between the two torsion methods, however, became much smaller. They were normally within a few to tens kcal/mol. The energies from the scaled method were consistently lower (about 22 kcal/mol on average). Since the combination of optimization methods is not the main subject in the paper, the properties and results will be studied elsewhere. The results shown in subsequent sections were all from a single optimization method unless otherwise specified.

Besides the comparison of the average energies of distributions, it is also important to check the energy distributions resulting from different optimizers. As an example, Figure 7 shows the accumulative percentage of conformers within the lowest 200 kcal/mol for the four systems in Figure 2.

As shown in Figure 7, scaled torsion optimization consistently recovers more low energy structures than direct torsion algorithm and most all atom CC optimization. The higher rate over all atom CC optimization is much less obvious because molecules had higher flexibility and greater numbers of DOF in all atom CC optimization. For 1E1X, scaled method, direct torsion, and all atom CC optimization behaved similarly. This can be explained by the simplicity of the ligand structure. Because most clusters are similar in size and there is only 1 rotatable bond that can affect its radius of gyration, the benefit of scaling torque is substantially reduced. However, the scaled method still recovered the global optimal structures almost twice as frequently as the direct torsion method, and its average total energy was also slightly lower than direct method. The difference between the methods is most evident for more complex

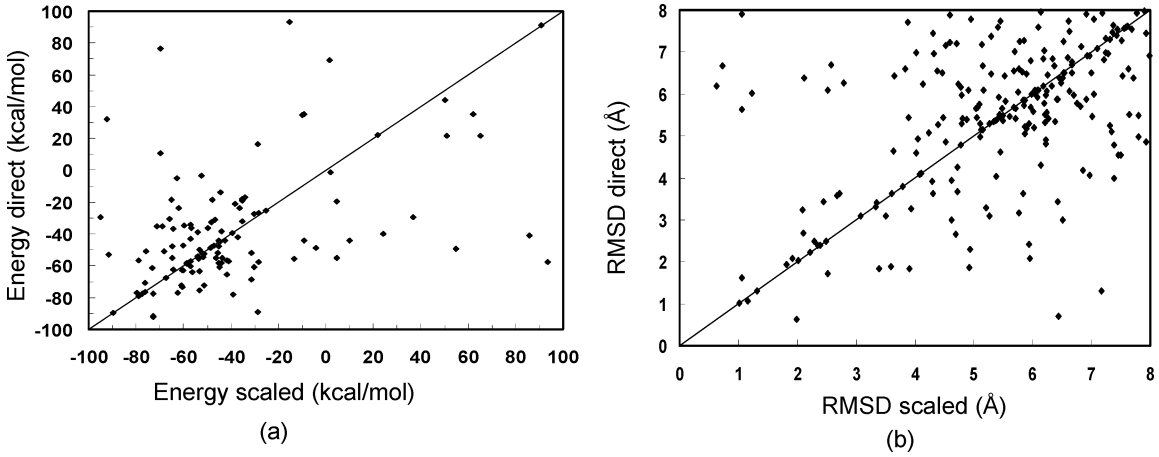

Figure 8. (a) Final energies obtained from direct torsion optimization against the ones from scaled torsion optimization. (b) RMSDs of final structures obtained from direct torsion optimization against the ones from scaled method.

Table 1. Number of Configurations that Reached Global Minima

	scaled torsion minimization	direct torsion minimization	all atom CC minimization
single optimization	69	47	60
reoptimized by all atom CC minimization	219	191	60

structures such as in Figure 7(d) for 1KV2, which showed the greatest advantage among the four examples. The comparison clearly demonstrated that the simple scaling improved the effectiveness of direct torsion optimization particularly for complicated ligand molecules. This trend was also evident in all 35 test cases that we studied here. Overall, the scaled torsion optimization found 34% more structures than the conventional torsion method and 16% more than the all atom CC method within the lowest 90 kcal/mol window. The numbers of structures that reached global optima showed similar behavior as well. In Table 1, we listed the total numbers of final structures that reached global optima for all 35 cases. These structures were chosen based on energies that were within 3 kcal/mol relative to the global optima. Their RMSDs were generally less than 1 Å relative to the global optimal structure. The numbers in the first row were from straightforward optimizations by three methods, respectively. The numbers in the second row were from further optimization by an all atom CC method after torsion optimization. It clearly showed that scaled torsion optimizer was able to recover more global minima structures than direct torsion minimization, particularly when only one level of optimization could be used due to other constraints (e.g. resources, time).

It is also interesting to compare how different the final structures were due to different optimization algorithms. In Figure 8(a), we showed the energy correlation for the lowest 200 kcal/mol window between the direct and scaled methods for complex 1G5S. Although there is a trend of positive correlation between the two methods, the correlation is very weak. This relation is also clear in the RMSD correlation plot of Figure 8(b).

Also, the initial structures that were able to reach the global optimal had very little overlap between all three methods. This behavior is the direct result of the rough energy surface. However, different algorithms have different levels of tolerance for starting structures from which the global optimal can be obtained. This can be inferred from Table 1. Clearly, the combination of the three methods achieved the best energy ensembles. It indicated that none of the optimization methods should be used exclusively. All three methods can be combined to maximize the likelihood of finding the global minimum. Nevertheless, if only one method can be used (e.g. due to the limitation of computing resource), we believe that the scaled torsion optimization is an optimal choice among the three.

IV. Summary

We have introduced a very simple and effective scaling step to improve torsion minimization of ligand molecules inside receptor binding sites. Unlike conventional torsion optimizations that directly use gradients from eq 5 as initial and restarting line search directions in CG, our method takes into account the molecule's 3-dimensional structural properties. This enhancement improved the effectiveness of an existing optimization program. The method can be combined with other functional optimization algorithms such as steepest descent or other versions of CG methods such as TNCG to further improve their performance. We demonstrated that the scaled torsion optimization was able to reach a lower energy ensemble than both direct torsion like all atom CC optimization while still maintaining the computational efficiency of direct torsional optimization. The scaling can also be applied to nongradient based optimization such as dihedral angle MC. It improves the efficiency of the MC algorithm by increasing the acceptance rate of dihedral angle movements.[28] Besides optimization, the principle is applicable to conformational sampling. In our docking software,[28] we found the scaling improved the quality of the conformational distributions.

Although the scaling method generally improves direct torsion optimization, it does not intend to replace existing optimization algorithms. The new method is optimally used in conjunction with other methods to improve the overall results of molecular modeling as we can see from Table 1. In summary, the scaling method is a simple but powerful algorithm that has many applications in molecular modeling.

We specifically emphasize the method for ligand molecules in receptor binding sites because of the roughness of the energy surface in general binding sites. For free small ligand molecules, there is no significant advantage. However, we believe that the scaled method can improve the optimization for large ligand molecules as well as proteins or polymers. It is also useful when simultaneously optimizing protein residues and ligand molecules.

Acknowledgment. This research was funded by the Novartis Research Foundation. We have made use of biochemical algorithm library (BALL) in our implementation.[27] The author especially thanks Dr. Andrew Su for the critical reading of the manuscript.

References

(1) Leach, A. R. *A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules*; VCH Publishers: New York, 1991.

(2) Blaney, J. M.; Dixon, J. S. *Perspect. Drug Discovery Des.* **1993**, *1*, 301.

(3) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.

(4) Meng, E. C.; Schoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 380.

(5) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelly, M.; Perry, J.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739.

(6) Totrov, M.; Abagyan, R. *Proteins: Struct., Funct., Genet., Suppl.* **1997**, *1*, 215.

(7) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.

(8) Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287.

(9) Brooks, B.; Bruccoleri, R. E.; Olafson, B. D.; States, O. J.; Swiminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(10) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.

(11) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.

(12) Xie, D.; Schlick, T. *SIAM J. Optim.* **1999**, 9.

(13) Derreumaux, P.; Zhang, G.; Brooks, B.; Schlick, T. *J. Comput. Chem.* **1994**, *15*, 532.

(14) Phillipss, A. T.; Rosen, J. B.; Walke, V. H. *Dimacs Ser. Discrete Math. Theor. Comput. Sci.* **1995**, *23*, 181.

(15) Fletcher, R.; Reeves, C. M. *Comput. J.* **1964**, *7*, 149.

(16) Nocedal, J.; Wright, S. J. *Numerical Optimization*; Springer: New York, 1999.

(17) Press: W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C++: The art of scientific computing*; Cambridge University Press: 2002.

(18) Polak, E. *Computational Methods in Optimization*; Academic Press: New York, 1971.

(19) Shanno, D. F. *Math. Oper. Res.* **1978**, *3*, 244.

(20) Watowich, S. J.; Meyer, E. S.; Hagstrom, R.; Josephs, R. *J. Comput. Chem.* **1988**, *9*, 650.

(21) Wenzel, W.; Hamacher, K. *Phys. Rev. Lett.* **1999**, 3003.

(22) Merlitz, H.; Wenzel, W. *Chem. Phys. Lett.* **2002**, *362*, 271.

(23) Schlick, T. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Eds.; John Wiley & Sons: 1997.

(24) Jain, A.; Vaidehi, N.; Rodriguez, G. *J. Comput. Phys.* **1993**, *106*, 258.

(25) Vaidehi, N.; Jain, A.; Goddard, W. A. *J. Phys. Chem.* **1996**, *100*, 10508.

(26) Schwieters, C. D.; Clore, M. G. *J. Magn. Reson.* **2001**, *152*, 288.

(27) Kohlbacher, O.; Lenhof, H. P. *Bioinformatics* **2000**, *16*, 815.

(28) Che, J. to be published.

(29) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.

(30) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 520.

(31) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 553.

(32) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 587.

(33) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 616.

(34) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720.

(35) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730.

(36) Cavasotto, C. N.; Abagyan, R. *J. Mol. Biol.* **2004**, *337*, 209.

(37) Dennis, J. E.; Schnabel, R. B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*; Prentice Hall: Englewood Cliffs, NJ, 1983.

(38) Katritch, V.; Totrov, M.; Abagyan, R. *J. Comput. Chem.* **2002**, *24*, 254.

CT0499433