Malaysian Journal of Mathematical Sciences 2(2): 29-39 (2008)

# Predicting Minimum Energy Structure of a Peptide via a Modified Potential Smoothing Kernel

**Leong Wah June and Malik Abu Hassan**
*Institute for Mathematical Research*
*and Department of Mathematics,*
*Universiti Putra Malaysia,*
*UPM 43400 Serdang, Selangor*

## ABSTRACT

A global optimization approach is proposed for finding the global minimum energy configuration of a peptide. First, the original nonsmooth total potential energy function of a peptide, composed using the AMBER model, is transformed to a smoother function (shifted-impulsive transformation) via a procedure performed for each pair potential that constitute the total potential energy function. Then, the Potential Smoothing and Search (PSS) procedure is used to provide the global minimum. Based on this procedure global optimum solution is generated for a synthesis peptide named Compstatin.

## INTRODUCTION

Global optimization of molecular structure is one of the classic problems of theoretical chemistry and biophysics. The problem can be stated simply: for a given molecule, what is the conformation with the lowest potential energy? Although simple to state, this problem is quite difficult to solve in a general way. Conformational space grows exponentially with molecular size, making systematic searching difficult for all but the smallest problems. Moreover, the number of local minima usually grows rapidly as well, with the result that straight-forward optimization methods tend to get trapped in local minima and fail to find the global minimum. A variety of approaches exist to deal with this difficulty. One of the most common is simulated annealing, which uses either Monte Carlo or molecular dynamics to sample conformational space at high temperatures, where large energy barriers are easily crossed [1]. By gradually lowering the temperature, the conformational sampling can be focused on lower energy regions of phase space.

In principle, as the temperature approaches 0K, only the global minimum energy structure remains. In practice, simulated annealing is not deterministic, because of the prohibitive computational cost of a sufficiently long trajectory. Rather, even long trajectories have a finite chance of finding

several different minima, depending on the topology and roughness of the potential energy surface. Thus, simulated annealing can be viewed as a rescaling of the energy surface; high temperatures effectively scale the energies and make the surface flatter and thus easier to search, at the cost of diminishing the favorability of the low energy states. However, the molecular dynamics and Monte Carlo methods used with simulated annealing generate a Boltzmann distribution at each temperature, rather than proceeding directly toward a minimum. Combined with the exponential dependence of the barrier-crossing time on barrier height and the exponential growth in the number of barriers with system size, this makes simulated annealing inefficient for global optimization.

Potential smoothing methods take an alternative approach to the problem. If global optimization is difficult because the potential surface has many local minima and large barriers, why not simply alter the potential energy surface to eliminate them? Smoothing methods transform the potential surface in such a way that the number of minima is greatly reduced and global optimization is easy. The degree of surface deformation is then gradually decreased, such that the global minimum on the deformed surface can be tracked back to the original undeformed surface. There are several distinct methods which proceed along these lines, including the diffusion equation method [2; 5], Gaussian density annealing [3; 7], and Gaussian packet annealing [6]. In all of these methods, discrete atoms are replaced by Gaussian distributions. The interactions between the distributions result in a smoother potential surface with fewer minima.

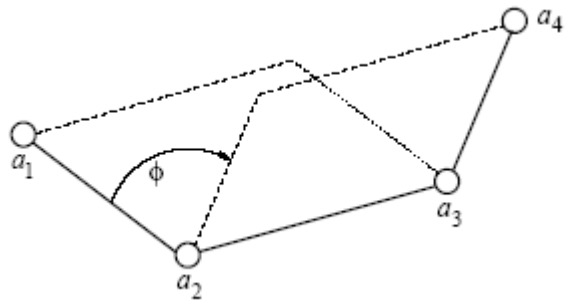
## Potential Energy Model

Molecular structure calculation is based on repeated calculation of a potential function, called force field, which depends on coordinates of the atoms of the molecule. A typical force field can be expressed by the following equation:

$$H = \sum_{bonds} S_k (r - r_0)^2 + \sum_{bond\ angle} K_\theta (\theta - \theta_0)^2 + \sum_{torsions/dihedrals} T_n [1 + \cos(n\phi - \gamma)]$$

$$+ \sum_{nonbonded\ ij-pair} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} + D \frac{q_i q_j}{r_{ij}} \right]. \qquad (1)$$

Here $r$ is the length of a bond, $\theta$ is the angle of a pair of consecutive bonds, $\phi$ is the value of a torsion, and $r_{ij}$ is the distance between the i-th and the j-th nonbonded atoms. A torsion is defined for every set of four atoms $\{a_1, a_2, a_3, a_4\}$ such that $a_k$ is bonded to $a_{k+1}$ for $k = 1,2,3$. The value of the torsion is equal to the angle between the two planes defined by coordinates of $\{a_1, a_2, a_3\}$ and $\{a_2, a_3, a_4\}$:




The symbols $r_0$ (equilibrium length of a bond), $\theta_0$ (equilibrium value of a bond angle), $S_k, K_\theta, T_n, A_{ij}$ and $C_{ij}$ in (1) are constants, which are specific to the types (orbital configurations) of atoms taking part in each particular interaction (types of atoms do not change during computation). $q$ is the value of electrostatic charge on an atom, which also does not change during the computation. $D$ represents electrostatic properties of the environment (solvent). Most frequently, $D = 1$, when solvent is explicitly included, or $D = cr_{ij}$, where $c \geq 1$, when solvent is not present.

**Impulsive Smoothing**

Many smoothing methods operate by replacing the potential (force field) function with an appropriately weighted average value. Mathematically, the smoothed function is defined via an integral of the potential function, multiplied by a suitable weighting function or kernel, $k(x)$:

$$\langle f \rangle_\lambda (y) = \frac{\int f(x) k(x, y, \lambda) dx}{\int k(x, y, \lambda) dx}. \qquad (2)$$

In this notation, $x$ is a $(3N-6)-$dimensional vector (for an $N$ atom system) representing all degrees of freedom for the system, the integral is over all phase space, and $\langle f \rangle_\lambda$ is the average of $f$, subject to weighting by $k$.

The most common choice of smoothing kernel is a Gaussian if width $\lambda$ centered at $y$.

$$k(x, y, \lambda) = \exp\left(\frac{-(x-y)^2}{\lambda}\right), \tag{3}$$

The degree of smoothing is controlled by varying $\lambda$; larger values of $\lambda$ result in more averaging, produce a smoother $\langle f \rangle_\lambda$, while $\lim_{\lambda \to 0} \langle f \rangle_\lambda = f$.
An alternative choice is an impulsive function. Specifically,

$$k(x, y, \lambda) = \begin{cases} 1, \text{ for } |x-y| \le \lambda \\ 0, \quad \text{otherwise} \end{cases}, \tag{4}$$

which select out a (hyper)sphere of radius $\lambda$ surrounding $y$. As above, each pairwise potential term will depend on three variables, allowing us to neglect the other dimensions in the integral. Substituting the kernel back into (2) and assuming $\lambda \le |y|$ gives

$$\langle f \rangle_\lambda (y) = \frac{3}{4\pi\lambda^3} \int_{y-\lambda}^{y+\lambda} dr \int_0^{os^{-1}\left(\frac{y^2+r^2-\lambda}{2yr}\right)} d\theta \int_0^{2\pi} f(r) r^2 \sin\theta \, d\phi \,. \tag{5}$$

Performing the integrals over $\theta$ and $\phi$ yields

$$\langle f \rangle_\lambda (y) = \frac{3}{2\lambda^3} \int_{y-\lambda}^{y+\lambda} r\left(\frac{\lambda^2 - y^2 - r^2}{2yr} + 1\right) f(r) dr \,. \tag{6}$$

If $\lambda > |y|$, the region of integration contains the origin, and the boundaries on the integral are different:

$$\langle f \rangle_\lambda (y) = \frac{3}{2\lambda^3}\left[ \int_0^{\lambda - y} 2r^2 f(r)dr + \int_{y-\lambda}^{y+\lambda} r\left( \frac{\lambda^2 - y^2 - r^2}{2yr} + 1 \right) f(r)dr \right]. \quad (7)$$

As with Gaussian smoothing, $f$ must diverge no faster than $r^{-1}$ at the origin.

**Shifted-Impulsive Smoothing**

To overcome the disadvantage of Gaussian smoothing as well as impulsive smoothing, we introduce the shifted-impulsive smoothing. The shifted-impulsive (*simp*) smoothing method is derived in much the same way as the impulsive smoothing, but rather than selecting a sphere centered at $y$ as in (4), we center it at $y + \lambda z$. Instead of (5), we obtain for all $\lambda$

$$\begin{aligned} \langle\langle f \rangle\rangle_\lambda (y) &= \frac{3}{4\pi\lambda^3} \int_y^{y+2\lambda} dr \int_0^{os^{-1}\left( \frac{(y+\lambda)^2 + r^2 - \lambda}{2(y+\lambda)r} \right)} d\theta \int_0^{2\pi} f(r) r^2 \sin\theta \, d\phi \\ &= \frac{3}{4\lambda^3} \int_y^{y+2\lambda} -r\left( \frac{y^2 + 2y\lambda + r^2 - 2r(y+\lambda)}{y+\lambda} \right) f(r)dr . \end{aligned} \quad (8)$$

Unlike Gaussian and impulsive smoothing, *simp* remains well-defined even for functions which diverge rapidly as $r \to 0$. This is because the integral contains no contribution from the potential function for $r < |y|$.

*Simp* smoothing is not strictly an average of the total potential, and cannot be derived from (2). While each individual energy term is averaged according to a impulsive kernel, the locus of integration varies for different atom pairs.

**Applications to Force Field Terms**

*Simp* smoothing is directly applied to the terms of standard force fields simply by choosing $f$ accordingly and solving the integral in (8).

For example, the *simp* smoothed Coulomb function $\frac{1}{r}$ is calculated as

$$\langle\langle f \rangle\rangle_{\lambda}(y) = \frac{3}{4\lambda^3}\int_{y}^{y+2\lambda} -r\left(\frac{y^2+2y\lambda+r^2-2r(y+\lambda)}{y+\lambda}\right)\frac{1}{r}dr$$
$$= \frac{1}{y+\lambda}.$$

TABLE 1: The *simp* smoothed force field terms.

| Forcefield Descriptions | Original Forcefield Terms | Transformed Forcefield Terms |
|---|---|---|
| Bonded terms | $(r-r_0)^2$ | Not necessary |
| Electrostatics | $\frac{1}{r}$ | $\frac{1}{r+\lambda}$ |
| Lennard-Jones | $\left(\frac{\sigma}{r}\right)^{12}-2\left(\frac{\sigma}{r}\right)^{6}$ | $\frac{\sigma^{12}\left[\begin{array}{l}15x^6+90\lambda x^5+288\lambda^2x^4\\+522\lambda^3x^3+648\lambda^4x^2\\+432\lambda^5x+128\lambda^6\end{array}\right]}{15x^9(x+2\lambda)^9}-2\sigma^6\left[\frac{1}{x^3(x+2\lambda)^3}\right]$ |
| Dihedrals/ Torsion | $\cos(n\phi-\gamma)$ | $\cos(n\phi-\gamma)\frac{\sin(n\lambda_\tau)}{n\lambda_\tau}$, $\lambda_\tau$ is the angular smoothing parameter in radian. |

## Potential Smoothing and Search Procedure (PSS)

The potential smoothing and search (PSS) protocol used here is essentially that of Pappu et al. [4], except the *simp* smoothing is used. So, we will merely give the algorithm here:

## PSS-SIMP ALGORITHM

**Phase 0**: Input biopolymer sequences

**Phase I**: Generation of Initial Configuration

a) Local minimization on an extended conformer using the smoothed potential function.
b) Local minimization on the output of step I-a, using the original energy function.

**Phase II**. Improvement of Local Minimizars

a) Select a subset of the best sample points from step I-b to be start points for local minimizations.
b) Perform a local minimization using the transformed functions, from each starting point selected in step II-a by using the backward smoothing scheme.
c) Collect some number of the best of these minimizers for improvement in step II-d.
d) Apply a global optimization algorithm (annealing simulated algorithm) to the original energy of the selected configuration using the selected small subset of minimizers in step II-c as starting point.

**Post-processing phase**: Merge the new lowest configuration into the existing list of local minimizers, then choose the lowest among them as the final configuration.

### Computational Results – Compstatin Peptide

Compstatin is a synthetic 14-residue ICVVQDWGHHRCTX cyclic peptide. The solution structure was previously identified using two-dimensional NMR techniques (accession number 1A1P at the RCSB Protein Data Bank (PDB), http://www.rcsb.org/pdb/) as well as the average Compstatin conformation. Here, we attempt to find the minimum energy structure of the Compstatin. A full global minimization of the Compstatin structure is performed using 2 procedures – annealing simulated cooling (ASC) protocol and PSS-SIMP algorithm outlined in the Section 6.

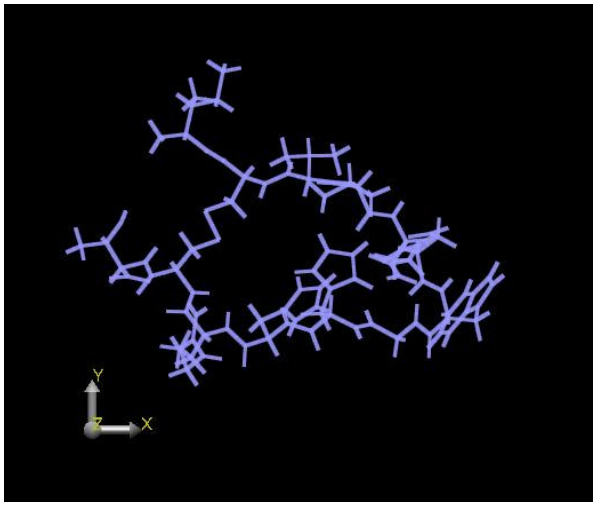



Figure 1: Direct transformation of PDB to AMBER[8] structure with energy 147.6129 kcal/mol)

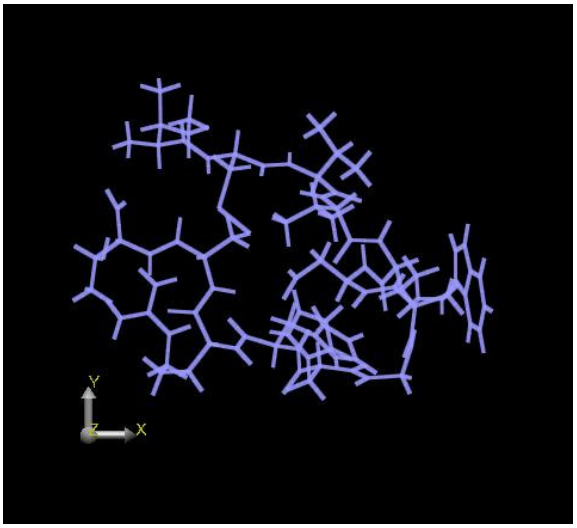



Figure 2: Minimum energy structure (-320.9031 kcal/mol) subjected to ASC protocol.

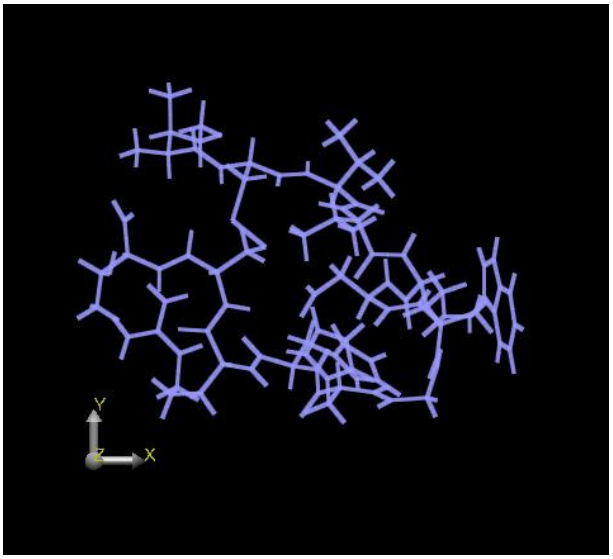



Figure 3:Minimum energy structure (-362.2145 kcal/mol) subjected to PSS-SIMP.


Figure 4: Backbone structure of Compstatin-ASC (left) VS Compstatin-PSS-SIMP (right)

The PSS-SIMP minimum energy structure satisfying distance constraint and dihedral angle bounds provided an AMBER99 energy of -362.2145 kcal/mol, which is lower in energy than the minimum structure of annealing simulated cooling protocol (-320.9031 kcal/mol).

## CONCLUSIONS

The solution technique described in this paper represents another enhancement over existing methods. The PSS-SIMP algorithm, based on the *simp* smoothing approach, has already been successfully applied to the identification of global minimum energy structures of peptides modeled by full-atom force fields. The application of this technique to the Compstatin structure prediction problem emphasizes the merits of the approach. The globally predicted structure using PSS-SIMP exhibits an improvement energy, which indicates better definition of structural details, in contrast, results obtained from ASC protocol.

## REFERENCES

[1] Kirkpatrick S., C. D. Gelatt, and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science*, **220**, 4598:671-680.

[2] J. Kostrowicki, and H. A. Scheraga. 1992. Application of the diffusion equation method for global optimization to oligopeptides, *Journal of Physical Chemistry*, **96**:7442-7449.

[3] J. Ma and J. E. Straub. 1994. Simulated annealing using the classical density distribution. *Journal of Chemical Physics*, **101**:533-541.

[4] R. Pappu; R. Hart; J. Ponder. 1998. Analysis and Application of Potential Energy Smoothing for Global Optimization. *Journal of Physical Chemistry*, **102**:25-9742.

[5] L. Piela, J. Kostrowicki, and H. A. Scheraga. 1989. The MultipleMinima Problem in the Conformational Analysis of Molecules: Deformation of the Protein Energy Hypersurface by the Diffusion Equation Method. *Journal of Physical Chemistry,* **93**:3339-3346.

[6] D. Shalloway, in C. A. Floudas, P. M. Pardalos, eds.,1992. *Resent Advances in Global Optimization*, Princeton University Press, 433-477

[7] J. E. Straub, J. Ma and P. Amara. 1995. Simulated annealing using coarse grained classical dynamics: Smoluchowski dynamics in the Gaussian density approximation. *Journal of Chemical Physics* **103**:1574-1581.

[8] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, G. Alagona, S. Profeta Jr. and P. Weiner. 1984. A new force field for the molecular mechanical simulation of nucleic acids and proteins. *Journal of American Chemistry Society*, **106**:765-784.