

# Scaling Behaviour of Minimization Algorithms in a Simple Model for Protein Folding

K. Hamacher and W. Wenzel

*Theoretische Physik I, Universität Dortmund, 44221 Dortmund, Germany*

*E-mail: wenzel@wap.physik.uni-dortmund.de*

We investigated the scaling behaviour of various minimization algorithms in a simple model, which exhibits key characteristics of the protein folding problem. For Monte-Carlo with minimization and a novel stochastic tunnelling method we established that the computational effort for the first passage time of the global minimum grows with approximately the eighth power of the system size. We further provide explicit evidence that this behaviour is matched in the growth laws of a derived lattice model, which can be studied at significantly reduced effort.

The large number of sequenced proteins of unknown structure motivates the search for theoretical methods to predict the native structure of proteins from their amino acid sequence alone<sup>1,2</sup>. In one widely pursued approach the native structure of the protein is sought as the global minimum of a suitable potential/free energy surface (PES)<sup>3</sup> of the molecule. Even if one assumes that appropriate models for the PES in the presence of solvent effects exist, the resulting multi-minima problem is among the most challenging computational problems for realistic PES. Its “trivial” solution, i.e. the direct simulation of the folding process, is far beyond the present-day computational resources for all-atom models of large proteins.

It has been argued that the underlying minimization problem is NP-hard<sup>4,5</sup> implying an exponentially growing numerical effort of the enumerative search for the true native state. For this reason most presently studied methods sacrifice the certainty of finding the native state for the faster convergence of a probabilistic solution, often using thermodynamically motivated stochastic minimization techniques, such as simulated annealing<sup>6</sup> and its derivatives<sup>7,8,9,10</sup>. It was argued that the underlying structure of the energy surface of polypeptides<sup>4,11</sup> exhibits a hierarchical funnel-like structure<sup>12</sup>. While only the presence of such a structure allows stochastic minimization methods to succeed in polynomial time, little is known about the dependence of the computational effort on the size of the molecule. The determination of such growth laws can serve as a strong criterion for the selection of promising algorithmic avenues and provide order-of-magnitude estimates for the overall computational effort for larger proteins than are presently accessible.

As a first step in this direction we have determined the scaling laws of the computational effort  $n_{\text{CPU}}$  with the system size ( $N$ ):

$$n_{\text{CPU}} \sim N^\alpha \quad \text{or} \quad n_{\text{CPU}} \sim \exp(\gamma \cdot N). \quad (1)$$

for a simple model that incorporates some key characteristics of the protein structure problem. For Monte-Carlo-with-minimization(MCM)<sup>13</sup> and a novel stochastic tunnelling method(STUN), we found that the overall computational effort grows with approximately the eighth power of the system size, while a naive multi-start-algorithm<sup>14</sup> yielded the expected exponential dependence. These results indicate a

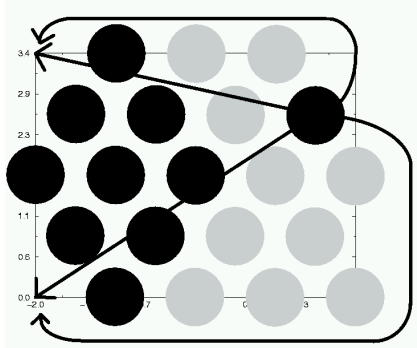


Figure 1: Typical metastable configuration of the model with one particle not in an optimal position. The arrows indicate possible demixing paths to reach the global minimum of the PES — which must either cross many sterically forbidden regions or diffuse long distances in the absence of a guiding force.

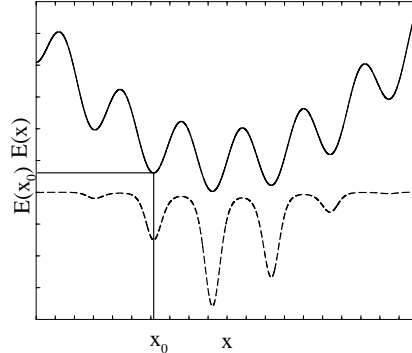


Figure 2: Schematic one-dimensional PES (full line) and its STUN effective potential (dashed line), where the indicated minimum  $E(\vec{x}_0)$  is used as the reference. All energies ranging from the best present estimate to infinity are mapped to the interval  $[0, 1]$ , while all lower energies are exponentially enhanced.

natural resolution of the Levinthal paradox in the presence of a hierarchical PES. We were further able to demonstrate that the Monte-Carlo results of a derived lattice model are in agreement with those of the continuum model, providing explicit evidence for equivalence of the two problems.

## 1 Model and Methods

In order to investigate sufficiently large systems to reliably determine the scaling behaviour, we chose a simple model that reflects only the most crucial characteristics of the original problem: (i) The interactions relevant to the folding process fall into two classes of significantly different strength: relatively small energies separate the various local minima (hydrogen bonding, solvent interactions), which are separated by large energy barriers (steric repulsion, covalent bonding). (ii) Recent work on the nature of the PES<sup>2</sup> supports the view of a funnel-like structure of the PES, in which a multitude of pathways of on-average decreasing energy lead to the native state.

Both these characteristics are replicated in the PES of an abstract model, consisting of a collection of  $N$  disks of two different types (“black” and “white”) interacting with a Lennard-Jones-potential of two different strength:

$$V_{\text{LJ}}(r) = A \cdot \left( \frac{1}{r^{12}} - \frac{2}{r^6} \right) \quad A = \begin{cases} 1: & \text{different colour} \\ 2: & \text{same colour} \end{cases} \quad (2)$$

The local minima of this NP-hard problem are characterized as slight distortions of a triangular lattice (Figure 1). The global minimum is reached by a demixing of the two types of particles into two adjacent clusters of particles of the same type.

Exploiting the one-to-one mapping of the local minima and arrangements of the disks on a triangular lattice, a natural lattice derivative of the continuum model can be defined. The interactions of the lattice model depend only on the type of the nearest neighbour bonds.

As one of the most straightforward techniques we have investigated Monte-Carlo-with-Minimization-method(MCM)<sup>13</sup>. This method differs from standard Monte-Carlo in that the newly generated configuration is relaxed locally before the application of the Metropolis criterion. Thus only the energies of local minima are compared in the simulation, which can proceed at a sufficiently low temperature to resolve the small energy differences between them. Attempts to investigate standard Monte-Carlo, simulated annealing and their recent generalizations<sup>7,8</sup> had to be abandoned since the CPU requirements for the study of sufficiently large systems with adequate statistics proved too large ( $\gg 10^4$  h).

Secondly we studied a novel stochastic tunnelling technique(STUN), which successively eliminates the high-energy regions of the PES: we applied a nonlinear transformation, which preserves the locations of the extrema of the original potential:

$$\tilde{E}(\vec{x}) = 1 - \exp(-\gamma[E(\vec{x}) - E(\vec{x}_o)]), \quad (3)$$

where  $\vec{x}_0$  designates the position of the best local minimum encountered so far. As illustrated in Figure 2, this transformation maps the entire energy region above  $E(\vec{x}_0)$  into the interval  $[0, 1]$ , while exponentially stretching the energy range below. Thus the particles can be said to “tunnel”<sup>15,16</sup> through the forbidden regions of the PES. To check, whether the range of systems studied was sufficiently large to differentiate exponential and power-law growth, we studied the multi-start method(MS), where a sequence of independent random initial conditions is subject to local minimization. This method is expected to scale exponentially for a NP-hard minimization problem.

## 2 Results and Discussion

For all methods we have optimized the parameters for system sizes  $N \leq 16$ . The optimal parameters were found to be robust and could be extrapolated to large systems, where a full parameter optimization proved too costly. For each optimal parameter set and system size we performed 150 – 500 independent runs to obtain sufficiently good statistics for the first-passage “time” of the optimal conformation. The details of the parameter optimization and data analysis will be published elsewhere<sup>17</sup>.

As illustrated in Figure 3 we find power-law growth of the computational effort for MCM and STUN, with exponents  $\alpha_{STUN} = 7.6(\pm 1.8)$  and  $\alpha_{MCM} = 6.4(\pm 1.5)$  for the continuum model. The slight curvature of the MCM results for larger system size is explained by the saturation of the computational effort of the local minimizer (see inset). For the lattice model we find  $\alpha_{MC} = 4.7(\pm 1.6)$  for the standard Monte-Carlo algorithm(MC), which is the natural equivalent of MCM on the lattice. When the linear increase of the computational cost of the local minimizer is taken into account, this exponent is consistent with MCM on the lattice. The exponential

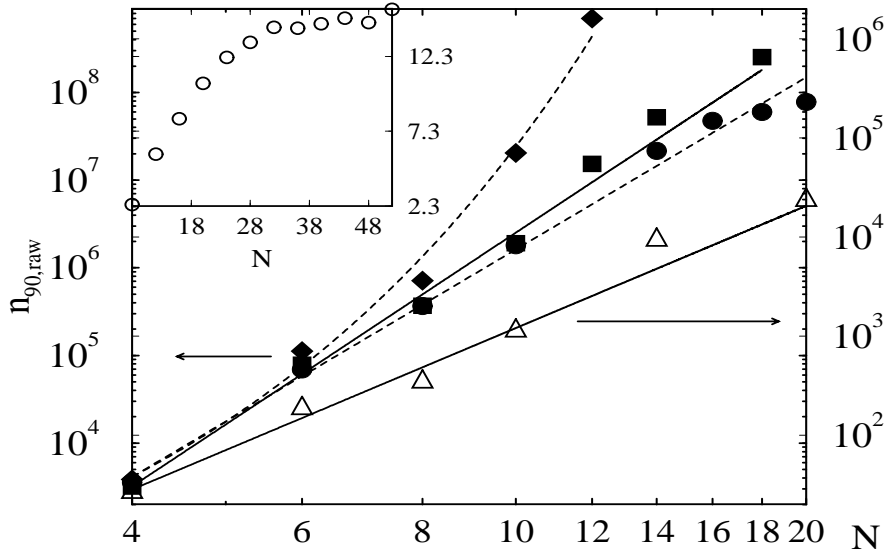


Figure 3: The scaling behaviour of the average number of function evaluations required for first passage of the optimal configuration with 90% probability as a function of system size for MS(diamonds) STUN(squares), MCM(circles) in the continuum (all left scale) and Monte-Carlo(triangles) on the lattice. The lines indicate exponential and power-law fits for MS and MCM/STUN/MC respectively. The inset shows the average number of function evaluations (in thousands) to reach a metastable state with the conjugate gradient minimizer as a function of system size.

growth of the numerical effort of the MS method is clearly distinguishable from the power-law growth observed for the other methods.

These results demonstrate that stochastically motivated minimization methods overcome the exponential complexity of the PES, provided that the PES has a hierarchical structure in which many pathways of on average decreasing energy lead to the globally stable configuration. In the presence of such a funnel like structure no specifically “designed” or evolved folding pathway is required to reach the global minimum in polynomial time. We note that the dynamic processes investigated here avoid sterically forbidden regions altogether, but these regions are also not visited in the folding process in nature, because the structural integrity of the protein is preserved in the folding process. There are, however, kinetic barriers on the small energy scale and the results of STUN indicate that the presence and height of such barriers does not affect the scaling behaviour of the algorithm.

In the context of the Levinthal paradox our observations indicate that an enumerative, exponentially expensive, search of all locally stable conformations of the protein can be avoided in the presence of a structured PES, which has been postulated on the other grounds for the folding pathway of naturally occurring proteins.

The “failure” of the MS algorithm illustrates that it is the correlation of successive configurations in MCM/STUN, which allows the exploitation of the structural features of the PES. It would be desirable to investigate more realistic models and a wider variety of models to determine whether there are classes of algorithms which differ in their scaling exponent. It is therefore encouraging that our results offer an explicit confirmation that the scaling behaviour of the continuum systems is consistent with the behaviour of the derived lattice model, while the numerical effort of the treatment of the latter is reduced by four orders of magnitude.

1. H. A. Scheraga. *Chemica Scripta*, 29A:3–13, 1989.
2. D. Shalloway P. M. Pardalos and G. Xue, editors. *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding*, volume 23 of *DIMACS*, Princeton, 1995.
3. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. *J. Comp. Chem.*, 4(2):187–217, 1983.
4. J. T. Ngo and J. Marks. *Protein Engineering*, 5(4):313–321, 1992.
5. L.T. Wille and J. Vennik. *J. Phys. A: Math. Gen.*, 18:L419–L422, 1985.
6. S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. *Science*, 220:671–680, 1983.
7. B. Berg, *Int. J. Mod. Phys. C***3**, 1083 (1992).
8. E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
9. U. Hansmann, this volume
10. P. Derreumaux, this volume
11. R. L. Baldwin. *Nature*, 369:183–184, 1994.
12. E. Shakhnovich A. Šali and M. Karplus. *Nature*, 369:248–251, 1994.
13. Z. Li and H.A. Scheraga. *Proc. Nat. Acad. Sci. U.S.A.*, 84:6611, 1987.
14. A. Törn and A. Žilinskas. *Global Optimization*. Number 350 in *Lecture Notes in Computer Science*. Springer Verlag, Berlin, 1989.
15. A. V. Levy and A. Montalvo. *SIAM J. Sci. Stat. Comput.* **6**, 15-29 (1985)
16. B. C. Cetin and J. Barhen and J. W. Burdick. *J. Optim. Theory Appl.***77** 97-126 (1993)
17. K. Hamacher and W. Wenzel, in preparation
18. L. Piela, J. Kostrowicki, and H. A. Scheraga. *J. Phys. Chem.*, 93:3339–3346, 1989.