Homology as a Tool in Optimization Problems: Structure Determination of 2D Heteropolymers

Chen Keasar and Ron Elber*

Department of Physical Chemistry, the Fritz Haber Research Center for Molecular Dynamics and the Israel Center for Structural Biology, Hebrew University, Givat Ram, Jerusalem 91904, Israel

Received: March 22, 1995[⊗]

A combination of two complementary approaches that are widely used to predict protein structure—homology and energy-based methods—is proposed. The algorithm is discussed in detail and is demonstrated to be a useful tool for optimization. The properties of the combination are exemplified on a simple model system: structural optimization of 2D heteropolymers on a square lattice. The algorithm is considerably more efficient than simulated annealing in the system studied. The relevance of the results of the protein folding problem is discussed.

Introduction

Most of the methods that predict protein structures from sequences belong to one of two broad categories: global optimization of an energy function or homology-based modeling. Both approaches are of considerable applicability and are complementary. Here we propose a combination of the two that is likely to benefit from the useful properties of both. The properties of the combination are exemplified on a simple model system: structural optimization of 2D heteropolymers on a square lattice (2DHP-s).

The global optimization approach to protein structure prediction is based on the assumption that the native structure of proteins is at the global minimum of the free energy. This assumption has been validated for several proteins.[1] The prediction scheme consists of an energy function that maps protein conformations to energy values and an optimization algorithm to find the global minimum of that function. The calculated minimum-energy conformation is considered an approximation to the native fold.

The energy function can be derived from experimental and computational data on small molecules[2−4] or from statistical analysis of known protein structures.[5,6] In both cases the energy surface is likely to be very rough, including a broad distribution of barrier heights and well depths. This poses a significant challenge on the optimization algorithm. The computational effort of exhaustive search requires time that grows exponentially with the number of amino acids. More efficient methods such as Monte Carlo (MC) annealing[7] and molecular dynamics annealing (MD)[8] tend to find and to get "stuck" at local minima.

A possible strategy to bypass this multiple minima problem, which is an important topic in the field of global optimization, is to modify the energy function. The energy function is modified in a way that will smooth the energy surface while still keeping the original location of the global minimum, or at least a reasonably good approximation to it. Smoothing approaches include methods like the diffusion equation method,[9] the Liouville equation approach,[10] imaginary time Schrödinger equation,[11,12] and the locally enhanced sampling (LES) method.[13]

Most optimization algorithms lack an internal test for the global energy minimum. At the end of the optimization process one cannot be certain that a deeper energy well does not exist. The simulation should be repeated a number of times, searching for alternative and possibly better minima. Convergence is assumed when the same minimum is repeatedly found regardless of the initial conditions. The LES algorithm[13] is an exception, since it provides an internal consistency check. The check in LES is a necessary but insufficient condition.

Another approach to predict protein conformations is structure building by homology. In spite of many documented successes it is still not applicable in many cases. The prediction scheme is based on the observation that proteins with similar sequences (homologous proteins) have the same approximate native fold. The relation between sequence and structural similarities can be quantified statistically based on known structures and sequences.[14] The working hypothesis is that the similarity of two or more sequences suggests considerable resemblance in structures. This assumption is employed in two ways. The most common application is to use an experimentally determined structure of a protein as a template in the modeling of other members of its family.[15] However, the applicability of this approach is greatly limited by the availability of known homologous structures.

Even when there is no structural data on a family of homologous proteins, insight may be gained by comparing homologous sequences. Conservation patterns may provide hints on the burial state of the amino acid residues,[16,17] on secondary structure elements,[18−21] and even on contacts between residues that are distant in sequence.[22,24] A very similar approach is used in the prediction of secondary structure and contacts in RNA.[25] Recently, it was suggested[26] that homology considerations may be useful in mean field conformational searches in which different side chains (corresponding to homologous proteins) are attached to the same protein backbone.

The current work is an attempt to bridge the above two approaches—homology and energy optimization—in a different and more general way. We suggest that the basic observation of the homology approach, the structural similarity of homologous proteins, can be used to smooth the energy function. In the modified energy function we enhance the difference between the global energy minimum and other local minima. At the same time the heights of the barriers are reduced as compared to the well depths.

Homologous proteins do not share the same energy function, and therefore they do not share all local minima. However, they do share a common fold which is the global energy minimum. Let the number of homologous proteins be N. The energies at the native structures are always negative (attractive).

[⊗] Abstract published in *Advance ACS Abstracts*, June 15, 1995.

0022-3654/95/2099-11550$09.00/0

The sum of the individual energies is therefore an extensive property (proportional to N in the limit of large N). On the other hand, at *an unfolded conformation* the energies of the homologous proteins can be positive or negative due (for example) to different van der Waals contacts. Roughly, the sum of the last energies is expected to be proportional to the fluctuations. Therefore, the sum of unfolded energies has a weaker dependence on the number of homologous proteins we use. In the limit of large N it is proportional to $N^{1/2}$.

We define "coupling" of the energy functions of separate proteins by an additional "energy" term that penalizes homologous proteins when they adopt structures that differ appreciably. By "coupling" two or more homologous energy functions that are similar near the global energy minimum but are different elsewhere, we expect to get a smoother energy function. In this function local minima that are not shared by all functions will tend to cancel. In practice, the coupling between the homologous energy functions can be done by performing the search for the global optimum on all of the protein series in parallel.

In order to demonstrate the feasibility of this approach, we consider a simple model system: a heteropolymer on a 2D square lattice (2DHP). Similar systems were used in the past to investigate guiding principles of protein thermodynamics and kinetics.[27-29] Remote as they are from the complexity of real proteins, 2DHP-s still have global minima that are not trivial to find. As a test case for a new optimization procedure, this system has a number of important advantages compared to real proteins: First, the global energy minimum can be unequivocally found by exhaustive search. Second, there is no problem in the accuracy of the energy function, since no attempt to model "real systems" is made. Third, sequence alignment is easy. Finally, an extensive investigation can be pursued with only modest CPU requirements. Thus, while our model does not have much biophysical significance by its own right, it does allow a rigorous test of the optimization scheme.

The resulting energy function, which is a sum of the individual energies of the 2DHP-s and of the coupling term, is (of course) more complicated than the uncoupled, individual energies of the homologous 2DHP-s. Nevertheless, as we shall show below, the sum has a smoother surface and is more convenient for global optimization. Furthermore, the coupled system shows stronger tendency for cooperativity and first-order phase transition at folding as compared to the uncoupled system. Cooperativity is an essential requirement from good folders.[30] The suggested method also contains an internal test for being at the global energy minimum, similar to the test used by the LES method.[13]

Unlike local energy minima that are not the same in different 2DHP-s, the global energy minimum is shared by both the system of the coupled homologous energy functions and each of the individual functions. If, at the end of the optimization, all of the functions are in the radius of convergence of the same minimum (which can be easily tested by a short minimization), then the self-consistency check is satisfied, and it is more likely that the computations end at the global energy minimum.

Being a feasibility test, the current work concentrates on comparing the suggested method with standard Monte Carlo (MC)[7] annealing. For simplicity, we choose to work on pairs of "homologous" 2DHP-s rather than on larger sets.

Methods

Our two-dimensional heteropolymers (2DHP-s) are defined as linear, self-avoiding chains (Figure 1). The monomers of the chains occupy successive points on a two-dimensional square lattice. Each monomer belongs to one of four types to be discussed below. Depending on the conformation, a monomer may be in contact with other nonconsecutive monomers. The contacts are associated with energy terms, and the sum of all these terms is the intrinsic energy of the 2DHP in that conformation. All the 2DHP-s of the same length have the same conformation space but differ in sequence. As a result, different 2DHP-s may have different mappings of conformations to energies. The following paragraphs define the monomer types, contacts, energy terms, conformation space, and distances between conformations.

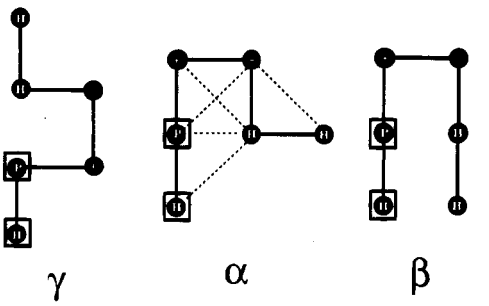

Figure 1. Three conformations of a 6-mers 2DHP. Solid lines represent interresidue bonds, and dashed lines represent interresidue contacts. The first two residues (boxed) are fixed to prevent rigid body rotations and translations. α and β are neighbor conformations ($d_{\alpha\beta} = 1$) since one 90° rotation is needed to move from α to β. α and γ are also neighbors since one 90° rotation is needed in order to move from γ to α′, the mirror image of α.

TABLE 1[a]

contact types	contact energies
(H,H)	−1
(+,−)|(P,H/+/−)	0
(H,+/−)|(+/+)|(−,−)	2

[a] Each contact is associated with a contact energy. A contact type is denoted by a pair that is enclosed in brackets. The energies are in arbitrary units.

The Model of the 2D Heteropolymers. (i) The conformation of a 2DHP of length l is uniquely defined by the set of $l - 2$ angles between intermonomer bonds, which may be 90°, −90°, or 180°. (ii) A conformation α is considered identical to its mirror image α′. Rigid body rotation and translation are prevented by fixing the first and second monomers to the lattice points (0,0) and (0,1), respectively. (iii) The distance $d_{\alpha\beta}$ between two conformations α and β is defined as the minimal number of 90° rotations of intermonomer bonds needed in order to move from α to β or to its mirror image β′. Two conformations α and β are considered neighbors if $d_{\alpha\beta} = 1$. (iv) Two nonconsecutive monomers are considered to be in contact if they are separated by less than two lattice spacings. Thus, any terminal monomer may have at most seven contacts while nonterminal monomers may have six. Every contact is associated with a contact energy term.

Monomer Types and Interactions. Four monomer types were used in this work. Their names mimic specific groups of amino acids: "hydrophobic" (H), "polar" (P), and "charged" (+ or −). The energy values associated with the intermonomer contacts were also chosen to have the flavor of interactions in proteins (Table 1). "Hydrophobic (H) monomers tend to cluster together, "charged" (+ or −) monomers tend to disperse, and "polar" (P) monomers possess intermediate properties. By choosing these names and interactions, we gain some intuitive feelings and expectations. However, the values of the energy terms are rather arbitrary and should not be considered quantitative estimates of the interactions in real proteins.

The intrinsic energy of a conformation α is the sum of all the contact energies.

The definitions of contacts and monomer types are somewhat different than those used by other authors on 2D lattices (for example, contact separation of only one lattice spacing and only two monomer types[28]). Our definitions were shown in a preliminary work to allow for a larger number of "homologous" sequences that have quite different energy functions, while still sharing the same global energy minimum. This is in contrast to monomer interactions that include only H and P. These definitions also result in a rough energy surface. Searches for the global energy minimum on the rough energy surface are more difficult to pursue. While being more suitable for our purposes, we believe that our definitions carry some features of the protein folding problem, as do other 2D models.

Foldicity of 2DHP-s. If a 2DHP has a nondegenerate global energy minimum (up to a mirror image), it is said to fold to that conformation, and the structure is called the native conformation for that 2DHP.

Two 2DHP-s with different sequences that fold to the same conformation are considered homologous.

Folding Sequences. In the current work 14-mers 2DHP-s were used, for which 110 186 different conformations are available. Fifty of them have the maximal number of contacts, 23. We called these structures the compact conformations.

By enumerating all possible conformations and all possible sequences, we identified all the sequences that fold to one of the 50 compact conformations. While some compact conformations are not the global energy minimum for any sequence, others are the global energy minimum of up to 513 sequences that may differ by 11 out of the 14 monomers.

Choice of Homologous Sequences. In a pair of homologous 2DHP-s, a site at which the two monomers are not identical is considered a mutation. Earlier investigations showed the H/P and +/− mutations do not have a strong effect on the shape of the energy function; we thus consider them as conservative mutations. The homology score of two sequences was defined as the fraction of identical monomers or conservative mutations.

The sequences used in this work were chosen in the following way: from the set of all the sequences the pairs with minimal homology score were used.

MC Steps. In a Monte Carlo step,[7] a trial conformation is chosen first, and ΔE, the difference between the energies of the current and the trial conformations, is calculated. A weight function, $\delta = e^{-\Delta E/KT}$, and R, a random variable distributed uniformly between 0 and 1, are compared. If $R \leq \delta$, the trial conformation is accepted, becoming the new conformation. Otherwise, the current conformation remains unchanged.

For each of the conformations available to 2DHP-s, a list of neighbors was calculated. Trial conformations were randomly selected from the neighbor list of the current conformation. In the coupled systems (defined below) another type of conformation can be used as a trial—the conformation of the other 2DHP ("jumping"). When "jumps" were introduced, two possibilities for the Monte Carlo step were considered: with a predefined probability (10% in the present study) the conformation of the second 2DHP was used as a trial conformation ("jump"). The other possibility was to use the usual neighbors as trial conformations.

Coupling: Penalty Function That Forces Different Homologous 2DHP-s To Look Alike. Consider two homologues 2DHP-s: A and B. We optimize the structures of A and B simultaneously. The optimization method that we used is simulated annealing (SA)[31] in conjunction with the Monte Carlo method (MC). In the ith MC step, A and B adopt conformations

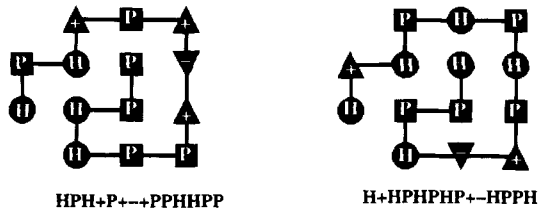

Figure 2. Two homologous 2DHP-s that share the same global minimum with significant variance in sequence.

α_i and β_i, respectively. We penalize differences between α_i and β_i by adding a new "energy" term that is growing with the distance, $d_{\alpha\beta}$, between the two conformations. MC is then employed to optimize the total energy, which is the sum of the intrinsic energies of the 2DHP-s and the coupling term. The specific functional form for the coupling term used in the current work is

$$E_{\text{cpl}} = \begin{cases} 0 & \text{if } d_{\alpha\beta} \leq 1 \\ K_{\text{cpl}}(d_{\alpha\beta} - 1) & \text{if } d_{\alpha\beta} > 1 \end{cases}$$

Simulations. Three types of simulations were performed. The first type of simulation was direct optimization by simulated annealing (SA).[31] In these simulations an increase in optimization efficiency is demonstrated once the coupling is introduced. In the second type of simulation we attempted to explain the origin of the enhancement in computational efficiency. Equilibrium properties of the system were calculated using simulations with constant temperature. Finally, the shape of the energy function was further characterized by computing the attraction radii of local minima in the coupled and the uncoupled cases.

In each of the simulations, three cases were considered: $K_{\text{cpl}} = 0$ (no coupling), $K_{\text{cpl}} = 1$ (coupling), and $K_{\text{cpl}} = 1$ with 10% of the MC trial conformations being the conformation of the other chain (coupling and jumps). We shall describe in detail the results for one conformation (Figure 2). The results of other conformations (we investigated three additional structures) are qualitatively similar and therefore not discussed.

Results

SA Simulations. At the beginning of every SA simulation the two 2DHP-s were assigned the same conformation. The initial conformation was selected at random from the set of conformations with maximal (12) distance from the native conformation. The temperature was set to 1 (arbitrary units). During the simulation the temperature was reduced linearly to 0.

The SA runs were followed by 100 MC steps at zero temperature (minimization) and without coupling. A simulation was considered a success if, at the end of the minimization, the 2DHP was in its minimal energy structure.

Eight sequence pairs with minimal homology score (0.5) were used. For each of these pairs, 100 simulations were performed starting from different random conformations. The success rate plotted against the simulation length (the number of MC steps) is presented in Figure 3. The results show a clear increase in SA efficiency when coupling is used. Further increase in efficiency is obtained by the modified MC step that allows for "jumps".

The additional minimizations can be employed as a self-consistency test for the quality of the annealing. If the two structures were minimized to different conformations, one may conclude that the simulation was not adequate and needs to be repeated. Only in 0.2% or less of the coupled (without jumps) simulations did the two 2DHP-s end the minimization in the same wrong minimum (false positives). However, false

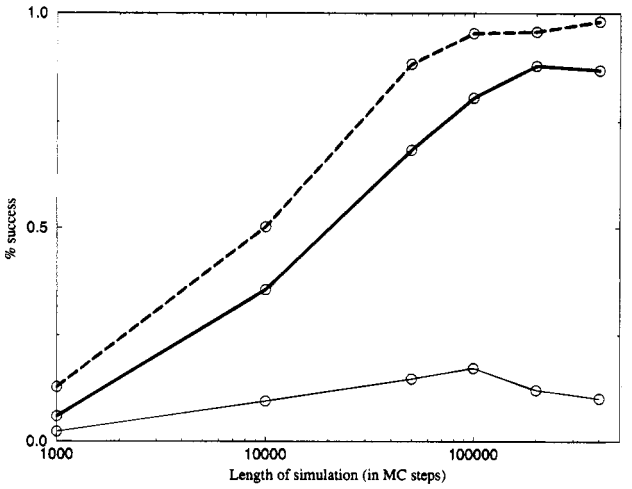

Figure 3. Success rates at different simulation lengths of the three optimization protocols. Every data point is an average of 800 simulations of 16 2DHP-s (8 sequence pairs, 100 simulations with each). Standard deviations are less than 0.1. For clarity error bars were omitted. When "jumps" were introduced, two possibilities for the Monte Carlo step were considered: with a probability of 10% the conformation of the second 2DHP was used as a trial conformation ("jump"). The other possibility was to use the usual neighbors as trial conformations. Thin line, cpl = 0 (no coupling); thick line, cpl = 1 (coupling); dashed line, cpl = 1 (coupling with "jumps").

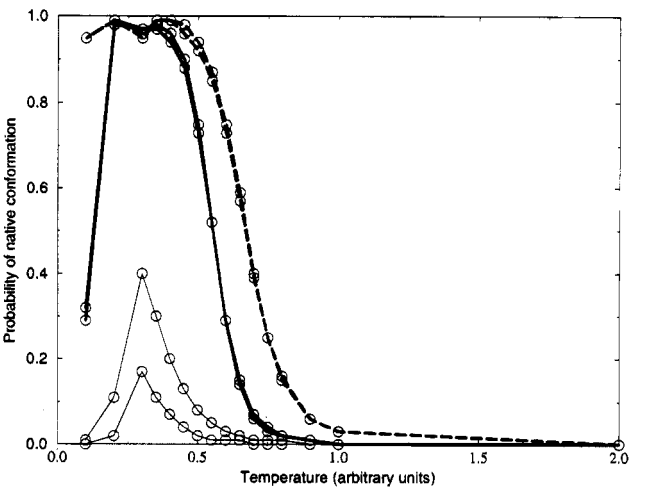

Figure 4. Probability of the 2DHP to be found in its native conformation at different temperatures. Data points are averages over 10 simulations, starting from different random conformations. Most simulations were of 10 000 000 MC steps. In an attempt to reach equilibrium, simulations 10 times longer were performed at temperatures below 3.5. However they are clearly not in equilibrium. Thin line, cpl = 0 (no coupling); thick line, cpl = 1 (coupling); dashed line, cpl = 1 and "jumps". See the legend of Figure 3 for more details.

negatives—when one of the 2DHP-s reached the native conformation while the other did not—occurred in up to 12% of the simulations. When jumps were introduced, false negatives disappeared while the false positive fraction increased to 3% in the shortest simulations. The internal check seems to be quite reliable, especially in long (50 000 MC steps and up) simulations.

Constant-Temperature Simulations. In order to investigate how the coupling between homologous 2DHP-s changes the shape of the energy surface, we pursued computations of equilibrium properties. Long MC simulations (10^7 or 10^8 MC steps) were carried out at constant temperatures. Each simulation was also repeated 10 times, starting from different random conformations, with a maximal distance from the conformation of the global energy minimum. We focused on the following equilibrium properties: (i) the probability of finding the 2DHP in the native conformation (Figure 4), (ii) heat capacity (Figure 5a), (iii) end-to-end distance (Figure 5b), and (iv) the first passage time—the average time that is needed to reach the native conformation for the first time (Figure 6).

In an equilibrium system, the probability of finding the 2DHP in the native conformation (Figure 4) is a measure of the depth of the global energy minimum, compared to the other minima. At the lowest temperatures that we examined, in which the relaxation times were too long to obtain a proper sampling of configuration space, the system is clearly not in equilibrium. However, the probabilities computed by the repeated trajectories are still useful as a measure of the radius of convergence of the global energy minimum. The coupling clearly makes the global energy minimum deeper, compared to the energies of other conformations, and also increases its radius of attraction.

Peaks in the heat capacity (Figure 5a) are indicators of first-order phase transitions. The coupled systems show clearer peaks. These spikes are associated with the cooperative collapse of the coupled 2DHP-s to the compact native conformation (Figure 5b).

The first passage time is a measure of the heights of the barriers on the pathway to the native structure. The higher the barriers the longer is the time needed to cross them at a given temperature. The first passage times computed for the uncoupled and coupled optimization models are very similar (Figure 6). This suggests that the additional coupling term did not create new barriers and therefore did not slow the optimization.

Radius of Attraction of Different Minima. Further insight into the shape of the energy surface was gained by examining the volume of attraction of the local minima. A series of 5000 energy minimizations (zero temperature MC simulations of 1000 steps) starting from random conformations sampled from a uniform distribution were used. The number of minimized conformations that were found in the same local minimum was assumed proportional to the radius of convergence of this minimum. The number of conformations that were mapped to a specific local minimum is proportional to attractive volume of this well.[32]

The minimum-energy conformations were classified according to their distance from the native fold and their energy. The probability of reaching each class by a minimization starting from a random conformation is plotted in Figure 7. The figure suggests that the coupling modified the distribution of the local minima as a function of energy and distance from the global minimum. In the uncoupled simulations most minimizations converge to low-energy minima that are considerably different from the native fold. A considerable portion of these local minima are not shared by the two 2DHP-s. A minimum for one 2DHP which is of high energy for the second 2DHP may be inaccessible when the coupling is introduced. As a result, the coupling shifts most minimizations to higher energy minima which are also closer to the native fold. Consistent with the other simulations this trend is more significant when "jumps" are allowed.

Discussion

In the present study we explored the possibility of merging homology-based modeling and modeling that is based on optimization of energy functions. For that purpose we introduced a new energy term. The new term penalizes the energy of a set of structures of homologous compounds when the structures differ appreciably from each other. That is, we used

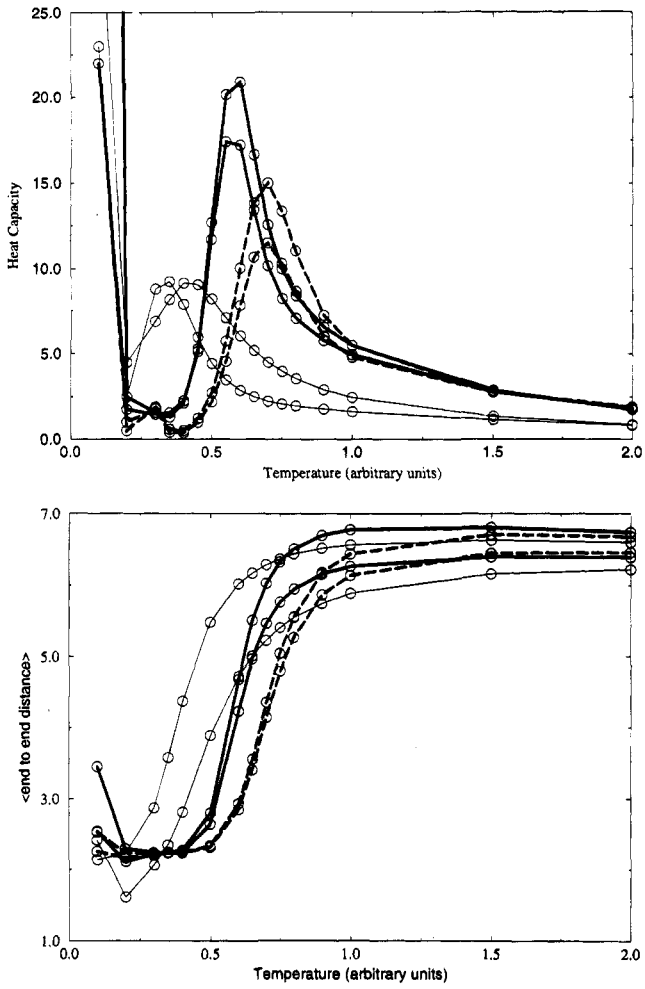

Figure 5. Cooperativity of coupled 2DHP folding is demonstrated by (a, top) heat capacity and (b, bottom) end-to-end distance (the distance between the first and the last monomers) as a function of temperature. The simulation parameters are as in Figure 4, except that in (b) all the simulations were of 10 000 000 MC steps. Thin line, cpl = 0 (no coupling); thick line, cpl = 1 (coupling); dashed line, cpl = 1 and "jumps". See Figure 3 for more details.

the experimental observation that homologous compounds tend to have similar structures, and we incorporated this information into our energy function.

We demonstrated the feasibility of the new computational approach on a model system that is solvable by other techniques and is accessible to comprehensive and exhaustive analysis. We considered structure optimization of two-dimensional heteropolymers (2DHP-s) on a square lattice. In the 2DHP-s example, the coupling cancels local minima that are not shared by the energy functions of the individual 2DHP-s and, as a result, biases the energy function toward the global energy minimum.

The present approach is clearly much better than the usual simulated annealing, as we demonstrated in the present paper. The improvement is obtained from the use of additional information (on the existence of homologous functions) that is typically not employed in optimizations. This additional information can be added to a variety of optimization algorithms and is not necessarily restricted to the Monte Carlo and simulated annealing protocols that we used here.

As an additional test of the importance of the coupling, we also optimized two coupled *identical* 2DHP-s. In this case, two identical copies of the system were run simultaneously, with penalty on a large separation of the two identical structures. This system is similar to the discretization of path integrals in which the kinetic energy term provides the "penalty" on the

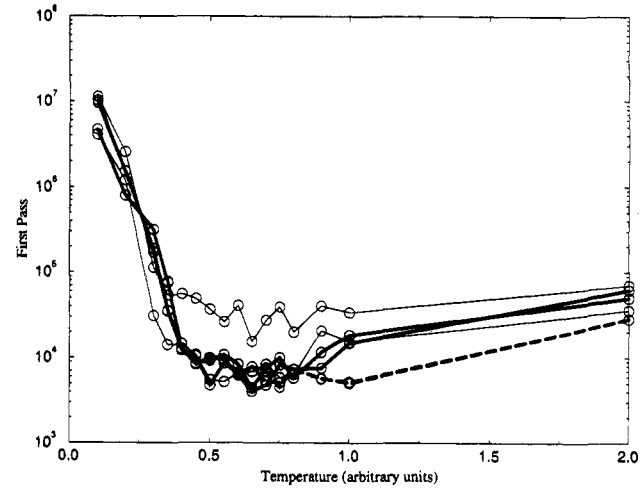

Figure 6. First passage time as a function of temperature is employed to test the influence of the coupling on the kinetics of folding. The first passage time is the average time (number of MC steps) needed to reach the global minimum for the first time. Simulation parameters are as in Figure 4. Thin line, cpl = 0 (no coupling); thick line, cpl = 1 (coupling); dashed line, cpl = 1 and "jumps". See Figure 3 for more details.

distance. We may therefore expect that this approach will lead to smoothing of the energy surface in a similar way to other smoothing algorithms.[9-13] However, compared to the homologous pair, the identical duo did not show considerable improvement.

The 2DHP-s made a useful test case. However, the interest in these particular systems beyond the scope of the present application is limited. We therefore devote the rest of the discussion to possible applications to other optimizations, emphasizing the questions and the issues that still remain to be solved when applying the present ideas to other global optimization problems.

The first and the obvious application of the present technique is for structure prediction of proteins. Does it work at present? The answer to this question for the test case that we pursued is unfortunately "no". We attempted to optimize the structure of the Kazal family of homologous proteins,[33-35] using standard force field (AMBER/OPLS,[2,4] as implemented in MOIL[36]) together with potentials derived from data bases.[5,37] Additional terms in the force field, of our own design, were also used. Unfortunately, it was easy to find many minima with energy values that were significantly lower than the native state using the above force fields, together or individually. The protein folding problem of the Kazal family therefore must wait for a better potential. The coupling can improve the global shape of the energy surface, making it more attractive for optimization; however, in this case the improvement was not sufficient.

On a more philosophical ground we expect homology to help when the individual energies are as different as possible while still maintaining the same global fold. If the energy functions of homologous proteins are very similar then one cannot expect to gain much from the coupling. If, however, they are very different, so that most of the local minima are not shared by all the homologous proteins, then coupling these functions will effectively eliminate the unshared local minima, leaving a smaller conformation space for the optimization algorithm to search.

We believe that the latter possibility is closer to reality. Even a single amino acid substitution alters, in many cases, the stability of proteins and/or its folding pathway without a profound effect on the native structure.[38,39] Thus, though the native structures of proteins seem very robust to mutations, the

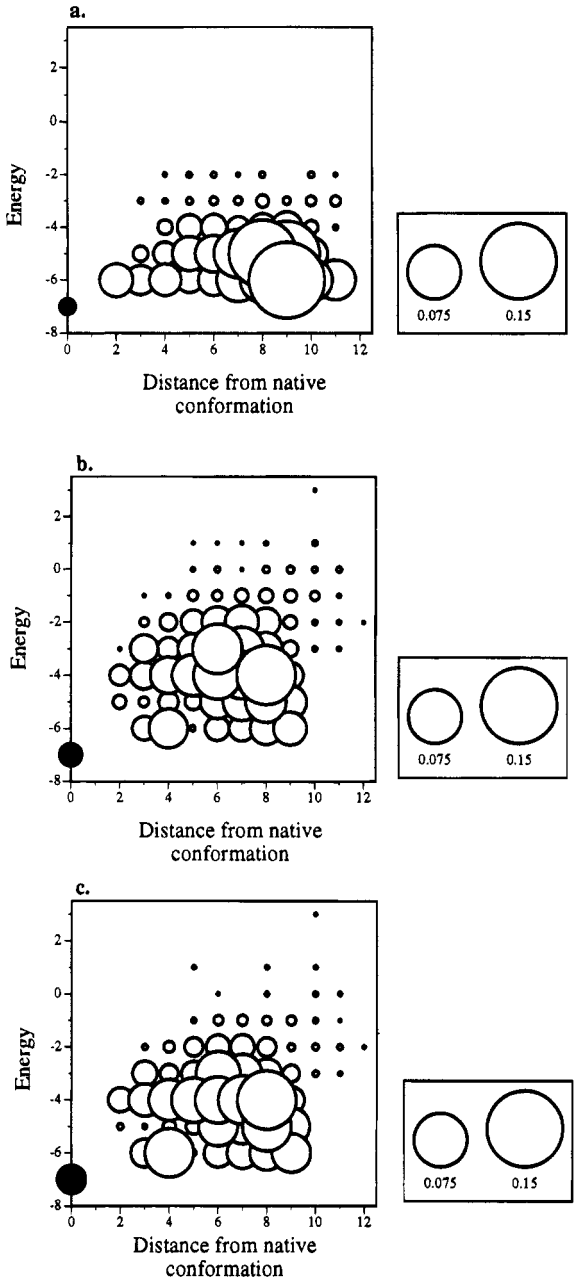

Figure 7. Attraction radii of local minima of (a) uncoupled 2DHP-s, (b) two coupled 2DHP-s (cpl = 1), and (c) two coupled 2DHP-s that may jump to the other (see text for more details). The local minima are classified according to their spatial distance from the global minimum and their energy. The position of the center of each circle in the figure corresponds to the energy of the conformation and its distance from the native state. The area of the circle is proportional to the number of optimization runs that end at the above state.

energy function is much more sensitive. The energy functions of homologous proteins with more then 50% substitutions are expected to be rather different. However, the global fold of these proteins is likely to be rather similar.[14] We therefore expect that the suggested method may benefit structure prediction of proteins when better potentials are available.

Another question is related to the best form of coupling between the homologous proteins. Obviously, the suggestion made in this paper is rather arbitrary, and better choices are most likely to be found when moving to the more complex protein problems. The coupling term should be able to ignore insertions/deletions and to adopt some differences in overall fold. In this paper, a distance of one between coupled conformations was allowed with no penalty. Only by using a flexible functional form for the energy term can one expect to use remote sequences. Remote sequences, however, are the most promising, since their energy functions are expected to be very different. Simple model systems like the 2DHP may serve as useful tools for the design of efficient coupling terms. The previously suggested, LES like, idea of binding different side chains (corresponding to homologous proteins) to the same backbone[26] is equivalent to a very stringent coupling. In our experience this coupling may lead to a significant increase in barrier heights and may limit the application to proteins with very similar sequences.

In the current work it was found to be beneficial to introduce the coupling not only to the energy function but also to the search mechanism. That is, allowing in the MC step trial conformations borrowed from the other 2DHP. This is similar in spirit to genetic algorithms, and this similarity is probably the reason for its success. This idea, at least in its current implementation, is less general than the coupling of energy functions and is probably limited to lattices. It may be very useful when applied to MC simulations using a reduced protein representation.[40,41] It is difficult, however, to see how it can be implemented for MD simulations of all atom models of proteins due to the high probability of van der Waals clashes.

The combination of homology and energy functions is clearly motivated by the protein folding problem. The applicability of coupling different energy functions with similar minima, for more efficient optimization, is wider than for optimization of polymers only. Another possible application is to locate the common *biologically-active* conformation of several flexible ligands that bind to the same receptor. A better chance of detecting the right conformation is expected by forcing all the ligands to look alike. This may succeed even if the active conformation is not the global energy minimum. An algorithm that forces different molecules to look alike is available in the program DISCOVER.[42]

Finally, the suggested method lends itself for parallelization, especially when used in all (or near all) atom systems. Each protein may be simulated on a different CPU, with a comparatively small amount of data (for example, the positions of the α carbons) needed to be transferred between the CPUs. With a flexible definition of the coupling energy term, the data transfer need not even be performed after each iteration. Since data transfer is currently the bottleneck of parallel computation, the very limited communication needed for the above parallelization scheme is a considerable advantage. A parallel computation on a cluster of workstations was used in the attempt to optimize the structures of the Kazal family.

Acknowledgment. R.E. is an Alon fellow. This research was supported by NIH Grant GM41905.

References and Notes

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223–230.
(2) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(3) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminthan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(4) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Sinh, U. C.; Ghio, C.; Alagons, G.; Profeta Jr., S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(5) Casari, G.; Sippl, M. J. *J. Mol. Biol.* **1992**, *224*, 725–732.
(6) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534–552.
(7) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *96*, 768–780. Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412–9426.
(8) McCammon, J. A.; Havey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(9) Kostrowicki, J.; Schraga, H. A. *J. Phys. Chem.* **1992**, *96*, 7442–7449.

(10) Ma, J.; Hsu, D.; Straub, J. E. *J. Chem. Phys.* **1993**, *99*, 4024–4035.
(11) Amara, P.; Hsu, D.; Straub, J. E. *J. Phys. Chem.* **1993**, *97*, 6715.
(12) Finnila, A. B.; Gomez, M. A.; Sebenik, C.; Stenson, C.; Doll, J. D. *Chem. Phys. Lett.* **1994**, *219*, 343–348.
(13) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277–9287.
(14) Sander, C.; Schneider, R. *Proteins* **1991**, *9*, 56–68.
(15) Greer, J. *Methods Enzymol.* **1991**, *202*, 239–252.
(16) Benner, S. A.; Badcoe, I.; Cohen, M. A.; Gerloff, D. L. *J. Mol. Biol.* **1994**, *235*, 926–958.
(17) Wako, H.; Blundell, T. L. *J. Mol. Biol.* **1994**, *238*, 682–692.
(18) Niermann, T.; Kirschner, K. *Methods Enzymol.* **1991**, *202*, 45–48.
(19) Zvelebil, M. J.; Barton, G. J.; Taylor, W. R.; Sterberg, M. J. *J. Mol. Biol.* **1987**, *195*, 957–961.
(20) Wako, H.; Blundell, T. L. *J. Mol. Biol.* **1994**, *238*, 693–708.
(21) Gerloff, D. L.; Jenny, T. F.; Knecht, L. J.; Benner, S. A. *Biochem. Biophys. Res. Commun.* **1993**, *194*, 560–565.
(22) Benner, S. A.; Garloff, D. *Adv. Enzyme Regul.* **1991**, *31*, 121–181.
(23) Shindyalov, I. N.; Kolchanov, N. A.; Sander, C. *Protein Eng.* **1994**, *7*, 349–358.
(24) Gobel, U.; Sander, C.; Schneider, R.; Valencia, A. *Proteins* **1994**, *18*, 307–309.
(25) Gutell, R. R. *Curr. Opin. Struct. Biol.* **1993**, *3*, 313–322.
(26) Zheng, Q.; Rosenfeld, R.; Kyle, D. J. *J. Chem. Phys* **1993**, *99*, 8892–8896.
(27) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6369–6372.
(28) Lue, K. F.; Dill, K. A. *Macromolecules* **1989**, *22*, 3986–3997.
(29) Shakhnovich, E. I.; Gutin, A. M. *Nature* **1990**, *346*, 773–775.
(30) Bryngelson, J. D.; Wolynes, P. G. *J. Chem. Phys.* **1989**, *93*, 6902.
(31) Kirkpatrick, K. S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(32) Stillinger, F. H.; Weber, T. A. *Phys. Rev. A* **1982**, *25*, 978–989.
(33) Bode, W.; Epp, O.; Huber, R.; Laskowski Jr., M.; Ardelt, W. *Eur. J. Biochem.* **1985**, *147*, 387.
(34) Bolognesi, M.; Gatti, G.; Menegatti, E.; Guarneri, M.; Marquart, M.; Mokos, E. P.; Huber, R. *J. Mol. Biol.* **1982**, *162*, 893.
(35) Williamson, M. P.; Havle, T. F.; Wuthrich, K. *J. Mol. Biol.* **1985**, *182*, 295.
(36) Elber, R.; Roitberg, A.; Simmerling, C.; Goldstein, R.; Verkhivker, G.; Li, H.; Ulitsky, A. In *Statistical Mechanics, Protein Structure and Protein Substrate Interactions*; Doniach, S., Ed.; Plenum Press: New York, 1994.
(37) Maiorov, V. N.; Crippen, G. M. *J. Mol. Biol.* **1992**, *227*, 876–888.
(38) Fersht, A. R.; Serrano, L. *Curr. Opin. Struct. Biol.* **1993**, *3*, 75–83.
(39) Sturtevant, J. M. *Curr. Opin. Struct. Biol.* **1994**, *4*, 69–78.
(40) Covell, D. G. *Proteins* **1992**, *14*, 409–420.
(41) Rabow, A. A.; Scheraga, H. A. *J. Mol. Biol.* **1993**, *232*, 1157–1168.
(42) *Discover User Guide, Version29*; Biosym Technologies: San Diego, 1993.

JP950832X