Simulation of Protein Folding

(With an emphasize on categorizing the different protein folding models)

By Hanif Bayat Movahed
Project for a course in Molecular Biophysics

1. Introduction:

Understanding the mechanism of protein folding is often referred as the second half of genetics. By solving this problem, there will be a revolution in drug industry (folding of peptide is an important issue in biotechnology) as well as finding cures for the diseases which are originated from the miss folding such as Alzheimer, mad cow , Parkinson, and many different cancers. The protein shape determines its biological activity. For example Enzymes do their work because parts of them match the shapes of the molecules whose reactions they control. [2, 14, 16]

I have tried to give a good overview on this topic by comparing different protein folding models while it was not easy to find exactly what I was looking for in the many different resources available in the internet about protein folding. And it was really hard when I wanted to combine all of these different methods together and try to categorize them besides comparing them. I have tried to use lots of references (21 different resources) and I have tried to put the similar methods in the same categories. I have mentioned in front of each paragraph about the references that I might use in that paragraph, while I have used those references in that paragraph directly or indirectly and it is clear that most of the time I concluded by myself about the result of what I read about that subject. I have tried to cover most of the significant points about protein folding models just in following as a brief essay.

In the following you will first find some basics about the property of protein and protein folding, such as energy minimization in protein folding. Then you will see the models and methods used for different kinds of protein simulation. In end I tried to mention a little about some techniques which are used in simulation of the described models.

2. What is protein folding:

2.1: Protein structure:

There are 4 levels for protein structural organization. They are called as primary, secondary, tertiary and quaternary structures.

1. The primary structure is defined as the linear sequence of Amino acids in a polypeptide chain. [3]
2. The secondary organizational level of macromolecules is related to the certain regular geometric figures of the chains. [3]
3. Tertiary structure results from long range contacts within the chains. Tertiary structure describes the folding of the polypeptide chain to assemble the different secondary structure. The three types of bonds occurring are as follows, 1. Hydrogen bonds, 2. Ionic bonds and 3. Disulphide bridges that occur between atomic groups, which are not so far from each other, play a major role in making the tertiary structure. [3, 8]
4. The quaternary protein structure involves the clustering of several individual peptide or protein chains into a final specific shape. Different kinds of bonds are involved in making this kind of structure such as: Hydrogen bonding, salt bridges, and disulfide bonds. There are two major categories of proteins with quaternary structure - fibrous and globular. [15]

Tertiary structure describes the folding of the polypeptide chain to assemble the various different secondary structure elements in a particular arrangement. As helices and sheets are units of secondary structure, so the domain is the unit of tertiary structure. We can predict the secondary structure of protein by accuracy of more than 70% by using different methods such as statistical prediction and artificial neural networks. [7]

2.2: Protein Folding Definition:

2.2.1: Medical definition:

Proteins consists of linear chains of amino acids, but they do not simply flop in your cells, instead, the regarding proteins "fold" up into a particular three-dimensional conformation in solution (tertiary structure) , and this conformation helps the proteins to carry out the functions, which are responsible for. Understanding the protein folding is the next step in deciphering the genetic code. [9, 10]

2.2.2: Biophysics Definition:

That is the equilibrium state under the influence of compensating energy differences; protein structure is formed to the folded structure. The equilibrium state is the conformation of greatest (or one of the greatest) thermodynamic stability. That is, the protein spontaneously assumes the conformation of lowest energy for a given environment (Anfinsen theory). It should be mentioned that many experts believe that the equilibrium state need not to be one of the absolute lowest free energy. It may rather represent the lowest free energy state that is kinetically available to the protein. [1, 3, 11]

2.3: Energy functions & folding forces:

The native structure of protein is a result of a delicate balance of energy terms. The energy function of protein consists of two main terms. First, E_P is the conformational energy of protein itself (without paying attention to the solvent) and second, E_S is the energy of interactions between protein and its solvent.

E_P consists of energy of electrostatics term (E_C), Len nard -Jones term (E_{LJ}), Hydrogen bonds term (E_{HB}) and torsion angles term (E_{tor}). Those terms are in Kcal/Mol as follows:

$$E_C = \sum_{i,j} \frac{332 q_i q_j}{\varepsilon r_{ij}} \quad , E_{LJ} = \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r^6_{ij}}\right), E_{HB} = \sum_{i,j} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r^{10}_{ij}}\right), E_{tor} = \sum_i U_i (1 \pm \cos(n_i \chi^i))$$

(In more sophisticated model you can put Stretching $\frac{k}{2 \times (R - R_0)^2}$ or Bending $\frac{H}{2 \times (\theta - \theta_0)}$ also) [16]

Here r_{ij} is the distance between atom i and j, ε is a dielectric constant and χ^i is the torsion angle for the chemical bond i. [1, 4]

E_S, the energy of protein and solvent interactions consists of three terms as following. First, the hydrophobic term that is related to work that is necessary to create a cavity of the shape of the solute molecule in solution. Second, the electrostatics term between the solute and solvent molecules. This energy includes the hydrogen-bond energy. And the last one is the Lennard- Jones energy between solvent and solute. There are different equations to represent the E_S and depends on how much accuracy we need in our simulation, the formula will change. [4]

Studies on the protein denaturation has a given us very useful information regarding the thermodynamics of unfolding. These measurements gave us standard ΔH, ΔS and ΔG regarding the properties of unfolding. We can separate these data's to two main categories. First, by the experiments on the transfer of liquid hydrocarbon to water we can determine the values of ΔH_{hyd}, ΔS_{hyd}. These values are regarding the hydration of all the groups previously embedded. The second one which is called as residual energy is just the difference between the two sets of values. It is important to note that the ΔG is nearly small for protein folding while it is just -5 to -15 kcal per mol. This energy is equal to the energy of 2 or 3 hydrogen bonding.

On one hand some believes that (like Anfinsen) the protein accepts spontaneously the lowest free energy for a given environment. (Depending on the solvent, temperature and etc). This means that it is not so important what combination of amino acids appears

while the final conformation totally depends on free energy consideration. On the other hand, it was suggested that the conformation assumed by the protein does not have to be the lowest free energy. The native conformation might actually be of a higher free energy than some alternative ones but they are large energy barriers, which are not available to the protein. In an enough time, most probably the protein would assume the lowest energy state but this condition is nearly non-existed under normal circumstances. Folding could be assumed similar to crystallization process also. [3, 11]

As you will see in the next chapters of this project, much other different energy have been given for lattice model or other kinds of model, depending on how complex the systems are. For example in the outmost simple lattice model the energy is just calculated between the adjacent hydrophobic beads and it is as follow:

$$E= \sum_{i+1<j \wedge b(i)=H \wedge b(j)=H} -Contact_{i,j} .$$

As you will see in the reduced potential model part (3.4.5), it is possible to define other kinds of energy functions and force fields that can satisfy the energy of reduced parts of protein. (They were reduced to decrease the complexity)

2.4: Some qualitative physical facts of protein folding:

If we assume protein as chains, chain quickly folds into a compact shape, similar to the way a piece of string will bunch up while the ends are over each other. The lower time limit for protein folding is 1μs. (Marginally stable for very small protein)[14, 17]

The two significant features of native protein structure in water are as following. First is, the overwhelming tenancy of charged or polar side chains to appear in the exterior part of protein which is exposed to the water. Beside that, there is a tendency of hydrophobic side chains to be located in the interior part of the protein that forms a compact globular structure. Second, the formation of secondary structure segments as Alpha helices and Beta sheets to satisfy the backbone hydration bonds. [1]

In a naturally global protein structure the hydrophobic residues are hidden in the interior while the hydrophilic residues are exposed to the water. A special feature of protein in comparison with other polymers is that the micellar formation must be compatible with 2 main secondary structure classes. [1]

The relativity large fraction of Protein (40%-70%) must be in either Alpha helix or Beta sheets. It should be mentioned that, because of forming the globular structure, each individual secondary structure region as well as the loop region between individual secondary structures can not be too long. [1]

It was proved by both theoretical and experimental results that the secondary structures in protein are not stable if it was removed from the total protein structure and studied as small peptide. In protein folding the final secondary structure and tertiary architecture are

determined by global optimization of hydrophobic effect and backbone hydrogen bonding.[1]

As a simple model we can define the folding as equilibrium between two opposite forces. On one hand the protein wants to minimize the solvation energy and to minimize the interaction between the hydrophobic groups with water and on the other hand the packing will decrease the conformational entropy and this will increase the ΔG that is unfavorable. [3]

A typical protein consists of a few salt bridges, several hundred hydrogen bonds and several thousand Van der Waalse interactions. By existence of all of this force the protein is marginally stable and the ΔG of the protein folding is just -5 to -15 Kcal/mol (Near the energy of 2 or 3 hydrogen bonds) [16]

We can assume that each protein consists of some domains (we can introduce some algorithms which give these domains). The internal dynamics of protein may involve the motion of these domains relative to each other and it should be mentioned that each two identical domains in two different Proteins have a similar functions. [3]

3. Protein folding models and their simulation methods:

It is obvious that if you want to simulate one physical phenomenon you should first select/design the appropriate models which satisfy the accuracy as well as saving the time/money. Theoretically, a computer could calculate all the possible shapes for one sample protein and select the lowest potential energy one. But in practice, however, it is possible that this process takes longer than the age of the universe to do all the calculations. It was shown for many kinds of model that the protein folding problem is NP hard complete problem (including lattice and off lattice models) [13, 14]

In this part some different models will be explained and compared with each other on the ground of accuracy and computational tractability. It is clear that we can have the spectrum of models. In one side we can use the most precise model which is so time/money consuming. (May be it is not possible at all!) And on the other side we face some oversimplified or minimal models that they are not useful enough to describe many complex phenomenon but they can help us to address general questions.

It should be mentioned that there are lots of different models for protein folding and I am not going to cover all of those models in this essay. I just want to take a glance on some of the well known ones and compare the different parts of the spectrum. In the end I will focus more on some specific parts of the spectrum.

3.1: Explicit solvent representation model and all atom models:

As it was said before, the native structure of a protein is related to the global free energy minimization of a physically accurate potential function. As it was said before it may not be the absolute minimization point but it can be another more accessible point. Proteins are required to adapt the suitable structures soon after being synthesized and being transported to their designated location. This gives us an upper limit for folding process and how many MD steps we need in the simulation of protein folding [1, 2]

The models, which are using the absolute explicit solvent representation in their computer model, are so expensive (time consuming) for determining the global energy minimum. In these models the starting point is an unfolded state and we need lots of averaging over solvent configurations. In all atom representation both the protein and the solvent will be represented. [1, 2]

On one hand this approach has a several advantages. It assures the generality and allow to be improved by more accurate quantum mechanical methods and it can be extended by parameterize polarization energy. On the other hand, the detailed model needs a large number of particles (typically more than 10000) and it needs small time step (10^{-15} seconds) while this process takes place in a microsecond or larger time. Therefore this method can be only used for small proteins where can be accelerated by raising the simulation temperature, changing the solvent condition, by applying the external force and by applying pressure. [1, 2]

I am not sure about the exact time of MD step (I think it should be near 10^{-15} s) but as I have observed in different references, the MD time step is <u>at least</u> six-nine orders of magnitude smaller in comparison with the protein folding time but encouraging results have been obtained by this approach. Folding using an all atoms model has been made wonderful development in the simulation process of small peptides. There is another problem with atomic level method. This method leads to very rough energy landscape that can be trapped in very high local minimum. One of the most important priorities of this method than most of the methods is that the folding path can be traced [1, 2]

3.2: Continuum solvent model:

In this model, there is a detailed atomic level models of protein coupled with a continuum model for solvation free energy. The history of molecular modeling molecular potential function is linked to that of physical chemistry of intermolecular interactions as well as the protein force fields. Lots of improvement has been made in this model by finding an alternatives potential function such as numerical solution of the Poisson-Boltzman equation. These models have priority to be improved systematically by a development of more precise physical-chemical potential functions. If this method combined by the all atom representations it can reach to wonderful results will be the input of all atom models in next step. [1]

3.3: Lattice models:

One of the earliest works which is done in the area of protein folding was an Ising model simulation on the unfolding and hydrogen exchanging of the proteins. [2]

One of the main starting works in this area is done by Ptitsyn and Rashin's work. In their work they studied folding of Myoglobin without the computer by treating each alpha helix as a uniform rigid body cylinder and then concluded that the folding was like a nucleation process such as crystal growth. In the late 1970s Levitt and Warshal studied the folding of "Bovine Pancreatic Trypsan Inhibitor" by more complicated model. In their model they use 2 atoms for each amino acid. The challenged and then-popular view, that the folding was always preceded by forming stable secondary structures, was first given by them. [2, 3]

In lattice models, amino acids are represented by connected beads in two dimensional lattices or three dimensional cubic lattices. Each site can be empty or have one bead. By this way the excluded volume will be considered. Bond angels can be ± 90 or ± 180 and the distances between two adjacent beads are assumed as one (fixed). The compact state in this lattice is the state that all the accessible positions are filled by beads. [3]

The advantage of lattice is that the simplified models allow efficient sampling of conformational space. You should consider the fact that the earlier computers are many orders of magnitudes slower than current computers. If these models designed properly they can give a valuable energy minimum and it is possible to enumerate the all possible conditions and therefore calculate the partition function and then find the corresponding free energy by using partition function. Depending on the how carefully it is parameterized, it can give the encouraging results. We can use a MC (Monte Carlo) method to find the ensemble averaging. On the other hand the oversimplified simple lattice model with nearest neighbor interaction is not enough to get valuable results and many details of atomic structure are lost. [1, 2, 3]

The minimal models such as lattice models can address the general questions such as: how can polymer chains fold quickly to unique (native) conformations or why some sequences fold while others do not. [21]

We can categorize the lattice models in to two different models: 1. Simple Lattice Models and 2. Lattice Models witch are parameterized by using realistic protein data. (Knowledge based Potential) [2]

3.3.1: Simple Lattice Model:

This model was suggested first by Go and co workers to understand the basic physics of protein folding. The main character of this model is its simplicity. The size of the lattice

can be from 3*3 to 5*5*5 points in lattice. We can use a binary map to increase the efficiency of computation of these kinds of model. [2, 3, 5]

These types of models (simple ones) are not designed for real proteins so these models are confined to study the general features of protein folding.

In one of these kinds of model, which is called as HP model, there are only two types of beads. H represents the hydrophobic beads and P is referred to polar ones. These beads are randomly distributed in the lattices and there are simple estimations about the lengths of α- helices and β-sheets in the lattice. 11 residues per α- helix and 6 residues per β-chain are found on average. These numbers are really close to experimental values. Conformational search in these kinds of model is NP complete problem. It is clear that in this model it is assumed that the main force in the folding process is the hydrophobic forces. There is one stablising force in this kind of model and it is the time that two residue interact in the lattice (they are special nearest neighbor) while they are not nearest neighbor in amino acids sequence. [3, 5]

In the first HP models they assumed the energy of two hydrophobic residues as negative energy, while they considered the rest of interaction, hydrophobic-polar, polar-polar and the interaction of solvent with any of those kinds as neutral and therefore the energy of those interactions was assumed as zero.

$$E=\sum_{i+1<j \wedge b(i)=H \wedge b(j)=H} -Contact_{i,j}$$

Most randomly generated sequences of H and P do not fold in the single state. For example by the work of Chan and Drill in 1996, they found for protein of 18 monomers in the square lattice (most probably 6*6) just a few of them 2.4% have a unique ground-state conformation. (6349 out of 2^{18}) To find more you can refer to Chan and Drill paper in 1989 by the title of Comparing folding codes for proteins and polymers. [3, 5, 12]

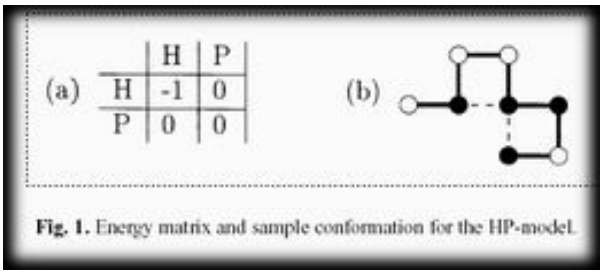

Fig. 1. Energy matrix and sample conformation for the HP-model.

In the left you can find the matrix of energy functions and in the right you can find the sample conformation for the sequence of PHPPHHPH with energy of -2.

Selected from: http://bioinformatics.oupjournals.org/cgi/reprint/15/3/234.pdf by R. Backofen and his colleagues [13]

In the next step Tang and coworkers has applied some more forces to previous model. They added the attraction force between H and P in the two letters lattice model. By calculating the energy of all 2^{27} HP sequence of 27 monomers, in 103346 completely compact cases (1/48 of all 4960608 maximally configuration cases. Because they did not want to count the similar conformation cases, which can be reached by the rotational and inversion from each others) in a 3*3*3 cubic lattice, they found the ground state energy of all those cases. This approach has been extended to 4*3*3 lattice for 36 monomers by the same researchers. [5]

This approach can be extended to more alphabetical-letters models. For example we can make matrix of potential energy value for each of pairs between 20 amino acids and then try to make a model of 20 alphabetical beads. One of the most famous extensions of HP model is HPNX model. In this model, polar monomers are splitted in to 3 different categories. Positive charge (P), negative charge (N) and the neutral one as (X)

	H	*P*	*N*	*X*
H	−4	0	0	0
P	0	1	−1	0
N	0	−1	1	0
X	0	0	0	0

E=-4. HH contacts+ PP contacts+ NN contacts- NP contact-PN contacts [13]

Selected from: http://bioinformatics.oupjournals.org/cgi/reprint/15/3/234.pdf by R. Backofen and his colleagues [13]

3.3.2: Realistic lattice models for realistic proteins:

One of the pioneers of these kinds of model is Skolnick and his coworkers. They began to use of increasingly complex lattices. This model emerged with fine grained representation of the torsion angle phase space of protein besides the more accurate mapping from the lattice to the 3-dimentional space. This model is parameterized by real proteins as templates by using known structures. This model is better in long polymers because of the detail representation of secondary structures. Simple examples of both four-helix bundles and beta-sheets have been gathered from random initial conditions. The 46-residue protein has recently been folded, that is less than 4 Å RMS deviations from its native structure [1, 2]

3.4.: Reduced models:

The specific character of any reduced models is for lowering the number of degrees of freedom in the system as well as discrimination the space of possible conformations for optimization purpose. This is necessary if we want to predict the protein folding in large proteins. It can be done by representing the molecule as coarse grain beads or combining two methods. The use of multiple levels of resolution is the key point in many of the

following models. **Bellow I try to pay attention to many different kinds of reduced models.** [1]

In these models the geometry of peptide bond and the various secondary structure elements can be represented well while the side chains and the intermolecular forces are treated in an approximate manner to reduce the computational time. Potential functions in this model can be built heuristically by using the statistical information from PDB (Protein Bank Data) and physical-chemical energies. This approach is limited to small resolution such as 4-6 A. [1]

3.4.1: Combination with detailed atomic level:

If we combine two methods by paying attention to a reduced model and the detailed atomic level respectively, we may reach the more precise and accessible method. We can use this model and the related method at first and then change the algorithm to the detailed atomic level. [1, 2]

3.4.2: Combination with simple lattice model:

For discretization purpose, it is possible to use the simple lattice models. It is clear that by using the simple lattice model at first step, we confine ourselves to work with the proteins that have 50-75 residues (at Max) because you just can put one residue in each site of lattice. [1]

3.4.3: Residue based off lattice model:

A group led by Scheraga simulated the folding of protein HDEA on Cornell's IBM supercomputer in 1998 by using the residue-based off lattice model. In their method they ignored the nitrogen and carbon atoms at the ends and worked with a simple model of central carbon and its side chains to generate structures, which it was used as starting points for a next step. In next step they considered all the forces between all the atoms. They predicted the existence of 5 spiral coils that match 80% of structure found by X-ray crystallography. It is nice to mention that the regarding calculation of the structure of the HDEA protein took 70 hours running on 64 parallel processors of the Cornell's IBM supercomputer. This method gave some wonderful results when compared to a known protein structures which their shape are determined by NMR and X-ray methods. [14]

The method that was used by Scheraga group is called as residue-based off lattice method. In residue based off lattice model, it is tried to develop methods that can be reached the native states from non-native ones. The most of structure prediction use the residue level model with the potential data obtained from PDB. They considered the continuum solvation in many of their methods to minimize the calculation in atomic level [2]

Wu and Sung proposed the use of the mean solvation force to represent the force. By using this technique on alanine-dipeptide they found the reasonable results. By using this method the number of particles involved in the simulation decreases significantly. [2]

3.4.4: Coarse grained model:

A question arises regarding the coarse grained model and that is about the prediction of secondary structure since the detailed structure of protein has been reduced from the model. This problem can be solved by using the specific potential interaction designed to stabilize the secondary structure or by holding it fix during the process of minimization. It is obvious that you need to know the secondary structure in advance. [1, 2]

As it was said, there are lots of different approaches to predict the secondary structures of proteins. These methods can rely on direct statistical analysis of the data bank by using mathematical procedure or using the neural network recognition. We can use some experimental results from NMR for understanding the secondary structure as well. So it is possible to use the reduced model simulation with fixed secondary structure. Cohn, Richmond and Richard initiated this method and Cohn and his coworkers have been using of this method. This method is based on the idea that if the secondary structure is known the phase space for the folding problem is qualitatively reduced so it makes the determination of correct tertiary rapidly. This approach was really successful for large protein such as myoglobin. [1, 3]

With the secondary structure held in its own place during the simulation there is a huge advantage in computational calculation of coarse grained model and we can also reach reasonable value for precision. For example in this model it is tried to show different structure motifs by “cylinder and sphere”(CS). If we want to reach the precision of RB (residue based) model in this method, a hierarchical method is needed. The CS model can be used as crude trial conformation in first and then those can be passed on as trial conformation at the residue based level. [1]

3.4.5: Energy functions of reduced models:

For each kind of reduce models different kinds of energy and force fields should be defined. From the computational point of view, evaluation of potential function is more costly than manipulation of the coordinates. [1]

The reduced potential functions have two main roles: 1. They should represent the true potential at a comparable level of coarse graining model. 2. They must compensate the omitted structure parts. This means that they should include average packing forces and the effects of solvent and hydrophobicity and hydrogen bonding. [1]

The tortional potential can indirectly include the Calpha pseudo-angle and pseudo-dihedral potentials or include them implicitly by using a biased selection of trial moves. Hydrogen bonding can be included by special multi-body interactions which are designed to recognize the backbone conformations in hydrogen bonded structures. [1]

There are several reduced contact potentials, which are used in literature. These potentials are determined empirically by distribution of known structures in the database. The interactions can be supplemented with single residue potentials that include the surface exposure of a residue. [1]

For the hierarchical model, it is possible to calculate directly the interaction between entire secondary structure elements. For CS model, these were done by taking a long range potential and expand it around the center-center vector between calendars. The "hydrophobic dipole" interaction is the second term in that expansion that depends on how the hydrophobic and hydrophilic residues distributed in the helix. [1]

4. Some algorithms techniques for simulation of the regarding models:

Many details about the simulation techniques for the specific models are said in the previous parts. In this part I just want to mention a little about MC and MD, which are used in the above models. Most of these techniques are used in efforts to effectively determine the global minimum of those models.

For the concerned purpose the following techniques are available: Monte Carlo simulated annealing, Molecular Dynamics and Langevin Dynamics. (I just describe the two first)[1]

4.1: Monte Carlo method:

Monte Carlo (MC) methods are very general methods in computational calculation of an arbitrary system. The name comes from the famous Monaco casino to emphasize the importance of randomness. MC is particularly used in a system with large number of degrees of freedom (and quantities of interest) that can be found in statistical physics. For example this approach can be used in thermal average and excluded volume calculation. [6]

The use of a biased distribution of trial moves as a means of improving the efficiency of the conformational sampling. In the very small chains all conformations can be enumerated and thermal averages or entropy can be computed exactly by the following equation.

$$\langle A \rangle = \frac{1}{Z} \sum_{\mathbf{x}} A(\mathbf{x}) e^{-\frac{E(\mathbf{x})}{T}}$$

Selected from http://www.unb.br/ib/cel/chico/artigos/thesis/node6.html [6]

But when the chain is long, complete enumeration of conformational space is nearly impossible with present day computers. In MC simulations this problem is solved by calculating just for M conformations (M is much smaller than the total number of conformations) using the following equation:

$$\langle A \rangle_{est} = \frac{\sum_{l=1}^{M} A(\mathbf{x}_l) e^{-\frac{E(\mathbf{x}_l)}{T}}}{\sum_{l=1}^{M} e^{-\frac{E(\mathbf{x}_l)}{T}}}$$

selected from http://www.unb.br/ib/cel/chico/artigos/thesis/node6.html [6]

And if we select the points biased towards conformations that are significantly populated at equilibrium, the equation will be

$$\langle A \rangle_{est} = \frac{\sum_{l=1}^{M} A(\mathbf{x}_l)}{M}$$

selected from http://www.unb.br/ib/cel/chico/artigos/thesis/node6.html [6]

Where the probability of occurrence of a given conformation (or sample) is proportional to its Boltzmann factor as in the above equation. They are generated by the Metropolis algorithm. By using this algorithm it is possible to make Markov chain of conformation that started form random X_1 and a proper probability of movement from X_i to X_{i+1} as W ($X_i \rightarrow X_{i+1}$) that it satisfies the following equation. [6]

$$\frac{W(\mathbf{x}_l \rightarrow \mathbf{x}_m)}{W(\mathbf{x}_m \rightarrow \mathbf{x}_l)} = \exp\left(-\frac{E(\mathbf{x}_m) - E(\mathbf{x}_l)}{T}\right)$$

Selected from http://www.unb.br/ib/cel/chico/artigos/thesis/node6.html [6]

4.2: Molecular dynamics methods:

The molecular dynamic simulation was initiated by "Alder: and "Wainwright" in the late 1950's. It was used to study the interaction of hard sphere model that was useful to find the behaviors of simple liquids. The first protein simulations were done in 1977 with the simulation of the bovine pancreatic trypsin inhibitor (BPTI) (McCammon, *et al*, 1977). [20]

One of the techniques in the theoretical study of biological molecules is molecular dynamics simulations. In this method you observe the time behavior of atoms of the system. In this method, the motion of atoms is traced out in the space. Fluctuations in the system are simulated by the trajectory of the points referred to atoms. [3, 20]

Molecular dynamics simulation gets information about the microscopic property of the system, including atomic positions and velocities. To emerge information regarding the macroscopic observable such as temperature and pressure, it is necessary to use statistical mechanic. [20]

In the classical systems the method is based on Newton's second law. From the force on each atom, which is completely related to your model, it is possible to find the acceleration on each atom and integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. [20]

The most important advantage of molecular dynamics in comparison with Monte Carlo method is its time dependence and your ability to find the folding path in this method.

References:

[1] Richard A. Friesner and John R. Gunn, "Computer Simulation of Protein Folding", Annu. Rev. Biol. Biomol. Struc., in press, http://www.smi.stanford.edu/projects/helix/bmi214/annrev.txt.html

[2] Y. Duan and P. A. Kollman, "Computational protein folding: From Lattice to all-atom", IBM research journal (2001), Science (1998), http://www.research.ibm.com/journal/sj/402/duan.html , www.public.iastate.edu/~leiyang/597Pres/**folding.ppt**

[3] Text Book, Molecular Biophysics (structure in motion), Michel Daune

[4] Yuko Okamoto, "Protein folding problem as studied by new simulation", Recent research development in pre & Applied Chemistery2, 1998, http://konf2.ims.ac.jp/review/review.pdf

[5] H.S. Chun, E.B. Bauer, Review, "Prospective on protein evaluation form simple exact model", http://arrhenius.med.toronto.edu/~chan/chan-bornberg-bauer-rev02.pdf

[6] Internet site, "Monte Carlo method and the Metropolis algorithm", http://www.unb.br/ib/cel/chico/artigos/thesis/node6.html

[7] Internet site, "Quaternary protein structures", http://elmhurst.edu/~chm/vchembook/567quatprotein.html

[8] Internet site, "Tertiary protein structure and folds", http://www.expasy.org/swissmod/course/text/chapter4.htm

[9] Internet site, "Proteomics & Bioinformatics at the Samuel Lunenfeld Research Institute", http://bind.mshri.on.ca/trades/index.html

[10] Internet encyclopedia, http://www.thefreedictionary.com/protein%20folding

[11] Internet site, "Protein folding", http://class.fst.ohio-state.edu/FST822/lectures/Fold.htm

[12] Chan HS, Dill KA, "Comparing folding codes for proteins and polymers", Proteins. 1996, March, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=8778780

[13] R. Backofen, S. Will, E. Bauer, "Application of constraint programming techniques for structure prediction of lattice protein with extended alphabets", Bioinformatics, Vol. 15, No. 31999, http://bioinformatics.oupjournals.org/cgi/reprint/15/3/234.pdf

[14] B. Steele, "Researchers use physical laws to simulate protein folding on supercomputer", http://www.news.cornell.edu/Chronicle/99/8.19.99/CASP-3.html

[15] Folding@home, distributed computing community related to Stanford university, http://www.stanford.edu/group/pandegroup/folding/

[16] Slide show, Molecular simulation of protein and peptide, http://chsfpc5.chem.ncsu.edu/Poznan/foldingI/sld001.htm

[17] Slide show Chemistry of Biotechnology, http://chsfpc5.chem.ncsu.edu/Poznan/chem_bio/sld001.htm

[18] From slide of course 134044 at Technion, Israel institute of Technology, http://biology.technion.ac.il/courses/134044/PartA/Lectures/19

[19] David Brown, "Deciphering the message of life's assembly", http://www.people.virginia.edu/~rjh9u/protfold.html

[20] "Theory of Molecular Dynamics Simulation" http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html

[21] Antonio Francisco Pereira de Araujo "The cubic lattice model with engineered sequence", http://www.unb.br/ib/cel/chico/artigos/thesis/node9.html