The nanostructures conformation changes play a crucial role in a signal transduction processes in nanonetworks. In the paper, the method for determining of the conformation of organic nanostructures based on random simulation approach is presented. The methodology for simulation of conformation processes of the nanostructures like the polypeptide chains is discussed. In order to determine the conformation of the backbone of the polypeptide chain, the Modified Monte Carlo method utilizing probability matrix is used. As a result, we get solution in a form of trajectory representing the backbone of the chain. The simulation of conformation processes is illustrated through the numerical examples for selected proteins.

1. Introduction

For simulation tasks, the synthesis programs generating the conformation of the main chain of amino-acid residues in polypeptide, use the amino acids library and need the information on the angles between α-carbons bonded with their side chains and contiguous peptide groups [6, 17, 18]. The backbone of a chain of amino-acid residues is determined by the set of torsion angles [7]. The backbone conformation of an amino-acid residue can be specified by listing the torsion angle ϕ (rotation around the nitrogen-α-carbon bond in the main chain), and ψ (rotation around the α-carbon-carbon bond in the main chain of the polypeptide).

The relationship between the peptide groups, α-carbons and torsion angles can be expressed in the following form

\[ \rightarrow PB \rightarrow \phi_1 \rightarrow C_{\alpha 1} \rightarrow \psi_1 \rightarrow PB \rightarrow \phi_2 \rightarrow C_{\alpha 2} \rightarrow \psi_2 \rightarrow PB \rightarrow \]

where PB denotes the peptide bond and \( C_{\alpha i} \) is the \( i \)-th α-carbon atom. The torsion angles play a crucial role in the conformation of proteins because the three-dimensional structure of protein determines its biological functions including activity in signal transduction cascades [2, 12, 14]. On the other hand not all combinations of torsion angles \( \phi \) and \( \psi \), in (1) for each amino acid are possible, as many leads to collision between atoms in adjacent residues exposed by Ramachandran restrictions [10, 12]. The possible combinations of \( \phi \) and \( \psi \) angles that do not lead to collision can be plotted on a Ramachandran map. The small changes of the torsion angles cause fundamental changes in conformation of the polypeptide in a case, when the chain of amino acids is very long. By assuming that atoms behave as hard spheres, allowed ranges of torsion angles can be predicted and visualized in contour diagrams called Ramachandran plots based on X-ray crystallography [3] (Fig. 1).

In case of nanostructures like amino acid chains, the flow of genetic information from DNA (data bank and replication) through the transcription of genetic code triplets (codons) into the mRNA chain finally is expressed in the material form as a synthesized proteins [12, 15, 16]. Synthesis of proteins is performed on ribosomes in the environment, where all aminoacylated tRNAs, activating enzymes and protein factors are accessible. When the translation process reaches the last, termination stage, the chain of amino-acid residues forming the polypeptide
is inserted into an environment which is usually the fluid. This fluid, cytosol, can be treated in first approximation as a solution. The polypeptide forms a chain of the molecules of residues connected together through the rotational bond realized by peptide bonds. The force powered changes, is a tendency of the chain of polypeptide to get minimal potential energy. This can be called free conformation of the polypeptide chain. Using Monte Carlo approach [8], we will find the free conformation for the selected proteins through the minimization of the energy [9] as a function of the torsion angles [4, 9]. In the proposed approach, we omit the problem of hydrogen and disulfide bonds creation [17].

The paper discusses Modified Monte Carlo (MMC) method (applying the tuning of the set of best solutions). The MMC simulation method of protein folding is illustrated through the numerical examples for selected proteins. A new Protein Simulator developed in Visual .NET™ environment utilizes the C# object oriented language to best facilitate abstraction of protein spatial structure descriptions [5, 13].

2. Determining the Free Conformation by Monte Carlo Method

The full algorithm of Modified Monte Carlo method of polypeptide model generation consists of two phases: first — the Monte Carlo Algorithm and second — the Angle Adjustment Algorithm.

First phase is the generation of given amino acid sequence with random $\phi$ and $\psi$ angles. The results of such generation are examined and best of generated conformations (those having lowest potential energy [17, 19]) are stored in the output array (to decide whether to allow the $\Delta \phi$ and $\Delta \psi$ changes or not, it is possible to use other criteria e.g. comparing the current value of exponential function of energy with random number generated [22]). Experiments have shown that results array should be proportional to the length of the amino acid chain in the protein, however practical dimension of such array should stay between 10 and 100. Second phase takes each of the proteins from the output array and manipulates the angles to obtain best (minimum) energy. The proteins in the output array are the initial condition for the angle 'twisting'. The larger number of proteins in the output array, the higher chances of finding the neighborhood of a global minimum.

Phase I — Monte Carlo Algorithm

The utilization of full search with 1° span in 10 amino acid chain requires years of computing, which makes such approach impractical. The Monte Carlo approach provides alternative, which with much smaller amount of tries can produce approximation of valid solution. Later such solution is adjusted. Because as mentioned previously, not all the angles are allowed in the amino-acid chain [3], a special modified technique is proposed for the angle generation.

Angle Selection Algorithm

Two maps have been utilized in the Angle Selection Algorithm. The maps limit the space in which the angles can be selected. The 360x360 space defining available angles (Fig. 1), is translated to linear vector of coordinates as it is presented in Fig. 2. The selection of angles is done with the standard random function with uniform distribution. Both angles $\phi$ and $\psi$ are selected in single attempt in the GetRandomAngles() function.

Example of such function is presented below:

```csharp
bool GetRandomAngles( ref double phi, ref double psi )
{
    return V[Rand.GetNext()];
}
```

After all angles are generated for the given polypeptide, calculation of energy allows for placing current solution in the Array representing best solutions so far. Whole process is repeated for a given amount of tries.

![Figure 2. Example of mapping of allowed angles. $V$ – vector of angle pairs.](image-url)
Current_Energy = MinEng
SaveSolution(Protein)
End if
End while

The algorithm is presented in a graphical form in Fig. 3 where the variables have the following meanings:

- **k** – current step (external loop),
- **d** – number of allowed loops,
- **i** – current step internal loop for each angle,
- **M** – minimum energy,
- **E** – current calculated energy,
- **P** – protein structure,
- **A[]** – angle array holding both $\varphi$ and $\psi$,
- **dA** – the current amount of degrees by which each of the angles in the protein is being adjusted (check if new $A[i] = A[i] + dA$ or $A[i] = A[i] - 2dA$ appears in the allowed area of $\varphi$ and $\psi$, if not, reject the change of $A[i]$).
- **sA** – size of the angle array $A$,
- **Q** – angle division quotient for $dA$, usually equals 1.2-2.0. Better accuracy is achieved with smaller quotient.

Whole the block presented in Fig. 3 is repeated separately for each of the best stored candidates resulting from Monte Carlo Algorithm of angles generation (described as the Phase I). The algorithm is looking for local minimum in the $N^6$ dimensional space (the angles) by applying of small changes to each dimension repetitively. After each of the iterations the best candidate is selected. In internal loop (i) of the algorithm each of the angles ($\varphi$ and $\psi$) is rotated slightly in both directions. Starting from arbitrary value, the angle adjustment value $dA$ is slowly decreased with quotient $Q$ (external loop $k$) and angles accessibility checking.

**Calculation of Energy**
Calculation of energy in presented examples requires several steps. The sequence of the steps is presented in Fig. 4. All steps beside last one require execution of external programs [17]. The overhead of the initialization of these programs is very well visible during analysis of tool performance.

**Numerical Examples**
First example presents Nine-Amino Acids structure [1, 6], which was calculated with utilization of both the Monte Carlo and the Angle Adjustment algorithms. 3-D representation was obtained from RasMol [11] program. The initial conformation of this structure:

has a spatial shape presented in Fig. 5. A molecular shape of simulated polypeptide with minimal potential energy is illustrated in Fig. 6.

**Figure 5.** Initial Nine-Amino Acids structure.

The corresponding angles calculated for nine amino acids chain are found to be:

<table>
<thead>
<tr>
<th>Nine amino acids chain</th>
<th>met</th>
<th>-46.719</th>
<th>130.250</th>
</tr>
</thead>
<tbody>
<tr>
<td>ala</td>
<td>-62.438</td>
<td>125.531</td>
<td></td>
</tr>
<tr>
<td>gly</td>
<td>177.000</td>
<td>175.750</td>
<td></td>
</tr>
<tr>
<td>pro</td>
<td>-65.876</td>
<td>126.659</td>
<td></td>
</tr>
<tr>
<td>gly</td>
<td>177.750</td>
<td>155.593</td>
<td></td>
</tr>
<tr>
<td>asn</td>
<td>-60.000</td>
<td>-60.499</td>
<td></td>
</tr>
<tr>
<td>ala</td>
<td>-100.000</td>
<td>173.875</td>
<td></td>
</tr>
<tr>
<td>his</td>
<td>-69.094</td>
<td>139.500</td>
<td></td>
</tr>
</tbody>
</table>

Second example is the EGF (Epidermal Growth Factor) structure containing 53 amino acids. Computing was performed with 5° minimum step (in Phase II). Spatial shape of the generated model of protein is presented in Fig. 7. The set of torsion angles for EGF protein is inserted in Appendix.

**3. Computation Requirements**

**Computation Time**

Protein Simulator developed in Visual .NET™ environment utilizes the C# object oriented language. Language C# was utilized in the implementation of all algorithms. The computation was performed on the PC computer with Intel Pentium 4 processor 2.4 GHz, 1GB of memory and 120GB free drive space. Fig. 8 presents average time of calculation of a single amino acid in function of chain length. The figure clearly indicates strong influence of the ‘worm-up’ procedures on the computation time, in case of chains with length <50.

Proposed Angle Adjustment Algorithm was able to decrease initial energy by a factor of 5 in case of 9 amino acids and by a factor of over 20 (5° minimum step) in case of EGF protein. Visibly (Fig. 9) the computation time is more directly dependent on the chain length for lengths >100. The picture also shows almost linear dependence on the chain length.

**Figure 6.** Final Nine-Amino Acids structure.

**Number of Random Trials**

Let \( V \) be a sample space. We draw out, for selected polypeptide containing \( n \) amino acids, \( N \) samples \( v_i \in V, i = 1, 2, ..., N \) with uniform distribution of each \( v_i \) over \( V \).

**Figure 7.** The EGF 3D simulated structure (final state).

Vector \( v_i \):

\[
v_i = [v_{i_1}, ... v_{i_k}, ... v_{i_N}]^T,
\]

determines the \( i \)-th sample of the set of accessible pairs of torsion angles \( \varphi \) and \( \psi \) (Fig. 2) for \( k \)-th amino acid in the considered polypeptide chain (\( k = 1, 2, ..., n \)). Symbol \( T \) denotes transposition. For the sequence \( v_1, v_2, ..., v_N \) we compute the values of corresponding sequence \( E_1, E_2, ..., E_N \) where \( E_i = E(v_i) \), and \( E(v) \) is a value of the potential energy function [19] of the polypeptide structure in the point \( v \).

Let the point \( v_{opt} \), denotes the argument value of the potential energy function \( E(v) \), where \( E(v) \) reaches its global minimum. Let the point \( v_{opt} \) be in the part of space \( V \) denoted by normalized volume \( \gamma \) (for all space \( V', \gamma = 1 \)).
average calculation time per single angle
in function of chain length

**Figure 8.** Average computation time of one amino acid in function of the length of amino acid chain.

In [20, 21] the minimal number $N$ of drawing from a sample space $V$ for finding the value $y$ with the probability $\alpha$, determines the following relation:

$$\alpha \leq 1 - (1 - y)^N.$$  \hspace{1cm} (3)

From (3) we get the minimal number $N$ of drawing in the form:

$$N \approx \frac{\log(1 - \alpha)}{\log(1 - y)}. \hspace{1cm} (4)$$

Referring to Eqs. (3) and (4), let us consider e.g. the polypeptide containing three amino acids ($n = 3$), and assume the accuracy of localization of $v_{opt}$ to be a half of range for each pair of the torsion angles with probability $\alpha = 0.9$. For this example, we have $y = 0.5^{n} = 0.125$, and from (4), $N = 18$. In case of nine-amino acids polypeptide ($n = 9$), with the same remaining assumptions $y = 0.5^{9}$ and $N = 1178$.

Examples indicate a low efficiency of Phase I (which is however necessary) of the method for longer chains, and in conclusion, this is a reason of modification introduced by implementation of the *Angles Adjustment Algorithm* (Phase II).

4. Conclusions

The paper presents Modified Monte Carlo method paradigm for modeling and simulation of spatial conformations processes of organic nanostructures illustrated by the polypeptide folding problem. **Firstly**, the modeling of these processes in nascent state is important from its cognitive nature, and existing software tools make possible to use them in experiments for better understanding the informatic nature of these complex biological processes. **Secondly**, the MMC method approach, promises the development of different modifications in random determining of the set of torsion angles in the folded polypeptides.

References


Appendix

The corresponding pairs of torsion angles calculated for EGF chain (Fig. 7) are found to be:

EGF 53 amino acids structure