CHAIN-LIKE STRUCTURES' MOTION - METAMODELING OF THE TRANSLOCATION TIME IN SOME SPECIFIC SITUATIONS

Kamila Bartłomiejczyk

Institute of Mathematics, Czestochowa University of Technology
Częstochowa, Poland
kamila.bartlomiejczyk@im.pcz.pl

Abstract. This article discusses an application of an algorithm for modeling the chain-like structures' motion. The basic and the most important assumption of the adopted approach is sequentialization of the moves. In the presented version of the algorithm there is a possibility of modeling the tension propagation mechanism. The main purpose of the paper is to investigate the impact of different algorithm parameters on final movement characteristics. Based on the Monte Carlo experiments metamodels describing the relationship between the translocation time and selected algorithm parameters are proposed and discussed.

Keywords: algorithm, chain-like structure, metamodel, movement statistics

Introduction

The motion of chain-like structures is encountered and studied in many physical, chemical, biological and technological processes. One of the most important features of CLS movement is their translocation time. It plays a crucial role in such processes as drug delivery process, gene therapy, rapid DNA sequencing or invasion of viruses into cells. Consequently, a chain-like structures motion analysis is very important and gains great attention in experiments, theory as well as in simulations.

In this paper we focus on simulation studies of the translocation time of the CLS. In our simulation analysis we use a relatively novel approach for modeling the CLS movement. The approach was introduced in paper [1] and then the concept was developed in [2]. The fundamental idea underlying this approach is sequentialization of the moves. This idea enables a researcher to easily model the tension propagation mechanism. The resulting algorithm is apparently faster than those that are typically used in the Kinetic (or Dynamic) Monte Carlo simulations, see e.g. [3, 4].

The main purpose of the presented research is to study the relationship between the translocation time of the CLS and various movement characteristics that can be incorporated into an adopted algorithm via its various parameters. Thus, in order to make this article easy to understand, in the next section we describe the algorithm

in detail, emphasizing the role of its different parameters which can be changed during the simulation experiment. In Section 2 a simulation framework which was used in our experiments is described. Then, in the same section, we present selected results of the simulations. We focus on the relation between algorithm parameters and the speed of CLS motion. In Section 3 metamodels describing this relationship are presented and discussed. These metamodels are generalizations of the metamodels developed in [5].

1. The simulation algorithm

To understand a formal description of the algorithm it is necessary to introduce some denotations and terminology. It is presented briefly in this section. More detailed information about this algorithm can be found in [2].

The CLS position in 2D space is given by a sequence $c = \{c_1, ..., c_n\}$ of n points $c_i = (x_i, y_i)$, where the number n is called a length of the chain. The elements of this sequence are called segments of CLS. The distance d between two successive elements of this sequence is limited by a given range $[L_{min}, L_{max}]$, so $L_{min} \leq d(c_i, c_{i+1}) \leq L_{max}$, $i = 1, 2, ..., n-1$.

One of the important algorithm assumptions (commonly found in literature [1, 2]) is the **discretization of the motion space**. To fill this assumption, coordinates of the structure segments should be integers. This means that the CLS moves along the integer lattice nodes. However, to achieve better approximation of the continuous space one can assume that the abstract unit of the length (distance) is equivalent to a given number of grid sides. Obviously, the greater the number of grid sides per unit (GSPU), the better the approximation of the continuous motion space.

In the description of the chain dynamic, two terms "step" and "move" are used. A step transfers a particular segment from one node to another, whilst a move carries the whole structure to a new location. It is possible (when the first step does not cause any tension in the chain) that some moves consist of only one step. The new position ($c_i *$) of a single segment (after a step) could be chosen randomly from a set of one step reachable nodes (OSRN). This set is specified by condition $d(c_i, c_i *) \leq R_{max}$, where R_{max} is maximum length of a step which a segment can do in case where other forces or restrictions are not present. The move of CLS is continued until the tension in the structure is created, i.e. until $d(c_{i-1}, c_i *) \geq T_d$ (or/and $d(c_{i+1}, c_i *) \geq T_d$), where T_d is a tension parameter.

Another important assumption of the algorithm is the **sequentialization of the CLS movement**. At the beginning of each move a segment which moves as the first one (FTM) is chosen. That segment initialized a move of the other segments. The tension propagation through the chain can be considered as a sequence of steps. The choice of FTM is made randomly and it is realized according to a given probability distribution defined on all segments of CLS. It can model physical aspects of the structure.

A movement trajectory is a sequence of consecutive structure positions stored in matrix C whose i-th row is interpreted as CLS position after $i-1$ moves. Thus element C_{ij} denotes the segment j in the CLS position at moment i.

The final positions of the chain can be influenced by various physical laws. To include these laws in the simulation model there is probability distribution implemented on OSRN set. It determines probabilities of different directions or lengths of each step. Such one step probability distribution may also depend on the segment's coordinates.

For any given segment, not all nodes from OSRN set can be actually reached. Some of them may be forbidden because of some assumed properties of the environment and/or assumed features of CLS itself. An exemplary restriction may be the upper limit for the distance between consecutive segments, which ensures the inseparability of the structure. Another possible restriction is the existence of different objects that already occupy some of nodes (e.g. cell's membrane). The subtraction of OSRN and forbidden nodes is called a set of actually accessible nodes (AAN).

Additionally there is the cost connected with the CLS structure defined. It is necessary because of additional constraints related to the biochemical nature of the system under consideration which occur in real-world physical problems (i.e. polymer behavior inside a tissue of a living organism). The cost of the structure and its location in the motion space is a function representing its fitness connected with its conformation or another external properties. The fitness of the chain-like structure position is the best when the cost is the lowest.

The last assumption of the considered algorithm is **CLS position acceptance**. This means that the new position of the chain should be accepted with a probability determined by its cost.
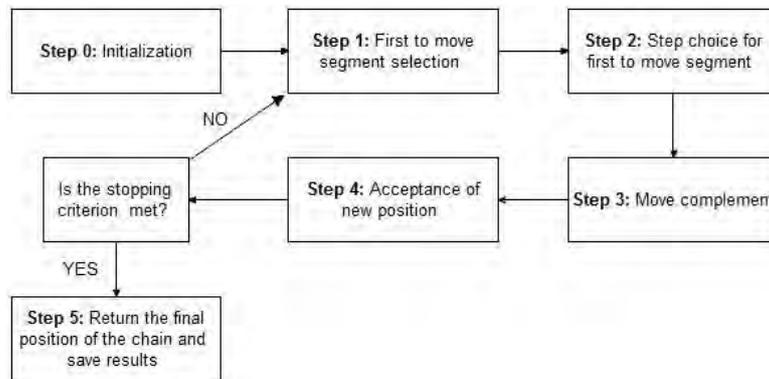

Fig. 1. Block scheme of algorithm for CLS movement simulation

Mentioned assumptions and ideas are implemented in the algorithm for CLS movement simulation. Steps that are taken during the algorithm are presented in scheme in Figure 1 (for more, see [1, 2]).

In Step 0 the space and movement parameters are set. Moreover, the initial position of the structure (c_{curr}) is set by the user (it can also be chosen randomly) and current cost is evaluated by the program. In the next step the FTM segment is selected according to given probability distribution. The new position of the FTM segment ($c_{curr,f}$) is chosen in Step 2. Move complement, Step 3, is the step at which movement of the CLS is performed. During this step successive steps of remaining segments are made to establish a new chain position c_{new}. In the first part of Step 3 the algorithm sequentially chooses segments $c_{new,i}$, $i = f-1,...,1$ and draws randomly neighboring nodes for their next position. It is done for the first k, $k = f-1,..,1$ for which condition $d(c_{curr,k}, c_{new,k+1}) \leq T_d$ is held. If $k > 1$, then $c_{new,i} = c_{curr,i}$ for $i = 1,...,k$. Next the algorithm chooses sequentially segments $c_{new,i}$, $i = f+1,...,n$ and draws at random their next position. This process is terminated for the first k, $k = f+1,..,n$ for which condition $d(c_{curr,k}, c_{new,k-1}) \leq T_d$ is held. If $k < n$, then we assume $c_{new,i} = c_{curr,i}$ for $i = k,...,n$. In Step 4 the new position can be finally verified and accepted (c_{curr} is replaced by c_{new}) or not (c_{curr} does not change). When new position of the chain is generated, the termination condition is verified. This condition depends on the investigated problems, but usually it refers to some requirements that should be fulfilled by the chain position, e.g. whether it reaches a membrane or not. When termination condition is satisfied, the simulation is terminated and the output is returned, otherwise the algorithm returns to Step 1.

2. Simulation framework and results

In this section the simulation framework that was used for the analysis of chain-like structure motion is presented. Movement of the CLS is carried out upwards in the direction of a boundary. The boundary can reflect e.g. a presence of a cellular membrane. The main aim of the simulation studies is to investigate the influence of the introduced above algorithm parameters on the number of steps which are needed to reach the boundary by the chain. This quantity is denoted as NSTB. The initial position of structure, in each case, is placed parallel to the boundary and below it in a distance equals 30 units. It is assumed that the probability of a step in upward direction is two times greater than the probability of a step in another direction. Thus it is assured that the movement is done in the direction of the boundary.

The parameters of movement whose influence on NSTB is investigated are L_{max}, R_{max} and T_d - a tension parameter. It is assumed that GSPU = 2 and L_{min} is equal to 1 unit (2 grid sides). To study the influence of movement parameters on the 'time' reaching the barrier (measured in terms of the NSTB), R_{max} and T_d take on values 3, 6 or 9 units, while parameter L_{max} takes on values 2, 4 or 6 units. In Tables 1-3 the results of the simulation experiment for CLS length equal to 10 units are presented.

Table 1

Mean values of NSTB in relation to the parameters L_{max} and T_d - results obtained for CLS length equal to 10 and R_{max} equal to 3 units

	$T_d = 3$	$T_d = 6$	$T_d = 9$
$L_{max} = 2$	316.5	424.9	419.8
$L_{max} = 4$	341.5	410.2	405.6
$L_{max} = 6$	364.6	424.9	424.0

Table 2

Mean values of NSTB in relation to the parameters L_{max} and T_d - results obtained for CLS length equal to 10 and R_{max} equal to 6 units

	$T_d = 3$	$T_d = 6$	$T_d = 9$
$L_{max} = 2$	130.3	153.1	155.2
$L_{max} = 4$	122.6	139.5	159.1
$L_{max} = 6$	115.8	135.6	165.0

Table 3

Mean values of NSTB in relation to the parameters L_{max} and T_d - results obtained for CLS length equal to 10 and R_{max} equal to 9 units

	$T_d = 3$	$T_d = 6$	$T_d = 9$
$L_{max} = 2$	100.5	96.0	76.5
$L_{max} = 4$	79.9	74.7	74.4
$L_{max} = 6$	69.8	65.1	76.5

It seems to be obvious that the length of the CLS (N) has a big influence on the NSTB too. In simulation three different lengths of structure: 10, 20 and 50 units were studied. In Tables 4-6 results of the simulations for different values of CLS length and $R_{max} = L_{max} + 1$ are presented.

Table 4

Mean values of NSTB in relation to the parameters L_{max} and T_d - results obtained for CLS length equal to 10

	$T_d = 3$	$T_d = 6$	$T_d = 9$
$L_{max} = 2$	316.5	424.9	419.8
$L_{max} = 4$	156.1	210.1	210.4
$L_{max} = 6$	95.3	103.9	122.9

Mean values of NSTB in relation to the parameters L_{max} and T_d - results obtained for CLS length equal to 20

	$T_d = 3$	$T_d = 6$	$T_d = 9$
$L_{max} = 2$	588.3	739.1	747.8
$L_{max} = 4$	238.1	340.5	340.9
$L_{max} = 6$	140.2	150.8	187.0

Mean values of NSTB in relation to the parameters L_{max} and T_d - results obtained for CLS length equal to 50

	$T_d = 3$	$T_d = 6$	$T_d = 9$
$L_{max} = 2$	1275.6	1720.4	1633.4
$L_{max} = 4$	461.2	697.8	682.2
$L_{max} = 6$	240.4	268.5	376.3

Based on results presented in Tables 3-6 one may conclude that the greater the length of structure, the more 'time' CLS needs to reach the boundary - for longer CLS the NSTB is greater in each case of the analogous values of the other considered parameters. However, in analyzing the above results, one can also notice some more interesting features of the investigated relationship. The most significant is the value of the tension parameter. It can be observed that the greater T_d, the slower is the movement (except for the situation when $R_{max} \geq T_d$, see Table 3). Moreover, it can be seen that the larger L_{max}, the less noticeable is the increase of time caused by growth of T_d. Another conclusion is that when $R_{max} > L_{max}$ (see Tables 3-6) increase of L_{max} results in acceleration of the CLS motion. In Tables 1-3 it can be seen that the growth of parameter R_{max} affects the reducing of 'time' which is needed to reach the barrier by the CLS.

To describe these important relations in a more general form, proper metamodels are built.

3. Metamodels for the relation between NSTB, R_{max}, L_{max} and T_d

Typically the data collected during the simulation experiments are possessed by the black-box structure. The simulations transform observable inputs into observable outputs, whereas the values of internal variables and specific functions implied by the simulation's computer modules are unobservable [6]. In the analysis of such input-output a main role is played by metamodels. Metamodeling is a technique which is used to build rather simple and computationally inexpensive models, which replicate the relationships that are observed when samples of a more complicated model or simulation are drawn. Therefore, a simple model that captures only

the relationships between the relevant input and output variables (not modeling any underlying process) is generated. Thus it reduces the burden of the computer [7]. A very general definition of the metamodel is the following: a metamodel f is an approximation of the input/output relating function ϕ that is defined by the underlying simulation model [6]. Various statistical techniques, first of all the regression analysis, are useful to build metamodels on the base of the simulation data. In this section metamodels relating the algorithm parameters N, R_{max}, L_{max}, T_d and the motion speed characteristic NSTB are presented. Such metamodels are generalizations of the metamodels developed in [5], where the impact of the tension parameter on the translocation time was not studied.

The first metamodel is built for $N = 80$. Its assumed form is the following:

$$NSTB = f(N, R_{max}, L_{max}, T_d) + Z, \tag{1}$$

where Z is a random variable (disturbance) with $E(Z) = 0$ and finite variance.

In the simulation experiment the parameters N, R_{max}, L_{max} and T_d were chosen from the following ranges of integers: $N \in [10,100]$, $R_{max} \in [4,20]$, $L_{max} \in [4,20]$, $T_d \in [4,20]$.

It appears that the relation given by f in (1) has a nonlinear character. It is important that the parameters of this function interact, i.e. the effect of one of these parameters depends on the values of the others. On the basis of preliminary data analysis, the following regression function is investigated:

$$f(N, R_{max}, L_{max}, T_d) = \beta_0 + \beta_1 R_{max} + \beta_2 L_{max} + \beta_3 \frac{R_{max}}{T_d} +$$
$$+ \beta_4 L_{max} T_d + \beta_5 N + \beta_6 \frac{N}{R_{max}} + \beta_7 N^2 \tag{2}$$

The estimates b_i of the regression coefficients β_i, $i = 0,...,7$ are as follows:
$b_0 = -103.301$, $b_1 = 14.6474$, $b_2 = -13.2565$, $b_3 = 16.2944$, $b_4 = 0.950054$,
$b_5 = -16.0132$, $b_6 = 302.632$, $b_7 = -0.0270474$

The built model has really good statistical characteristics. For example its coefficient of determination is equal to $R^2 = 0.9781$ and all explanatory variables have significance below $p = 10^{-6}$ while usually $p = 0.05$ is assumed to be good enough.

The metamodel (2) illustrates well the character of the relations between parameters of the algorithm and their mutual influence on the CLS motion speed. For example it can be seen that impact of the T_d depends on values of parameters R_{max} and L_{max}. Such a good description of the parameters' effect on studied phenomena allows for a proper selection of the algorithm parameters to improve the quality of the final model in its real-world application.

Conclusions

This paper was devoted to a simulation study of the CLS translocation time. The main result is the metamodel (2) which is based on the results of the simulation experiment. This metamodel is more general than those developed in [5]. It allows one to relate the translocation time not only with the algorithm parameters R_{max}, L_{max} but also with the parameter T_d that reflects the tension propagation mechanism. Consequently, (2) can be used for studies on the dependencies between the elasticity of the CLS and its movement speed. This relationship is very important, because it allows the researchers better modeling of a real processes that can be observed in studying the behavior of the CLS under various conditions.

References

[1] Grzybowski A., Domański Z., A sequential algorithm for modeling random movements of chain-like structures, Sci. Res. Inst. Math. 2011, 1(10), 5-10.

[2] Grzybowski A.Z., Domański Z., A sequential algorithm with built in tension-propagation mechanism for modeling the chain-like bodies dynamics, arXiv:1312.4206 [cond-mat.soft].

[3] van Leeuwen J.M., Drzewiński A., Stochastic lattice models for the dynamics of linear polymers, Physics Reports 2009, 475, 53-90.

[4] Żurek S., Kośmider M., Drzewiński A., van Leeuwen J.M.J., Translocation of polymers in a lattice model, The European Phys. J. E: Soft Matter and Biological Physics 2012, 35, 47.

[5] Grzybowski A.Z., Domański Z., Bartłomiejczyk K., Algorithmization and simulation of the chain-like structures' dynamics - interrelations between movement characteristics, Acta Electrotechnica et Informatica, 14, 3, in print.

[6] Kleijnen J.P.C., Design and Analysis of Simulations in Statistical Physics, Cambridge University Press, Cambridge 2005.

[7] Knowles J., Nakayama H., Meta-modeling in multiobjective optimization, [in:] Multiobjective Optimization. Interactive and Evolutionary Approaches, (eds.) J. Branke, K. Deb, K. Miettinen, R. Słowiński, Springer, 2008, 245-284.

[8] Rubinstein M., Discretized model of entangled-polymer dynamics, Phys. Rev. Letters 1987, October, 59, 1946-1949.