A FAST PROCEDURE FOR MODELING OF PROTEASOMAL PROTEIN DEGRADATION IN VITRO

Antyufeev V.S.*1, Nikolaev S.V.2

1 Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia;
2 Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: nikolaev@bionet.nsc.ru
* Corresponding author: e-mail: ant@osmf.sscc.ru

Keywords: proteasome, protein degradation, mathematical model, computer analysis

Summary

Motivation: Modeling of complex processes with stochastic elements frequently involves computer-based simulation models. Calculation of the statistical characteristics (parameters) of the modeled processes often is time consuming. For this reason, we attempted to analytically calculate the distribution of the required parameters of the complex processes.

Results: A simple model for proteasomal proteolysis has been constructed as a Markov chain. Formulas were derived for estimating the distribution of the fragments resulting from proteolysis. To do it, we built the appropriate phase space which allowed us to transform the modeled process into Markov chain, and to calculate the limit probability of its absorbing states. This enabled us to reduce by hundreds of times the time for modeling of the proteolytic process and to achieve good calculation accuracy in an acceptable time.

Introduction

A living cell contains a variety of proteins with a broad range of lifetime. If a protein exists, but the cell does not need it any longer, the cell gets rid of it by proteolysis (degradation, or enzymatic cleavage of proteins). About twenty years or so ago, a high molecular protein complex working as a proteolytic enzyme was found in the cell. The proteasome is cylinder-shaped with a channel along its axis. The channel diameter suffices to accommodate the protein molecule within the channel and to allow its movement. The “cleaving unit” of the proteasome is located at the inner surface of the channel, and it can split the protein molecule or its fragments between the amino acids.

Many details of the protein degradation process remain unknown. There is a host of intriguing questions: Does protein degradation start from the N- or the C-terminus? Is degradation processive, or the proteasome-protein complex keeps dissociating-reassociating in a protein undergoing degradation? What mechanism provides the movement of a degrading protein relative to a proteasome? The relevant experiments are carried out *in vitro* using 20S proteasomes and unfolded proteins (a protein with a linear spatial structure or a protein unfolded by a denaturating agent). Apparently, the protein substrate-proteasome relative movement is not provided by a known molecular motor driven by the ATP cleavage. The distribution of the length of the resulting fragments is described by a unimodal curve with a maximum in the range of 6–10 amino acids (Kisselev et.al., 1998).

Our goal was to model proteasomal proteolysis. Attempts were made to answer questions, such as: Can the diffusive relative protein substrate-proteasome movement provide the splitting of the entire protein molecule into fragments? What cutting parameters produce a distribution of fragment length similar to the one experimentally observed?

The developed simulation model took too much time for computer-based experiments aimed at finding the distribution the fragments resulting from the proteolysis. For this reason, we turned to derivation of formulas expressing the distribution of the characteristics of the process.
Model

The following model of the proteasomal protein degradation process is considered.

We assume that there are numerous protein copies and some proteasomes in a solution at the beginning of the degradation process. As protein degradation proceeds, the fragments of protein copies appear in the solution. To model the degradation process, the points between neighboring amino acids in a protein sequence are considered rather than the amino acids as such. Every point is a possible point of protein molecule cleavage.

Because there are no universally recognized models of protein substrate movement relative to the proteasome, we postulate that this movement is a one-dimensional diffusion effected by thermal Brownian movement of the protein substrate along the proteasomal channel. During such Brownian movement, intermolecular interaction causes random gluing of the amino acids of the protein substrate and the ones of the proteasome. This gluing is the necessary but not a sufficient condition for proteasome to cleave the substrate molecule into two fragments. We suggest that the proteasome may cleave a protein or a protein fragment only if there are substrate-proteasome gluings at both sides of the “cleaving unit” of the proteasome, and the cleavage is random.

Algorithm for modeling

First we recall the basic steps of the proposed model. Suppose there is a pool of \(N\) protein fragments in a solution at a time. A fragment of the pool inputs in a proteasome with the probability \(1/(2N)\) (the coefficient \(1/2\) reflects the random choice of the N- or C-terminus of a fragment), and moves randomly in the proteasomal channel along it axis. During this movement, the fragment may be cleaved or it may exit from the proteasomal channel, being not cleaved. It is noteworthy that modeling of a fragment movement relative to the proteasome is equivalent to modeling of the movement of the proteasomal “cleaving unit” relative to the fragment. Modeling of the entire process is stepwise: 1. Choice of a number and of a terminus of the fragment, which is set to be in contact with the proteasome. 2. Model of random discrete steps of the proteasomal “cleaving unit” in either direction along the chosen fragment until the fragment becomes detached. 3. Model of gluing of two or more substrate amino acids to the proteasome (see the model description above). 4. If two or more substrate amino acids fixed to the proteasome at both sides of the proteasomal “cleaving unit”, then the division of the fragment into two fragments at the point under the “cleaving unit” is modeled. 5. If the fragment is cleaved into two smaller fragments, then the fragment that would leave the proteasome and the one that would remain in it are modeled. The detached fragment joins the substrate pool. Run the algorithm from point 2 with the remaining fragment. 6. If a fragment will not be divided during its movement through the proteasomal channel and the fragment detaches from the proteasome, then the fragment joins the substrate pool. Return to point 1 of the algorithm.

The choice of the division point is two-step: (i) the choice of the fragment to be divided, and (ii) of the division point in the preferred fragment. The fragments in the pool and the points of the possible division in the fragments are numbered. Then, the task is to sequentially model two discrete random values: the divided fragment number, and, next, the division point number. Once the discrete distributions are found, the time for modeling of the process in its entirety becomes much reduced. The two random processes can be expressed as absorbing Markov chains (Kemeny, Snell, 1969). For this purpose, one has to build a phase space of the process involving absorbing and non-absorbing states, transition and absorption probabilities. Let us consider modeling of the fragments resulting from division. The phase space consists of fictitious elements of the \([k,a,s]\) type, where \(k\) relates to the fragment number, \(a\) to the absorbance index; \(a=1\) or 0, depending on whether the \(k\)-th fragment will be divided or not respectively; \(s\) denotes the particular (the right or left) side of the fragment penetrating into the proteasomal channel (note, after a next fragment has
been divided, the phase space changes, and the next division points are chosen in it; this means that the entire process is not Markov chain. Because each subsequent step depends on all the preceding, the choice of the next division point depends on how they all were chosen). Suppose that there are $N$ fragments, the phase space consists of $4N$ elements.

The transition matrix of the chain falls into blocks:

$$P = \begin{bmatrix} A & B \\ 0 & E \end{bmatrix}.$$ 

All the square submatrices are of size $2N \times 2N$, $E$ – unitary matrix,

$$A = \begin{bmatrix} p_1q_1 & p_2q_1 & \cdots & p_{2N}q_1 \\ p_1q_2 & p_2q_2 & \cdots & p_{2N}q_2 \\ \vdots & \vdots & \ddots & \vdots \\ p_1q_{2N} & p_2q_{2N} & \cdots & p_{2N}q_{2N} \end{bmatrix} \quad B = \begin{bmatrix} 1-q_1 & & & \\ & 1-q_2 & & \\ & & \ddots & \vdots \\ & & & 1-q_{2N} \end{bmatrix}.$$ 

Here, $(1-q_i)$ stands for the probability of division of $i$–th fragment, all of the $p_i = 1/(2N)$.

The idea is to calculate the limit distribution for the probabilities of the absorbing states for the chain. The vector of the limit distribution is

$$u^* = \lim_{n \to \infty} uP^n.$$ 

By induction, we prove that

$$P^n = \begin{bmatrix} A^n & (E + A + K + A^{n-1})B \\ 0 & E \end{bmatrix}.$$ 

Since $\|A\| < 1$, $\lim_{n \to \infty} (E + A + K + A^{n-1}) = (E - A)^{-1}$.

Hence

$$P^* = \begin{bmatrix} 0 & (E - A)^{-1}B \\ 0 & E \end{bmatrix},$$

and $uP^* = u \begin{bmatrix} 0 & (E - A)^{-1}B \\ 0 & E \end{bmatrix}$.

To calculate the above limit distribution, the following steps should be performed: (i) calculation of the vector $x$, as a solution of the linear equation set $(E - A)x = u$, where $u$ – vector of the initial probability distribution, and (ii) calculation of the vector $xB$. In addition, we found a formal solution of the linear equation set. Thus, calculation of the distribution and modeling of the discrete random value with known distribution can be used instead of computer simulation of the fragment choice. The calculation of the probability distribution for the division points in the fragment is similar. Thus, the time consuming algorithm 1–7 can be replaced by fast calculation, followed by modeling of the discrete distributions. The consequence of this improvement is a decrease in the calculation time by two orders of magnitude.

**Results and Discussion**

The appropriate phase space is the major result of the work. The expansion of the phase space of the process by including the “artificial” elements, such as $[k,a,s]$, along with the “usual” elements (such as fragment and division point numbers) allowed us to express the random process as an absorbing Markov chain. In this way, the limit probabilities of the absorbing states were calculated. Based on the defined probabilities, the modeling time was reduced by two orders of magnitude, and calculation accuracy was achieved in an acceptable time.
Acknowledgements

The work is supported by Russian Ministry of Industry, Sciences and Technologies (grant No. 43.073.1.1.1501), RAS Presidium Program “Molecular and Cellular Biology” (No. 10.4).

References