BIOMOLECULAR CRYSTALLIZATION IN SILICO AS A SOURCE OF THE COLOUR NOISE*

JACEK SIÓDMIAK

Department of Physics, Institute of Mathematics and Physics
University of Science and Technology
Al. prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland

(Received June 16, 2016; accepted November 7, 2016)

Here are presented results of the coarse-grained Metropolis Monte Carlo simulation of the random walk performed by the biomolecular growth unit at the crystal surface during the protein crystallization process. The surface-to-solution energy exchange that occurs during the movement of the “walker” can be considered in terms of colour noise. The presence of the noise can impede the interpretation of the results obtained e.g. by the dynamic scattering methods. To characterize the noise, the power spectra are generated and analysed. The characteristic non-universal noise exponent is temperature- and surface morphology-dependent.

DOI:10.5506/APhysPolB.47.2399

1. Introduction

Crystallization of biomolecules is a complex process. In the initial phase, spontaneous aggregations occur which results in the formation of the disordered nucleus. Subsequently, the self-organization of the nucleus is followed by appropriate crystal growth. When the size of a crystal reaches a few micrometers and the crystal surfaces are well-shaped, due to surface defects, terraces may emerge [1–4]. The model of biomolecular crystal growth in silico assumes that mesoscopic macromolecular unit, that migrated from the solution onto the crystal surface, has to move to positions localized at the foot of the terraces which are the most favourable from the thermodynamical point of view [5].

Biomolecular crystals are of the non-Kossel type (complex structure with several molecules per unit cell in non-equivalent positions [6]), therefore, their surface is not uniform and the movement performed by biomolecules

* Presented at the XXVIII Marian Smoluchowski Symposium on Statistical Physics, “Thermodynamics in a Nanoscale”, Kraków, Poland, September 14–17, 2015.
is not a simple random walk. As in the vein of the material presented in [7], also in our simulation we do not judge unequivocally whether our random walk takes on normal or abnormal behaviour. Each step involves breaking of bonds and the creation of entirely new ones [8], and is executed with probability proportional to the Boltzman factor, cf. Eq. (1) in Sec. 3. Therefore, in contrast to the classical random walk, each step is performed at a different rate. As a consequence, the crystal growth rate is also altered (perturbed accordingly) [9]. While the classical random walk generates white noise, this type of motions can be a source of colour noise, see [10, 11] and references therein. In the model presented here, there is a number of factors that determine the colour of the noise. This may include (i) direct factors: hydrophobicity and hydrophilicity of the surface, the #H to #P ratio for the given morphology, the presence of the “sliding” pathways and (ii) indirect factors: temperature, pressure, pH, etc.

There are many methods of measurement that allow tracking of individual molecules, e.g. single molecular tracking, which are very sensitive to any fluctuations of the measured parameters [12]. In the model presented here, the walker-to-surface interacting energy varies at most in the range from −9.2 to 9.2 in $k_B T$ units. As a result of miniscale energy changes, the temperature fluctuations are also limited, but these fluctuations are repeated many times in one terrace’s place (flipping around local energetic minimum) and can be detected by measuring devices. Similar situations can be observed in the case of position tracking. If the molecule is trapped in the local energetic minimum, its position can be changed very fast. The molecule attempts to get away from the trap and jumps from one to neighbouring position and back many times before it gets “free”.

The paper is constructed as follows: first, an introduction to the protein surface diffusion problems is carried out, see Sec. 1. Next, the coarse graining procedure is explained, Sec. 2. In Section 3, the simulation procedure is described. In Section 4, the results are presented and, finally, in Section 5 a summary is carried out.

2. The walker and terrace model

2.1. The walker preparation

The simulation has been performed on a three-dimensional square lattice, and for this reason, the walker has been modelled as a cube. To keep in the model some amphiphilic properties of the living molecules, each wall (side) of the cubic walker must have different level of hydrophobicity and/or hydrophilicity.
The walker preparation procedure was similar to that described in [13]. For this simulation, we have chosen one of the most popular model protein, i.e. lysozyme. The structural information on this protein was read out from the Protein Data Bank (PDB) [14]. The full atom view of tetragonal hen egg white lysozyme (PDB ID: 193L) is shown in Fig. 1 (second from the left). At the beginning, the protein molecule has been inserted into the cube of which linear size \( d \) is equal to mean diameter of the protein. In this case, \( d = 3.22 \times 10^{-9} \) m.

Fig. 1. The walker preparation procedure: from full atom representation to coarse grained “toy-brick” unit. H and P symbols indicate how many H-type and P-type monomers appeared on each wall of the box.

Next, the molecule has been reoriented inside the cube adequately to the orientation of the molecules in a lysozyme crystal of tetragonal form [6]. Once the orientation has been succeeded, the HP (Hydrophobic/Hydrophilic) representation of the molecules was carried out [15, 16]. Based on the amphiphilic properties of the amino acids forming proteins [17], all of them were divided into two groups: H (hydrophobic) and P (hydrophilic). Because the molecules are flexible and interact (in first approximation) through the surface, we assume that the interaction takes place as if it were within the protein “skin” (peel) the thickness of which is taken to be equal to 25% of the molecule radius. Such “isolated peelings” of the protein molecule have been projected on the walls of the surrounding cube (like a blowing up balloon in a cubic box). Next, we have counted how many H-type and P-type monomers appeared on each wall of the box. The excess in a number of one of the monomer types (H or P) unambiguously determines the type of the wall of the surrounding box [13].

Here, in contrast to the previously presented model [13, 18, 19] (where two oppositely located and purely hydrophilic walls were neglected), all walls have been taken into consideration.

2.2. The terrace

In this study, we assumed that the crystal is in its mature growing stage. It signifies that the crystal has well-formed crystallographic planes and its growth is controlled by surface motion of biomolecules [1, 4, 6]. In the mature state of growth, as a result of some structural defects (e.g. screw
dislocations), as well as some impurities, terraces can be formed. Moreover, we assumed that the linear size of the terrace is much larger than the linear size of the walker and, therefore, in some cases the influence of kinks on the walker (vicinity of the Ehrlich–Schwoebel barrier [20, 21], see Fig. 6) can be neglected. According to the structure of the non-Kossel crystals and the data obtained from the PDB, the modelled surface has been built using eight biomolecules (cubic walkers) located at positions providing the minimum energy within a single unit cell.

The intra (unit cell) energy has been calculated based on the HP-based “toy brick” model described in [22]. In this model, the interaction energy between monomers of H-type has the value $E_{HH} = -2.3$, between monomer of H- and monomer of P-type $E_{HP} = -1$ and between monomers of P-type $E_{PP} = 0$. All energies are given in the $k_B T$ units. Because of the flexibility of the biomolecular chain, we have assumed that all possible contacts between H and P monomers can be created. All contacts have been formed in the following order: first, all H-to-H contacts have been counted (if any, because some walker’s walls are purely hydrophilic), next, if one of the walls still had free hydrophobic monomers, they have formed a bond with free hydrophilic monomers of the second walker and, finally, all free hydrophilic monomers have formed non-energetic contacts. Unpaired monomers do not contribute to the total energy of the unit cell, see Fig. 2. The same procedure must be repeated for all walls in contact. The total energy is equal to the sum of energies calculated for all interacting walls.

![Fig. 2. An example of the wall-to-wall contact energy calculation. In this case, $E = -5.6$ in $k_B T$ units.](image)

The final structure of the unit cell characterized by minimum energy $E_{UC} = -64.4$ is shown in Fig. 3 (right). It must be mentioned that this procedure gives four different configurations with the same minimum energy. Moreover, it is possible to build the terrace (surface) using the same unit cell reoriented (as all) into different six positions. The total number of possible minimum energy surface morphologies is 24. For all of the obtained structures, four types of the morphology have been distinguished, see Fig. 4: stripes alternating with bars, uniform, stripes and chessboard-like.
3. Random walk procedure

In this work, the square, $N \times N$ terrace has been used, where $N = 2000$. The maximum number of Monte Carlo steps has been set at $2 \times 10^6$. The simulation starts with putting probe particle in the middle of the terrace. Next, one of the 24 possible orientations has been randomly chosen and assigned to the walker. The initial arrangement (bottom wall and orientation) is very important because it determines the present and future interactions between the walker and the terrace. For the same posture (the same bottom wall) but different orientation, interaction energy after the roll over the same direction will be different\(^1\).

\(^1\) Let us consider classical right-handed dice. At the beginning, the dice is placed on “2” and oriented with “1” to us. In this position, the “top” number is “5”. Now, let us roll it over to the right. After this transition, bottom number is “3”, and top is “4”. Now arrange cubes in a way that the bottom number is again “2” but front number is “3”. After rolling it to the right, the bottom number is “6” and top number is “1”. Therefore, the same initial posture but different orientation after the same sequence of rolling events will give different energetic pathway and different final state. And likewise, it is possible to obtain the same final configurations after different sequence of roles.
In the next step, the movement direction has been randomly chosen. If the influence of the Ehrlich–Schwoebel barrier on the walker is taken into consideration, a small drift in the vicinity of the kinks/steps (V-type drift) or the constant drift can be added, see Fig. 5. After each draw of the direction of motion, the movement probability, $p_m$, has been calculated according to the Metropolis Monte Carlo (MMC) acceptance rules [23]:

$$p_m = \begin{cases} 1 & \text{if } \Delta E \leq 0, \\ e^{-\frac{\Delta E}{k_B T}} & \text{if } \Delta E > 0. \end{cases}$$

In MMC algorithm, downhill transitions that lower the energy ($\Delta E \leq 0$, where $\Delta E$ is a difference in the interaction energy between walker and the site on the surface in present and new locations) are accepted with probability one and uphill transitions with probability proportional to the Boltzmann constant, $k_B$. The simulation can be performed for different temperatures $T$. Mentioned above interaction energies are proper for $T = 300$ K. For different temperatures, the values of interaction energies must be rescaled.

![Fig. 5. Drifts’ models: no drift, constant drift and V-type drift. The arrow length indicates movement probability: longer arrow means that the movement to the given direction is more probable.](image)

The models of the drift that have been used in the simulation are presented in Fig. 5. If we assume that the terrace is small enough to take into consideration the influence of the Ehrlich–Schwoebel barrier, a small constant drift is to be added. When the terrace is big enough to neglect the long-range influence of the Ehrlich–Schwoebel barrier, a V-type drift (short-range repulsion from the edge of the lower terrace and short-range attraction to the kinks of the upper terrace) is added to the system. The scenario without drift has been also studied.
4. Results

The random walk of the probe particle has been realized for the temperature range from $T = 260$ K $\div 340$ K and for the three models of superimposed drift. To avoid the application of the boundary conditions, the length/width of the terrace has been equal to 2000 single units, i.e. $1000 \times 1000$ unit cells, what indicates that the terrace is in the mature state of growth (the minimum length/width of the terrace is 17 single units and the average distance between two neighbouring steps on the mature crystal surface is equal to 34 single units [4]).

4.1. The energy exchange

As seen in Fig. 6, each step binds to the exchange of energy between the walker–surface interface (in principle, the crystal) and the bulk. The energy exchange $\Delta E$, cf. Eq. (1), can be positive as well as negative. If $\Delta E > 0$, this means that to perform a step, a small amount of energy must be taken from the solution to the walker–crystal interface. If $\Delta E < 0$, this means that a small amount of energy was dissipated into the solution. This amount of energy causes a thermodynamical disturbance in the solution surrounding the walker–crystal interface. This phenomenon indicates that the movement

Fig. 6. The model of terrace with moving probe particle (walker). The possible path performed by the walker is shown. Each move is realized by random choice of the direction and the roll over of the walker from one to the neighbouring site with probability given by Eq. (1). The periodic Ehrlich–Schwoebel potential characterizing the terrace structure is shown (black curve on the side of the terrace), but in the model, its influence can either be neglected or can be added in the form of superimposed drift.
of the biomolecules and, therefore, the crystal growth, can be described in terms of an out-of-equilibrium (local) thermodynamic process. The $\Delta E$ changes during the simulation are presented in Fig. 7. It can be seen that the energy exchange fluctuates and the recorded signal exhibits noisy character of the process. The values of $\Delta E$ are from the range corresponding to the minimum and maximum values of the walker-to-surface interaction energy. Moreover, because the model presented here is coarse grained and is spread on the square lattice, $\Delta E$ takes discrete values. It can be also seen that in this plot occasionally occur quite long periods in which signal fluctuates in a small range (e.g., in the interval from 2000 to 3000 time steps). This means that the walker remained in the local energetic minimum.

![ENERGY CHANGES](image)

Fig. 7. The walker-to-surface energy change during the simulation.

4.2. The power spectrum of the energetic passes

It is known that in the case of the Monte Carlo simulations, there is a problem with estimation of the time [7]. In our case, the problem of the time estimation can be steered clear. To this end, the surface diffusion coefficient of the single biomolecule must be obtained from the experimental observation. Next, the time step can be defined as follows:

$$ t_s = \frac{\langle x^2 \rangle}{4D_S} , $$

(2)

where $D_S$ is an experimentally determined surface diffusion coefficient. The mean square displacement $\langle x^2 \rangle$ is usually evaluated by the second central moment of the probability distribution function $P(x, t)$ [24]

$$ \langle x^2 \rangle = \sum_i x_i^2 P(x_i, t) . $$

(3)
Based on the experimental data, the above parameters take the following values (see [13] and references therein): the mean square displacement $\langle x^2 \rangle = 4R_H^2 = 1.44 \times 10^{-17}$ m$^2$ (where $R_H = 1.9 \times 10^{-9}$ m is a hydrodynamic radius of the biomolecule), $D_S = 4.97 \times 10^{-13}$ m$^2$/s and, finally, the real time step $t_s = 7.25 \times 10^{-6}$ s. It means that for the random walk of the lysozyme protein at the crystal surface, one Monte Carlo step in the simulation corresponds to $7.26 \times 10^{-6}$ s in real (clock) time.

To characterize the noise-like signal presented in Fig. 7, the distribution of power into frequency components composing the resulting time series (the power spectra) has been generated. The spectral density (PSD), $S(f)$, of such a stochastic process, has the form of

$$S(f) = \text{constant}/f^\alpha,$$

where $f$ is frequency on an interval bounded away from both zero and infinity.

The PSD plots and $\alpha$-exponents were obtained using a Matlab package. An exemplary power spectrum of the energy change signal is presented in Fig. 8 (grey/green curve). This power spectrum has been obtained under the assumption that the energy change $\Delta E$ was related to the unit cell bonding energy. It means that the $\Delta E$ values presented in Fig. 7 represent deviations of the unit cell to “walker” bonding energy.

![POWER SPECTRUM + c/f^\alpha](image)

Fig. 8. (Colour on-line) The power spectrum of the energy change signal (grey/green) and three approximations of the power spectral density proportional to $1/f^\alpha$ for $\alpha = (0.9, 1.0, 1.1)$; $c$ — constant.
In the same figure, three additional curves have been plotted as the approximation of the power spectral density proportional to $1/f^\alpha$ for $\alpha = (0.9, 1.0, 1.1)$. Already in the first approximation, it can be seen that the power spectral density obtained for the signal presented in Fig. 7 is proportional to the $1/f^\alpha$, where $\alpha \approx 1$. It means that the walker–surface interaction energy changes can be characterized in terms of colour noise. For $\alpha \approx 1$, the noise is called pink. If energy change $\Delta E$ is related to the solution, the noise characteristics will be moved towards blue $\sim f$ or even violet noise $\sim f^2$. As was mentioned above, the temperature can influence the crystal growth process. So, it can be expected that the power spectrum of the observed walker–surface interaction energy changes will be also altered. Figure 9 presents the power spectrum exponent $\alpha$ as a function of temperature for three models of the superimposed drifts.

![Fig. 9.](image)

In Fig. 9, it is shown that the shape of the curves representing the temperature dependency of the $\alpha$-exponent and $c$-constant changes in a similar way when no drift is added and when drift is kept constant. In the case of V-type drift, however, the bending of the curve is in the opposite direction. The visible plateaus for the case with no drift and constant drift correspond to the crystallization window obtained in [13].

As was mentioned above, it is possible to build up the terrace using 24 equienergetic unit cells. For all of them, the value of the power spectrum exponent (calculated with respect to the unit cell energy) varies around one. The differences are becoming visible after analysing histograms, Fig. 10.
Fig. 10. The comparison of the histograms for six different equienergetic unit cells constituting crystal’s surface. The number of allowed energetic passes (including those for which $\Delta E = 0$) — energetic resolution $\Delta E_{\text{res}}$ — is surface morphology-dependent, namely: for stripes alternating with bars (a) $\Delta E_{\text{res}} = 33$ and (b) $\Delta E_{\text{res}} = 23$, for uniform surface (c) $\Delta E_{\text{res}} = 13$ and (d) $\Delta E_{\text{res}} = 11$, for striped morphology (e) $\Delta E_{\text{res}} = 25$ and for chessboard-like morphology (f) $\Delta E_{\text{res}} = 9$.

The main difference is in energetic resolution $\Delta E_{\text{res}}$ (the number of allowed energetic passes). It was observed that different morphologies can be characterized by different energetic resolution. But within one morphology, it can be observed that $\Delta E_{\text{res}}$ varies and takes different, sometimes very spread, values. For the stripes alternating with bars morphology Fig. 10 (a) $\Delta E_{\text{res}} = 33$ and Fig. 10 (b) $\Delta E_{\text{res}} = 23$, for uniform surface Fig. 10 (c) $\Delta E_{\text{res}} = 13$ and Fig. 10 (d) $\Delta E_{\text{res}} = 11$, for striped morphology Fig. 10 (e) $\Delta E_{\text{res}} = 25$ and for chessboard-like morphology Fig. 10 (f) $\Delta E_{\text{res}} = 9$. Jux-
tapping all passes for the aforementioned cases it can be seen that the total number of passes is also different. The efficiency varied from 10 to 30 percent. The comparison of $\Delta E_{\text{res}}$ and efficiency for different morphologies is presented in Table I.

### Table I

<table>
<thead>
<tr>
<th>Surface morphology</th>
<th>$\Delta E_{\text{res}}$</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stripes alternating with bars I</td>
<td>33</td>
<td>10</td>
</tr>
<tr>
<td>Stripes alternating with bars II</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>Uniform</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td>Stripes I</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>Stripes II</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>Chessboard-like</td>
<td>9</td>
<td>30</td>
</tr>
</tbody>
</table>

It can be seen that low energetic resolution favours mobility (higher efficiency). Similar effect was observed in [18] and [19], where striped morphology of the surface causes that during the crystal growth one of the crystal sides grows faster. As a consequence, elongated crystals were detected.

### 5. Summary and discussion

It has been demonstrated that (i) using course-grained model of the biomolecule and (ii) taking into account the basic amphiphilic interactions, the movement of protein at the crystal surface can be modelled. The first observation was that it is possible to construct four different unit cells characterized by the same bonding energy. Employing the same unit cell type reoriented (as all) into six positions, 24 different surfaces were examined. The surfaces were characterized by different morphologies, see Fig. 4. In some cases, certain “sliding” pathways (stripes of the same oriented units in the top layer of the surface/crystal) were observed. The presence of the uniform pathways tends to promote higher mobility, see Table I. In comparison to the previously presented two-dimensional unit cell model [13], where purely hydrophilic sides of the walker were neglected, the number of equienergetic morphologies is reduced from 42 to 24. Such formal improvement of the algorithm reduces the number of virtually redundant exploited configurations by a factor of 40%.

According to the MMC rules, the surface mobility of the biomolecule contributes to the energy exchange between protein–surface system and the bulk (solution). The analysis of the energy exchange as a function of Monte Carlo steps, and after applying some experimental assessed data as a function of time, indicates that the $\Delta E(t)$ presents the characteristics of the
noise. The power spectra of these signals can be described in terms of the colour noise. For all examined surface morphologies, the PSD exponent $\alpha$, cf. Eq. (4), varies around one, i.e. $S(f) \sim 1/f$. The $1$ over $f$ characteristics of the noise is characteristic for the pink noise or that of the flicker noise [26, 27]. Referring $\Delta E$ changes to the infinite bulk, it is possible to obtain the PSD characteristics $\sim f$ of even $\sim f^2$, i.e. the noise moves to the blue and violet colour.

It has been also shown that the surfaces with different morphologies feature various energetic resolutions, see Fig. 10. So far, there is no clearly seen correlation between specific surface morphology and the energetic resolution (some surfaces in the same group of morphologies reveal different $\Delta E_{\text{res}}$). Account must be taken of such parameters as: hydrophobicity and hydrophilicity of the surface, the $\#H$ to $\#P$ ratio for the given morphology, the presence of the “sliding” pathways. It is possible that the use of Principal Component Analysis (PCA) [28, 29], successfully used for the analysis of multidimensional data, will allow us to find some important correlations.

An indication that the protein crystal growth can be accompanied by the noise generated by the moving biomolecules can provide useful information in the case if you are using more and more sensitive measuring instruments.

Future work on macromolecular crystallization in silico should be focused on the reducing of the number of redundant configurations. This can be also accomplished by changing of the unit type from brick-like (6 faces) to the (semi-spherical) fullerene-like symmetry (32 faces).

The author gratefully acknowledges the Organizers of the XXVIII Marian Smoluchowski Symposium on Statistical Physics, Thermodynamics in Nanoscale, Kraków, Poland, September 14–17, 2015 for the opportunity of the poster presentation of his work. The author thanks Prof. Ewa Gudowska-Nowak (Jagiellonian University in Kraków) and Prof. Adam Gadomski (UTP Bydgoszcz) for preliminary encouraging and fruitful discussions. The author would like to thank Ms. Tahlia Casey from Delaware State University for the proofreading of the manuscript. This work was supported by an internal grant No. BS39/14 of the Institute of Mathematics and Physics of the University of Science and Technology in Bydgoszcz.

REFERENCES
