

EPJ B

Condensed Matter
and Complex Systems

EPJ.org
your physics journal

Eur. Phys. J. B (2019) 92: 253

DOI: [10.1140/epjb/e2019-90741-6](https://doi.org/10.1140/epjb/e2019-90741-6)

Investigation of DNA denaturation in Peyrard-Bishop-Dauxois model by molecular dynamics method

Ilya Viatcheslavovitch Likhachev and Victor Dmitrievitch Lakhno

edp sciences



 Springer

Investigation of DNA denaturation in Peyrard-Bishop-Dauxois model by molecular dynamics method

Ilya Viatcheslavovitch Likhachev^a and Victor Dmitrievitch Lakhno

IMPB RAS, 1, Professor Vitkevich St., 142290 Pushchino, Moscow Region, Russia

Received 18 December 2018 / Received in final form 15 April 2019

Published online 13 November 2019

© EDP Sciences / Società Italiana di Fisica / Springer-Verlag GmbH Germany, part of Springer Nature, 2019

Abstract. The phase transition of (PolyA/PolyT)₁₀₀ duplex into the denaturated state is studied in the Peyrard-Bishop-Dauxois model by the method of direct molecular-dynamical modeling. The temperature dependencies of the total energy and heat capacity of the duplex are calculated. The convergence of the dependence of heat capacity on the number of realizations is shown. The approach applied can be used to calculate the statistical properties of the duplexes of any length and nucleotide composition.

1 Introduction

In modeling molecular dynamics of biomacromolecules use is made of interatomic potentials which are always limited in value in the case of large distances between the atoms (Lennard-Jones, Coulomb, Morse, etc. [1]). This involves some problems when calculating the thermodynamic properties of biomacromolecules. Indeed, the statistical sum always diverges for bounded potentials. This means that macromolecules should always be in denaturated state at arbitrarily low temperature. However actually this is not the case since the concentration of molecules in a solution is finite which determines a characteristic cutoff scale in the calculations of the statistical sum.

If we formally deal not with an ensemble, but with a single molecule in an infinite solvent and the temperature is much lower than the denaturation temperature then potential energy always reaches a stationary quasi-equilibrium state which exists far longer than any experiment concerned. This equilibrium state will naturally coincide with the state which we would obtain dealing with an ensemble of molecules.

Hence the problem of modeling is not a formal calculation of the statistical sum over infinite time corresponding to the decomposed system, but the calculation leading to just the quasi-equilibrium state and finding of its thermodynamic characteristics.

A problem, however arises when the temperature appears to be close to the value for which two characteristic times become equal (i.e. the time during which potential energy reaches the stationary state and the time of the system's decomposition).

This situation arises, for example, in calculations of the thermodynamic characteristics of DNA whose

decomposition temperature not far exceeds the room one. In this case a direct calculation of the statistical sum is impossible, since it leads to a decomposition of the molecule during a finite time.

This paper just deals with consideration of this problem. We describe a method which can be effectively used to calculate the thermodynamic characteristics of the molecule in the vicinity of the transition and the transition *per se*.

Melting (denaturation) of DNA [1] occurs as a result of amplification of temperature fluctuations. In this case both the chains move away from each other breaking the hydrogen bonds. The chains *per se* remain nonseparable.

A popular dynamical model of DNA describing such behavior of the duplex – the Peyrard-Bishop-Dauxois (PBD) model – simulates both the behavior of breaking hydrogen bonds (Peyrard-Bishop model [2]) with the use of the Morse potential and the inseparability of each chain due to Peyrard-Dauxois addition [3,4] and steel popular [5]. The thermodynamics behavior of this model is in good agreement with experimental data. By this model the melting temperature is calculated by transfer-integral (TI) method and molecular dynamics simulation method – 361.5 K. This model is also used for bubbles search [6]. The nature of bubbles research and the nature of heat capacity peak are the same – destruction of hydrogen bonds.

The work studies in detail the dependence of the heat capacity on the number of realizations and the convergence of the maximum heat capacity to a smooth peak.

2 Methods and results

In this model a chain of DNA nucleotides is presented as a system of material points in a one-dimensional space whose motion is described by classical motion

^a e-mail: ilya.likhachev@mail.ru

equations:

$$m_i \frac{d^2 x_i}{dt^2} = - \frac{dU}{dx_i}$$

where $i = 1, \dots, N$; N is the number of particles; m is the reduced mass of a particle; x is the coordinate of deviation from the equilibrium position between the nucleotides; U is the potential; t is the time.

According to the PBD model, each particle occurs in a potential field:

$$U = U_{\text{Morze}} + W.$$

In the PBD model the interaction of the nucleotides in a pair is described by the Morse potential:

$$U_{\text{Morze}}(x_i) = D(1 - e^{-\alpha x_i})^2.$$

The interaction of neighboring pairs has the form:

$$W(x_i, x_{i-1}) = \frac{k}{2} \left(1 + \rho e^{-a(x_i + x_{i-1})} \right) (x_i - x_{i-1})^2.$$

Consideration is given to a chain of 100 AT base pairs for the parameter values: $D = k = 0.92$ kcal/mole, $\alpha = 4.45 \text{ \AA}^{-1}$, $a = 0.35 \text{ \AA}^{-1}$, $\rho = 0.5$, $m = 300$ a.m.u, used in [3].

In order to integrate the motion equations in the PBD model we will use the velocity Verlet algorithm [7].

The temperature is taken into account in the model in the same manner as in the method of all-atom molecular dynamics [8], with the use of a collisional thermostat, where the medium in which the modeled system occurs is simulated by point particles demonstrating the Maxwell velocity distribution [9–11]. The velocity distribution corresponds to a certain temperature T_{ref} . At random times virtual particles of the medium elastically collide with the particles of the system. The motion equations have the form:

$$m_i \frac{dv_i}{dt} = F_i + \sum_k f_{ik} \cdot \delta(t - t_{ik})$$

$$F_i = - \frac{dU}{dx_i}.$$

Here $\delta(t)$ is the Dirac delta function, f_{ik} is a stochastic source of force leading to jumps of the i th atom velocity at random times t_{ik} . The value of the velocity jump is calculated as a result of collision of two point particles which had the velocities v (for a modeled particle) and v_0 (for a virtual particle) before the collision:

$$\Delta v(t) = \frac{2m_0}{m_0 + m} (v_0(t) - v(t)).$$

Here m is the mass of a modeled particle, m_0 is the mass of a virtual particle. The velocities v_0 are chosen from the Maxwell distribution. This equation follows from the law of conservation of energy and momentum.

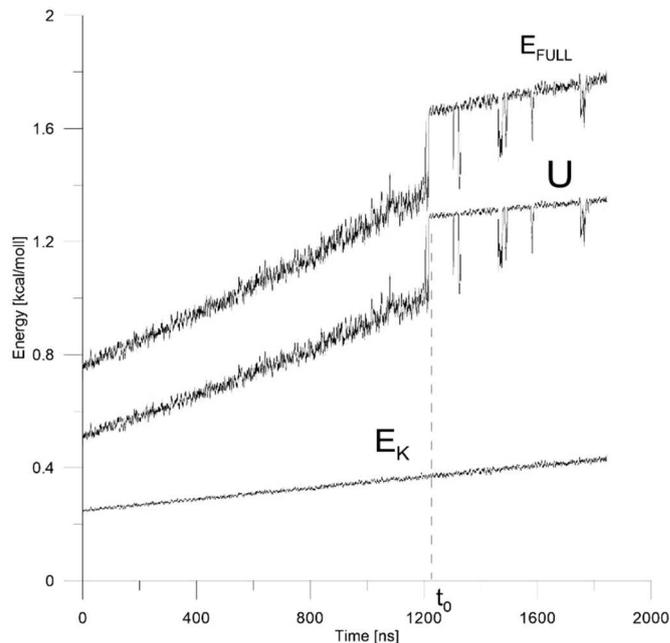


Fig. 1. Time dependence of the specific full (E_{FULL}) potential (U) and kinetic (E_K) energies for a single realization in the numerical experiment as the thermostat temperature linearly increases $T_{\text{REF}} = 250 \text{ K} + 0.1 \text{ K/ns} \times t$.

Collisions take place in accordance with the Poisson process which is determined by the only parameter λ_0 – the mean value of the collisions of an atom with medium particles per a unit of time, i.e. the frequency of collisions.

Brownian dynamics is a special case of collisional dynamics (collisional thermostat) when the mass of virtual particles tends to zero and collision frequency tends to infinity. This conclusion was shown in [9–11]. Analog of friction is the product of mass of virtual particles and collision frequency.

In the course of modeling the temperature of the collisional thermostat T_{current} changed linearly at a constant rate $V_t = 0.1 \text{ K/ns}$, $T_{\text{current}} = T_{\text{initial}} + V_t \times t$; which yields the graph of the dependence of the specific potential, kinetic and full energy on the time (see Fig. 1).

The initial temperature is chosen to be demonstrably less than the lowest point of DNA denaturation. In our example $T_{\text{initial}} = 250 \text{ K}$.

The velocity of heating per a unit time is rather small so that we can believe that on any rather small time interval the system occurs in equilibrium. According to numerical experiments, the system's relaxation – leveling off of the potential energy from temperature 0 K to 250 K – occurs for less than 1 ns.

Figure 1 illustrates the temperature dependence of the energy of a polynucleotide chain consisting of 100 base pairs in a single computational experiment (single realization). The kinetic energy changes linearly in the range from 300 K to 400 K. The potential energy and, as a consequence, the full energy have two near-linear regions and a transition region where the energy changes sharply. The time-averaged kinetic energy depends linearly on time due to the collisional thermostat.

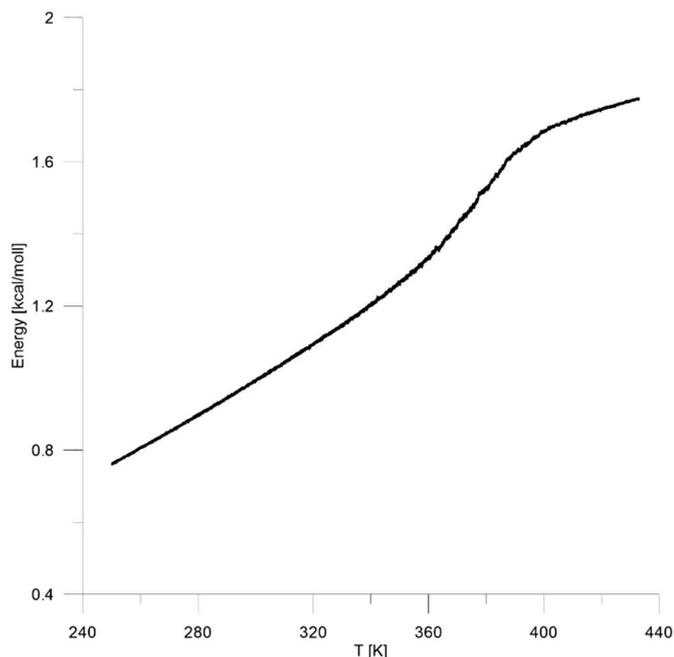


Fig. 2. Temperature dependence of the specific full energy averaged over 512 realizations.

The time dependence of the full and potential energy shown in Figure 1 describes the transition from the quasi-harmonic regime to the denaturated one. The time t_0 corresponds to the thermostat temperature $T_{REF} = 250\text{ K} + 0.1\text{K/ns} \times t_0$. Avaraging of these times over realizations yields the transition temperature (see Fig. 2).

Also we increase the time of simulation (decreasing the constant rate V_t) for different temperature and obtain the results which are close to our previous result.

Actually Figure 1 describes the energy gap whose order of magnitude is $k_B T$, where T is the temperature corresponding to the time of the jump onset t_0 . It separates the denaturated state (to the right of the jump) from the non-denaturated one (to the left of the jump). In different realizations the jump occurs at different times since the external force (thermostat) acts randomly. The early transition illustrates a larger energy gap, the late transition – a smaller one. The availability of such a gap, as was mentioned above, is caused by the divergence of the statistical sum for bounded potentials which presents computational difficulties for the description of the melting process.

To analyze the macroscopic properties such as the heat capacity, averaging over a large number of the systems (realizations) is required.

Figure 2 illustrates the temperature dependence of the specific full energy $E(T)$ averaged over $N = 512$ realizations which accounts for one site of a chain of 100 nucleotide pairs.

Figure 3 illustrates the temperature dependence of a specific heat capacity $C_V = \frac{dE}{dT}$ for four different systems averaged over 512 realizations: a system of 25 particles ($\rho = 0.5$), a system of 50 particles ($\rho = 0.5$), a system of 100 particles ($\rho = 0.5$) and a system of 100 particles

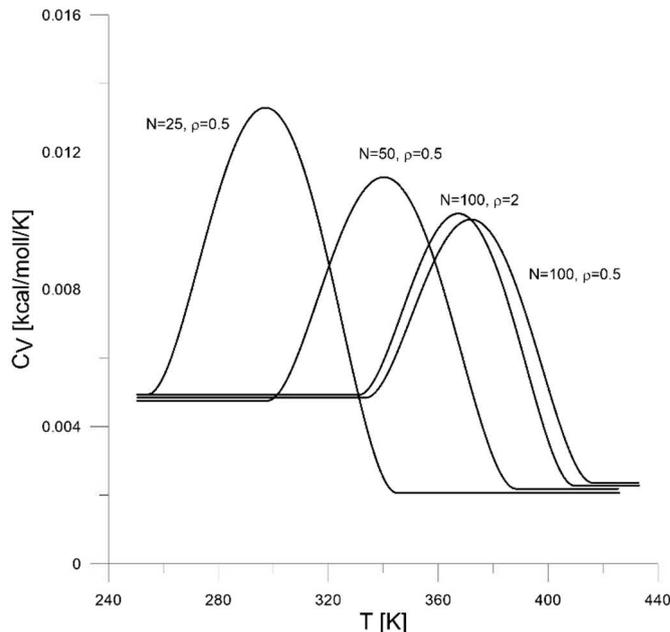


Fig. 3. Specific heat capacity as a derivative of the full energy for four systems by approximation.

($\rho = 2$). In order to find a derivative, the full energy function was approximated by two straight lines (before and after the transition) and by the polynomial of degree 5 in the vicinity of the transition. After approximation the derivative was found analytically. The boundaries of the approximation region were found by the minimization method so that the piecewise analytical function would have the smallest squared deviation from the data of modeling.

To search for the derivative, the total energy function was approximated by two straight lines (before and after the transition), as well as a fifth-degree polynomial in the transition region. After approximation, the derivative was found analytically. The boundaries of the approximation areas were found using the minimization method, so that the obtained piecewise-specified analytical function has the smallest square of deviation from the simulation data.

The curves in Figure 3 describe the process of DNA melting which corresponds to a wide peak of the heat capacity. The value of the heat capacity to the left of the peak is larger than that to the right of the peak since on the left, the heat supplied to the chain is spent on the enhancement of the potential energy between neighboring Watson-Crick pairs and on the enhancement of the potential energy of the chain plus the kinetic energy. To the right of the peak when the chain is completely denaturated its further heating is not spent on the potential energy between the nucleotides in pair. Numerically, the value to the left of the peak of the heat capacity exceeds two universal gas constants ($2R$), and to the right of the peak tends to R . Prior to the peak, we have two potentials close to parabolic: the Morse potential at low temperatures is close to parabolic, and the stacking interaction potential is close to parabolic due to the tendency of the exponent to zero.

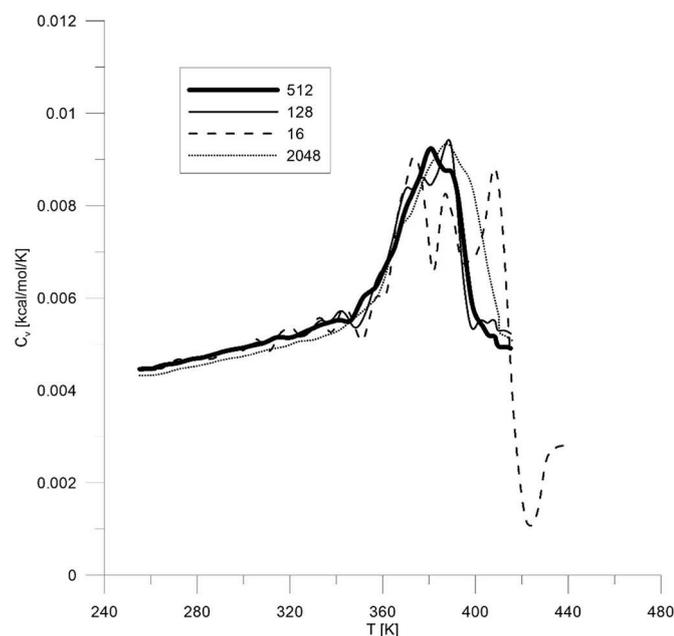


Fig. 4. Heat capacity with a different number of realizations. — 16 realizations, - 128 realizations, - 512 realizations, ... 2048 realizations.

The temperature at which the peak is maximum corresponds to the melting point of DNA. At this point the full energy of a homogeneous duplex is equal to ND , where D is a constant of the Morse potential.

The choice of realizations number is not accidental. As the number of realizations increases, the heat capacity curve tends to an even hill.

Figure 4 shows the convergence of the calculation results from the number of implementations. With the number of implementations 16, there is a jagged curve. With 128 implementations, the number of local maxima decreases. At 512 implementations, the curve takes the form. The shape of the curve at 2048 realizations qualitatively and quantitatively resembles the previous case. Therefore, we assume that 512 realizations are sufficient for calculating the heat capacity curve.

3 Discussion

In papers [3,12,13], as distinct from Figure 3 and paper [14], stepwise curves were obtained for the heat capacity. In the region of near-critical temperatures in the PBD model the kinetic energy obtained from the thermostat is spent on the enhancement of the potential energy of internucleotide interaction. This explains the smooth decrease of the heat capacity after the peak. The transfer integral technique used in [2] to calculate $C_V(T)$ yields a δ -shaped peak which corresponds to the results of modeling of very long chains. As distinct from the transfer integral technique, our approach can be used for the chains of any length and any nucleotide composition. Besides the transfer integral technique is limited by the Peyrard-Bishop model. Even simplest sophistication of the model, such as consideration of the PBD model, does not allow any

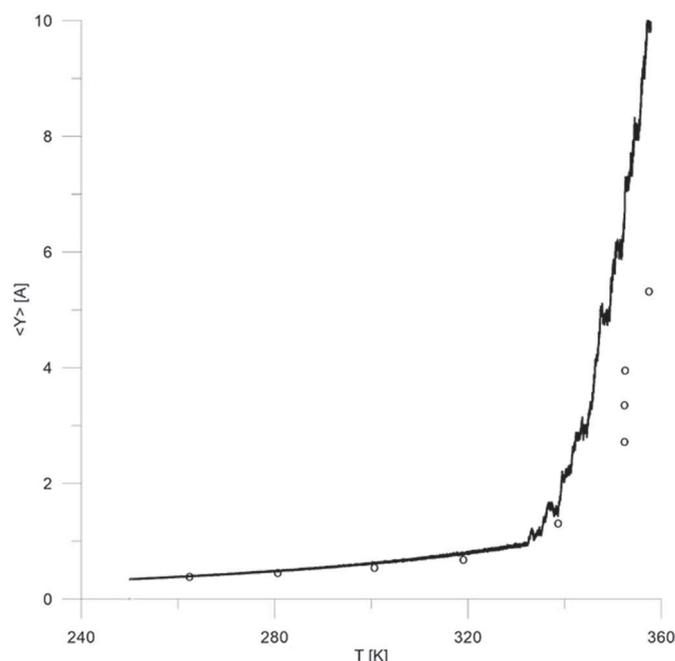


Fig. 5. Temperature dependence of the mean coordinate value (black curve). The length of the chain is 100 nucleotides, the number of realizations is 100. The circles represent the data of molecular modeling with the use of the Hose-Hoover thermostat from paper [3].

analytical solutions of relevant integral equations which renders calculation of actual systems impossible.

In paper [15] the thermodynamical functions were calculated by the direct method for calculation of the statistical sum as a multidimensional integral. As a rule, this method is essentially limited by the chain length. The most serious limitation of the method, however, lies in the fact that it enables investigations of only statistical thermodynamic states, while our method makes possible investigations of dynamical and statistical properties of the system.

The peak of the heat capacity received by us depends on the type of a nucleotide sequence. This conclusion is consistent with paper [16]. When the number of physical particles is smaller the phase transition occurs at lower temperature. This conclusion is consistent with paper [17] which deals with a slightly modified model which is similar to our model in the region of small oscillations.

Comparison of the results obtained by our method and those obtained by TI-method from paper [2] is given in Figure 5 which illustrates the temperature dependence of the mean coordinate value of a nucleotide pair displacement from the equilibrium position over the realizations. The extent of coincidence is rather high which suggests the adequacy of the calculations performed. Some distinctions can be explained by the fact that in [2] use was made of the Nose-Hoover thermostat while we used a collisional thermostat.

The thermostats by Berendsen [18], Nose and Nose-Hoover [19] can introduce some artifacts in the investigations. All these thermostats set the mean temperature

well. However as one very hot (or very cold) particle emerges these thermostats in an effort to maintain the mean temperature cool off (or heat) the system which occurs at a temperature close to a certain predetermined value. After some time the temperatures of all the particles of the system become balanced due to the interparticle interaction but not due to the thermostat. As regards the Morse potential, cold particles can emerge while attempting to get out of the potential well. If a particle does not fall back to the well its velocity is lower than that in a quasi-harmonic regime.

As distinct from the Berendsen and Nose-Hoover thermostats, the collisional thermostat acts on each particle individually irrespective of the temperature of the rest of the particles in the system which is more consistent with reality.

Referring to paper [3] we notice that our mean coordinate values coincide with their results, while the heat capacity does not. The reason is that in paper [2] the coordinates are calculated by the molecular dynamics method until the moment of denaturation and the heat capacity is given only by the transfer integral technique. We use the molecular dynamics method to calculate the coordinates both before the denaturation and after it.

The results obtained seem to be rather general and can be generalized to other systems. In particular, a wide peak of the heat capacity was observed experimentally in proteins [20] and in simulations of lattices with motion of particles in a plane [21].

Also we can see the PBD model is not applicable to short DNA chains. This result is consistent with [22].

The authors express their gratitude to N.K. Balabaev for his proposals on the technique of carrying out the experiments and to N.S Fialko for her critical remarks. The research was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [23] and supercomputers k100 and k60 of the KIAM RAS. The work was done with the support from Russian Science Foundation (RSF project N16-11-10163).

Author contribution statement

The authors contributed equally to this article.

References

1. R.M. Wartell, A.S. Benight, *Phys. Rep.* **126**, 67 (1985)
2. M. Peyrard, A.R. Bishop, *Phys. Rev. Lett.* **62**, 2755 (1989)
3. T. Dauxois, M.J. Peyrard, *Phys. Rev. E* **47**, 44 (1993)
4. T. Dauxois, M. Peyrard, A.R. Bishop, *Phys. Rev. E* **47**, 684 (1993)
5. B.S. Alexandrov, L.T. Wille, K.Ø. Rasmussen, A.R. Bishop, K.B. Blagoev, *Phys. Rev. E* **74**, 050901 (2006)
6. F. De los Santos, O. Al Hammal, M.A. Muñoz, *Phys. Rev. E* **77**, 1 (2008)
7. W.G. Swope, H.C. Andersen, P.H. Berens, K.R. Wilson, *J. Chem. Phys.* **76**, 637 (1982)
8. W.G. Hoover, A.J.C. Ladd, B. Moran, *Phys. Rev. Lett.* **48**, 1818 (1982)
9. N.K. Balabaev, A.S. Lemak, *J. Phys. Chem.* **69**, 28 (1995)
10. A.S. Lemak, N.K. Balabaev, *Mol. Simul.* **15**, 223 (1995)
11. A.S. Lemak, N.K. Balabaev, *J. Comput. Chem.* **17**, 1685 (1996)
12. T. Dauxois, M.J. Peyrard, *Phys. Rev. E* **51**, 4027 (1995)
13. M.J. Zoli, *Phys. Rev. E* **79**, 041927 (2009)
14. P. Vaitiekunas, C.-R. Colyn, P.L. Privalov, *Nucleic Acids Res.* **43**, 8577 (2015)
15. S. Srivastava, N. Singh, *J. Chem. Phys.* **134**, 21 (2011)
16. J. Marmur, P. Doty, *J. Mol. Biol.* **5**, 109 (1962)
17. M. Joyeux, S. Buyukdagli, *Phys. Rev. E* **72**, 051902 (2005)
18. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, J.R. Haak, *J. Chem. Phys.* **81**, 3684 (1984)
19. W.G. Hoover, *Phys. Rev. A* **31**, 1695 (1985)
20. P.L. Privalov, N.N. Khechinashvili, *J. Mol. Biol.* **86**, 665 (1974)
21. W. Ebeling, V. Podlipchuk, A. Valuev, *Physica A* **217**, 22 (1995)
22. A. Campa, A. Giansanti, *Phys. Rev. E* **58**, 3585 (1998)
23. V. Sadovnichy, A. Tikhonravov, V.I. Voevodin, V. Opanasenko, *Computational Science* (CRC Press, Boca Raton, USA, 2013)