Theor Chem Acc (2011) 130:1155–1167
DOI 10.1007/s00214-011-1022-9

REGULAR ARTICLE

The application of condensed matter methods to the study of the conformation and elastic properties of biopolymers and the transport of DNA through cell membranes

C. C. Matthai · N. H. March

Received: 8 December 2010 / Accepted: 11 August 2011 / Published online: 10 September 2011

Abstract The review draws on earlier interests of the authors and especially in two areas. The first of these consists of modelling polymers in different solvent conditions using Langevin molecular dynamics. For fully flexible polymers, it was found that in good solvents the polymer conformation had extended structure while in poor solvents, a globular-like conformation resulted. Stiffness in semi-flexible polymers allows for the description of a polymer with non-zero persistence length, thereby allowing biological polymers to be modelled. The second area focuses on the transport of DNA through cell membranes and in the related area of using molecular dynamics simulations in understanding biomolecular processes. Finally, under future directions, some of the areas in which techniques from condensed matter have been used in recent years point to how they may be employed in the future.

Keywords Biopolymers · Protein folding · DNA transport · Biomembranes

Dedicated to Professor Akira Imamura on the occasion of his 77th birthday and published as part of the Imamura Festschrift Issue.

C. C. Matthai (✉)
Department of Physics and Astronomy,
Cardiff University, Cardiff, UK
e-mail: matthai@astro.cf.ac.uk

N. H. March
Department of Physics, University of Antwerp,
Groenenborgerlaan 171, 2020 Antwerpen, Belgium

N. H. March
Abdus Salam International Centre
for Theoretical Physics, Trieste, Italy

N. H. March
University of Oxford, Oxford, UK

1 Background and outline

It is a pleasure to contribute to this festschrit issue honouring Professor Imamura. His early contributions to the calculation of the electronic structure of polymers had, of course, a substantial influence on the development of that subject. This major interest of Professor Imamura has prompted us to contribute this article in relation to polymers, proteins and biomolecular systems. In particular, we shall offer some insights from condensed matter theory that can be applied in molecular biophysics.

The mechanical properties, deformation behaviour and relaxation of biopolymer chains and proteins have been the subject of much interest in recent years because of their importance in understanding the structure and function of cells and muscles. Molecular modelling techniques are generally based on ab initio electronic structure calculations, semi-empirical methods or molecular mechanics. While the ab initio methods (e.g. DFT) are capable of providing accurate and consistent results, they are computationally expensive and so prohibitive for carrying out simulations on very large biophysical systems. Semi-empirical approaches (like the Extended Huckel Approximation) are based on quantum mechanics but parametrized with parameters determined from experimental data. They do not take account of electron transfer processes which can be crucially important. By contrast, molecular mechanics methods (Molecular Dynamics and Monte Carlo) rely on the use of parametrized interatomic force fields which are assumed to be transferable. Although electron reorganization effects are not considered, they can be usefully applied to studying the conformation and dynamics of large systems over long time scales. When the systems are too large even for such approximate methods to be applied, there may be a need to invoke coarse-grained

models where groups of atoms which are not altered appreciably during the course of a simulation are treated as a single unit (eg., a monomer).

In this article, we present some of the results of computer simulation studies carried out by our groups on the conformation and dynamics of biopolymers using both atomic and coarse-grained descriptions and employing the molecular dynamics method. We also show how computer simulations can give information about molecular motors and biomechanical processes. Finally, we shall also make some observations, particularly with regard to suggestions for future work, which should prove fruitful for further studies. Whilst the present article is about theory, it is designed to promote subsequent interaction between theory and experiment.

2 Conformation and force–extension curves of biopolymers and proteins

The problem of the equilibrium configurations of biological polymers is an important one, as a knowledge of the factors that influence it can provide the basis for an understanding of the structure and functions of cells and muscles. Thus, for example, the functioning of a protein is thought to depend on its structure. The designing of novel folds and the prediction of the most stable fold are therefore of great significance in the production of new enzymes with specific catalytic activities [1]. There have, consequently, been numerous attempts, using a wide variety of techniques, to investigate this general problem. For instance, there have been efforts directed at solving the problem using a statistical-mechanical approach [2, 3], as well as by modelling the protein structures as a copolymer chain [4, 5]. Mean field theories of compact phases and first order freezing transitions have also been constructed [6]. There have also been an increasing number of computer simulation studies of the equilibrium configurations of polymer structures [7–12]. Computer simulations afford a relatively simple approach with the added benefit that useful information about the mechanical response of polymers may be obtained by investigating the effect of stretching forces on a typical polymer chain. This allows a means of checking the validity of the inter-monomer forces employed in the description of the polymer. However, the complexity of the problem is such that polymer models tend to be minimal in the sense that they represent only the gross features of the polymer chain.

The dynamics of protein folding and unfolding is determined by the inter(and intra)-monomer forces. Information about these forces can be gained experimentally from the force–extension curves using the AFM and other techniques. Much of the early work on the computer simulations of polymer chains was done on ideal or near to ideal systems represented by self-avoiding walks. Biological macromolecules like proteins, DNA and actin have relatively large persistence lengths L_p [13] (e.g. DNA has $44 < L_p < 81$ nm while vimentin has $L_p \approx 1{,}050$ nm) and exhibit markedly different properties depending on their environment. Hence, in recent years, the focus has shifted to less than ideal polymers. These systems have, in the main, been studied using the so-called bead-spring and semi-flexible chain models.

In investigating the role of the environment in determining the conformation and structure of a biomolecule or polymer, there are in general two approaches that are taken in the treatment of the solvent. In the macroscopic approach, the solvent may be assumed to be a continuous medium characterized by a bulk dielectric constant. Such an approach can be used to modify hydrophobic and hydrophilic interactions. Other methods within this framework have included the treatment of the solvent as effecting an entropic interaction on the molecule. This latter method has the advantage of being able to include temperature effects but does not modify the inter- and intra-molecular interactions.

It is clear, however, that in order to arrive at a better understanding of the structure and electronic properties of molecules in solution, it is necessary to invoke quantum mechanics. The quantum mechanical mechanical approach can be used to provide the conceptual framework for constructing continuum models to describe solvent effects [14]. Alternatively, in the microscopic approach, the solvent may be considered to comprise of individual molecules, and for example, a hydrated molecule is taken to be a molecule in the presence of other water molecules. Even though the addition of extra water molecules greatly increases the computational time and can result in a complicated potential energy landscape, the importance of introducing such explicit water molecules to model the influence of the solvent effects in determining the stabilized structures of biomolecules has been demonstrated by, for example, the Suhai group [15].

Indeed, the motivation for developing such models came from the realization that the experimental data on the vibrational spectra of alanine dipeptide could not be explained without the inclusion of explicit water molecules in their calculations [16]. This recognition of the need to incorporate the effect of the environment in such investigations has led to further studies in a similar vein. Thus, it was shown that explicit water molecules were also necessary for the determination of the structure of the zwitterionic form of hydrated L-alanine and other biomolecules. By applying quantum mechanical DFT methods, Jalkanen et al. [17, 18] found that the structure of the hydrated molecules was stabilized by a network of water molecules. Furthermore, this structure was different from that of the

isolated molecule thereby demonstrating the need to include the effect of the environment in an explicit manner. In extending this approach, Jalkanen et al. [19] have used a combined approach of computational methodologies (molecular mechanics, quantum chemistry and DFT) together with the experimental measurements (vibrational spectra) to get a better understanding of the structure, function and electronic properties of biomolecules in their various environments.

In parallel with these ab initio calculations, much work has also gone into the investigating the effect of hydrophobic interactions on biological processes. In their reviews, Blokzijl et al. [20] and Chandler [21] assert that most of the data relating to the hydrophobic interaction have been determined by the study of the hydration of non-polar molecules, their transfer from organic liquids to water, and the association of amphiphilic molecules in aqueous solution. It is therefore important to carry out computer simulations to test various theoretical interpretations [22, 23]. Recently, Graziano [24] investigated that the salts may have role in affecting the strengthening of pairwise hydrophobic interactions on adding salt to water and found that the structural findings from computer simulations [25, 26] were consistent with a number of neutron scattering studies. Thus, it is clear that the simulations of polymer conformations and dynamics must take into account the solution in which the biopolymer resides and the model should also be able to represent the hydrophobic interactions.

In this section, we focus on work done on determining the factors in the inter- and intra-molecular forces that influence polymer conformations. Using coarse-grained models and treating the solvent through its entropic contribution, Maurice and Matthai [27, 28] carried out a series of computer simulations aimed at arriving at an understanding of the dependence of various interactions (including solvent) on the conformation of polymers. In their molecular dynamics simulations, a single polymer chain was represented by N spherical beads interconnected by (N − 1) springs in a three-dimensional space. The effective monomer–monomer interactions between all the monomers (bonded or not) were described by a potential comprising of three contributions

$$V = V_M + V_{\text{bond}} + V_{\text{stiff}}. \tag{1}$$

The Morse potential, V_M, between two monomers labelled i and j given by

$$V_M(r_{ij})/\epsilon = \exp(-2\alpha(r_{ij} - a)) - 2\exp(-\alpha(r_{ij} - a)); \quad r_{ij} < r_c$$
$$V_M(r_{ij})/\epsilon = 0; \quad r_{ij} > r_c \tag{2}$$

represents the effect of the excluded volume interaction and monomer–monomer attractions through the solvent. The latter exists between all pairs of monomers which is of special significance when the polymer collapses to a globule state. The equilibrium bond length, a, defines the minimum of the potential of depth ϵ. The bead radius, σ, is defined through the condition $V_M(\sigma) = 0$, and the cut-off radius, r_c, was taken to be 1.5σ. The parameter α, which is a measure of the restoring force on a bead displaced from equilibrium, was taken to be $10/\sigma$. With the energy minimum, ϵ, in units of the thermal energy, k_BT, and all the distances in units of the bead radius, based on the van der Waals energies for hydrogen bonds in biopolymer systems [30], the values of ϵ are in the range between 7 and 16meV. The strong bonding force between adjoining beads was taken to be

$$V_{\text{bond}}(r_{ij}) = -0.5\,\gamma R_{\max}^2 ln[1 - (r_{ij}/R_{\max})^2], \quad r_{ij} < R_{\max}$$
$$V_{\text{bond}}(r_{ij}) = \infty, \quad r_{ij} > R_{\max} \tag{3}$$

The maximum bond length over which this interaction is non-zero was taken to be the same as for the Morse potential; $R_{\max} = 1.5\ \sigma$ and the strength of the potential, represented by γ may be estimated from the various studies of binding energies of proteins and ligands [1]. In these simulations, it was taken to be $30\epsilon/\sigma^2$. The bond stiffness potential, modelling bond bending, was taken to be of the form suggested by Keating [31] for the tetragonal covalently bonded semiconductor structures,

$$V_{\text{stiff}}(r_{ijk}) = \epsilon_k(\theta_{ijk} - \theta_0)^2 \tag{4}$$

where θ_0 and θ_{ijk} are the equilibrium and normal angles between bonds r_{ij} and r_{jk}. For all the simulations, the stiffness parameter ϵ_k was taken be ten times the energy minimum in Eq. 2. Because most polymer systems have carbon chain backbones, θ_0 was taken as the tetrahedral bond angle in crystalline carbon, viz 109°.

The influence of the solvent was taken into account by applying the method proposed by Grest and Kremer [32] based on Langevin molecular dynamics. In this approach, the test system is coupled to a heat bath to keep it at constant temperature. The equation of motion for the ith bead/monomer becomes

$$\ddot{\boldsymbol{r}}_{\mathbf{i}} = -\nabla \boldsymbol{U}_{\boldsymbol{i}} - \boldsymbol{\Gamma}\underline{\boldsymbol{r}}_{\boldsymbol{i}} + \mathbf{w}_{\boldsymbol{i}}(\boldsymbol{t}) \tag{5}$$

where Γ represents the viscosity of the solvent and $\mathbf{w}_i(t)$ describes the random, temperature-dependent force of the heat bath acting on the ith monomer. The advantage of this approach is that the heat bath allows for work to be done on the system (eg. extension) without a resulting change in the temperature, as the thermal agitation and the viscosity cancel out each other. Note that if the viscosity was included without the thermal fluctuation term, the system would simply dissipate and no temperature effects could be addressed. The effect of having large random forces

$\mathbf{w}(t)$ (high temperature) corresponds to good solvents as these forces are entropic in nature and lead to extended polymer configurations.

2.1 Flexible biopolymers

The results of the simulations on fully flexible polymers, in which there are bond stretching but no bond-bending terms (ϵ_k in Eq. (4) is set to zero) in the potential energy description, showed that in good solvents, the polymer conformation was an extended structure, while in poor solvents, the polymers took on a globular-like conformation [27]. A consequence of this is that force–extension curves in these two regimes display markedly different behaviour. In good solvents, these polymers behave very much like a freely jointed chain (FJC) [29].

By contrast, in poor solvents, because of the initial closed globular conformation, the applied force needs to reach a critical value before there is a near spontaneous uncoiling of the polymer.

2.2 Semi-flexible biopolymers

With the bond-bending term, $\epsilon_k \neq 0$, the chains are termed semi-flexible. The flexibility of the chains is characterized by the magnitude of the size of the bond-bending energy which in turn determines the Kuhn, or persistence, length.

For these semi-flexible polymers, Maurice [28] investigated their conformation as a function of temperature (T) and bond stiffness (ϵ_k). He found that at low temperatures (corresponding to poor solvent conditions), as the stiffness parameter is increased, the conformation of the polymer underwent a transition from a globular phase to a folded configuration, with the length of each segment being a multiple of the persistence length (see Fig. 1).

The stiffness at which the folds begin to appear is, rather unsurprisingly, temperature dependent, suggesting that it is the finite persistence length which is responsible for the observed folding. The chain conformation is determined by the competition between the stiffness parameter which increases the persistence length and the temperature which tends to reduce it. This is best seen in the phase diagram of the polymer conformation in Fig. 1.

The critical temperature below which folded configurations are formed was also found to depend on the chain length with the longer chains only folding at higher temperatures.

To summarize the results of Maurice et al. [27, 28], at high temperatures or in good solvents, both flexible and semi-flexible chains behave as freely jointed chains with extended conformations. At low temperatures, whereas flexible polymers form compact globule type structures, semi-flexible polymers take upfolded chain structures with the number of folds varying linearly with the length of the chain.

While the folding may be attributed to the model, the stiffness term should favour tetrahedral bonds and it is not at all obvious that this model polymer would equilibrate into the observed folded conformation. Other groups have reported the observation of toroidal structures [9–11]. This could be due either to the potential energy interactions that were employed or to their using the lattice based Monte Carlo method in their simulations. It is ought to be stressed that the Langevin dynamics method used in this study is particularly adept at ensuring that the system is not trapped in local minima. Also, the off-lattice nature of the simulations allows a full exploration of the configuration space and the structure is not constrained by the underlying lattice structure which can result in artificial conformations. The influence of different potential energy interactions on the final conformation requires further study. The presence of a critical temperature for a prescribed length at which this folding occurs suggests that there is some kind of *phase transition* from one state to another similar to the globule to extended chain transition observed by us and by other groups [6, 9, 10, 12, 28].

2.3 Force–extension relations

The force–extension relation of a polymer is intimately connected with the nature of the various interactions between the monomers and the solvent. Thus, a knowledge of this relation can yield information about the nature and strength of the bonds in a chain. This is of particular interest in the study of protein molecules and structures. The recent studies of Perkins et al. [33], Smith et al. [34], Rief et al. [35] and Oesterhelt et al. [36] on the elastic properties of DNA, spectrin and PEG molecules have shown that it is possible to represent such polymers by the FJC model. The WLC model, where the polymer can be modelled as a string with local stiffness, has been used to describe the elastic properties of DNA [37, 38].

The results of the simulations (Fig. 2) show that for flexible polymers in good solvents, or semi-flexible polymers at high temperatures, the force–extension curves could be fitted by the form arising from the FJC model [29]

$$\frac{\langle R_N \rangle}{N} = a\left(\coth\frac{fb}{kT} - \frac{kT}{fb}\right) \tag{6}$$

Here, $\langle R_N \rangle$ is the end-to-end distance of a chain comprising of N monomer units subject to a force f at temperature, T. The parameter b is taken to be the monomer bond length (flexible) or the persistence length (semi-flexible). The force–extension curves derived from the worm-like-chain (WLC) model [39] were also found to

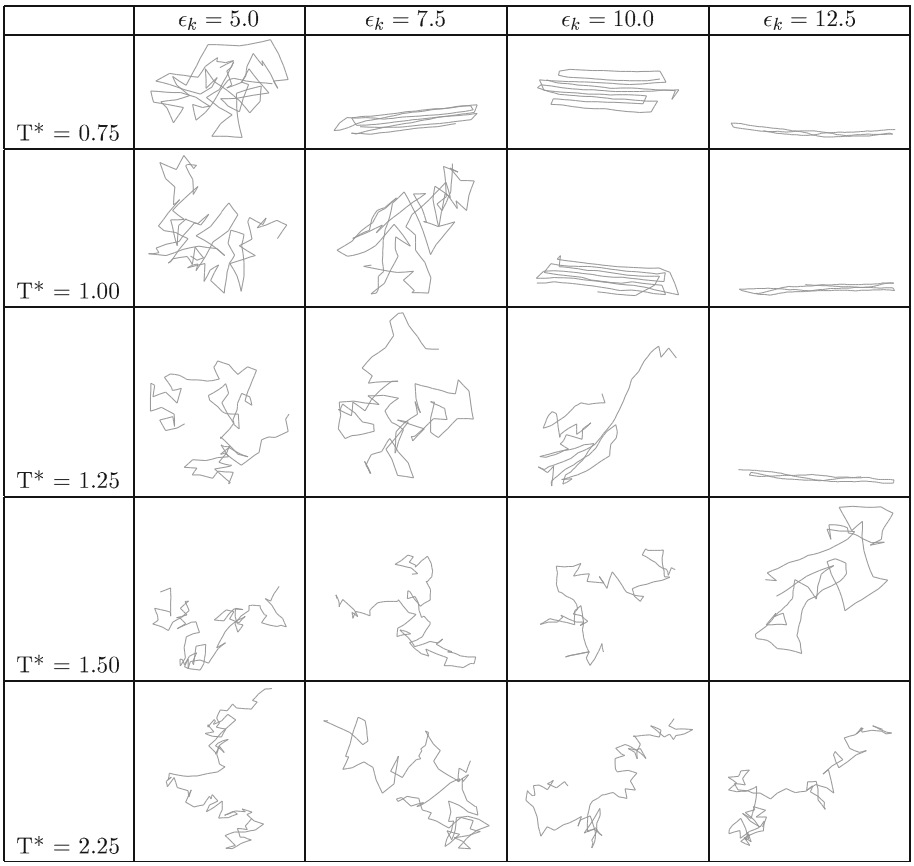

Fig. 1 Equilibrium configurations of semi-flexible polymers as a function of the reduced temperature, $T^* = k_B T/\epsilon$, for varying stiffness constants, ϵ_k

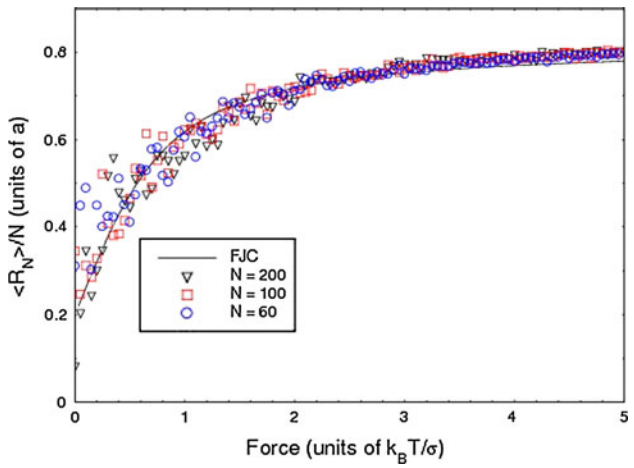

Fig. 2 Force–extension curves for flexible polymers of varying length compared with that expected from the FJC model. For each value of the applied force, multiple simulations were carried out

give a good fit, but with a shorter persistence length than that used for the FJC fit.

When the force–extension simulations were performed on the low temperature folded chain configurations, there was virtually no extension of the chain until the magnitude of the force reached a critical value. There is, then, a rapid increase in the extension, so that $\langle R_N \rangle$ becomes equal to a multiple of the fold length, L_f. This is indicative of the start of the unfolding of the chain. It remains at this extended length for further increases in the applied force until the force reaches another critical value whereupon the chain experiences another jump in the extension (Fig. 3). This process continues until the chain is fully unfolded and its length is equal to that of the chain in a good solvent. The critical values of the applied forces turn out to be dependent on N as is the amount of the extension at each step. This is not surprising as the number of folds varies with N and the critical force is a measure of the energy required to break 'bonds' between those segments close to each other in the folded configurations.

The force–extension curves provide another means by which the details of these forces may be determined. When the polymer is in a folded state, a force applied to the ends does not result in a significant increase in its length as the force is not propagated 'around' the fold. Once the force is of sufficient magnitude to unfold one segment, this results in an immediate doubling of the chain length. While these simulations do not model any specific biopolymer system, for the values of the ϵ used, the calculated force–extension curves correspond very well to the experimental results on single molecules of DNA reported by Wang et al. [40].

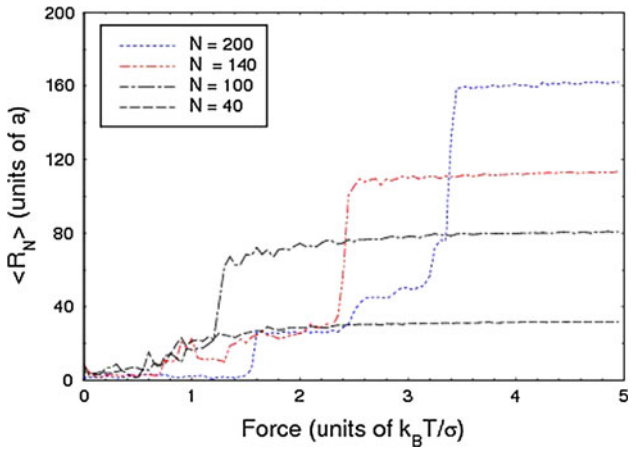

Fig. 3 Force–extension curves for semi-flexible polymers at low temperatures

3 Modelling biological processes

Biopolymers are transported through channels across membranes in a number of biological processes. These include the transport of RNA across nuclear pore complexes and protein transport across mitochondrial membranes and the endoplasmic reticulum. The biological channels are also made up of proteins and are of nanometre scale. Following the experimental investigations of Kasianowicz et al. [42] on the translocation times of the ssDNA segments across the α-haemolysin membrane, there have been many studies aimed at using this approach to study various aspects of macromolecular translocation across nanopores including the determination of a DNA sequence. More recently, there have been studies of DNA translocation through single-walled carbon nanotubes [43] and graphene nanopores [44].

Although there has been some efforts at understanding the translocation process from a theoretical standpoint, these have, in the main, been based on polymer diffusion models and restricted to one- dimensional systems. For example, Randel et al. [45] applied the Monte Carlo (MC) method to simulate the experimental findings of Kasianowicz et al. [42]. Using a minimalist approach with the single-stranded DNA (ssDNA) modelled as a simple homopolymer and with the lipid bilayer membrane represented by an impenetrable wall with the pore modelled as a cylindrical tube through the wall, this model was able to reproduce the correct form of the temperature dependence of the translocation time across the membrane. However, to model different types of DNA sequences or indeed protein chains, a more sophisticated approach is required. This includes modelling the haemolysin pore in a more realistic fashion, to investigate the effects of the electrostatic interactions between the ssDNA and the pore and also the effects of drag during the translocation process may be

Table 1 Parameter values used for the heteropolymer DNA model

Parameter	Adenine	Cytosine	Guanine	Thymine
ε_i	1.4	1.0	1.0	1.0
σ_i (nm)	0.7	0.1	0.9	0.5
α_i	10	10.0	10.0	10.0
q_i	1.0	0.7	0.7	0.7

taken into account. By using monomers with different charges, it may be possible to understand the effects of these factors on the way in which ssDNA translocates through the pore. Loebl and Matthai [46] (LM1) modelled the ssDNA as a chain of N monomers, each representing a nucleotide. All monomers were allowed to interact with each other through the Morse potential (Eq. 2), which for different types of monomers, can be written as

$$V_M(r_{ij}) = \epsilon_{ij}\epsilon_0\left[\exp\left(-2\alpha_{ij}\frac{(r_{ij}-a)}{\sigma_{ij}}\right) - 2\exp\left(-\alpha_{ij}\frac{(r_{ij}-a)}{\sigma_{ij}}\right)\right]$$

where r_{ij} is the distance between different types of monomers i and j. σ_i is the hard sphere radius of monomer, and ϵ_i is the strength of the interaction between monomers of the same type, i. ϵ_0 is the unit of energy in the model so that the temperature is in units of ϵ_0/k_B. For interactions between different monomers, the composite values were obtained from the relations, $\epsilon_{ij} = \sqrt{\epsilon_i\epsilon_j}$, $\alpha_{ij} = \sqrt{\alpha_i\alpha_j}$ and $\sigma_{ij} = \frac{\sigma_i+\sigma_j}{2}$. The parameters assigned to the monomers are only designed to make the monomers different so as to investigate the effects of different types of ssDNA or other biochemical chains. They are not designed to specifically model adenine, cytosine etc, although the monomers that are assigned the larger values of ϵ_i or σ_i are adenine and guanine, which are the nucleotides with the larger bases made up of two rings. The monomers are also assigned a charge, q_i, which determines the strength of the pulling force on monomer i, due to the applied electric field, which drives the translocation (Table 1).

The three-dimensional potential energy map of the electrostatic and van der Waals potential energy due to the protein pore interacting with a model DNA nucleotide was constructed by using the NAMD software package [47] with the CHARMM22 force-field parameters [48]. This was done by determining the energy of a single test ion interacting with the α-haemolysin protein pore. The test ion was chosen with suitable parameters to represent a single DNA nucleotide and corresponds to the monomers used in the heteropolymer model. The electrostatic energy could be scaled for any charge. The minimum energy configuration of the α-haemolysin pore was determined by performing a NAMD simulation. The starting configuration of the simulation was taken from the X-ray structure of the α-haemolysin pore (Fig. 4), resolved by Song et al. [49],

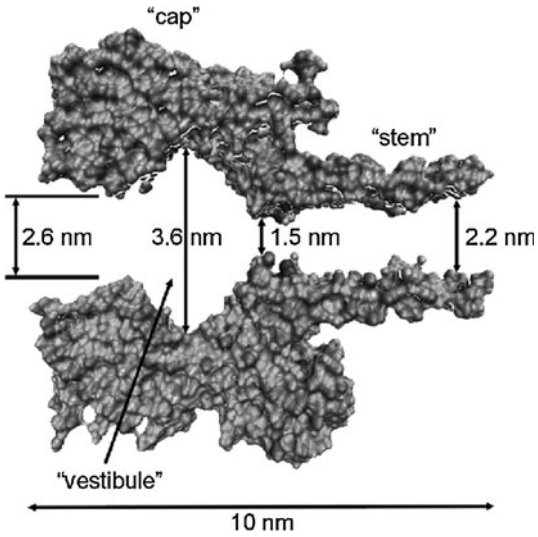

Fig. 4 A slice through the X-ray structure of the haemolysin pore (after Song et al. [49])

which is available in the file 7AHL from the Protein Data Bank. The initial positions of the H-atoms in this structure (not resolved by X-ray diffraction) were guessed and the energy of the structure minimized such that there were no steric interactions.

NAMD was then used to calculate the energy of the interaction between the test ion and the protein pore. The central part of the protein pore, the region containing the channel, was mapped in detail. This is the most important part of the protein because the interactions with the central channel have the greatest effect on translocation. The dummy ion was placed at different positions on a 0.3-nm grid, which extended 6 nm at either end of the protein and covered an area of 0.48 nm^2 in the centre of the protein. The regions outside this central core were mapped by calculating the energy as a function of the distance from the centre of the core for radii up to 2.4 nm. Outside the protein, at either end, the van der Waals interactions tail off to zero and the electrostatics are almost radially independent. These regions were again mapped in less detail than the central channel. The interaction energy values at points between grid points were obtained by interpolation.

The resulting electrostatic potential (Fig. 5) explains some of the behaviour of translocating DNA observed in experiment. It may be noted that the depth of the electrostatic potential (Fig. 5a) $\approx$4.6 kcal/mole corresponds to $\approx$200 mV, which is of the order of magnitude generated by a biological ion pump [50]. As the DNA moves near to the pore, the decreasing electrostatic energy attracts the negatively charged DNA into the pore until it reaches the energy minimum. The DNA then sits at that point unless there was some external driving force to

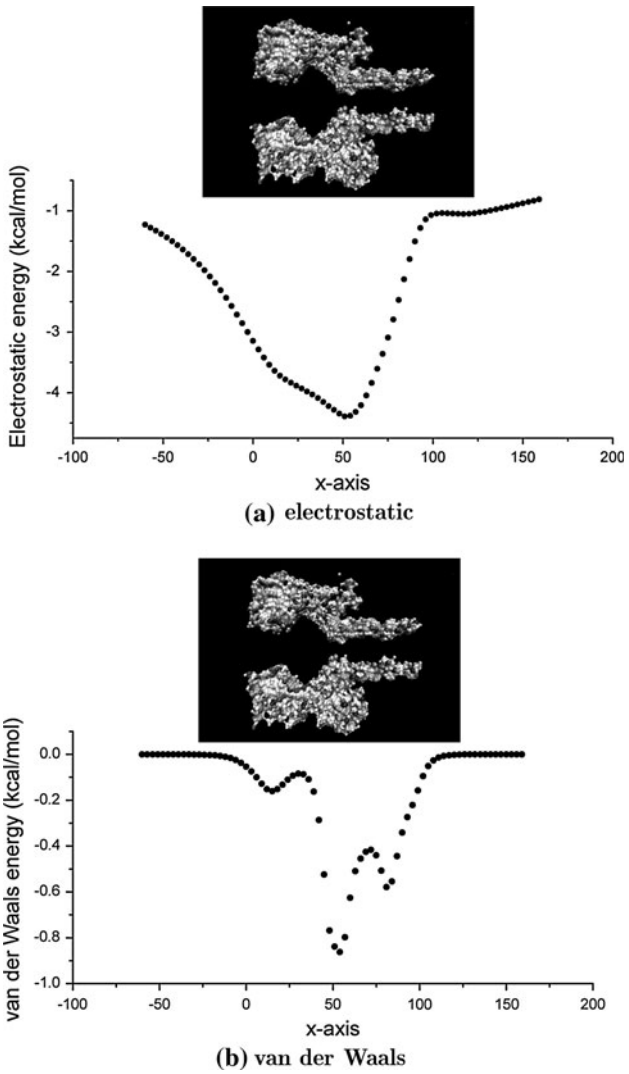

Fig. 5 The profiles of **a** the electrostatic and **b** the van der Waals potentials for a line running through the centre of the pore

overcome the energy minima and release the DNA from the other side of the pore. This driving force is the electric field applied to the system in experiment. Thus, there is some threshold field strength, below which translocation does not occur. The van der Waals profile (Fig. 5b) is about an order of magnitude weaker and in the main acts as a hard wall interaction and defines the shape of the protein.

In using this potential energy map for the different monomers, a simple scaling procedure was adopted. The electrostatic potential simply scales as the charge on the monomer as the test ion was taken to have unit charge.

To simulate the heteropolymer translocating through the α-haemolysin pore, LM1 placed the first monomer inside the α-haemolysin pore and the rest of the chain was generated as a self-avoiding random walk on a cubic lattice, with the constraint that no monomers can be placed within

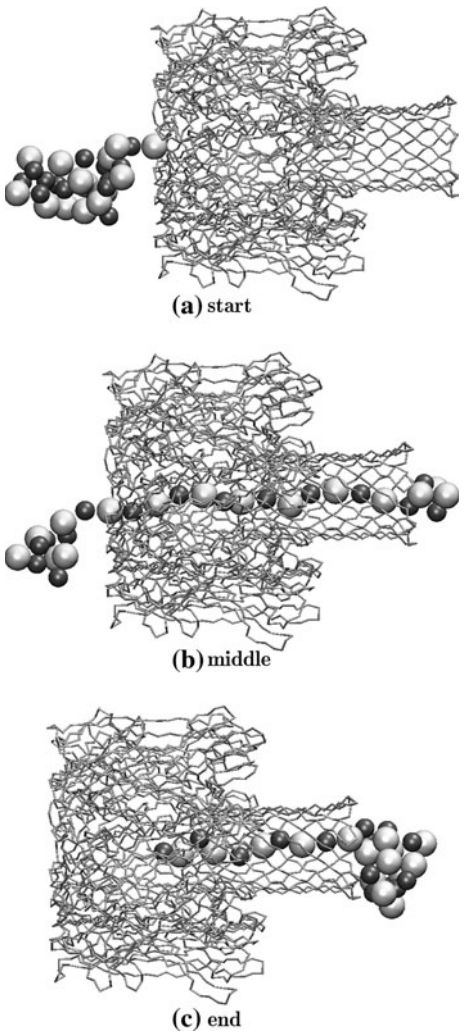

Fig. 6 Snapshots of a typical translocation run for an alternating guanine and cytosine heteropolymer

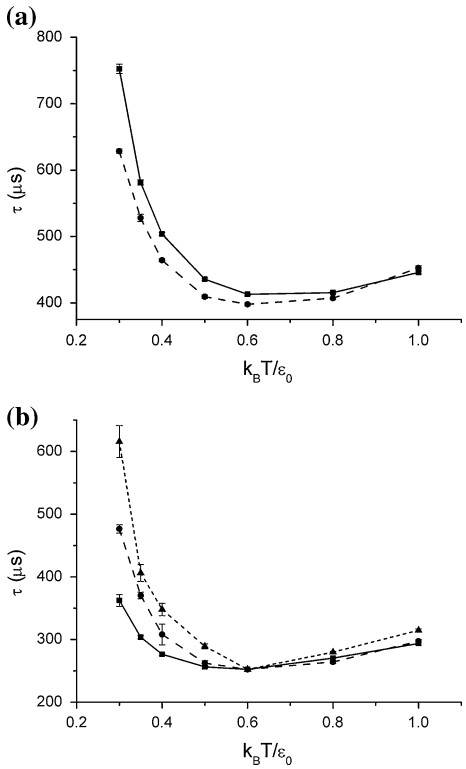

Fig. 7 The dependence of temperature for **a** two different heteropolymers, made up of alternating guanine and cytosine monomers (*black square*) and two blocks of 15 guanine and 15 cytosine monomers (*bullet*); **b** three different homopolymers, thymine (*black square*), cytosine (*bullet*) and guanine (*black triangle*)

the pore. The heteropolymer chain is then allowed to equilibrate using the standard Monte Carlo method, but with the initial monomer still fixed just inside the pore.

Once the equilibrium conformation has been reached (Fig. 6a), the chain is released and the translocation process begins. The number of working moves from the release of the chain, to the last monomer leaving the other end of the pore is taken as the time for one translocation. Figure 6 shows a typical translocation run for a heteropolymer. At least 500 of these times were measured and plotted as a distribution, and the peak in the distribution is taken as the translocation time. Translocation runs were performed for thymine, cytosine and guanine homopolymers for a range of temperatures. The translocation times, determined from the peaks of the distributions, as a function of temperature for the three different homopolymers of thymine, cytosine and guanine are shown in Fig. 7. The slowest of the chains to translocate is the guanine homopolymer, which has the largest value, while the thymine homopolymer is the fastest.

To model the experimental results showing the dependence of the translocation times on base type, simulations on two types of heteropolymers, each with 15 adenine monomers and 15 cytosine monomers, were also carried out. The heteropolymers were either made up of alternating adenine (A) and cytosine (C) monomers, i.e. ....ACACACACAC… or a block of 15 adenine monomers followed by a block of 15 cytosine monomers, i.e., …AAAACCCC… Again, the simulations were performed at over the same temperature range. The results for the two sets of simulations were almost identical suggesting that this model cannot distinguish between these two different heteropolymer types in its current form. Further, simulations on a second pair of heteropolymers, this time made up of cytosine (C) and guanine (G) monomers arranged in the same sequences as above, were carried out. Figure 7 shows the temperature dependence of for the two heteropolymers. It shows that the alternating guanine–cytosine chain translocates slower than the block heteropolymer, which is in agreement with the results of Meller et al. [51] (given in Fig. 7 of their paper). Thus, the pore model seems better able to distinguish between monomers shapes than

monomer–pore interaction strengths. This method could thus provide a useful tool in future studies on DNA sequencing. By constructing model electrostatic potentials using test charges, LM1 demonstrated that it was possible to model high complex biological processes involving several hundred atoms.

4 Modelling molecular motors and biomechanical processes

Mitochondria are organelles contained within animal cells, which among their other functions perform the important task of converting glucose and oxygen into ATP (adenine 5′-triphosphate) [52]. Although mitochondria have their own DNA, which encodes some of the proteins required by them, most of the 1000 or so proteins in mitochondria are imported into the organelle. The mechanism of the protein import is not fully understood, but it is clear that it is a highly complex system which ensures that the proteins to be imported are recognized, sorted and transported to their different destinations.

Mitochondria have two membranes, the outer membrane which defines the smooth shape of the mitochondrion, and the inner membrane which has a much larger surface area and contains many folds or cristae. The area between the membranes is called the intermembrane space and that within the inner membrane is known as the matrix. The outer membrane contains mitochondrial porin, which is a channel protein allowing the passage of ions and small molecules. The inner membrane, which is made up of 76% protein, many of them transport proteins, allows otherwise impermeable molecules and large proteins to move across it [53–56]. The production of ATP causes a constant exchange of ions across the inner membrane and results in setting up chemical and potential gradients across it, which in turn becomes important for proteins getting into the matrix [53]. Not all proteins from the cytosol are destined for the matrix as some of these will become new import pores in either of the two membranes [55–57]. In this paper, we focus on those proteins which transport across the two membranes through to the matrix.

Figure 8 shows a schematic representation of the mechanism of preprotein import [54]. When the preprotein is first synthesized and released to the cytosol, it has an extra signal sequence attached to the N-terminus. This sequence is what makes the preprotein attracted to the correct organelle, in this case a mitochondrion. The signal sequence is first attracted to the import proteins of the outer membrane, called the TOM (translocase of the outer membrane) complex. The main protein in this complex is Tom40, which is the actual transport channel through the outer membrane. Once through this channel, the signal sequence is attracted to the TIM (translocase of the inner membrane) complex, and it moves through the channel made by Tim23 due to the potential gradient, $\Delta\psi$ across the inner membrane. The signal sequence, or presequence, is positively charged and so is attracted across the inner membrane and pulls the rest of the preprotein through behind it. Once inside the matrix, the presequence is removed from the preprotein and the next phase of the import takes place.

There is much debate over the nature of the next part of the import process and two mechanisms have been proposed. Both involve the matrix heat-shock protein, Hsp70 and another part of the TIM complex, Tim44. As the preprotein comes through the inner membrane and into the matrix, an Hsp70 molecule binds onto the preprotein by using ATP and also onto the Tim44. In the ratchet model of import, the binding Hsp70 acts as a molecular ratchet; i.e., once the preprotein has bonded onto the Hsp70, it cannot then move back into the membrane [58]. Thus, with repeated attachment and release of Hsp70 molecules, this system will result in the preprotein moving into the matrix by catching most of the forward motion due to thermal fluctuations but not allowing it to move back into the membrane. This mechanism relies on the frequent attachment of Hsp70, with a minimal time where the preprotein

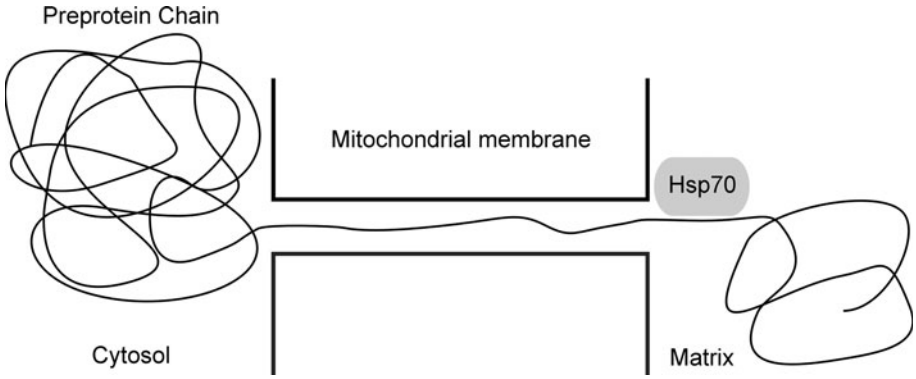

Fig. 8 A schematic representation of the Hsp70 acting as a motor or as a ratchet

is free to drop back out of the matrix. The ratchet model has been the subject of some investigation, and Simon et al. [59] have advocated it on the basis that (i) it is non-specific; and (ii) it gives the 'right' order of the translocation rate.

The second proposal is that the Hsp70 acts as a molecular motor pulling on the preprotein to encourage it through the channel in the inner membrane [60]. In this scenario, the Hsp70 attaches to both the preprotein and Tim44, but it undergoes a conformational change while pivoting on Tim44 levering the protein forwards. Although there has been much experimental work done on investigating this process, there is still no consensus about the actual mechanism. Indeed, it is not even clear whether there is a discernible difference between the mechanisms.

Loebl and Matthai [61] (LM2) sought to discrimate between these two mechanisms by investigating the temperature dependence of the time taken for the protein to translocate through the membrane. This method had been successfully employed to study biopolymer transport across biological nanopores [62]. The preprotein and Hsp70 proteins were both treated as homopolymers modelled by a chain of identical beads interconnected by springs. The beads representing the monomers were allowed to interact between themselves via a force-field based on that developed for semi-flexible polymer chains [63].

In this representation, each homopolymer chain comprises of N_μ monomers, with equilibrium separation σ_μ (μ = 1 for preproteins or 2 for the Hsp70 protein) giving a total chain length $L_\mu = N_\mu\sigma_\mu$. Changes in the bond energy due to the stretching and bending were modelled by empirical potentials. In addition, the interaction between all non-adjacent neighbours, i and j, on a particular chain and between monomers on different chains (e.g. between those on the preprotein and the Hsp70) was represented by a Morse potential. The interaction strengths were so chosen that the Hsp70 molecule bonded strongly onto the preprotein.

4.1 Modelling the motor

In this simplistic representation of the motor mechanism, the Hsp70 molecule attaches on to part of the preprotein closest to the membrane as it emerges out of the import tube and pulls it away from the wall for a specified period of time and then detaches itself. This process is then repeated until all of the protein is pulled into the matrix. In this simulation scheme, this action was modelled by first bonding the Hsp70 on to the preprotein and the motor mechanism leveraging the protein forwards was effected by the introduction of a linear spatial potential gradient seen by the Hsp70 molecule over a specified distance. With this potential gradient, there is a constant force pulling the Hsp70 and attached preprotein away from the inner wall. Once the Hsp70 is maximally displaced, it detaches from the preprotein. The reattachment of Hsp70 to the preprotein monomers closest to the inner wall is a measure of the availability of ATP and as such, was taken to be a variable of the simulation.

Because of the pulling force, the Hsp70 moves away from the wall taking the attached preprotein with it. Once the Hsp70 has moved a specified distance, it is released from the preprotein which in turn is allowed to evolve for a period of time before another Hsp70 attaches itself to that part of the preprotein chain nearest the membrane wall. This period of time is determined by the availability of ATP and was taken as a parameter in the simulations. The pulling process is then repeated until the whole chain is driven into the mitochondrion matrix. The translocation time was determined in the way described in Sect. 5.

The concentration of ATP within the mitochondrion would affect its ability to import proteins. In simple terms, this would be manifest in the time period over which the Hsp70 molecule is attached to the preprotein. Two cases were investigated. In the first, the Hsp70 molecule is continually attached to the preprotein. So, as soon as one Hsp70 detaches itself at distance s_z from the membrane wall, then another attaches itself at the membrane wall. This corresponds to a 'hand-over-hand' pulling [64]. In the second scenario, once the Hsp70 has moved the distance s_z and detached itself, the same time period is allowed to elapse before anther molecule is allowed to attach to the preprotein at the membrane wall. Thus, in this case, the preprotein is detached from the Hsp70 for roughly half the time. Such a situation could correspond to a mitochondrion with a damaged ATP production system. The translocation time dependence on motor strength and temperature were then investigated.

4.1.1 Hand-over-hand pulling

An attachment probability of unity represents a fully working mitochondrion, i.e., one producing plenty of ATP for its needs. In this 'hand-over-hand' pulling, there is no time delay between preprotein being released and attached again.

Having determined an optimum motor strength for performing the simulations, LM2 studied the temperature dependence of the translocation process for this case. A strong inverse square temperature dependence, similar to that found in the simulation of polymer translocation through a channel [62], was seen confirming the idea that in this region, the reduction in the translocation time with temperature was simply a reflection of the reduced uncoiling barrier at higher temperatures.

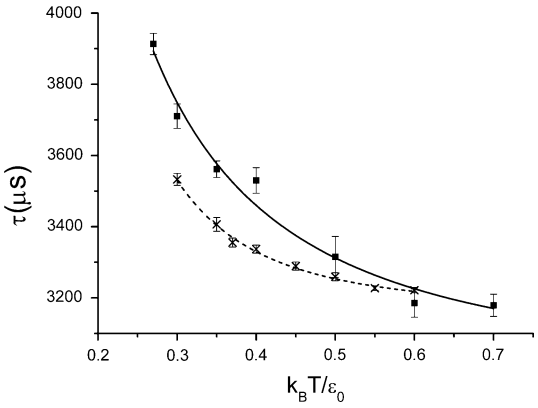

Fig. 9 The dependence of τ on temperature, with motor strength, $m = \epsilon_0/\sigma$ for the hand-over-hand pulling (*crosses*) and $m = 9\epsilon_0/\sigma$ for the intermittent pulling (*squares*)

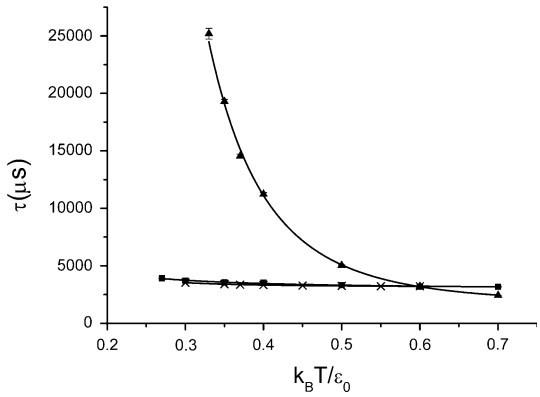

Fig. 10 The dependence of τ on temperature for both motor models (*crosses* for the hand-over-hand) and (*squares* for the intermittent pulling), and the for the ratchet model (*triangles*)

4.1.2 Intermittent pulling

In the intermittent pulling scenario, the preprotein is pulled by the Hsp70 until the latter is displaced a some specified distance, followed by the detachment of the Hsp70. The preprotein is allowed to undergo Brownian motion for the same period of time,, and with it, there is the possibility of moving backwards into the tube.

From the results, shown in Fig. 9, it may be seen that while both are broadly comparable and similar, when the motor mechanism is not in continuous operation, the motor strength needs to be increased quite substantially to achieve comparable translocation times.

4.2 Hsp as a ratchet

When the Hsp70 acts as a ratchet, instead of pulling the preprotein chain across the IM pore, it acts as a bias on the Brownian motion of the translocating chain. This ratchet model of the process can be modelled by allowing an Hsp70 protein to bond to the 10 preprotein monomers closest to the inner membrane as the latter is making its way through the pore. However, since the Hsp70 cannot enter the import tube, it also prevents the preprotein chain from dropping back into the tube.

As might be expected from a model based on Brownian motion, the temperature was found to have a great effect on the translocation process in this ratchet model—especially when compared with that of the motor models (see Fig. 10).

It is clear that the motor process (with the ATP production system fully working) results in translocation times which do not vary greatly as a function of temperature and thus would it make it a good candidate to describe protein transport across the inner membrane in mitochondria. Although the ratchet model benefits from the non-specificity of its mechanism, it exhibits a large variation in the translocation times with temperature which would be a major shortcoming for a biological process.

5 Summary and future directions

In this article, we have attempted to show how methods of condensed matter theory may be used effectively in the study of biopolymers and biological processes. We have given examples of how molecular dynamics simulations have been employed in the study of the conformation and force–extension curves of biopolymers and proteins. We have also demonstrated how the application of such simulation techniques can give useful information about the dynamics of biological molecules and, further, may even be able to shed light on the validity of particular molecular processes. Although the examples of the studies outlined above are based on coarse-grained models, as computational resources have increased it has become more feasible to undertake ab initio calculations of biopolymer conformations and other properties. Now, many methods from condensed matter theory and quantum chemistry, like DFT, are being used to study biophysical properties and processes [15, 65]. Another important aspect in the study of biological molecules is the necessity to consider the environment surrounding the molecule. While the treatment of the environment can be considered in different approximations, it is clear that it has profound effects on the molecular conformation, as outlined in Sect. 2. In parallel with these theoretical methods, experimental methods from condensed matter are also being used to probe the structure and dynamics of biological macromolecules [66, 67].

It has been proposed [68] that Alzheimer's and Parkinson's diseases are connected to the aggregation of proteins in the human body, since the proteins are polymers, Ma and

Hu [69], have recently considered what they term the minimum form of the above aggregation problem, namely the clustering of chains in the presence of repelling background fluid. Quite specifically, Ma and Hu employed molecular dynamics simulations to investigate the aggregation of polymer chains in the environment with or without Lennard-Jones (LJ) fluid. Within the framework of a polymer chain connected by rigid bonds with small or zero bending and torsion angle–dependent potentials and with fluid present, they found that a small number of linker sites along a polymer chain are randomly assigned to be fluid attractive, in contrast to the remaining fluid-repelling monomers. They concluded that the polymer chains tend to aggregate with or without linker sites and suggested that their results might prove useful for building further models for understanding the key factors which affect the aggregation of peptides or proteins. This could result in being able to design a strategy, or drugs, to avoid the aggregation of such peptides or proteins in the human body.

Biological motion from cellular mobility to replication and segregation of DNA, reflects molecular-scale forces. Newmann and Nagy [70], in a recent review, have emphasized the development of techniques that allow measurement of force and displacement generated by single molecules ranging from cells to proteins. These authors stress, in particular, that current single-molecule manipulation capacity spans six orders of magnitude in length (from 10^{-10}–10^{-4} m) and forces (10^{-14}–10^{-8} N). They note that present applications cover a range from the manipulation of cells ($\approx$100 μm) to the measurement of RNA polymerase advancing a base pair (0.34 nm) along DNA [71] and from the mechanical disrupting of covalent bonds [72] to the study of nucleic acid folding kinetics [73]. Neumann and Nagy also note the ability to manipulate and control unlabelled proteins has recently been recorded [74], but stress that this and other techniques that are emerging have, as yet, to be fully realized. Neumann and Nagy finally emphasize that, additional to these advances, new techniques should enable the continuing development of novel single-molecule force spectroscopy techniques.

In a similar vein, Shafran et al. [75], by adapting a scanning fluorescence correlation spectroscopy technique to measure the structure factor $S(q)$ [76] of semi-dilute solutions of long DNA polymers, have demonstrated a direction in which liquid state theory can offer a way forward furthering our understanding the structure of DNA solutions. Another area ripe for investigations relates to the growth kinetics of clusters which can be useful for controlling crystal growth to achieve increased purity of crystals [77–80] which is in turn required for X-ray diffraction investigations. Another area of much interest is that charge transfers through biopolymers and in particular DNA [81] because of their implications in the development of electronic devices based on biopolymers.

Acknowledgments One of us (NHM) acknowledges partial financial support from the University of Antwerp (UA) through BOF-NOI. Thanks are due to Professors D. Lamoen and C. Van Alsenoy for thereby continuing the affiliation of NHM with the UA. Finally, NHM completed his contribution during a stay at ICTP, Trieste. His thanks are due to Professor V.E.Krovtsov for generous hospitality at ICTP. Both authors also acknowledge the very useful comments of all the referees which has resulted in a much improved article.

References

1. Fersht A (1999) Structure and mechanism in protein science. W.H.Freeman, New York
2. Ramanathan S, Shakhnovich E (1994) Phys Rev E 50:1303
3. Sfatos CD, Gutin AM, Shakhnovich EI (1994) Phys Rev E 50:2898
4. Stillinger FH, Head-Gordon T (1995) Phys Rev E 52:2872
5. Yue K, Dill KA (1993) Phys Rev E 48:2267
6. Doniach S, Garel T, Orland H (1996) J Chem Phys 105:1601
7. Doye JPK, Sear RP, Frenkel D (1998) J Chem Phys 108:2134
8. Bastolla U, Grassberger P (1997) J Stat Phys 89:1061
9. Noguchi H, Yoshikawa K (1998) J Chem Phys 109:5070
10. Noguchi H, Yoshikawa K (1997) Chem Phys Lett 278:184
11. Ivanov VA, Paul W, Binder K (1998) J Chem Phys 109:5659
12. Lai PY (1998) Phys Rev E 58:6222
13. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Science 265:1599
14. Tomasi J, Menucci B, Cammi R (2005) Chem Rev 105:2999
15. Jalkanen KJ, Suhai S, Bohr HG (2009) In: Bohr HG (eds) Handbook of molecular biophysics methods and applications. WILEY-VCH, pp 1–66
16. Jalkanen KJ, Suhai S (1996) Chem Phys 208:81
17. Tajkhorshid E, Jalkanen KJ, Suhai S (1998) J Phys Chem B 102:5899
18. Jalkanen KJ, Bohr HG, Suhai S (1997) Proceedings of the international symposium on theoretical and computational methods in genome research, (S Suhai). Plenum Press, New York pp 255–277
19. Jalkanen KJ, Jurgensen VW, Claussen A, Rahim A, Jensen GM, Wade RC, Nardi F, Jung C, Degtyarenko IM, Nieminen RM, Herrmann F, Knapp-Mohammady M, Niehaus TA, Frimand K, Suhai S (2006) Int J Quantum Chem 106:1160
20. Blokzijl W, Engberts Angew JBFN (1993) Chem Int Ed Engl 32:1545
21. Chandler D (2005) Nature 437:640
22. Rank JA, Baker D (1998) Biophys Chem 71:199
23. Southall NT, Dill KA, Haymet ADJ (2002) J Phys Chem B 106:521
24. Graziano G (2009) Chem Phys Lett 483:67
25. Jonsson M, Skepo M, Linse P (2006) J Phys Chem B 110:8782
26. Thomas AS, Elcock AH (2007) J Am Chem Soc 129:14887
27. Maurice RG, Matthai CC (1999) Phys Rev E 60:3165
28. Maurice RG (1999) PhD Thesis, Cardiff University, UK
29. Flory F (1969) Statistical mechanics of chain molecules. Springer, New York
30. Cornell WD, Cieplak P, Bayly CJ, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5170
31. Keating PN (1966) Phys Rev 145:637
32. Grest GS, Kremer K (1986) Phys Rev A 33:3628

33. Perkins TT, Smith DE, Larson RG, Chu S (1995) Science 286:83
34. Smith SB, Finzi L, Bustamante C (1992) Science 258:1122
35. Rief M, Pascual J, Saraste M, Gaub HE (1999) J Mol Biol 286:553
38. Oesterhelt F, Rief M, Gaub HE (1999) New J Phys 1:6.1
37. Ahsan A, Rucnick J, Bruisma R (1998) Biophys J 74:132
38. Marko JF, Siggia E (1995) Macromolecules 28:209
39. Bouchiat C, Wang MD, Allemand J-F, Strick T, Block SM, Croquette V (1999) Biophys J 76:409
40. Wang MD, Yin H, Landick R, Gelles J, Block SM (1997) Biophys J 72:1335
41. Wittkop M, Kreitmeier S, Goritz D (1996) Phys Rev E 53:838
42. Kasianowicz JJ et al (1996) Proc Natl Acad Sci USA 93:13770
43. Liu H, He J, Tang J, Liu H, Pang P, Cao D, Krstic P, Joseph S, Lindsay S, Nuckolls C (2010) Science 327:64–67
44. Schneider GF, Kowalczyk SW, Calaldo VE, Pandraud G, Zandbergen HW, Vandersypen LMK, Dekker C (2010) Nano Lett 10:3163–3167
45. Randel R, Loebl HC, Matthai CC (2004) Macromol Theory Simul 13:387–391
46. Matthai CC, Loebl HC (2004) Advances in science and technology: modelling and simulating materials, Nanoworld, pp 337–344
47. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) J Comput Phys 151:283–312
48. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comp Chem 4:187–217
49. Song 1L, M R Hobaugh, C Shustak, S Cheley, H Bayley, J E Gouaux (1996) Science 274:1859–1866
50. Buch-Pedersen MJ, Pedersen BP, Veierskov B, Nissen P, Palmgren MG (2009) Pflugers Arch 457:573–579
51. Meller A, Nivon L, Brandin E, Golovchenko J, Branton D (2000) Proc Natl Acad Sci USA 97:1079–84
52. Hoogenraad NJ, Ward LA, Ryan MT (2002) BBA-Mol Cell Res 1592:97–105
53. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2001) Molecular cell biology, 4th edn. W.H. Freeman and company, New York
54. Pfanner N, Truscott KN (2002) Nat Struct Biol 9:234
55. Chacinska A, Pfanner N, Meisinger Chris (2002) TRENDS Cell Biol 12:299
56. Pfanner N, Wiedmann N (2002) Curr Opin Struct Biol 14:400
57. Matouschek A, Glick BS (2001) Nat Struct Biol 8:284
58. Okamoto K, Brinker A, Paschen SA, Moarefi I, Hayer-Hartl M, Neupert W, Brunner M (2002) EMBO 21:3659
59. Simon SM, Peskin CS, Oster GF (1992) Proc Natl Acad Sci USA 89:3770
60. Matouschek A, Pfanner N, Voos W (2000) EMBO reports 1:404
61. Loebl HC, Matthai CC (2004) Physica 342:612–622
62. Loebl HC, Randel R, Goodwin SP, Matthai CC (2003) Phys Rev E 67:041913
63. Sakaue T, Yoshikawa K (2002) J Chem Phys 117:6323
64. Bauer MF, Hofmann S, Neupert W, Brunner M (2001) TRENDS Cell Biol 10:25
65. Serdyuk IN, Zaccai NR, Zaccai J (2007) Methods in molecular biophysics: structure–dynamics–function. Cambridge University Press, Cambridge
66. Zaccai G (2010) Acta Cryst D 66:1224
67. Jasnin M, van Eijck L, Koza MM, Peters J, Laguri C, Lortat-Jacob H, Zaccai G (2010) Phys Chem Chem Phys 12:3360
68. Dobson CM (2003) Nature 426:994
69. Ma WJ, Hu CK (2010) J Phys Soc Japan 79:054001
70. Neumann KC, Nagy A (2008) Nat Methods 5:491–505
71. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM (2005) Nature 438:460–465
72. Grandbois MG, Beyer M, Rief M, Clausen-Schaumann H, Gaub HE (1999) Science 283:1727–1730
73. Hohng S, Zhou R, Nahas MK, Yu J, Schulten K, Lilley DMJ, Ha T (2007) Science 318:279–283
74. Cohen AE, Moerner WE (2006) Proc Natl Acad Sci USA 103:4362–4365
75. Shafran E, Yaniv A, Krichevsky O (2010) Phys Rev Lett 104:128101
76. March NH, Tosi MP (2002) Introduction to liquid state physics world scientific, Singapore
77. Streets AM, Quake SR (2010) PRL 104:178102
78. Wolde PRt, Frenkel D (1997) Science 277:1975
79. Velikov PG, Pan W, Gliko O, Katsonis P, Galkin O (2008) Aspects of physical biology. Springer, Heidelberg, pp 65–95
80. Velikov PG (2007) Cryst Growth Des 7:2796–2810
81. Lee MH, Avdoshenko S, Gutierrez R, Cuniberti G (2010) Phys Rev B 82:155455