

Automating the search of molecular motor templates by evolutionary methods

Jose D. Fernández*, Francisco J. Vico*

*Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga
Severo Ochoa 4, 29590 Málaga, Spain
Phone: +34 952137037*

Abstract

Biological molecular motors are nanoscale devices capable of transforming chemical energy into mechanical work, which are being researched in many scientific disciplines. From a computational point of view, the characteristics and dynamics of these motors are studied at multiple time scales, ranging from very detailed and complex molecular dynamics simulations spanning a few microseconds, to extremely simple and coarse-grained theoretical models of their working cycles. However, this research is performed only in the (relatively few) instances known from molecular biology. In this work, results from elastic network analysis and behaviour-finding methods are applied to explore a subset of the configuration space of template molecular structures that are able to transform chemical energy into directed movement. While using methods based on elastic networks limits the scope of our results, it enables the implementation of computationally lightweight methods, in a way that evolutionary search techniques can be applied to discover novel molecular motor templates. The results show that molecular motion can be attained from a variety of structural configurations. Additionally, these methods enable a new computational way to test hypotheses about molecular motors.

Keywords: Molecular motor, elastic network model, behavior-finding

*Corresponding author

Email addresses: josedavid@geb.uma.es (Jose D. Fernández), fjvico@geb.uma.es (Francisco J. Vico)

1. Introduction

Cells can be deemed as complex, highly organized molecular assemblies. To achieve such a degree of complexity and organization, cells use a wide array of means. Molecular motors, being able to transform chemical energy into (harnessable) mechanical work, are one of such means. These devices underpin most vital functions related to motion and cellular rearrangement at all scales, in all domains of life (Schliwa and Woehlke, 2003). Molecular biology is the natural domain of scientific enquiry for these devices, but interest in them spans most of life sciences, from medicine (Goldstein, 2001) to cladistics (Vale, 2003), to cite just two disparate examples, and also nanotechnology (van den Heuvel and Dekker, 2007).

There are many kinds of biological molecular motors. This work focuses on motors able to push or walk along cytoskeletal filaments. There are three classes of such motors: myosins, kinesins and dyneins. These molecular motors perform many tasks: moving cargo around the cell and changing its shape, inducing cilia and flagella beatings, and muscle cells contractions (Schliwa and Woehlke, 2003; Vale, 2003; Vale and Milligan, 2000). As they are so crucial for most cell functions, they elicit a great deal of scientific interest. They have been extensively studied in order to understand their biochemical and structural features (Vale and Milligan, 2000), and their *processivity*, i.e., the way some of them are able to walk along filaments taking many steps in a row (Gennerich and Vale, 2009; Warshaw, 2005; Yildiz et al., 2004). There are also other instances of biological molecular motors which travel along other cellular filaments, as DNA helicases and RNA polymerases, though mechanical work is not their primary function.

As computers have grown more and more powerful, they have become able to tackle ever more computationally heavy and complex tasks, and computational methods have come to a paramount position in molecular biology research. Specifically, many computational studies based on simulations of kinesin and myosin molecules have been performed. The most detailed models require minute molecular dynamics simulations with explicit solvent (Liu et al., 2006; Koppole et al., 2006; Kawakubo et al., 2005), but the large amount of computations needed to simulate reasonable time spans, even in the nanosecond scale, renders this method impractical for performing many consecutive simulations. More lightweight approaches have been devised, such as rule-based models (Gruenert et al., 2010; Wilson, 2008), classical coarse-grained models (Parker et al., 2009; Taylor and Katsimitsoulia, 2010; Bowling and Palmer, 2009; Craig and Linke, 2009), and even very simple models based on reaction kinetics (Schilstra and Martin, 2006) and ratchet potentials (Ciudad et al., 2006; Jülicher et al., 1997). These lightweight approaches vary over their degree of detail, explanatory power, and rigorous foundation.

The present study can be deemed as a contribution to

this scientific endeavour, as a new computational method to provide new perspectives into this area. We define the concept of *molecular motor template* as an elastic network model of an abstract molecular structure, with one or more catalytic cores and docking sites, which is able to undergo a cycle of conformational changes. This cycle allows the structure to processively move through a filament, if it is suited to the task. The chemical realization or realizability of a given molecular motor template is not the subject of this study. Instead, we focus in gaining insights on evolutionary and structural features of actual molecular motors by studying molecular motor templates as simplified abstract models.

The design of processive molecular motor templates has been considered as a behaviour-finding problem (Lobo, 2010). In this framework, structures are selected when their behaviour (resulting from the interaction of its form with an environment) verifies a set of restrictions. Usually, the evaluation of the behaviour involves the physical simulation of some task (Rieffel et al., 2010). In our case, the task consists in travelling along a filament, and evolutionary computation is used to explore a search space of candidate structures.

Some other questions about proteins have been addressed by evolutionary search techniques, as protein folding (Ortiz, 2000; Taverna and Goldstein, 2002; Mirny et al., 1998; Wroe et al., 2005; Xia and Levitt, 2004) and protein evolutionary dynamics (Xia and Levitt, 2004; Munoz and Deem, 2008). Another work (Togashi and Mikhailov, 2007), which has been a source of inspiration for this study, has applied simulated annealing to study structural properties of peptides. A related work whose approach is very similar to ours has been published recently (Flechsigs and Mikhailov, 2010). However, as far as we can tell, this is the first study targeting questions about molecular motors through evolutionary search techniques and behaviour-finding methods.

1.1. Elastic network models

X-ray crystallography, NMR and related techniques have improved as computer machinery has empowered. The elucidation of the 3D structure of proteins and other macromolecules has changed from an extremely slow, time-consuming process to a largely automated one, enabling a rapid growth of repositories of 3D structure data of proteins. However, the raw data provided by these techniques is still difficult to interpret, and complementary mathematical tools are needed to extract useful knowledge from the data.

In this regard, normal modal analysis (NMA) has long been used to understand the dynamics of protein structures. In NMA, a structure is modelled as a collection of objects (atoms or groups of atoms) subject to a number of energy potentials modelling their interactions. Then, the normal modes of motion of the model are analysed to understand its dynamics. At first, very detailed NMA models were developed, explicitly modelling individual atoms

and detailed quantum potentials (Case, 1994). However, researchers gradually realized that normal mode analysis was able to yield useful results and insights even using simpler and more coarse-grained NMA models (Tirion, 1996). Eventually, a consensus developed: for many proteins, their dynamics are largely dictated by their overall structure rather than by the specific shape and intensity of the energy potentials (Lu, 2005; Doruker and Jernigan, 2003; Takano et al., 2004; Hinsen, 1998; Zheng et al., 2006). Thus, while complex and detailed models retained their utility and are still being developed (Hills et al., 2010; Moritsugu and Smith, 2007), simpler and more elegant NMA models were soon formulated, as the Gaussian Network Model (GNM) (Bahar, 1997; Haliloglu et al., 1997) and the Anisotropic Network Model (ANM) (Atilgan et al., 2001; Eyal et al., 2006).

Figure 1 shows how an elastic network model is obtained from the 3D structure of a protein. In the GNM, proteins are modelled as elastic networks, each residue (i.e., each aminoacid in the peptide chain) represented by a vertex placed at the position of its C^α atom, where vertices are connected if they are within a given cutoff distance. Two conditions are also required: (a) the vertices are assumed to fluctuate around their default positions with random isotropic Gaussian distributions, and (b) the network potential depends on the magnitude of the displacement of the vertices around their default positions (e.g., isotropically). If these conditions are met, the fluctuations of the vertices depend just on the topology of the network, up to a factor (Rader et al., 2006).

Despite its simplicity, this model consistently yields good estimates of the empirically measured fluctuations for a wide range of proteins. The normal vibrational modes predicted by GNM are computed from the Kirchhoff matrix of the underlying graph. The Kirchhoff matrix $\Gamma_{n \times n}$ (also known as the Laplacian matrix) for a graph with vertices $v_1 \dots v_n$ is defined as:

$$\Gamma_{ij} = \begin{cases} \deg(v_i) & \text{if } i = j \\ -1 & \text{if } v_i \text{ and } v_j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Where $\deg(v_i)$ represents the number of connections of the i -th vertex. The normal modes of vibration are given by the spectral decomposition of Γ : each eigenvector X_i represents a vibrational mode, its components representing the amplitude of the movement for each vertex, and the associated eigenvalue λ_i is proportional to the square of the frequency of the mode. In this way, the slowest vibrational modes of many proteins are fairly well predicted by GNM, a fact which has been exploited to explain with reasonable accuracy many structural features of proteins, as their unfolding pathways (Su, 2008), their domain decomposition (Kundu et al., 2004), their conformational changes and the position of their catalytic cores (Yang and Bahar, 2005).

However, GNM has a drawback that becomes problematic for some applications: it tells nothing about the directionality of the fluctuations, because of the isotropic assumption (displacements from the default position affect the potential function in the same way, regardless of direction).

The Anisotropic Network Model (ANM) is one of the simplest models dropping this assumption, and, after linearization of the equations of motion, the potential of the elastic network becomes dependent on the directions of the connections between the vertices, just like in a mass-spring system. For this reason, the ANM is able to predict the directionality of the fluctuations of the vertices, as well as the geometry of the conformational changes. However, this presents a downside: the level of agreement between experimental data and ANM predictions is lower than for the GNM, as the predicted fluctuations are higher than in the GNM (because the motion constraints are weaker) (Rader et al., 2006). This can be overcome by lengthening the cutoff distance (up to 10 Å, and sometimes even higher), at the price of a lower level of realism.

Despite these drawbacks, ANM and related models can be successfully applied to a number of problems: interpolation methods to decompose proteins in structural domains (Zheng and Thirumalai, 2009), describing conformational transitions (Kim, 2002; Petrone and Pande, 2006; Zheng, 2008), describing structural differences between related proteins (Leo-Macias et al., 2005; Echave, 2008), analyzing the deformation of proteins under mechanical stress (Eyal and Bahar, 2008), and protein vibrational dynamics (Yang et al., 2009; Micheletti et al., 2004; Ming and Wall, 2005).

However, it is worth mentioning that normal mode analyses based on elastic network models fail for some proteins, remarkably for the motor protein kinesin (Togashi et al., 2010; Zheng and Doniach, 2003).

2. Methods

Our evolutionary search method evaluates, selects and mutates elastic networks. However, these elastic networks are not themselves the molecular motor templates: rather, each elastic network deterministically encodes for a template. The details will be presented in the following subsections.

2.1. Definition and generation of elastic networks

Elastic networks representing raw structures are created as randomly folded chains of N vertices, roughly modelling a folded peptide. Each vertex has a diameter D . The vertices are laid down in sequence: after setting the first one in place, the successive vertices are placed at random, one after the other, such that each vertex is at a distance L ($L_{min} \geq L \geq L_{max}$) from the preceding vertex. They also must be placed at a distance M

($M_{min} \leq M \leq M_{max}$) from at least K vertices, out of the C preceding vertices in the chain ($K \leq C$). After all the vertices have been placed, elastic links are instantiated between every pair of vertices at a distance lower than M_{max} . The natural length of each elastic link is set to the distance between its corresponding vertices.

In our experiments, $N = 50$, $K = 3$, $C = 10$, $D = 3.8$, $L_{min} = 0.9 \cdot D = 3.42$, $L_{max} = 1.1 \cdot D = 4.18$, $M_{min} = L_{min}$ and $M_{max} = 10$. This parameterization is very similar to the one used in (Togashi and Mikhailov, 2007), but, by setting $K = 3$ and $C = 10$, we ensure that (almost always) no point has any rotational degree of freedom. Once these parameters are fixed, a generated elastic network can be specified by a sequence $[\mathbf{P}_1, \dots, \mathbf{P}_N]$, where each \mathbf{P}_i is the position of the vertex i of the structure.

The value chosen for D is reminiscent of other studies, where each vertex represents an aminoacid, and the distance between the locations of two consecutive aminoacids in a peptide chain is supposed to be 3.8 Å. However, this study is not concerned with the exact scale of the elastic network models. In fact, the direct interpretation (i.e., a correspondence between vertices and aminoacids) is not possible due to the small number of vertices. Instead, the structures generated with this method should be regarded as rough templates of plausible proteins.

2.2. Mutation operator

As evolutionary search methods are to be applied to optimize the structural templates, at least a mutation operator must be defined. A mutation operator must take one elastic network and transform it in some way to produce a new one.

The most simple way to mutate a network is to add a small and random displacement to one or more vertices (Togashi and Mikhailov, 2007), but this straightforward choice makes the fitness landscape too difficult to explore, since to get from some given individual to a better one, many correlated mutations might be required. For example, if an individual might become more efficient by translating or rotating some part of its structure relative to other part, this change would require many precisely coordinated mutations (perhaps tens of them) to happen in sequence, preferably in a short amount of evolutionary time.

Therefore, we need a new kind of mutation making many small, coordinated changes to several vertices in the elastic network (from a few ones to all of them). To fulfil these requirements, a physics-based mutation has been developed (Figure 2). As each network is a spatial configuration of vertices connected by relaxed elastic links, a mutation consists of changing the natural length of one or several elastic links, each one by an independent, random amount. The perturbation of the natural lengths introduce potential energy in the elastic network.

If it is allowed to relax, the relative positions of many vertices will change coordinately (just as originally intended) to relieve the stress. After the relaxation process, the network is validated in order to ensure that all restrictions described in Section 2.1 still apply to the chain of vertices. If it is so, new elastic links are calculated for the new spatial configuration (that is to say, the distance between each pair of vertices is calculated for the new spatial configuration, and elastic links are instantiated if the vertices are near enough).

2.3. Evaluation of molecular motor templates

In this section, we describe how each elastic network deterministically defines a molecular motor template, and how these templates are evaluated.

2.3.1. From elastic networks to templates

As the elastic networks are intended to represent molecular motors, it is necessary to define their catalytic cores and docking sites.

The catalytic core (or cores) comprises the parts of an enzyme carrying out the chemical reaction mediated by it. More specifically, it is the place where the substrate (ligand) of the enzyme adheres by intermolecular forces and steric effects, and subsequently undergoes a chemical reaction. Molecular motors travelling along filaments also have well defined docking sites, the places where they attach to the filament. Each catalytic core, together with its associated docking site (consisting of a set of labelled vertices) is termed a *motor head*.

Catalytic cores are characterized by a precise spatial arrangement of residues, cooperating to stabilize the ligand and to catalyze the chemical reaction. The chemical reaction is often preceded, accompanied and/or followed by conformational changes, which in many cases can be described in terms of deformations, rotations and/or translations of some subunits of the enzyme, relative to other subunits.

The catalytic core is often located in or near hinge sites (i.e., the sites experiencing minimal translation, rotation or distortion during a conformational change). For many enzymes, GNM can explain most of their conformational changes as combinations of a few slow normal modes. Also, the hinge sites (and so the probable location of catalytic cores) can often be explained by GNM as co-located with the local minima (in vibrational amplitude) of a slow normal mode.

For motor proteins, the catalytic core is the pocket (or pockets) where an ATP molecule binds to the protein. The binding and/or the hydrolysis of the ATP molecule, possibly in conjunction with other circumstances, triggers a conformational change in the protein. This conformational change, in turn, performs mechanical work.

The present work is restricted to the structural level: it is not concerned with the larger (and far more difficult) question of designing or modelling catalytic cores *ab initio*, but just with the geometry of the molecule. Thus, in

our method, a catalytic core is just a site where ATP is placed, plus the connections (elastic links) between the bound ATP and some of the nearby residues, just like in (Togashi and Mikhailov, 2007).

This method is inspired by GNM-based protein domain decomposition (Kundu et al., 2004): given that elastic networks are used, it is quite natural to use GNM theory to predict the location of the core or cores. Specifically, we use the third eigenvector X_3 of the Kirchhoff matrix. Each component x_i of X_3 is associated with a vertex v_i in the structure. For most elastic networks with an elongated shape, the normal mode associated with X_3 splits them in three clusters. Each cluster is characterized as a connected subnetwork whose vertices v_i have associated components x_i of the same sign, while the magnitude of the component x_i correlates with the vibrational amplitude of the vertex v_i in that normal mode (see Figure 3, middle). There are two interfaces (hinges) between the clusters, such that two of the clusters are distal (they share just one interface) while the other one is central (it shares two interfaces, each one with one distal cluster).

As predicted by GNM models of actual proteins, the sites where ATP is placed in our elastic networks are near hinge sites, as in (Yang and Bahar, 2005), which can be identified as the interface between two clusters. In this way, two catalytic cores are defined for each elastic network.

Each catalytic core is defined as a pair of vertices (v_i, v_j) at the interface, each one from a different cluster. When an ATP molecule binds to the core, it is placed exactly in the middle of the two vertices, connected by an elastic link to each vertex in the pair. These links are stretched to model the change in potential energy brought by the ATP molecule. If \mathbf{P}_i and \mathbf{P}_j are the positions of the vertices v_i and v_j , $\mathbf{P}_j - \mathbf{P}_i$ is the vector from \mathbf{P}_i to \mathbf{P}_j , and the ATP molecule is placed at a distance $q_{ij} = 0.5 \cdot \|\mathbf{P}_j - \mathbf{P}_i\|$ of each one of the two vertices of the catalytic core (see Section 2.3.2 for further use of the value q_{ij}).

However, in each interface there are many possible pairs, in general. In each interface, the pair of vertices (v_i, v_j) selected as catalytic core is determined by the following heuristics:

- The vertices should not be too close (to avoid pairs of vertices already connected by an elastic link) nor too distant (to avoid unrealistically large elastic links with the catalytic core). The allowed range is $M_{max} = 10 \leq \|\mathbf{P}_j - \mathbf{P}_i\| \leq 11.4 = 3 \cdot D$. If no pair fulfils these requisites, no motor template can be defined from the elastic network.
- The power stroke (as described in Section 2.3.2) of the motor head should maximize the movement along the direction defined by the filament. Heuristically, this criterion can be expressed as finding a pair whose vector $\mathbf{P}_j - \mathbf{P}_i$ is both, perpendicular to,

and coplanar with the direction of the filament. Additionally, for the motor head to be efficient, the pair should not be too close to the filament. These three heuristics are elaborated in the following points.

- The centers of mass (\mathbf{C}_1 and \mathbf{C}_2) of each distal cluster is calculated. The line defined by the vector $\mathbf{C}_2 - \mathbf{C}_1$ is taken as the approximate direction of the filament (see Section 2.3.3). To measure the orthogonality of $\mathbf{C}_2 - \mathbf{C}_1$ and $\mathbf{P}_j - \mathbf{P}_i$, we take the absolute cosine of the angle between them:
$$u_{ij} = \frac{|(\mathbf{P}_j - \mathbf{P}_i) \cdot (\mathbf{C}_2 - \mathbf{C}_1)|}{\|\mathbf{P}_j - \mathbf{P}_i\| \cdot \|\mathbf{C}_2 - \mathbf{C}_1\|}.$$
- Let v_i be the closest vertex in the pair (v_i, v_j) to the line defined by the vector $\mathbf{C}_2 - \mathbf{C}_1$, and let \mathbf{C}_3 be a point in the line defined by $\mathbf{C}_2 - \mathbf{C}_1$, such that the vector $\mathbf{P}_i - \mathbf{C}_3$ is perpendicular to $\mathbf{C}_2 - \mathbf{C}_1$. Let \mathbf{Z} be a vector defined as $\mathbf{Z} = (\mathbf{P}_i - \mathbf{C}_3) \times (\mathbf{P}_j - \mathbf{P}_i)$. Then, the more coplanar $\mathbf{C}_2 - \mathbf{C}_1$ and $\mathbf{P}_j - \mathbf{P}_i$ are, the more perpendicular $\mathbf{C}_2 - \mathbf{C}_1$ and \mathbf{Z} are. To measure the orthogonality of these vectors, we take the absolute cosine of the angle between them:
$$v_{ij} = \frac{|\mathbf{Z} \cdot (\mathbf{C}_2 - \mathbf{C}_1)|}{\|\mathbf{Z}\| \cdot \|\mathbf{C}_2 - \mathbf{C}_1\|}.$$
- Also, if the length of the vector $\mathbf{P}_i - \mathbf{C}_3$ is too short, the motor head will probably be inefficient, as the catalytic core will be too near the filament to produce an efficient power stroke. Heuristically, we take the value $D \cdot 0.05 = 3.8 \cdot 0.05 = 0.19$ as a too short distance. We use a sigmoid function whose value will be near 1 if the distance is not short enough:
$$w_{ij} = 1 - \left(1 + e^{-500 \cdot \left(\frac{\|\mathbf{P}_i - \mathbf{C}_3\|}{3.8} - 0.05 \right)} \right)^{-1}.$$
- The selected pair (v_i, v_j) will be the one with minimal $u_{ij} + v_{ij} + w_{ij}$.

For each interface, a pair of vertices is selected as catalytic core, and the associated docking site is defined as the vertices of the nearest distal cluster (see Figure 3). Thus, two motor heads are defined for each elastic network.

2.3.2. Working cycle and interaction with the filament

The walking behaviour of all known instances of molecular motors using filaments can be described as a periodic sequence of steps, comprising some combination of: attachment and detachment from the filament, switching between two or more well-defined shapes, one or more phases of advancement over the filament, binding of ATP, hydrolyzation of ATP to ADP and a P_i group, and release of the ADP and the P_i group. This is usually called the *working cycle* of the molecular motor. From a computational point of view, a working cycle can be modelled as a state machine, that is to say, a description of the different states of the motor. In fact, kinesin motion has been simulated in this way (Wilson, 2008, 2009). As the details of the working cycles vary a great deal between

different molecular motors (for example, between myosin and kinesin), different state machines are needed for each one.

For simplicity, we will use just one state machine for all our molecular motor templates. Each motor head comprising a catalytic core along with its associated docking site has a separated state machine. The details of this state machine are loosely based on the working cycles of myosin and kinesin, as this work is not concerned with the accurate simulation of any concrete instance of molecular motor. The proposed state machine describes how the catalytic core is coordinated with the docking site. For each state in the machine, the characteristics of the catalytic core, the docking site and the transition to the next state are described (see also Table I for a summary):

1. *Sticky* state: The docking site is detached from the filament, and the catalytic core is without ligand (Figure 4, *a*). This state ends when any vertex of the docking site touches the filament: then, the vertex is fixed to the filament. After this event, the catalytic core presents high affinity for ATP. In all our simulations, we assume that ATP is present in saturated concentration, so a molecule of ATP, represented as a vertex, gets immediately placed in the catalytic core, in the middle of the two vertices of the core, bound to them with elastic links (see Section 2.3.1 and Figure 4, *b*). To model the change in potential energy brought by the binding, the links are stretched by giving them a natural length of $0.5 \cdot q_{ij}$, where q_{ij} is the ideal distance from the ATP molecule to each vertex of the catalytic core, as calculated in Section 2.3.1). Then, the motor head transitions to the next state.
2. *Bound* state: The stretched links introduced in the transition to this state induce a conformational change (Figure 4, *c*), while the docking site remains firmly attached to the filament. The result is the power stroke of the motor head. Meanwhile, if any new vertex of the docking site touches the filament, it gets also fixed to it. After a fixed amount of time T_b passes (see Section 2.3.3), the ATP is hydrolyzed to ADP, and the motor head transitions to the next state.
3. *Nonsticky* state: The change from ATP to ADP induces a change in the docking site. Some of the vertices of the docking site were fixed to the filament in the previous state, but all of them lose the ability to be fixed in this state (in biochemical terms, the docking site now presents low affinity for the filament). However, the unfixed vertices remain in touch with the filament, as long as no force pushes them away from it. The needed force might come, for example, from the activity of the other motor head, or from residual elastic forces in the motor head. When no vertex from the docking site remains in touch with the filament (Figure 4, *d*), this information changes

the state of the catalytic core: the ADP is unbound from the core (in biochemical terms, the core loses affinity for ADP). The elastic links binding the ADP to the catalytic core are severed (Figure 4, *e*). Then, the motor head transitions to the next state.

4. *Relaxing* state: After ADP unbinds, the absence of the severed elastic links triggers another conformational change. In this state, the docking site still presents low affinity for the filament, that is to say, its vertices do not get fixed to the filament even if they become in touch. After a fixed amount of time T_r passes (see Section 2.3.3), the vertices of the docking site regain the ability to get fixed to the filament, and the motor head transitions to the initial state (Figure 4, *f*), completing the cycle.

It is worth mentioning that, while actual ligands like ATP and ADP have been named in the description, this working cycle represents a simplification not relying on the actual chemical interaction between phosphorylated nucleotides and actual molecular motors, while still being a reasonable abstraction of a process similar to actual working cycles (for example, nucleotide binding and phosphorylation are known to change affinity for the filament in kinesins and myosins, respectively (Sablin, 2001)). It can be replaced by any other cycle adequate to enable processivity in the motor templates.

Conformational changes are induced by the attachment and detachment of vertices with stretched links, as in (Togashi and Mikhailov, 2007). Although this mechanism to induce conformational change can be argued to lack justification in biochemical considerations, it fits neatly in the paradigm we are using, i.e. representing proteins as elastic networks. Besides, our proposal does not pursue an accurate simulation of actual macromolecules, but an abstract and computationally tractable model of molecular motor. Furthermore, we expect the evolutionary search techniques to perform well at finding good molecular motor templates even if different mechanisms for conformational changes are implemented.

2.3.3. Simulation

Section 2.3.1 and Section 2.3.2 present the way to define a molecular motor template out of an elastic network. To evaluate this motor template, it is simulated in the following setting:

- A straight filament is defined, composed of consecutive, touching spherical beads, each one with a diameter $D = 3.8$ (the same as the vertices of the motor template).
- The motor template is placed over the filament, such that the line defined by the centers of mass of its two docking sites is parallel to the filament, and at least one vertex from each docking site touches the filament.

- One motor head is set to state *sticky*, while the other is set to state *relaxing* (see Section 2.3.2). The election of which head is set to which state is arbitrary but deterministic.

Then, the function of the motor template is physically simulated. For each vertex v_i in the elastic network, the equation of motion is

$$\ddot{\mathbf{P}}_i = \mathbf{F}_i^e + \mathbf{F}_i^v + \mathbf{F}_i^f + \mathbf{F}_i^x$$

where:

- \mathbf{F}_i^e is the sum of elastic forces on the vertex v_i :

$$\mathbf{F}_i^e = - \sum_{j=1}^N A_{ij} \cdot \frac{\mathbf{P}_j - \mathbf{P}_i}{\|\mathbf{P}_j - \mathbf{P}_i\|} \cdot (n_{ij} - \|\mathbf{P}_j - \mathbf{P}_i\|)$$

where $A_{ij} = 1$ if there is an elastic link between vertices v_i and v_j , and $A_{ij} = 0$ otherwise; and n_{ij} is the natural length of the elastic link.

- \mathbf{F}_i^v is a linear damp force on the vertex v_i :

$$\mathbf{F}_i^v = -\mu \cdot \dot{\mathbf{P}}_i$$

- \mathbf{F}_i^f is a sum of elastic forces to avoid the crossing of the vertex v_i through the filament:

$$\mathbf{F}_i^f = - \sum_j \frac{\mathbf{G}_j - \mathbf{P}_i}{\|\mathbf{G}_j - \mathbf{P}_i\|} \cdot (D - \|\mathbf{G}_j - \mathbf{P}_i\|)$$

where the G_j are the positions of the beads in the filament that are too close to the vertex:

$$\|\mathbf{G}_j - \mathbf{P}_i\| \leq D = 3.8$$

(of course, this force is 0 if the vertex is not too close to any bead of the filament).

- If the vertex v_i is fixed to the filament (see Section 2.3.2), both $\ddot{\mathbf{P}}_i$ and $\dot{\mathbf{P}}_i$ are set to 0, overruling the integration of the equations of motion.

As implied by the equations of motion, the mass of each vertex and the stiffness constant of each elastic link are set to 1, while the linear damping constant is set to $\mu = 0.01$. The state machine of each motor head is also simulated. The constants T_b and T_r (defined in Section 2.3.2) are set to $T_b = 500$ and $T_r = 200$ time units, and the whole system is simulated using a classical fourth-order Runge-Kutta integrator with a time step of 0.1. The simulation runs for 10000 time units, which is enough for the motor template to make 10-15 steps over the filament in optimal conditions. While it might seem that the simulation of just one step is needed to measure the effectiveness of the motor template, this is not enough, since it may malfunction after this initial step

(for example, by permanently losing contact with the filament).

The value of T_b is much larger than T_r , following the hypothesis that processivity in some molecular motors like myosin V is favoured by spending most of its working cycle bound to the filament (Sellers and Veigel, 2006). The actual magnitude of the time unit is not relevant to our study, since we are not concerned with the exact scale of the motor templates, as stated in Section 2.1.

2.4. Evolutionary algorithm

Randomly generated elastic networks induce motor templates that usually do no walk at all. Because of this, molecular templates able to walk are generated using the following evolutionary algorithm:

1. An initial set of K randomly generated elastic networks is created, and evaluated, and the fitness of each one is calculated. This is the initial population, and each elastic network is an individual of the population.
2. For T times:
 - (a) Individuals from the previous population are selected, according to their fitness, to generate a new population of K individuals. Generally, in the i^{th} time that this loop is executed, the previous population is named the $(i-1)^{th}$ generation, and the new population the i^{th} generation.
 - (b) Mutate some of them.
 - (c) Each mutated individual is evaluated and its fitness calculated.

In all the executed instances of the evolutionary algorithm (each one named an *evolutionary run*), $K = 100$, while T ranges from 100 to 200. The evaluation method for elastic networks consists of creating its associated motor template and simulating it. However, for some elastic networks, the definition of the motor template from the elastic network fails at some point (see Section 2.3.1):

- For some networks, the third eigenvector X_3 does not split the vertices in three clearly delimited clusters.
- For other networks, no suitable pair of vertices can be found to represent a catalytic core.

In any of these cases, the elastic networks cannot be evaluated: they are discarded from the evolutionary process. They are classified as *nonevaluable*.

The most straightforward way to measure the fitness of an evaluated motor template is to measure the displacement d of its center of mass along the direction of the filament during the simulation. However, motor templates from randomly generated elastic networks present very low to low displacements almost always (see Figure 5), so this measure does not discriminate for templates that potentially might become good walkers after some mutations. This can be solved by taking into account the number of working cycles completed by each

motor head of the template, measuring it as the number of times n_i that the motor head i transitions to the *bound* state. In this way, templates whose geometrical configuration is good for processivity have an advantage. Finally, the fitness is measured as: $f = d + n_1 + n_2$.

The procedure to select individuals from an existing population A to generate a new population B is the following:

- Let A' be the population A without the nonevaluable elastic networks.
- Let m and M be the minimal and maximal fitness in A' , and f_i the fitness for the individual A'_i . Each A'_i is given a rank $r_i = 1 + p \left(\frac{f_i - m}{M - m} - 1 \right)$, where p is the selective pressure. Effectively, fitness values f_i in the range $[m \dots M]$ are mapped to ranks r_i in the range $[1 - p \dots 1]$.
- For each individual B_j to be in population B , a random number $d_{ij} \in [0 \dots 1]$ is generated for each individual A'_i in A' . Then, B_j is defined as the A'_i with maximal $d_{ij}r_i$: $B_j := A'_k \mid k = \underset{x}{\operatorname{argmax}} (d_{xj}r_x)$.

Therefore, the competitive advantage of an individual with good fitness over other with a bad one is modulated by the selective pressure. We use a relatively high selective pressure $p = 0.01$.

After being selected, individuals from a population B may be mutated. For each individual B_i , a sample s_i is drawn from a Poisson distribution with $\lambda = 0.5$. If $s_i = 0$, the individual is not mutated. Otherwise, the individual is mutated by the procedure described in Section 2.2, and s_i is the number of elastic links e_j whose natural lengths n_j will be perturbed. The elastic links to be perturbed are selected at random, and their perturbed natural lengths are $n'_j = n_j c_j$, where each factor c_j is a random number drawn from a uniform distribution with interval $[0.5 \dots 1.5]$.

3. Results

To measure the effectiveness of the evolutionary algorithm, it was performed 30 times; each one of these times is named an evolutionary run. For each run, 100 random elastic networks have been generated for the corresponding initial population, 3000 in total. Almost all of them are able to walk no farther than a distance $d = 20$ or are nonevaluable (see Section 2.4 for a definition). However, taking as the result of each evolutionary run the longest distance walked by any individual from it, significantly improved individuals (when compared to the randomly generated ones, see Figure 6) can evolve. The profile of one of the evolutionary runs is shown in Figure 5.

In many cases, relatively minor modifications to the elastic networks trigger significant increases in the distance covered by the corresponding motor templates,

hinting that good templates need to be precisely tuned to the working cycle and to the details of the simulation.

Several walking strategies, associated with wide collection of shapes, have evolved. Several examples are provided (see supplementary material). In several instances, the evolutionary algorithm has yield templates using just one catalytic core and one docking site to move, as some forms of myosin (Lister et al., 2004), dynein (Hirakawa et al., 2000) and kinesin (Okada and Hirokawa, 1999). It is worth pointing out that templates that use both motor heads do not use them symmetrically, i.e., in most instances, one motor head is responsible for the motion, and the other is used to secure the template to the filament and/or to maintain a correct orientation. Some of the evolved molecular motor templates present varying degrees of rotation around the filament as they travelled along it, mainly because they are asymmetric. However, it should be noted that some studies have also reported rotation in actual molecular motors (Ali et al., 2002).

Several instances of motor templates walking a relatively large distance have emerged, whose motor heads complete their working cycles in a coordinated fashion. However, it must be noted that in all these cases, they move in an inchworm mode (i.e., one head was always ahead of the other one). As most instances of actual motor proteins with two motor heads have been shown (or are strongly suspected) to move in a hand-over-hand fashion (Gennerich and Vale, 2009; Warshaw, 2005; Yildiz et al., 2004), this discrepancy must be taken into account when analyzing the results.

The methods presented in this work enable a new computational way to test hypotheses about molecular motors. To provide an example, we consider the *spectral gap* question raised by Togashi and Mikhailov (2007). In that work, two motor proteins were studied: myosin and F₁-ATPase. Taking the ANM models of these proteins, the nonzero eigenvalues (squared frequencies) $\lambda_1, \dots, \lambda_n$ of their normal modes were studied in normalized logarithmic scale (the numerical cutoff to consider an eigenvalue as 0 was set to 10^{-12}). They were found to feature what was called a spectral gap: the slowest normal modes (from λ_1 to λ_3) were much slower than all the other modes, thus producing a noticeable gap in the spectral signatures of the elastic networks (see Figure 7). This gap was argued by Togashi et al to significantly influence their structural properties, rendering them robust against external perturbations. As random elastic networks were generated and their spectral signature and structural properties examined, most of them were found to lack both a wide spectral gap and the same robustness as myosin and F₁-ATPase. Therefore, a wide spectral gap was hypothesized to be a salient feature of motor proteins.

We have checked that hypothesis by examining the presence or absence of a wide spectral gap in molecu-

lar motor templates. To be rigorous, the n^{th} spectral gap is defined as the distance between the first and the $(n + 1)^{\text{th}}$ normal modes in normalized logarithmic scale. Both myosin and F₁-ATPase present a wide third spectral gap (~ 0.5 and ~ 1.2 , respectively), so we will use it in our study.

In Figure 8, data from the 30 evolutionary runs are lumped together in two curves, measuring the distribution of the third spectral gaps of elastic networks: one distribution for randomly created ones, the other for the best individuals from each evolutionary run. From both histograms it is apparent that neither very small nor very large spectral gaps are frequent, neither in randomly generated examples nor in the result of evolutionary processes. This represents an inherent bias for moderately large gaps in randomly generated elastic networks. On the other hand, the distribution for evolved elastic networks seems to be biased towards larger gaps than the distribution for random elastic networks, though the data must be interpreted with care, as the sample size of the former is very small in comparison with the latter.

Additional evolutionary runs have been performed, with the elastic networks constrained to have either short or large spectral gaps. Runs with very large spectral gaps have evolved molecular motor templates as easily as unconstrained runs (data not shown). The required large spectral gap usually means that structures have a hinge in a nearly degenerate configuration (i.e., a joint almost having a degree of freedom), thus being able to bend easily in some direction, but that does not affect their performance as molecular motor templates.

However, evolutionary runs requiring elastic networks to have extremely low spectral gaps have performed noticeably worse (data not shown), because networks with this constraint are extremely hard to evolve: most of the elastic networks are nonevaluable (see Section 2.4 and Section 2.3.1), because networks with low spectral gap tend to have globular shapes which are not well partitioned by the algorithm presented in Section 2.3.1. Although some motor templates are able to move a distance comparable to some results from unconstrained evolutionary runs, they move through unacceptable mechanisms, exploiting glitches in the simulation algorithm.

In conclusion, these results shed new light over previous studies (Togashi and Mikhailov, 2007), confirming that a wide spectral gap is a very important (and very probably necessary) feature of molecular motors. However, elastic networks representing randomly folded peptides present an inherent bias for relatively wide spectral gaps, and this fact might constitute a substantial contribution to explaining wide spectral gaps in actual molecular motors.

4. Conclusions and Discussion

In this work, a set of methods to simulate and evolve structural templates of motor proteins have been de-

veloped. By using them, research questions about motor proteins can be tackled from a new angle, providing new insights. An example in this regard has been provided: while it is apparent that the shape of an actual molecular motor is exquisitely adapted to the way it moves through the filament, the evolutionary processes driving this adaptation cannot be directly examined, and researchers must instead rely on comparative studies of related motors (Kull et al., 1998). However, analyzing the results of our evolutionary runs, we can make insights about this evolutionary process. Specifically, the results suggest that, while the shape of a molecular motor must be exquisitely tuned to its working cycle, the characteristics of the filament and the interaction with it, the most difficult aspect of the evolution of motor proteins is not dominated by this tuning process, which seems to be relatively straightforward in our evolutionary runs, but by the emergence of a functional working cycle, which has been taken for granted in this work.

A related argument can be suggested: as it can be safely assumed that the computational resources available to perform our evolutionary runs are vastly inferior to the resources available for motor proteins to emerge through natural selection, the structure of motor proteins can be assumed to emerge from evolutionary processes without much difficulty, probably even if biological and chemical considerations are taken into account. However, while we know that tens of proteins able to walk over cytoskeletal filaments have evolved in eukaryotes, all of them belong to just three families: dyneins, kinesins and myosins. Furthermore, kinesins and myosins are supposed to share a common ancestor (Kull et al., 1998), while dyneins belong to the AAA+ protein superfamily. Although these three families are assumed to have evolved early and thus to have hindered the evolution of further families of walking proteins in a winner-takes-all evolutionary scenario, a stronger hypothesis is that walking proteins are, in fact, hard to evolve. In this scenario, the present work shows that structures able to walk can evolve very easily. Therefore, the difficulty must be in other considerations, most probably in the evolution of a viable working cycle.

Another research question about molecular motors has been studied: the spectral gap of motor proteins. While our results mainly confirm that a wide spectral gap is a very important feature of motor proteins (Togashi and Mikhailov, 2007), they also suggest that, if actual peptides share the same bias as our structures, they might be inherently biased to have wide spectral gaps.

The methods presented here are based on several results from elastic network analysis, a mature and well-proved collection of tools to understand biological macromolecules, as it has been stated in the introduction. The definition of the structures studied here is based on ANM, so the presented method shares some characteristics with it, particularly, the long cutoff distance relative to the

separation between vertices in the chain. For practical considerations, the size of the structures has been kept small, with few vertices, so they can be regarded as very rough and coarse-grained models of hypothetical motor proteins.

The GNM model provides theoretical justification to the way catalytic cores and docking sites are placed within our structures. The basic outline of our model of interaction between the protein structures and ATP molecules has been inspired by previous studies on conformational changes of elastic models of peptides by Togashi and Mikhailov (2007). A very similar approach by Flechsig and Mikhailov (2010) has been developed independently, but it concentrates on modelling just one protein: HCV helicase, an enzyme which moves through DNA strands unzipping them.

In a way, our methods can be regarded as reverse-engineering applications of structural research based on GNM. In this regard, only structural considerations have guided the method, avoiding to tackle extremely complex issues, like the shape, arrangement, coordination and working cycles of catalytic cores and docking sites based on actual biochemical considerations.

It can be argued that elastic network models, and more precisely NMA techniques, are not valid universal models of molecular structures. They are not valid for some proteins (Petroni and Pande, 2006), and, specifically, for the motor protein kinesin (Togashi et al., 2010; Zheng and Doniach, 2003). The conformational changes of the working cycle of kinesin cannot be accounted for by models of interaction (between ATP and a catalytic core) as simple as the one proposed here, but need to be modelled using other approaches (Wang et al., 2007). It is apparent that somewhat important portions of the configuration space of possible molecular motors are unreachable by our method. However, this issue is not as troublesome as it might appear, as the dynamics of many other proteins (among them other molecular machines as myosin (Zheng and Thirumalai, 2009; Zheng and Doniach, 2003) and DNA helicase (Flechsig and Mikhailov, 2009)) can be described to varying degrees of accuracy by elastic network models and NMA techniques. Besides, any model taking into account the complex dynamics of kinesin would probably not be as computationally lightweight as the one we use. And, as evolutionary algorithms require many structures to be evaluated, these performance considerations become very important.

There is another point of possible concern: the structures are not immersed in a thermal bath (simulation of solvent with Brownian fluctuations), which is not negligible at the scale of molecular motors. This must be explained in a satisfactory way. The justification can be provided by examining the work presented in (Cressman et al., 2008). In that work, the simulation of the working cycle of a molecular machine (Togashi and Mikhailov, 2007) was coupled to a multiparticle collision method

(Malevanets and Kapral, 1999, 2000), a computationally costly hydrodynamic model of thermal bath. The resulting simulations were fairly similar to simulations without the solvent. These results can be interpreted as supporting the view that the dynamics of the working cycle can be described (up to a given level of realism) without taking into account the solvent. Furthermore, when the hydrodynamic solvent model was substituted by a less costly (but also less realistic) simulation of thermal bath, the working cycle became very erratic, as several conservation laws were violated.

Therefore, the lack of a (necessarily very costly) model of thermal bath stems from performance considerations. Of course, a thermal bath study might be important for some walking strategies; specifically, it has been argued that the hand-over-hand movement of myosin and kinesin includes biased diffusive steps greatly aided by the random component added by the thermal bath (i.e., a Brownian ratchet mechanism) (Houdusse, 2001; Jülicher et al., 1997). On the other hand, in (Flechsig and Mikhailov, 2010), the molecular motor HCV helicase is simulated with methods very similar to ours, and also without taking into account Brownian fluctuations.

4.1. Future work

Much work remains to be done. A possibility is to expand the model to allow the co-evolution of the working cycle together with the shape of the structure, based on biochemical considerations. A related issue, not considered here, is the configuration of docking sites through the filament: the model presented here assumes that molecular motor templates can attach to the filament at any point and orientation, but this is a simplification whose implications should be explored.

Another interesting research question is the modelling of more types of molecular motors: for example, motors with just one motor head, or motors configured as dimers formed by two identical and attached motor heads, which would resemble more closely actual motors like myosin, kinesin and dynein.

A deeper analysis of the dynamics of the evolutionary process is also of interest, as well as the structural features contributing to processivity. Several ways can be suggested to obtain the characteristic hand-over-hand gait of many molecular motors as kinesin and myosin V: adding a simulation of a thermal bath and tuning the fitness function.

Acknowledgements

The first author is supported by a FPU grant (AP2007-03704) from the Ministerio de Educación of the Spanish Government, and has been supported by the BioEmergences project (code 28892) of the Sixth Framework Programme of the European Union. Our research group has been partially supported by the local government (Junta de Andalucía) through grant P09-TIC-5123.

Supplementary material

Four examples of molecular motor templates are provided. Each one is the best individual from one of the evolutionary runs. For each example, there is a video file showing the simulation of the motor template, and one text file providing the specification of the elastic network and its corresponding motor heads, and the fitness resulting from the evaluation of the structure.

References

- Ali, M. Y., Uemura, S., Adachi, K., Itoh, H., Kinoshita, K., Ishiwata, S., June 2002. Myosin V is a left-handed spiral motor on the right-handed actin helix. *Nature structural biology* 9 (6), 464–467.
URL <http://dx.doi.org/10.1038/nsb803>
- Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., Bahar, I., January 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal* 80 (1), 505–515.
URL [http://dx.doi.org/10.1016/S0006-3495\(01\)76033-X](http://dx.doi.org/10.1016/S0006-3495(01)76033-X)
- Bahar, I., June 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design* 2 (3), 173–181.
URL [http://dx.doi.org/10.1016/S1359-0278\(97\)00024-2](http://dx.doi.org/10.1016/S1359-0278(97)00024-2)
- Bowling, A., Palmer, A. F., June 2009. The small mass assumption applied to the multibody dynamics of motor proteins. *Journal of Biomechanics* 42 (9), 1218–1223.
URL <http://dx.doi.org/10.1016/j.jbiomech.2009.03.017>
- Case, D., 1994. Normal mode analysis of protein dynamics. *Current Opinion in Structural Biology* 4 (2), 285–290.
URL [http://dx.doi.org/10.1016/S0959-440X\(94\)90321-2](http://dx.doi.org/10.1016/S0959-440X(94)90321-2)
- Ciudad, A., Sancho, J., Lacasta, A., November 2006. Dynamics of an inchworm nano-walker. *Physica A: Statistical Mechanics and its Applications* 371 (1), 25–28.
URL <http://dx.doi.org/10.1016/j.physa.2006.04.099>
- Craig, E. M., Linke, H., October 2009. Mechanochemical model for myosin V. *Proceedings of the National Academy of Sciences* 106 (43), 18261–18266.
URL <http://dx.doi.org/10.1073/pnas.0908192106>
- Cressman, A., Togashi, Y., Mikhailov, A. S., Kapral, R., May 2008. Mesoscale modeling of molecular machines: Cyclic dynamics and hydrodynamical fluctuations. *Physical Review E* 77 (5), 050901+.
URL <http://dx.doi.org/10.1103/PhysRevE.77.050901>
- Doruker, P., Jernigan, R. L., 2003. Functional motions can be extracted from on-lattice construction of protein structures. *Proteins: Structure, Function, and Genetics* 53 (2), 174–181.
URL <http://dx.doi.org/10.1002/prot.10486>
- Echave, J., May 2008. Evolutionary divergence of protein structure: The linearly forced elastic network model. *Chemical Physics Letters* 457 (4-6), 413–416.
URL <http://dx.doi.org/10.1016/j.cplett.2008.04.042>
- Eyal, E., Bahar, I., May 2008. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophysical Journal* 94 (9), 3424–3435.
URL <http://dx.doi.org/10.1529/biophysj.107.120733>
- Eyal, E., Yang, L.-W., Bahar, I., November 2006. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 22 (21), 2619–2627.
URL <http://dx.doi.org/10.1093/bioinformatics/btl1448>
- Flechsig, H., Mikhailov, A., 2009. Elastic network modeling of molecular motor HCV helicase: inchworm translocation and DNA unzipping cycles. In: *Dynamics Days Europe 2009 (session on Networks and Time Series)*. pp. 6+.
URL <http://www-dcf.ds.mpg.de/dd09/network.pdf>
- Flechsig, H., Mikhailov, A. S., 2010. Tracing entire operation cycles of molecular motor hepatitis c virus helicase in structurally resolved dynamical simulations. *Proc Natl Acad Sci U S A* 107 (49), 20875–80.
URL <http://www.biomedsearch.com/nih/Tracing-entire-operation-cycles-molecular/21081697.html>
- Gennerich, A., Vale, R. D., February 2009. Walking the walk: how kinesin and dynein coordinate their steps. *Current opinion in cell biology* 21 (1), 59–67.
URL <http://dx.doi.org/10.1016/j.ceb.2008.12.002>
- Goldstein, L. S. B., December 2001. Molecular motors: from one motor many tails to one motor many tales. *Trends in Cell Biology* 11 (12), 477–482.
URL [http://dx.doi.org/10.1016/S0962-8924\(01\)02143-2](http://dx.doi.org/10.1016/S0962-8924(01)02143-2)
- Gruenert, G., Ibrahim, B., Lenser, T., Lohel, M., Hinze, T., Dittrich, P., 2010. Rule-based spatial modeling

- with diffusing, geometrically constrained molecules. *BMC Bioinformatics* 11 (1), 307+.
 URL <http://dx.doi.org/10.1186/1471-2105-11-307>
- Haliloglu, T., Bahar, I., Erman, B., Oct 1997. Gaussian dynamics of folded proteins. *Physical Review Letters* 79 (16), 3090–3093.
 URL <http://dx.doi.org/10.1103/PhysRevLett.79.3090>
- Hills, R. D., Lu, L., Voth, G. A., June 2010. Multiscale coarse-graining of the protein energy landscape. *PLoS Comput Biol* 6 (6), e1000827+.
 URL <http://dx.doi.org/10.1371/journal.pcbi.1000827>
- Hinsen, K., March 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins: Structure, Function, and Genetics* 33 (3), 417–429.
 URL [http://dx.doi.org/10.1002/\(SICI\)1097-0134\(19981115\)33:3<417::AID-PROT10>3.0.CO;2-8](http://dx.doi.org/10.1002/(SICI)1097-0134(19981115)33:3<417::AID-PROT10>3.0.CO;2-8)
- Hirakawa, E., Higuchi, H., Toyoshima, Y. Y., March 2000. Processive movement of single 22S dynein molecules occurs only at low ATP concentrations. *Proceedings of the National Academy of Sciences of the United States of America* 97 (6), 2533–2537.
 URL <http://www.pnas.org/content/97/6/2533.abstract>
- Houdusse, A., April 2001. Myosin motors: missing structures and hidden springs. *Current Opinion in Structural Biology* 11 (2), 182–194.
 URL [http://dx.doi.org/10.1016/S0959-440X\(00\)00188-3](http://dx.doi.org/10.1016/S0959-440X(00)00188-3)
- Jülicher, F., Ajdari, A., Prost, J., Oct 1997. Modeling molecular motors. *Reviews of Modern Physics* 69 (4), 1269–1282.
 URL <http://dx.doi.org/10.1103/RevModPhys.69.1269>
- Kawakubo, T., Okada, O., Minami, T., May 2005. Molecular dynamics simulations of evolved collective motions of atoms in the myosin motor domain upon perturbation of the ATPase pocket. *Biophysical Chemistry* 115 (1), 77–85.
 URL <http://dx.doi.org/10.1016/j.bpc.2004.12.049>
- Kim, M., October 2002. Elastic models of conformational transitions in macromolecules. *Journal of Molecular Graphics and Modelling* 21 (2), 151–160.
 URL [http://dx.doi.org/10.1016/S1093-3263\(02\)00143-2](http://dx.doi.org/10.1016/S1093-3263(02)00143-2)
- Koppole, S., Smith, J. C., Fischer, S., August 2006. Simulations of the myosin II motor reveal a nucleotide-state sensing element that controls the recovery stroke. *Journal of Molecular Biology* 361 (3), 604–616.
 URL <http://dx.doi.org/10.1016/j.jmb.2006.06.022>
- Kull, F. J., Vale, R. D., Fletterick, R. J., November 1998. The case for a common ancestor: kinesin and myosin motor proteins and G proteins. *Journal of Muscle Research and Cell Motility* 19 (8), 877–886.
 URL <http://dx.doi.org/10.1023/A:1005489907021>
- Kundu, S., Sorensen, D. C., Phillips, G. N., Jr, December 2004. Automatic domain decomposition of proteins by a gaussian network model. *Proteins: Structure, Function, and Bioinformatics* 57 (4), 725–733.
 URL <http://dx.doi.org/10.1002/prot.20268>
- Leo-Macias, A., Lopez-Romero, P., Lupyan, D., Zerbino, D., Ortiz, A. R., February 2005. An analysis of core deformations in protein superfamilies. *Biophysical Journal* 88 (2), 1291–1299.
 URL <http://dx.doi.org/10.1529/biophysj.104.052449>
- Lister, I., Schmitz, S., Walker, M., Trinick, J., Buss, F., Veigel, C., Kendrick-Jones, J., March 2004. A monomeric myosin VI with a large working stroke. *The EMBO Journal* 23 (8), 1729–1738.
 URL <http://dx.doi.org/10.1038/sj.emboj.7600180>
- Liu, Y., Scolari, M., Im, W., Woo, H.-J., 2006. Protein-protein interactions in actin-myosin binding and structural effects of R405Q mutation: A molecular dynamics study. *Proteins: Structure, Function, and Bioinformatics* 64 (1), 156–166.
 URL <http://dx.doi.org/10.1002/prot.20993>
- Lobo, D., 2010. Evolutionary development based on genetic regulatory models for behavior-finding. Ph.D. thesis, Universidad de Malaga.
- Lu, M., October 2005. The role of shape in determining molecular motions. *Biophysical Journal* 89 (4), 2395–2401.
 URL <http://dx.doi.org/10.1529/biophysj.105.065904>
- Malevanets, A., Kapral, R., May 1999. Mesoscopic model for solvent dynamics. *J. Chem. Phys.* 110, 8605–8613.
 URL <http://www.chem.utoronto.ca/~rkapral/Papers/mesosolv.pdf>
- Malevanets, A., Kapral, R., April 2000. Solute molecular dynamics in a mesoscale solvent. *J. Chem. Phys.* 112, 7260–7269.
 URL <http://www.chem.utoronto.ca/~rkapral/Papers/hybrid.pdf>

- Micheletti, C., Carloni, P., Maritan, A., March 2004. Accurate and efficient description of protein vibrational dynamics: Comparing molecular dynamics and gaussian models. *Proteins: Structure, Function, and Bioinformatics* 55 (3), 635–645.
URL <http://arxiv.org/pdf/cond-mat/0405145>
- Ming, D., Wall, M. E., Nov 2005. Allostery in a coarse-grained model of protein dynamics. *Physical Review Letters* 95 (19), 198103+.
URL <http://dx.doi.org/10.1103/PhysRevLett.95.198103>
- Mirny, L. A., Abkevich, V. I., Shakhnovich, E. I., April 1998. How evolution makes proteins fold quickly. *Proceedings of the National Academy of Sciences of the United States of America* 95 (9), 4976–4981.
URL <http://dx.doi.org/10.1073/pnas.95.9.4976>
- Moritsugu, K., Smith, J., November 2007. Coarse-grained biomolecular simulation with reach: Realistic extension algorithm via covariance hessian. *Biophysical Journal* 93 (10), 3460–3469.
URL <http://dx.doi.org/10.1529/biophysj.107.111898>
- Munoz, E., Deem, M. W., May 2008. Amino acid alphabet size in protein evolution experiments: better to search a small library thoroughly or a large library sparsely? *Protein Engineering, Design and Selection* 21 (5), 311–317.
URL <http://dx.doi.org/10.1093/protein/gzn007>
- Okada, Y., Hirokawa, N., February 1999. A processive single-headed motor: kinesin superfamily protein KIF1A. *Science* 283 (5405), 1152–1157.
URL <http://dx.doi.org/10.1126/science.283.5405.1152>
- Ortiz, A., October 2000. Sequence evolution and the mechanism of protein folding. *Biophysical Journal* 79 (4), 1787–1799.
URL <http://lola.cbm.uam.es/downloads/pdfs/11023886.pdf>
- Parker, D., Bryant, Z., Delp, S., September 2009. Coarse-grained structural modeling of molecular motors using multibody dynamics. *Cellular and Molecular Bioengineering* 2 (3), 366–374.
URL <http://dx.doi.org/10.1007/s12195-009-0084-4>
- Petrone, P., Pande, V., March 2006. Can conformational change be described by only a few normal modes? *Biophysical Journal* 90 (5), 1583–1593.
URL <http://dx.doi.org/10.1529/biophysj.105.070045>
- Rader, A. J., Chennubhotla, C., Yang, L.-W., Bahar, I., 2006. The Gaussian Network Model: theory and applications. *Chapman & Hall/CRC*, Ch. 3, pp. 41–63.
URL <http://www.ccbb.pitt.edu/people/chakra/pubs/gnmBook.pdf>
- Rieffel, J. A., Valero-Cuevas, F. J., Lipson, H., Apr. 2010. Morphological communication: exploiting coupled dynamics in a complex mechanical structure to achieve locomotion. *Journal of The Royal Society Interface* 7 (45), 613–621.
URL <http://dx.doi.org/10.1098/rsif.2009.0240>
- Sablin, E., December 2001. Nucleotide switches in molecular motors: structural analysis of kinesins and myosins. *Current Opinion in Structural Biology* 11 (6), 716–724.
URL [http://dx.doi.org/10.1016/S0959-440X\(01\)00265-2](http://dx.doi.org/10.1016/S0959-440X(01)00265-2)
- Schilstra, M. J., Martin, S. R., February 2006. An elastically tethered viscous load imposes a regular gait on the motion of myosin-V. simulation of the effect of transient force relaxation on a stochastic process. *Journal of The Royal Society Interface* 3 (6), 153–165.
URL <http://dx.doi.org/10.1098/rsif.2005.0098>
- Schliwa, M., Woehlke, G., April 2003. Molecular motors. *Nature* 422 (6933), 759–765.
URL <http://dx.doi.org/10.1038/nature01601>
- Sellers, J., Veigel, C., February 2006. Walking with myosin V. *Current Opinion in Cell Biology* 18 (1), 68–73.
URL <http://dx.doi.org/10.1016/j.ceb.2005.12.014>
- Su, J., June 2008. Protein unfolding behavior studied by elastic network model. *Biophysical Journal* 94 (12), 4586–4596.
URL <http://dx.doi.org/10.1529/biophysj.107.121665>
- Takano, M., Higo, J., Nakamura, H. K., Sasai, M., December 2004. On the model granularity to simulate protein dynamics: A biological physics view on biomolecular computing. *Natural Computing* 3 (4), 377–393.
URL <http://dx.doi.org/10.1007/s11047-004-2639-6>
- Taverna, D. M., Goldstein, R. A., January 2002. Why are proteins marginally stable? *Proteins* 46 (1), 105–109.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11746707>
- Taylor, W. R., Katsimitsoulia, Z., July 2010. A coarse-grained molecular model for actin-myosin simulation. *Journal of Molecular Graphics and Modelling*.
URL <http://dx.doi.org/10.1016/j.jmgm.2010.06.004>

- Tirion, M. M., Aug 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical Review Letters* 77 (9), 1905–1908.
URL <http://dx.doi.org/10.1103/PhysRevLett.77.1905>
- Togashi, Y., Mikhailov, A. S., May 2007. Nonlinear relaxation dynamics in elastic networks and design principles of molecular machines. *Proceedings of the National Academy of Sciences* 104 (21), 8697–8702.
URL <http://dx.doi.org/10.1073/pnas.0702950104>
- Togashi, Y., Yanagida, T., Mikhailov, A. S., June 2010. Nonlinearity of mechanochemical motions in motor proteins. *PLoS Comput Biol* 6 (6), e1000814+.
URL <http://dx.doi.org/10.1371/journal.pcbi.1000814>
- Vale, R. D., February 2003. The molecular motor toolbox for intracellular transport. *Cell* 112 (4), 467–480.
URL [http://dx.doi.org/10.1016/S0092-8674\(03\)00111-9](http://dx.doi.org/10.1016/S0092-8674(03)00111-9)
- Vale, R. D., Milligan, R. A., April 2000. The way things move: Looking under the hood of molecular motor proteins. *Science* 288 (5463), 88–95.
URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.5025&rep=rep1&type=pdf>
- van den Heuvel, M. G. L., Dekker, C., July 2007. Motor proteins at work for nanotechnology. *Science* 317 (5836), 333–336.
URL <http://dx.doi.org/10.1126/science.1139570>
- Wang, Z., Feng, M., Zheng, W., Fan, D., November 2007. Kinesin is an evolutionarily fine-tuned molecular ratchet-and-pawl device of decisively locked direction. *Biophysical Journal* 93 (10), 3363–3372.
URL <http://dx.doi.org/10.1529/biophysj.107.108233>
- Warshaw, D., May 2005. Differential labeling of myosin V heads with quantum dots allows direct visualization of hand-over-hand processivity. *Biophysical Journal* 88 (5), L30–L32.
URL <http://dx.doi.org/10.1529/biophysj.105.061903>
- Wilson, R. J., 2008. Simulating the kinesin walk: A small step towards understanding dementia. *Computer Modeling and Simulation, UKSIM European Symposium on* 0, 226–231.
URL <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04625276>
- Wilson, R. J., May 2009. Kinesin’s walk: Springy or gated head coordination? *Biosystems* 96 (2), 121–126.
URL <http://dx.doi.org/10.1016/j.biosystems.2008.12.002>
- Wroe, R., Bornberg-Bauer, E., Chan, H. S. S., January 2005. Comparing folding codes in simple heteropolymer models of protein evolutionary landscape: robustness of the superfunnel paradigm. *Biophysical Journal* 88 (1), 118–131.
URL <http://dx.doi.org/10.1529/biophysj.104.050369>
- Xia, Y., Levitt, M., April 2004. Simulating protein evolution in sequence and structure space. *Current Opinion in Structural Biology* 14 (2), 202–207.
URL <http://dx.doi.org/10.1016/j.sbi.2004.03.001>
- Yang, L., Song, G., Jernigan, R. L., July 2009. Comparisons of experimental and computed protein anisotropic temperature factors. *Proteins* 76 (1), 164–175.
URL <http://dx.doi.org/10.1002/prot.22328>
- Yang, L.-W. W., Bahar, I., June 2005. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure (London, England : 1993)* 13 (6), 893–904.
URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1489920/pdf/nihms-9262.pdf>
- Yildiz, A., Tomishige, M., Vale, R. D., Selvin, P. R., January 2004. Kinesin walks hand-over-hand. *Science* 303 (5658), 676–678.
URL <http://dx.doi.org/10.1126/science.1093753>
- Zheng, W., May 2008. A unification of the elastic network model and the gaussian network model for optimal description of protein conformational motions and fluctuations. *Biophysical Journal* 94 (10), 3853–3857.
URL [http://www.cell.com/biophysj/abstract/S0006-3495\(08\)70387-4](http://www.cell.com/biophysj/abstract/S0006-3495(08)70387-4)
- Zheng, W., Brooks, B. R., Thirumalai, D., May 2006. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proceedings of the National Academy of Sciences of the United States of America* 103 (20), 7664–7669.
URL <http://dx.doi.org/10.1073/pnas.0510426103>
- Zheng, W., Doniach, S., November 2003. A comparative study of motor-protein motions by using a simple elastic-network model. *Proceedings of the National Academy of Sciences of the United States of America* 100 (23), 13253–13258.
URL <http://dx.doi.org/10.1073/pnas.2235686100>

Zheng, W., Thirumalai, D., March 2009. Coupling between normal modes drives protein conformational dynamics: Illustrations using allosteric transitions in myosin II. *Biophysical Journal* 96 (6), 2128–2137.
URL <http://dx.doi.org/10.1016/j.bpj.2008.12.3897>

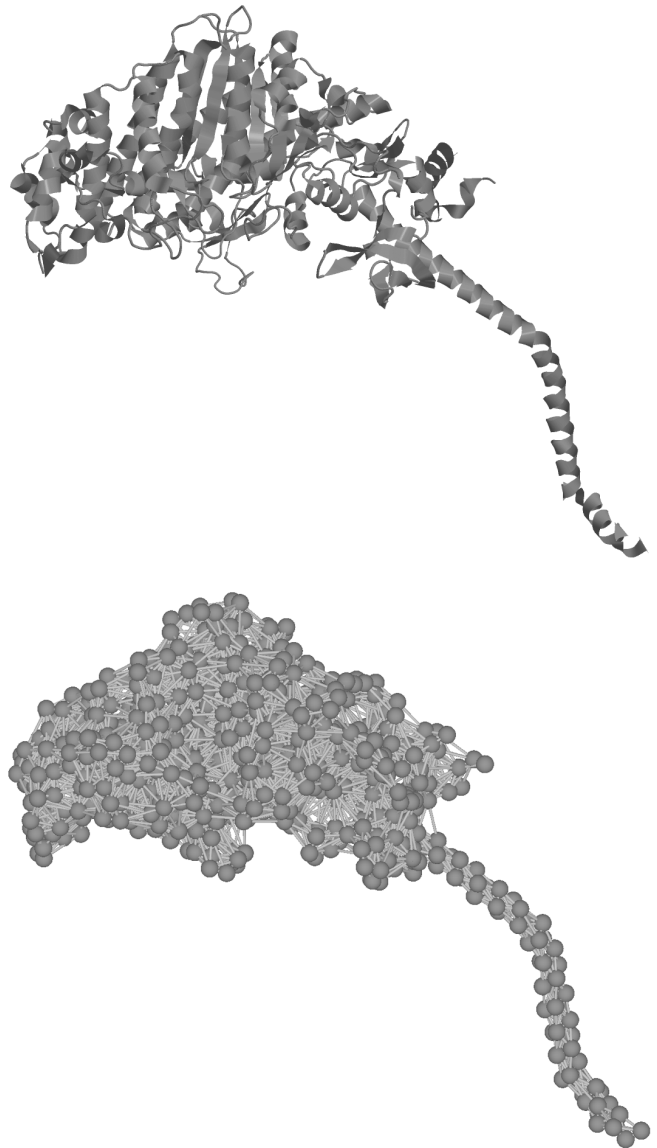


Figure 1: **Elastic network model of myosin.** In this picture, an elastic network model of a single molecule of myosin is represented (below), where each amino acid is replaced by a vertex (with a diameter of 3.8 Å), and vertices are connected if they are within a given cutoff distance (10 Å in this example). For comparison, the ribbon diagram of the protein is also shown (above). The data used in this example and the ribbon diagram are taken from the main peptide chain in the entry 1KK8 of the Protein Data Bank (<http://www.pdb.org/pdb/explore/explore.do?structureId=1KK8>).

<i>State</i>	<i>Description</i>	<i>Vertices in the docking site get fixed to the filament if they touch it</i>	<i>Transition trigger</i>	<i>Transition actions</i>	<i>Next state</i>
sticky	Ready to attach to the filament.	yes	Any vertex from the docking site touches the filament.	ATP with stretched elastic links is bound to the catalytic core, prompting a conformational change.	bound
bound	Power stroke (conformational change).	yes	A fixed amount of time T_b passes.	Vertices in the docking site are unfixed from the filament, though they still may touch it. ATP is hydrolyzed to ADP.	nonsticky
nonsticky	Lose contact with the filament.	no	No vertex from the docking site touches the filament.	ADP is unbound from the catalytic core, prompting a conformational change.	relaxing
relaxing	Recovery (conformational change).	no	A fixed amount of time T_r passes.		sticky

Table I: States of the working cycle of a motor head (catalytic core and associated docking site). In each row, a state is described. The first two columns display its name and description. the third says, for each state, if vertices from the docking site get fixed to the filament upon touching it. The remaining columns describe the transition to the next state: the condition that triggers it, the actions executed at the transition, and the next state.

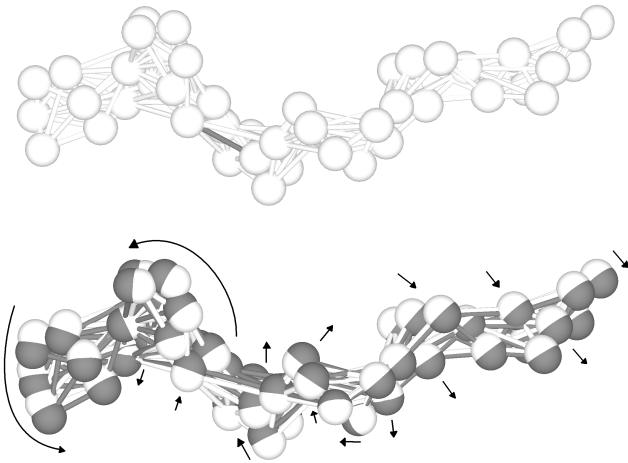


Figure 2: **A mutation example.** A structure is to be mutated (above) by perturbing the natural length of an elastic link (dark gray). In this case, the link is lengthened. The resulting structure after relaxation is shown (below) along with the original structure, in dark gray. The vertices have been displaced, and arrows point towards the main direction of displacement in each part of the structure. The configuration of elastic links is recalculated (some links may disappear, and new ones may appear, if the distances have changed beyond the cutoff value), although in this case no great changes have been done; only natural lengths are changed. After the mutation, the catalytic cores and docking sites of the corresponding molecular motor template are also recalculated (see Figure 3).

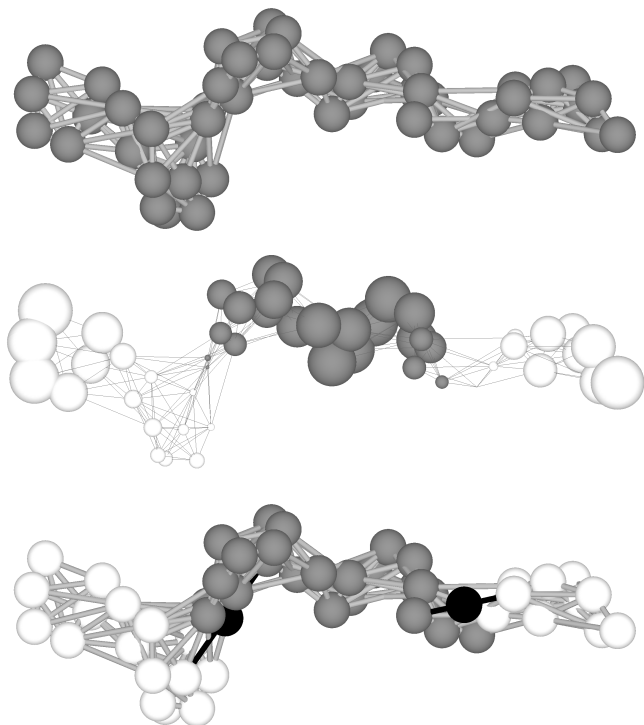


Figure 3: **From structure to molecular motor template.** A structure (above) is processed to determine its proposed catalytic cores and docking sites. The normal mode associated to the third eigenvector X_3 of its Kirchhoff matrix is shown (middle). Each vertex v_i is associated to a component x_i of X_3 , whose magnitude (size) and sign (white positive, gray negative) conveys information about the vibration of the vertex v_i in that normal mode. This normal mode splits the structure into three clusters, characterized as connected substructures whose vertices have components of the same sign. The resulting molecular motor template (below) has two motor heads, each one composed of a catalytic core (ATP and binding links shown in black) placed between a distal cluster and the central one, and a docking site (white) composed of the vertices in the distal cluster. The exact location of the catalytic core is heuristically determined to maximize the effectiveness of the conformational changes brought about by ATP ligands.

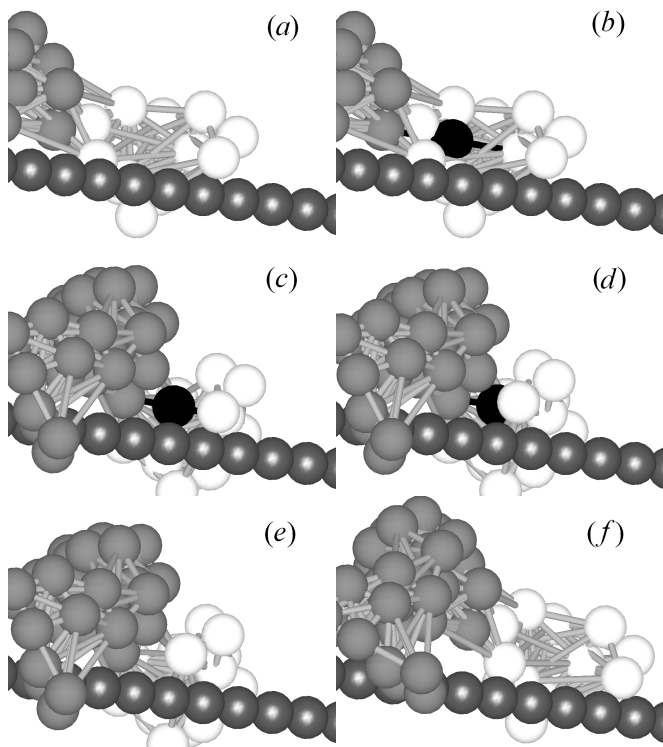


Figure 4: **Working cycle of a molecular motor template.** To show it clearly, the motor has just one motor head: a catalytic core (ATP and binding links in black) and the corresponding docking site (white). The filament is in shiny dark gray. In subfigure (a), the motor head is in the *sticky* state: there is no ligand and the vertices of the docking site are almost touching the filament (shiny gray spheres). In (b), as the docking site is contacting the filament, the motor head transitions to the *bound* state: an ATP molecule attaches to the catalytic core, the docking site gets fixed to the filament, and the conformational change begins. In (c), after a fixed amount of time T_b , the conformational change is assumed to be done, ATP is hydrolyzed to ADP, and the motor head transitions to the *unsticky* state. In this state, the vertices of the docking site are no longer fixed to the filament. In (d), the docking site has drifted so no vertex touches the filament; this event induces the catalytic core to have low affinity for ADP, severing its elastic links, and triggering a transition to the *relaxing* state (e). In (f), after a fixed amount of time T_r (see Section 2.3.3 for the values of T_b and T_r), the conformational change is assumed to be finished, and the motor head transitions to the *sticky* state, returning to (a).

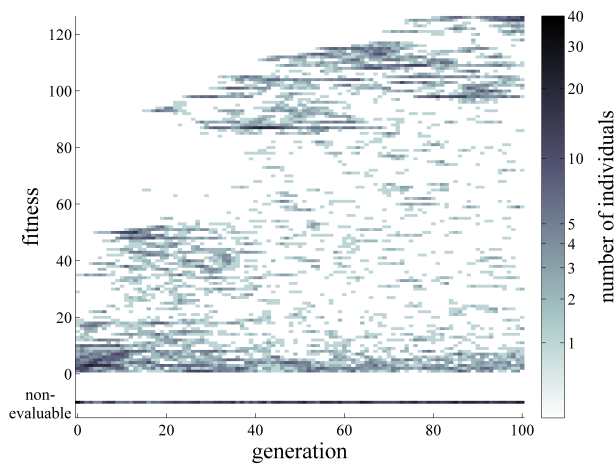


Figure 5: **Evolution of fitness in one evolutionary run.** This is a bidimensional histogram showing the fitness for each individual evaluated in an evolutionary run (one instance of the evolutionary algorithm described in Section 2.4). The X axis represents evolutionary time (*generations*), while the Y axis represents the fitness. The color at each point with coordinates (i, f) represents the number of individuals at the i^{th} generation which have a fitness f (see Section 2.4 for details). As f cannot be calculated for nonevaluable individuals (see Section 2.3.1 and Section 2.3.3 for an explanation), they are binned separately. A quick fitness improvement can be observed in the first generations, followed by a gradual refinement as the evolution progresses.

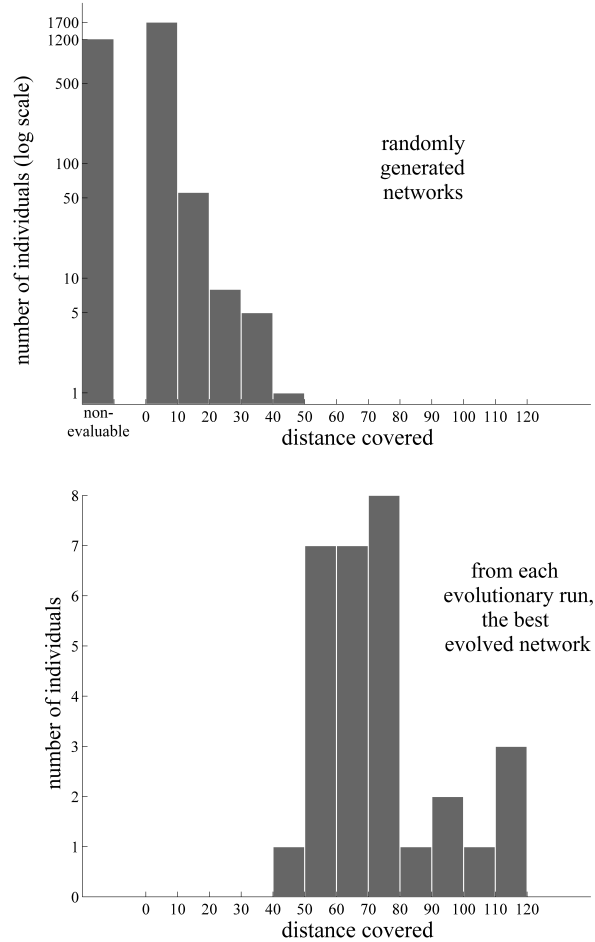


Figure 6: **Comparison of results from randomly generated elastic networks and evolved ones.** 30 evolutionary runs have been executed. The initial population for each one consists of 100 randomly generated elastic networks (the generation procedure is given in Section 2.1), so 3000 random networks have been generated. The results of the simulation of the motor templates associated to these networks is shown in the first histogram (above, Y axis in logarithmic scale): a significant fraction of them (~ 1200) were nonevaluable, that is to say, no motor template could be generated for them (see Section 2.3.1 and Section 2.3.3 for details). Most of the rest (~ 1700) were no able to cover a distance d (that is to say, to move along the filament) longer than $d = 10$. In the second histogram (below, Y axis in linear scale), the covered distance of the best individual in each evolutionary run is shown. The improvement over randomly generated networks is significative. The scale in the Y axis is not the same for both histograms.

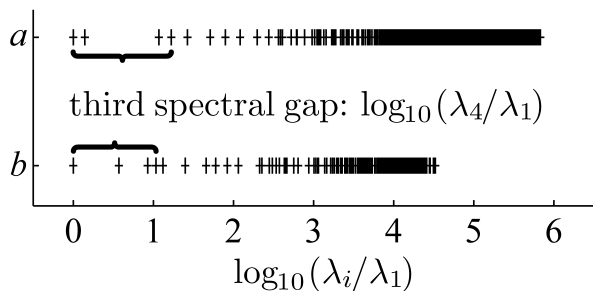


Figure 7: **Spectral gaps.** Spectra for two elastic networks, myosin as shown in Figure 1 (row *a*), and the elastic network shown in Figure 3 (row *b*). The n^{th} - spectral gap is the normalized distance in logarithmic scale between the first (λ_1) and the $(n+1)^{th}$ (λ_{n+1}) normal modes of the structure in ANM. The third spectral gap for myosin (1.224) and the other structure (1.033) are highlighted. Actual molecular motors and most molecular motor templates shown in this work usually have wide spectral gaps.

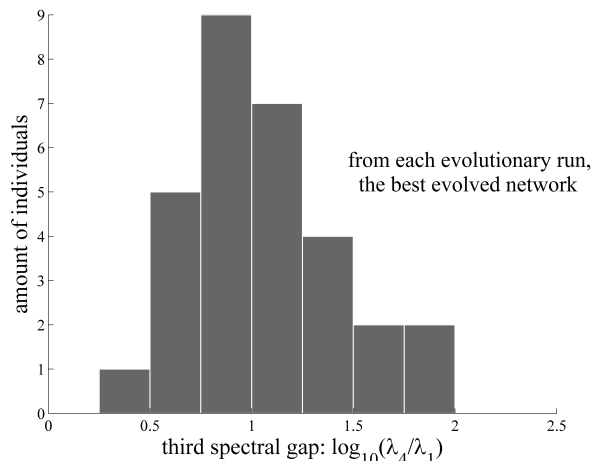
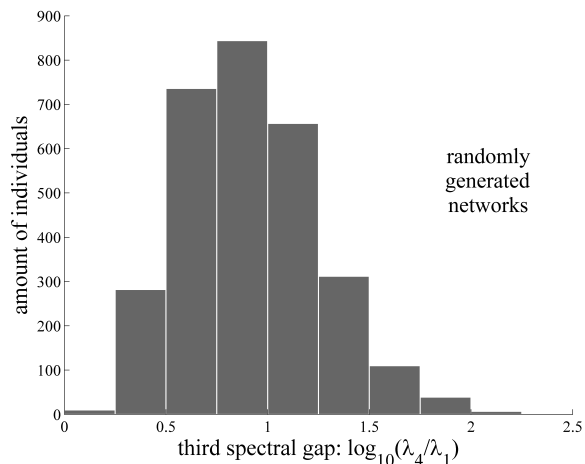


Figure 8: **Spectral gap distributions for random and evolved individuals.** Distributions of the third spectral gap (see Figure 7) for randomly created and for evolved individuals in the 30 evolutionary runs from Figure 6. The third spectral gap is the normalized distance in logarithmic scale between the nonzero eigenvalues (squared frequencies) λ_1 and λ_4 of the first and the fourth normal modes of the structure in ANM. It is a good implementation of the concept of spectral gap to be used with molecular motor templates, as the first three normal modes have usually frequencies in the same order of magnitude. It is also a good measure of the spectral gap for the motor proteins analyzed in (Togashi and Mikhailov, 2007). *Above.* Distribution of spectral gaps for 3000 randomly generated elastic networks. Both small and large spectral gaps are very rare. *Below.* Distribution of spectral gaps for the best individual from each evolutionary run (30 in total). While the interpretation must be cautious because of the small sample size of the distribution, it is similar to the distribution for randomly generated networks, though skewed towards larger gaps. The scale in the Y axis is not the same for both histograms.