

Computational Studies of Viral Protein Nano-Actuators

A. Dubey,^a G. Sharma,^a C. Mavroidis,^{a,*} M. S. Tomassone,^b
K. Nikitzuk,^c and M. L. Yarmush^c

^aDept. of Mechanical & Industrial Engineering, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

^bDept. Chemical & Biochemical Engineering, Rutgers University, 98 Brett Road, Piscataway, NJ 08854, USA

^cDept. of Biomedical Engineering, Rutgers University, 98 Brett Road, Piscataway, NJ 08854, USA

Delivered by Ingenta to

Dynamic and kinematic analyses are performed to predict the performance of a new nanoscale biomolecular motor: The Viral Protein Linear (VPL) Motor. The motor is based on a conformational change observed in a family of viral envelope proteins when subjected to a changing pH environment. The conformational change produces a motion of about 10 nm, making the VPL a basic linear actuator which can be further interfaced with other organic/inorganic nanoscale components such as DNA actuators and carbon nanotubes. This paper presents the principle of operation of the VPL motor, the development of dynamic and kinematic models to study their performance, and preliminary results obtained from the developed computational tools.

Keywords: Bio-Nano-Actuators, Molecular Motors, Molecular Dynamics, Molecular Kinematics.

1. INTRODUCTION

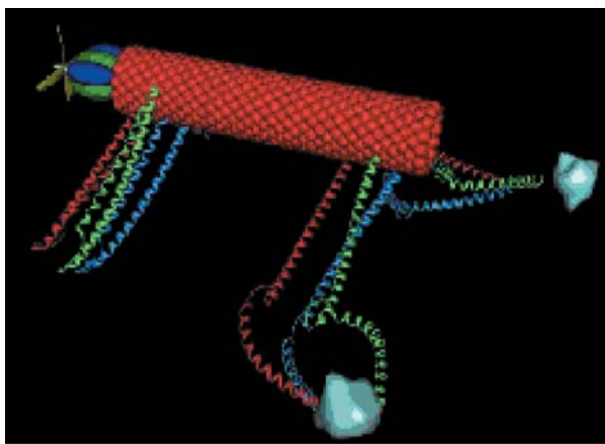
The recent explosion of research in nanotechnology, combined with important discoveries in molecular biology, has created a new interest in biomolecular machines and robots. The main goal in the field of biomolecular machines is to use various biological elements—whose function at the cellular level creates a motion, force or a signal—as machine components that perform the same function in response to the same biological stimuli but in an artificial setting. In this way, proteins and DNA could act as motors, mechanical joints, transmission elements or sensors. If all these different components were assembled together, they could potentially form nanodevices with multiple degrees of freedom, able to apply forces and manipulate objects in the nanoscale world, transfer information from the nano to the macroscale world and even travel in a nanoscale environment.

Just as conventional macromachines are used to develop forces and motions to accomplish specific tasks, bio-nanomachines can be used to manipulate nano-objects, to assemble

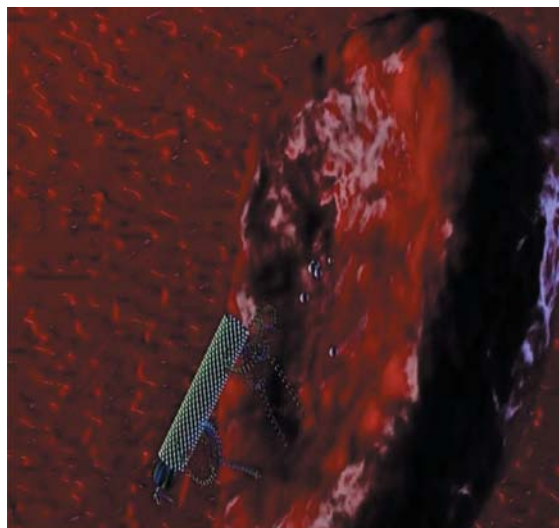
and fabricate other machines or products, and to perform maintenance, repair and inspection operations. The advantages in developing bio-nanomachines include: a) energy efficiency due to their intermolecular and interatomic interactions; b) low maintenance needs and high reliability due to the lack of wear and also due to nature's homeostatic mechanisms (self-optimization and self-adaptation); c) low cost of production due to their small size and natural existence. Figure 1a shows one such concept of a nano-organism, with its 'feet' made of helical peptides and its body using carbon nanotubes, while the power unit is a biomolecular motor. Figure 1b shows an example of a biomolecular nanorobot repairing an infected cell in a blood vessel. Our goal is that just as conventional macromachines are used to develop forces and motions to accomplish specific tasks, bio-nanomachines can be used to manipulate nano-objects, to assemble and fabricate other machines or products, to perform maintenance, repair and inspection operations.

In this project, we are studying the development of Viral Protein Linear (VPL) nanomotors and their integration as actuators in bio-nanorobotic systems. The project consists

*Author to whom correspondence should be addressed.



(a)



(b)

Figure 1. (a) A vision of a nano-organism: carbon nanotubes form the main body; peptide limbs can be used for locomotion and object manipulation. A biomolecular motor located at the head can propel the device in various environments; (b) A “nanorobot” flowing inside a blood vessel finds an infected cell. The nanorobot attaches on the cell and projects a drug to repair or destroy the infected cell.

of three research phases: 1) Development of concepts for novel bio-nanomotors and devices; 2) Performance of computational studies to develop models and design procedures that will predict and optimize the performance of the proposed bio-nanomotors and systems; and 3) Execution of experimental studies to demonstrate the validity of the proposed concepts, models and design methodologies. In this paper, we present the current activities and results for the first two phases. More specifically, we present the principle of operation of the VPL motor, the development of dynamic and kinematic models to study their performance, and preliminary results obtained from the developed computational tools.

2. BACKGROUND

There is a novel engineering interest in utilizing machines that have always been an integral part of all life. These motors, which are called *Biomolecular Motors*, have attracted a great deal of attention recently because they have high efficiency, they could be self-replicating, and hence cheaper in mass usage, and they are readily available in nature. A number of enzymes such as kinesin,^{1, 2} RNA polymerase,³ myosin,⁴ dynein⁵ and adenosine triphosphate (ATP) synthase⁶ function as nanoscale linear, oscillatory or rotary biological motors. Other machines that have been extensively studied include the flagella motors⁷ and the rotaxanes,⁸ which are examples of purely chemical motors. In addition, there are compliance devices such as spring-like proteins called fibronectin⁹ and vorticellids,¹⁰ as well as synthetic contractile plant polymers.¹¹

In addition to protein-based machines, several researchers are exploring the use of DNA in nanoscale mechanisms. Compared with protein structures, DNA is small, simple and homogeneous, and its structure and function are well understood. The generally predictable self-assembling nature of the double helix makes it an attractive candidate for engineered nanostructures. This property was exploited to build several complex geometric structures, including knots, cubes and various polyhedra.¹² Mathematical analyses of the elastic structure of DNA using energy minimization methods were performed to examine its molecular stability, wherein short DNA strands were treated as an elastic rod.¹³ Initial physical experiments on DNA visualization and manipulation using mechanical, electrical and chemical means have been underway for a decade.^{14, 15} A dynamic device providing atomic displacements of 2–6 nm was proposed in,¹⁶ wherein the chemically induced transition between the B and Z DNA morphologies acts as a moving nanoscale device. A method for localized element-specific motion control was seen in the reversible transition between four stranded topoisomeric DNA motifs (PX and JX₂), thereby producing rotary motion.¹⁷ A very important, though simple DNA machine that resembles a pair of tweezers was successfully created. Its actuation is also fueled by adding DNA fuel strands.¹⁸

Application of kinematics is a relatively new idea in molecular simulations. However, efforts to describe 3-D molecular conformations have been made for the past three decades. There has been work on problems such as molecular docking,¹⁹ protein folding, receptor-ligand binding,²⁰ alpha-helices packing^{21, 22} and applications in drug design.²³ There have also been efforts to generate algorithms to perform conformational search—i.e., search for feasible (low energy) conformations^{24–28} and derivation of molecular conformations.²⁹ Efficient maintenance of molecular conformations can greatly impact the performance of conformational search procedures, energy minimization procedures, and all computations that involve large molecules and require frequent recalculation of conformations.

3. THE VPL MOTOR AND ITS TYPES

In this project, we are focusing on the mechanical properties of viral proteins to change their 3-D conformation depending on the pH level of environment. Thus, a new linear biomolecular actuator type is obtained that we call: Viral Protein Linear (VPL) motor.

So far, we know, to a great degree of accuracy, the role of envelope glycoproteins of various retroviruses for the process of membrane fusion, which is a process necessary for the virus to be able to infect a cell. During the process of membrane fusion, there is a distinct conformational change in the peptide on the viral surface as it 'readies' itself for infecting the cell. This change is due to the pH change associated with the vicinity of the cell. Given similar conditions, it is proposed to use this conformational change to produce VPL motors. We have selected the following viral peptides for further consideration in this project: 1) The Influenza virus protein Hemagglutinin (HA) peptide HA2; 2) The Human Immunodeficiency Virus type 1 (HIV-1) peptide gp41; 3) The Human Respiratory Syncytial Virus (HRSV) protein subunit F1; 4) The Simian Immunodeficiency Virus (SIV) protein gp41; 5) The Human T cell Leukemia virus type 1 protein gp21; 6) The Simian Parainfluenza Virus peptide unit SV5; 7) The Ebola virus protein gp2.

Each one of these peptides can result in a different VPL motor that can have different properties in regard to weight, volume, range of motion, force and speed capabilities. However, the principle of actuation is the same. Studies have shown that the common characteristic in these viruses is the structure of a portion of the surface protein (envelope glycoprotein) and the mode of infection. The envelope glycoproteins of these viruses can be divided into two subunits, which are a result of proteolytic cleavage of a common precursor protein. The two subunits have different functions. For example, in the case of HIV-1 the precursor glycoprotein is gp160, which is proteolytically cleaved into gp120 and gp41 subunits. The gp120 is the surface subunit and the gp41 is the transmembrane (TM) subunit. The surface subunit, with the help of receptors located on the cell surface, serves to recognize the cell to be infected when it comes in the vicinity of the virus. The gp41 mediates membrane fusion between the viral and cellular membranes. It has been found that gp41, and corresponding TM subunits in other (above listed) viruses, acquire an alpha-helical conformation when the virus is in its active or fusogenic state. The structure is like a hairpin composed of three coils, having one C terminal (carboxy-end) and the other an N terminal (amino-end). This coiled coil structure undergoes a conformational change induced by mildly acidic conditions (i.e., pH around 5). This change is required for the process of membrane fusion, i.e., the fusion of viral and cellular membranes essential for infection of the cell. With the change in pH, the N-terminals pop out of the inner side and the peptide acquires a straightened position or fusogenic state.

In the first stages of this project, computational and experimental studies are performed using the Influenza virus protein Hemagglutinin (HA) as the basis for forming a VPL motor. The reason for making this peptide selection is that, based on current literature, this peptide seems to be able to perform repeatable motion controlled by variation of the pH.

The X-ray crystallographic structure of bromelain-released soluble ectodomain of Influenza envelope glycoprotein hemagglutinin (BHA) was solved in 1981.³⁰ BHA and pure HA were shown to undergo similar pH-dependant conformational changes which lead to membrane fusion.³¹ HA consists of two polypeptide chain subunits (HA1 and HA2) linked by a disulfide bond. HA1 contains sialic acid binding sites, which respond to the cell surface receptors of the target cells and, hence, help the virus to recognize a cell.³² Out of the various theories proposed to explain the process of membrane fusion, the spring-loaded conformational change theory³³ is the most widely accepted. According to this model, there is a specific region (sequence) in HA2 that tends to form a coiled coil. In the original X-ray structure of native HA, this region is simply a random loop. It further states that, upon activation, a 36 amino-acid residue region makes a dramatic conformational change from a loop to a triple-stranded extended coiled coil along with some residues of a short α -helix that precedes it. This process relocates the hydrophobic fusion peptide (and the N-terminal of the peptide) by about 10 nm. In a sense of biomimicking, we are engineering a peptide identical to the 36-residue long peptide mentioned above, which we call *loop36*. Cutting out the *loop36* from the VPL motor, we obtain a peptide that has a closed length of about 4 nm and an extended length of about 6 nm, giving it an extension by two-thirds of its length. Once characterized, the peptide will be subjected to conditions similar to what a virus experiences in the proximity of a cell, that is, a reduced pH. The resulting conformational change can be monitored by fluorescence tagging techniques and the forces can be measured using Atomic Force Microscope.

Figures 2a and 2b show a schematic of the VPL motor supporting a moving platform. The motor is shown in its initial, "contracted" phase that corresponds to the virus' native state (Figure 2a) and at its extended, fusogenic state (Figure 2b). The total outward protrusion is measured to be 10 nanometers. To augment the force capabilities of the VPL motors, several VPL actuating elements could be attached in parallel as it is shown in Figure 3. Such parallel attachment of multiple VPL motors could result in extremely powerful, micro-, meso- or even macro-actuators that will be able to apply ultra large forces while their dimensions are extremely small. To increase the displacement capability of the VPL motors, several VPL elements could be connected in a series (Figure 3).

To develop nanomechanical assemblies such as mechanisms that will be powered by the proposed VPL motors, we need to investigate the possibility for the VPL motors to be interfaced with other machine components so that they form

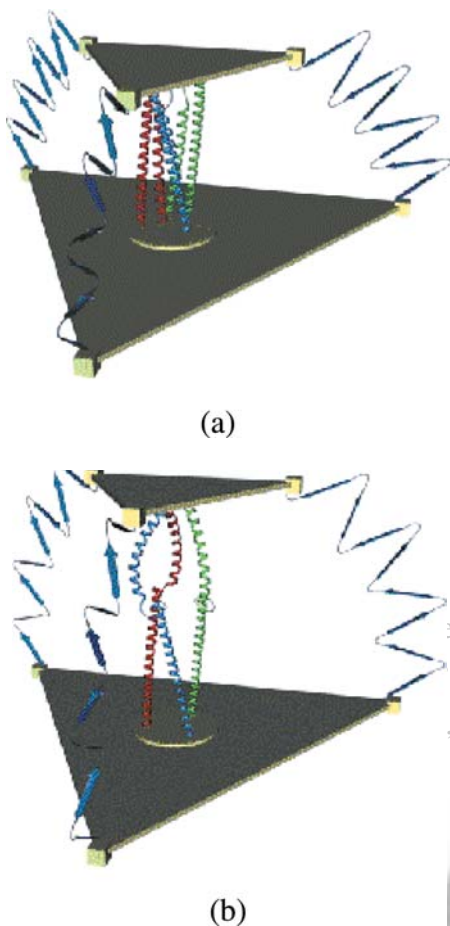


Figure 2. (a) Three titin fibers can be used as passive spring elements to join two platforms and form a single-degree-of-freedom parallel platform that is actuated by a Viral Protein Linear (VPL) actuator (center). (b) The VPL actuator has stretched out, resulting in the upward linear motion of the platform. The three titin fibers are also stretched out.

multi-degree-of-freedom bio-nanodevices. The idea is to use the VPL motors as the actuators of such nanodevices, where the structural elements are carbon nanotubes, while the joints are formed by appropriately designed DNA elements. A futuristic example of such a bio-nanomachine is shown in Figure 4. This is a 3-degree of freedom, 3-legged parallel manipulator. The top and bottom platform are made from carbon nanotubes. They are connected to each other using three legs made from the VPL motors. The interface of the VPL motors to each one of the carbon nanotube platforms is made using DNA-based universal joints.

4. MOLECULAR DYNAMICS

To predict the dynamic performance of the proposed VPL motors (i.e., energy and force calculation), we are performing Molecular Dynamics (MD) Simulations based on the calculation of the free energy that is released during the transition from native to fusogenic state. We used MD software called CHARMM (Chemistry at Harvard Molecular

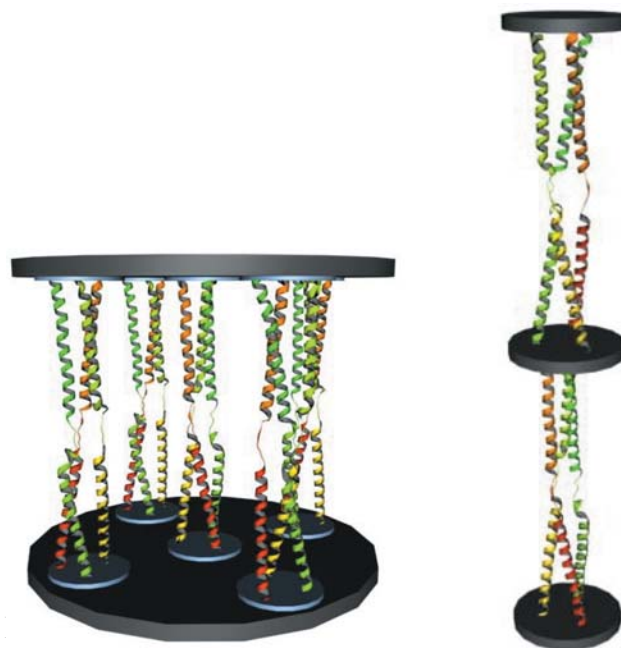


Figure 3. Several VPL motors placed in parallel (left) and series (right) to multiply force and displacement, respectively.

Mechanics).³⁴ In MD, the feasibility of a particular conformation of the biomolecule in question is dictated by the energy constraints. Hence, a transition from one given state to another must be energetically favorable, unless there is an external impetus that helps the molecule overcome the energy barrier. When a macromolecule changes conformation, the interactions of its individual atoms with each other—as well as with the solvent—compose a very complex force system. The CHARMM energy function is given by:

$$E = E_b + E_\theta + E_\varphi + E_\omega + E_{vdw} + E_{el} + E_{hb} + E_{cr} + E_{c\varphi} \quad (1)$$

The individual components of the total energy E are the bond potential E_b , the bond angle potential E_θ , the dihedral

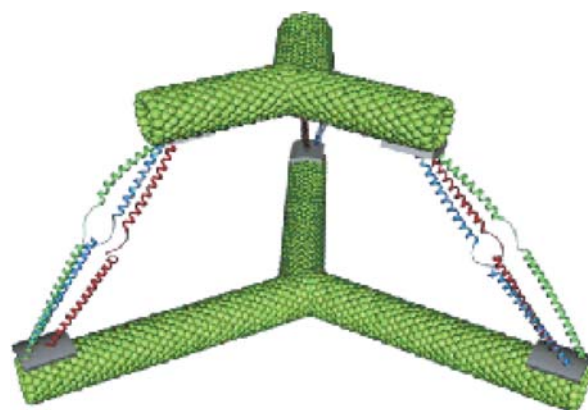


Figure 4. Three-degree-of-freedom parallel platforms can be formed using controllable VPL actuators attached at the three legs of the platform.

angle potential E_φ , the torsion potential E_ω , the Van der Waals Potential E_{vdW} , the electrostatic potential function E_{el} , the hydrogen bond potential E_{hb} , the constraint term E_{cr} , and the dihedral constraint potential $E_{c\varphi}$. The functional forms of these potentials are as follows:

Bond potential is given by

$$E_b = k_b \sum (r - r_0)^2 \quad (2)$$

The variable r is an instantaneous bond length due to an extension or contraction of a bond and r_0 is its original length. Similarly, the bond angle potential is given by

$$E_\theta = k_\theta \sum (\theta - \theta_0)^2 \quad (3)$$

Angle θ is the deviation from the equilibrium value θ_0 of the bond angle for a given set of atoms. Under normal temperature conditions, these two potential values are very small, as the bond lengths and angles do not change appreciably. However, there is rotation around certain bonds (torsion or bond twist). The dihedral angle or torsion potential is thus formulated, which is given by

$$E_\varphi = \sum |k_\varphi| - k_\varphi \cos(n\varphi) \quad (4)$$

This term is a four-atom term based on the dihedral angle φ about an axis defined by the middle pair of atoms. The energy constant in this term may be negative (with a maximum at cis-conformation) and it can consist of several contributions with different values of the constant k_φ . The improper energy term is given by

$$E_\omega = \sum k_\omega (\omega - \omega_0)^2 \quad (5)$$

It maintains chirality about tetrahedral atoms joined to heavy atoms such as α -carbons and planarity about planar atoms such as carbonyl carbon. Some of the force constants (k_i) are obtained from literature and some are obtained by fitting vibrational data. The geometric constants such as r_0 , θ_0 , n , ω_0 are obtained from crystallographic data. All constants are defined and stored in the parameter files, which are called upon every time a structure needs to be built.

The *electrodynamic* effect of Van der Waals' forces is included by using the Lennard-Jones potential as:

$$E_{vdW} = \sum_{\text{excl}(i,j)=1} \varepsilon_{ij} \left(\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}}{r_{ij}} \right)^6 \right) \quad (6)$$

In the above equation, ε_{ij} represents the energy of the minimum (deepest) point on Van der Waals curve for the atom pair $i-j$, R_{ij} is the separation distance between the atom pair $i-j$ at the energy minima, and r_{ij} is the current actual distance between atom i and atom j . The attractive forces between two proximal atoms are proportional to r_{ij}^{-6} and the repulsive forces due to their nuclear repulsions are

proportional to r_{ij}^{-12} . The other important factor in non-bonded interaction is the electrostatic interaction, which for atoms i and j , is given by:

$$E_{el} = \sum_{\text{excl}(i,j)=1} \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon_r r_{ij}} \quad (7)$$

with q_i and q_j being the partial charges on atoms separated by a distance ε_0 and ε_r are the permittivity of free space and the relative permittivity of the medium separating the atoms in question, respectively.

The hydrogen bond potential has the general form as:

$$E_{hb} = \sum \left(\frac{A}{r_{AD}^i} - \frac{B}{r_{AD}^j} \right) \cos^m(\theta_{A-H-D}) \times \cos^n(\theta_{AA-A-H}) \times sw(r_{AD}^2, r_{hon}^2, r_{hoff}^2) \times sw[\cos^2(\theta_{A-H-D}), \cos^2(\theta_{hon}), \cos^2(\theta_{hoff})] \quad (8)$$

The abbreviations AA, A, H and D represent acceptor-acceptor, acceptor, hydrogen atom and donor atoms, respectively, with angles and distances designated accordingly. The indices i and j are positive integers with $i < j$. Function sw is the cubic spline switching function used to achieve smooth cut-offs.

The atom harmonics constraints take the form

$$E_{cr} = K_i (r_i - r_{i0})^2 \quad (9)$$

This constraint potential is used to maintain the position r_i of certain chosen atoms (denoted by the index i) about their mean positions r_{i0} and to avoid large displacements when minimizing energy. Constant rigid distance constraints between two atoms i and j can be placed as $\delta r_{ij} = 0$. Dihedral constraints can be used to maintain certain local conformations by restricting the dihedral angle of an atom i by

$$E_{c\varphi} = \sum K_i (\varphi_i - \varphi_{i0})^2 \quad (10)$$

where φ_i and φ_{i0} are instantaneous and equilibrium dihedral angle values, respectively.

With CHARMM, it is possible to model a protein based on its amino acid sequence and allow a transition between two known states of the protein using Targeted Molecular Dynamics (TMD).³⁵

Targeted Molecular Dynamics (TMD) is a toolbox of CHARMM that is used for approximate modeling of processes spanning long timescales and relatively large displacements. Because the distance to be traveled by the N-terminal of the viral protein is relatively large, we cannot let the protein unfold by itself. Instead of 'unfolding' we want it to undergo a large conformational change and 'open' up. To achieve this, the macromolecule will be 'forced' toward

a final configuration 'F' from an initial configuration 'I' by applying constraints. The constraint is in the form of a bias in the force field. If we define the $3N$ position coordinates corresponding to N atoms in the molecule as

$$\mathbf{X} = (X_1, X_2, \dots, X_{3N})^T \quad (11)$$

where $3N$ are the Cartesian coordinates of the position vectors $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ of each individual atom, then for each configuration \mathbf{x} , its distance, ρ to the target configuration 'F' is defined as:

$$\rho = |\mathbf{x} - \mathbf{x}_F| = \left[\sum (x_{i_f} - x_{F_i})^2 \right]^{1/2} \quad (12)$$

The distance ρ is a purely geometric control parameter here, which will be used to force the macromolecule to undergo the desired transformation. The constraint applied for this is equal to:

$$f(\mathbf{x}) = |\mathbf{x} - \mathbf{x}_F|^2 - \rho^2 \quad (13)$$

This results in an additional constraint force:

$$\mathbf{F}_c = \lambda \frac{df}{d\mathbf{x}} = 2\lambda[\mathbf{x} - \mathbf{x}_F] \quad (14)$$

where λ is the Lagrange parameter.

The TMD algorithm steps are:

1. Set $\rho = \rho_0 = |\mathbf{x}_I - \mathbf{x}_F|$ where I is the initial and F is the final conformation.
2. Choose initial coordinates $\mathbf{x}_i(0) = \mathbf{x}_{iI}$ and appropriate initial velocities.
3. Solve, numerically, the equations of motion with the additional constraint force \mathbf{F}_c .
4. After each time step Δt diminish ρ by $\Delta\rho = (\rho_0 - \rho_f)\Delta t/t_s$ where t_s is the total simulation time.

At the end of the simulation, the final distance ρ_f is reached. In this way, a monotonous reduction of ρ forces the system to find a pathway from \mathbf{x}_I to a final configuration \mathbf{x}_F .

In this project, the two known states are the native and the fusogenic states of the 36-residue peptide of HA. The structural data on these two states was obtained from the Protein Data Bank (PDB).³⁶ These PDB files contain the precise molecular make up of the proteins, including the size, shape, angle between bonds and a variety of other aspects. We used the PDB entries 1HGF and 1HTM, respectively, as sources for the initial and final states of the peptide.

In a representative simulation, the "open" structure was generated arbitrarily by forcing the structure away from the native conformation with constrained high-temperature molecular dynamics. After a short equilibration, these two "closed" and "open" structures are then used as reference end-point states to study the transformation between the open and closed conformations. The transformation is enforced through a root mean square difference (RMSD) harmonic constraint in conjunction with molecular dynamics

simulations. Both the forward (closed to open) and the reverse (open to closed) transformations are carried out. The RMSD between the two end-point structures is about 9Å, therefore, the transformation is carried out in 91 intermediate steps or windows with a 0.1 Å RMSD spacing between each intermediate window. At each intermediate window, the structure is constrained to be at the required RMSD value away from the starting structure. It is minimized using 100 steps of Steepest Descent minimization, and then equilibrated with 0.5 picoseconds of Langevin dynamics with a friction coefficient of 25 ps on the non-hydrogen atoms. The harmonic RMSD constraint is mass-weighted and has a force constant of 500 kcal/mol/Å² applied only to the non-hydrogen atoms. The decision to calculate the RMSD only for non-hydrogen atoms is usually done by convention since the conformation of the protein is more directly dependent on the heavy atoms than on the hydrogen atoms. Spontaneous transformation between the two conformations using unconstrained molecular dynamics may occur in the microsecond timescale. In the present case, however, the transformation is sped up through the use of the artificial RMSD constraint such that a conformation close to the final state is approached successfully. Figure 5 shows the RMSD values of the structures at the end of each intermediate window from both the native closed (non-fusogenic) and the open (fusogenic) conformations. The two curves correspond to the forward and reverse transformations, respectively, and the difference between the two curves is due to hysteresis. A small RMSD value on the x-axis indicates that the structure is close to the native conformation while a small RMSD value on the y-axis indicates that the structure is close to the closed conformation. In

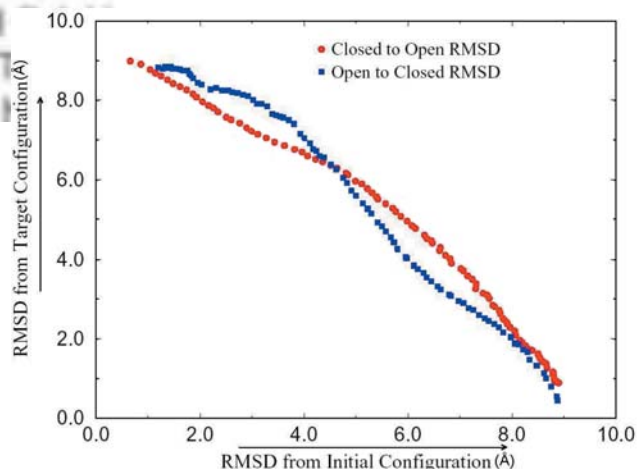


Figure 5. The non-hydrogen RMSD values of the structures at the end of each window in the constrained molecular dynamics transformation between the closed (native, non-fusogenic) and the open (fusogenic) states of the 36 amino acid peptide test system. The RMSD values (in Å) in comparison to the closed state are on the x-axis, those in comparison to the open state are on the y-axis. The two curves correspond to the forward (closed to open) and the reverse (open to closed) transformations, respectively.

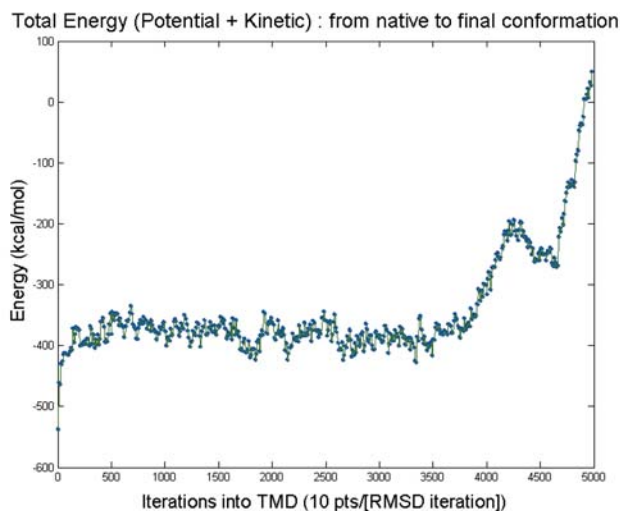


Figure 6. Energy variation for *loop36* peptide with solvation model EEF1.

both curves, it can be seen that structures close to the endpoint states (RMSD < 1 Å) are obtained. Increase in the amount of sampling for each window and decrease in the value of the force constant used in this transformation can lead to a progressively better description of the transformation. The energy plot using solvation function EEF1³⁷ is shown in Figure 6 and the simulation snapshots for the initial and final states are shown in Figures 7 and 8.

By using TMD we have been able to prove that the final state of the VPL shown in Figure 8 is feasible, and it is the result of the action of the applied potentials. However, from the energy graph in Figure 6, we learn that even though the 36-residue long protein is forced to undergo a conformational change, it would require an energy jump to overcome the barrier that appears at approximately 4000 iterations. Unless an external force induces that jump, the protein would not go naturally to that state. TMD works well in the first stages of the change in structure since the protein does not require much provocation to transform into an intermediate conformation. Only after 4000 iterations more energy is required. In fact, the opening of the helical region, followed by the adjustment of the remaining loop into an α -helical



Figure 7. Ribbon drawing of the closed conformation of the 36-residue peptide as obtained from PDB entry 1HGF.

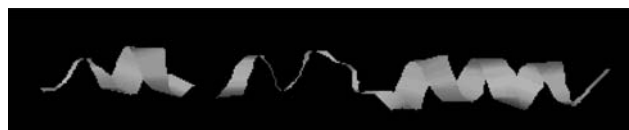


Figure 8. Ribbon drawing of the open conformation as obtained by TMD simulations. There is a noticeable increase in alpha-helical content and the peptide openings.

form, is what requires more energy. This process occurs solely due to the pH drop in the natural setting of the VPL since it is the rest of the large protein attached to the ends of this *loop36* that affects its behavior.

A more realistic environment of the VPL requires the inclusion of the effect of pH on the protein. For this, 10 titratable amino acids (Glutamic Acid—GLU, Aspartic Acid—ASP and Histidine—HIS) were chosen out of the VPL sequence to be protonated. New topology files were used to replace them with their protonated counterparts, and the effect on the structure was observed. So far, there has been very little observable effect; however, new efforts are underway to simulate water molecules explicitly around the protein using stochastic and periodic boundary conditions. We are also considering a model that takes into account the real-time effect of pH on the ionic stability of the protein so as to affect a conformation change.

5. MOLECULAR KINEMATICS

Molecular kinematic simulations are being developed to study the geometric properties and conformational space of the VPL motors. The kinematic analysis is based on the development of direct and inverse kinematic models and their use in the workspace analysis of the VPL motors. In this section, we present the derivation of the direct and inverse kinematic modules that have been incorporated into a MATLAB toolbox called BioKineLab that has been developed in our laboratory to study protein kinematics.

Proteins are macromolecules that are made up from 20 different types of amino acids, also called residues. For kinematic purposes, we consider these residues to be connected in a serial manner to create a serial manipulator. The “back bone” of the chain is a repeat of the Nitrogen – Alpha Carbon–Carbon ($-N-C\alpha-C-$) sequence. A side chain (R), which is different for each residue, is also attached to the $C\alpha$ atom. These side chains are passive 3-D structures with no revolute joints. Hydrogen atoms are neglected because of their small size and weight. The C-N bond joins two amino acid residues and has a partial double-bond character and, thus, is non-rotatable. There are, however, two bonds which are free to rotate. These are the N- $C\alpha$ and $C\alpha$ -C bonds and the rotation angles around them are known as *phi* (φ) and *psi* (ψ), respectively. The value of these angles determines the 3-D structure of the protein and makes it perform its function. Therefore, a protein is considered to be a serial linkage with K+1 solid links connected by K

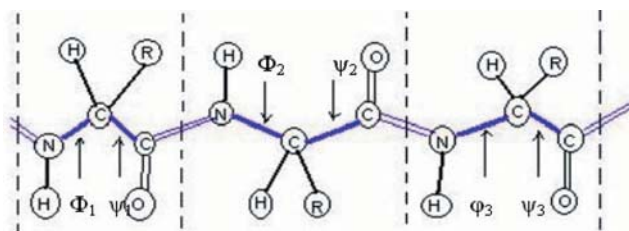


Figure 9. Rotational degrees of freedom along a residue chain. Adjacent residues are separated by dashed lines; side chains are denoted by R; and the purple line represents the back-bone.

revolute joints (Figure 9). In case of *loop36*, K takes the value 72. In most kinematic studies, bond lengths and bond angles are considered constant, while the torsional angles (φ and ψ) are allowed to change.²⁵

5.1. Direct Kinematics

The direct kinematic problem calculates the VPL motor's final configuration. When an initial configuration is given, all constant parameters of the chain are specified and a specific set of rotations for the torsional angles is defined. Frames are affixed at each backbone atom (Figure 10). Let b_i be a bond between atoms Q_i and Q_{i-1} . A local frame $F_{i-1} = \{Q_{i-1}; \mathbf{x}_{i-1}, \mathbf{y}_{i-1}, \mathbf{z}_{i-1}\}$ is attached at bond b_{i-1} as follows: \mathbf{z}_{i-1} has the direction of bond b_{i-1} ; \mathbf{x}_{i-1} is perpendicular to both b_{i-1} and b_i ; and \mathbf{y}_{i-1} is perpendicular to both \mathbf{x}_{i-1} and \mathbf{z}_{i-1} (Figure 10). Similarly, a local frame $F_i = \{Q_i; \mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i\}$ is attached to bond b_i .²⁹ The Protein Denavit-Hartenberg (PDH) parameters are defined to facilitate the geometric representation of one frame to another³⁸ as follows:

a_i is the distance from \mathbf{z}_{i-1} to \mathbf{z}_i measured along \mathbf{x}_{i-1} and physically represents the perpendicular distance between two successive bonds;

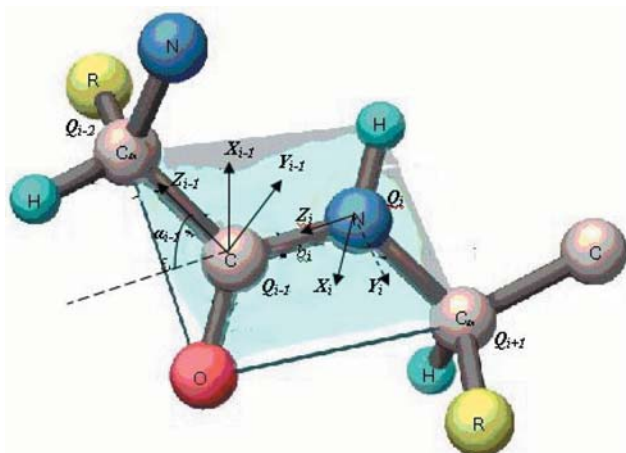


Figure 10. Frames F_{i-1} and F_i are attached to parent atom Q_{i-1} , and Q_i and bond rotation angle is α_{i-1} .

α_i is the angle between \mathbf{z}_{i-1} and \mathbf{z}_i measured about \mathbf{x}_{i-1} and represents the angle between two bonds;
 b_i is the distance from \mathbf{x}_{i-1} to \mathbf{x}_i measured along \mathbf{z}_i ; and represents the length of the bond joining Q_{i-1} to Q_i ;
 θ_i is the angle between \mathbf{x}_{i-1} and \mathbf{x}_i measured about \mathbf{z}_i and represents the torsional angle of the bond.

The coordinates of the origin and of the unit vectors of frame F_i with respect to frame F_{i-1} are represented using the following 4×4 homogeneous transformation matrix,³⁹ where $c\theta_i$ is the $\cos(\theta_i)$, $s\theta_i$ is the $\sin(\theta_i)$, $c\alpha_{i-1}$ is the $\cos(\alpha_{i-1})$ and $s\alpha_{i-1}$ is the $\sin(\alpha_{i-1})$.

$$R_i = \begin{bmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i c\alpha_{i-1} & c\theta_i - c\alpha_{i-1} & -s\alpha_{i-1} & b_i s\alpha_{i-1} \\ s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & -c\alpha_{i-1} & b_i c\alpha_{i-1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (15)$$

If an atom Q_i is connected to the root atom Q_0 by a sequence of bonds b_i, \dots, b_1 , then the coordinates of Q_i with respect to frame F_0 are:

$$(x' \ y' \ z' \ 1)^T = R_1 \dots R_i \cdot (0 \ 0 \ 0 \ 1)^T \quad (16)$$

In a similar way, the position of any atom on a side chain can be calculated with respect to the root frame F_0 .

A representative result of the direct kinematic module of the BioKineLab toolbox is shown in Figure 11. The results are obtained by running direct kinematic simulations on the native state of *loop36* as shown in Fig. 11a. The final state of the same protein obtained from NMR (Nuclear Magnetic Resonance) experiments is shown in Fig. 11b. Note that the random coil portion has turned into an α -helix after transformation, giving us a linear motion of the end-effector. The goal was to achieve the final *loop36* conformation using direct kinematic techniques. For this, the torsional angles corresponding to the final state were determined using the Accerlys Viewer ActiveX software. These angles, along with the initial state of *loop36*, were given as an input to the direct kinematic module of BioKineLab. Figure 11c shows the final structure generated by BioKineLab, which is a very good approximation of the actual output and shows the relevance of using molecular kinematics for predicting and generating protein conformations quickly.

5.2. Inverse Kinematics

The inverse kinematic problem calculates the VPL motor's torsional angles given in initial and final conformation and when all constant parameters of the chain are specified. A modified version of the Cyclic Coordinate Descent (CCD) method is used here. The CCD algorithm was initially developed for inverse kinematic applications in robotics.⁴⁰ For the inverse kinematics of protein chains, the torsional angles must be adjusted to move the C-terminal (end-effector) to a given desired position. The CCD method involves adjusting one torsional angle at a time to minimize the sum of

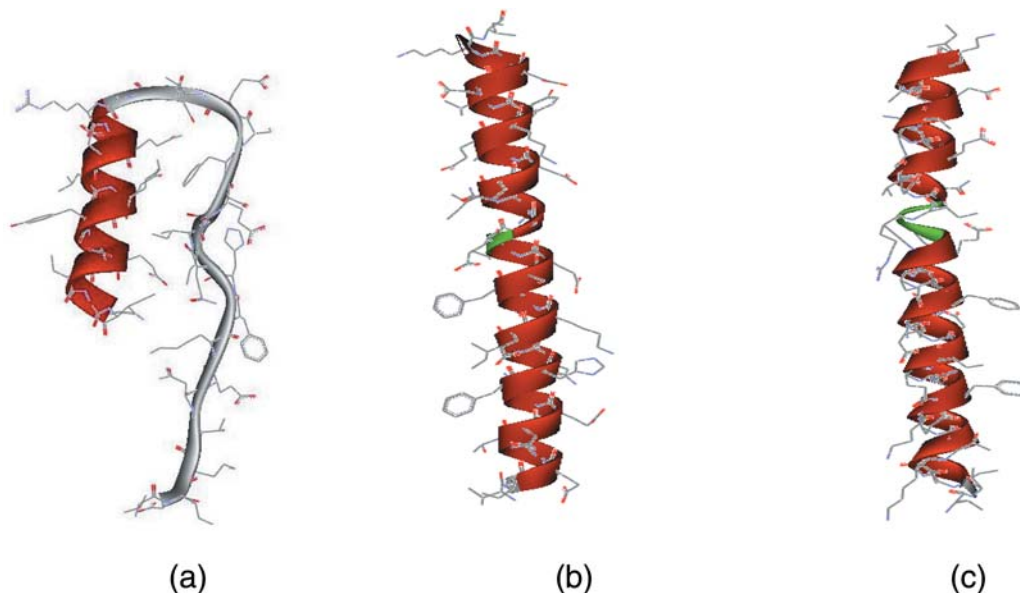


Figure 11. (a) *loop36* protein in the native state; (b) open state generated by NMR experiments which is similar to that generated by MD, computation time for MD is about 2 hours; (c) open state generated by molecular kinematics, computation time is less than 40 seconds.

the squared distances between the current and the desired end-effector positions. Hence, at each step in the CCD method, the original n -dimensional minimization problem is reduced to a simple one-dimensional problem (Figure 12). The algorithm proceeds iteratively through all of the adjustable torsional angles from the C-terminal to the base N-terminal.

At any given CCD step, the bond around which the rotation is being performed is called the pivot bond and its preceding atom is called the pivot atom. The torsional angle corresponding to the pivot bond is to be determined.

Let P_{id} be the position vector between the pivot atom and desired end-effector position; P'_{ih} be the position vector between the pivot atom and an intermediate stage reached by the end-effector during transition; and P_{ih} be the position

vector between the pivot atom and the current end-effector position. Then, δP denotes the error vector between the desired and the current end-effector positions and φ is the angle by which the pivot bond is to be rotated.

The errors between the current and the desired end-effector locations can be described by the following positive scalar function of φ that represents the scalar position error (distance) ΔP that is used as the objective function to be minimized:

$$\Delta P(\varphi) = \delta P(\varphi) \cdot \delta P(\varphi) = (P_{id} - P'_{ih}(\varphi)) \cdot (P_{id} - P'_{ih}(\varphi)) \quad (17)$$

where the symbol \cdot represents the scalar product of two vectors.

Vector P'_{ih} is obtained by rotating P_{ih} by an angle φ along the pivot bond:

$$P'_{ih}(\varphi) = [R(z_i, \varphi)]P_{ih} \quad (18)$$

where $[R(z_i, \varphi)]$ is a 3×3 rotation matrix and z_i is the unit vector along z_i . Expanding Equation (17) by using Equation (18), the following equation is obtained:

$$\Delta P(\varphi) = P_{id} \cdot P_{id} + P_{ih} \cdot P_{ih} - 2P_{id} \cdot ([R(z_i, \varphi)]P_{ih}) \quad (19)$$

Since the products $P_{id} \cdot P_{id}$ and $P_{ih} \cdot P_{ih}$ are positive constants, minimizing ΔP would be the same as maximizing:

$$g(\varphi) = P_{id} \cdot ([R(z_i, \varphi)]P_{ih}) \quad (20)$$

The right-hand side of Equation (20) can be expanded as:

$$[R(z_i, \varphi)]P_{ih} = z_i(P_{ih} \cdot z_i)(1 - \cos(\varphi)) + P_{ih} \cos(\varphi) + (z_i \times P_{ih}) \sin(\varphi) \quad (21)$$

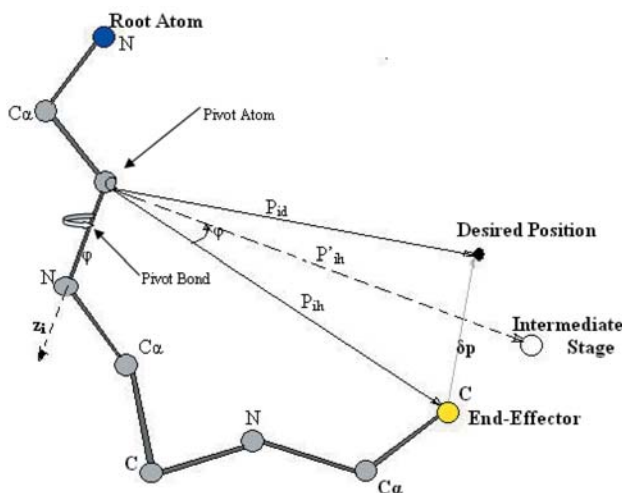


Figure 12. One step of the CCD method.

Substituting Equation (21) into (20), $g(\varphi)$ becomes:

$$\begin{aligned} g(\varphi) &= (\mathbf{P}_{id} \cdot \mathbf{z}_i)(\mathbf{P}_{ih} \cdot \mathbf{z}_i)(1 - \cos(\varphi)) + (\mathbf{P}_{id} \cdot \mathbf{P}_{ih})\cos(\varphi) \\ &+ \mathbf{P}_{id} \cdot (\mathbf{z}_i \times \mathbf{P}_{ih})\sin(\varphi) = k_1(1 - \cos(\varphi)) \\ &+ k_2\cos(\varphi) + k_3\sin(\varphi) \end{aligned} \quad (22)$$

Where k_1 , k_2 and k_3 are constant coefficients given by

$$\begin{aligned} k_1 &= (\mathbf{P}_{id} \cdot \mathbf{z}_i)(\mathbf{P}_{ih} \cdot \mathbf{z}_i), k_2 = \mathbf{P}_{id} \cdot \mathbf{P}_{ih}, \\ k_3 &= \mathbf{z}_i \cdot (\mathbf{P}_{ih} \times \mathbf{P}_{id}) \end{aligned} \quad (23)$$

If there are no boundary constraints imposed on φ , then $g(\varphi)$ is maximized if the following conditions are satisfied:

$$\frac{dg(\varphi)}{d\varphi} = (k_1 - k_2)\sin(\varphi) + k_3\cos(\varphi) = 0 \quad (24)$$

$$\frac{d^2g(\varphi)}{d\varphi^2} = (k_1 - k_2)\cos(\varphi) - k_3\sin(\varphi) < 0 \quad (25)$$

A unique value of φ can be determined from these equations. From Equation (24), angle θ can be calculated:

$$\varphi = \tan^{-1} \left[\frac{k_3}{k_2 - k_1} \right] \quad (26)$$

This determines a candidate value φ_c in the range $-\pi/2 < \varphi_c < \pi/2$. However, since function \tan is periodic, there is potentially one other value to consider: $\varphi_c + \pi$. Of these candidate values, those which pass the second derivative test in inequality (25) are the maximizing values for the objective function (18). If there is more than one value, the objective function is evaluated with each of them to determine which yields the true maximum.

Once φ has been uniquely determined, the corresponding pivot bond is rotated to reflect the change and the end-effector position is computed using direct kinematic techniques. Then, the CCD method proceeds to the next iteration. Figure 13 is the ball and stick model of a segment of the protein before and after the inverse kinematic simula-

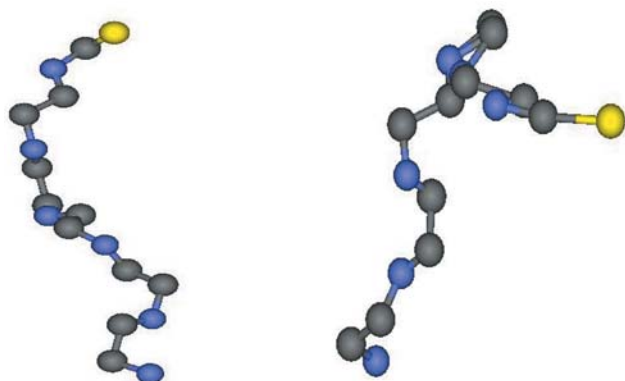


Figure 13. Initial conformation of the protein (left) and one of the solutions found by CCD simulations (right).

Table 1. Results of Inverse Kinematics with *loop36*.

	X(Å)	Y(Å)	Z(Å)
End-effector Initial Position	23.3	81.7	228.0
Desired Position	26.0	75.0	224.0
Position Reached	26.0	74.8	223.9
Error (Å)		0.1566	
Number of Iterations		31	

tion. Side chains are not shown for clarity. Table I shows the inverse kinematic results of the CCD simulations.

6. CONCLUSIONS

In this paper, the concept of the Viral Protein Linear nanomotor was presented. Dynamic and kinematic analysis methods were described to calculate important properties of the motor. Preliminary results from the application of these computational methods in the VPL motor were shown. The dynamic analysis, though slower, attempts a more realistic representation of the system. Each intermediate conformation is energy minimized to make sure that it is stable and feasible. Targeted molecular dynamics studies show that a large impetus is needed to make the protein undergo the desired conformational change, unless there are other environmental factors present, due to the presence of the remaining part of the protein not taken into account in this study. However, it assures the stability of the two end states of the system predicted by the kinematic analysis and experimental observations. The kinematic analysis can suggest the geometric paths that could be followed by the protein during the transition, while dynamics will narrow down the possibilities by pointing at the only energetically feasible paths. A combination of the two approaches—kinematics to give quick initial results, and dynamics to corroborate and select the feasible solutions—can prove to be an indispensable tool in bio-nanorobotics.

The future of bio-nanomachines is bright. We are at the dawn of a new era in which many disciplines will merge, including robotics, mechanical, chemical and biomedical engineering, chemistry, biology, physics and mathematics, so that fully functional systems will be developed. However, there remain many hurdles that must be overcome to reach this goal. Developing a complete database of different biomolecular machine components, the ability to interface or assemble different machine components, and the development of accurate models are some of the challenges to be faced in the near future. The problems involved in controlling and coordinating several bio-nanomachines will come next.

7. ACKNOWLEDGMENT

This work was supported by the national Science Foundation (DMI-0228103 and DMI-0303950). Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. S. M. Block, *Cell*, 93, 5–8 (1998).
2. M. J. Schnitzer and S. M. Block, *Nature*, 388, 386 (1997).
3. M. D. Wang, M. J. Schnitzer, H. Yin, R. Landick, J. Gelles, S. M. Block, *Science*, 282, 902 (1998).
4. K. Kitamura, M. Tokunaga, A. H. Iwane, T. Yanagida, *Nature*, 397, 129 (1999).
5. C. Shingyoji, H. Higuchi, M. Yoshimura, E. Katayama, and T. Yanagida, *Nature*, 393, 711 (1998).
6. C. D. Montemagno and G. D. Bachand, *Nanotechnology*, 10, 225 (1999).
7. H. C. Berg, *Nature*, 249, 77–79 (1974).
8. C. A. Schalley, K. Beizai, and F. K. Vogtle, *Accounts of Chemical Research*, 34, 465 (2001).
9. H. P. Erickson, *Science*, 276, 1090 (1997).
10. L. Mahadevan and P. Matsudaira, *Science*, 288, 95 (2000).
11. M. Knoblauch, G. A. Noll, T. Muller, D. Pruffer, I. Schneider-Huther, et al. *Nature Materials*, 2, 600 (2003).
12. N. C. Seeman, *Annual Review of Biophysics and Biomolecular Structure*, 27, 225 (1998).
13. I. Tobias, D. Swigon, B. D. Coleman, *Physical Review E*, 61(1), 747 (2000).
14. J. Yuqiu, C-B. Juang, D. Keller, C. Bustamante, D. Beach, T. Houseal, and E. Builes, *Nanotechnology*, 3, 16 (1992).
15. J. Hu, Y. Zhang, H. Gao, M. Li, U. Hartmann, *Nanoletters*, 2, No. 1, 55 (2002).
16. C. Mao, W. Sun, Z. Shen, and N. Seeman, *Nature*, 397, 144 (1999).
17. H. Yan, X. Zhang, Z. Shen, and N. Seeman, *Nature*, 415, 62 (2002).
18. B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, and J. L. Neumann, *Nature*, 406, 605 (2000).
19. M. Teodoro, G. N. Phillips Jr., and L. E. Kaviraki, *Proc. of the 2001 IEEE International Conference on Robotics and Automation (ICRA 2001)*, pp. 960, (2001).
20. P. W. Finn, D. Halperin, L. E. Kaviraki, J.-C. Latombe, R. Motwani, C. Shelton, and S. Venkat, *LNCS Series Applied Computational Geometry—Towards Geometric Engineering*, pp. 67.
21. K. Moon Kim, S. Gregory Chirikjian, and L. Robert Jernigan, *Journal of Molecular Graphics and Modeling*, 21, 151 (2002).
22. K. Moon Kim, L. Robert Jernigan, and S. Gregory Chirikjian, *Biophysical Journal*, 83 (2002).
23. P. W. Finn and L. E. Kaviraki, *Algorithmica*, 347 (1999).
24. S. M. LaValle, P. W. Finn, L. E. Kaviraki, and J. C. Latombe, *Journal of Computational Chemistry*, 21(9), 731 (2000).
25. E. J. Gardiner, P. Willett, and P. J. Artymiuk, *Journal of Chemical Information and Computer Sciences*, 40, 273 (2000).
26. G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *ACS Symposium Series (Rational Drug Design: Novel Methodology and Practical Applications)*, 719, 271 (1999).
27. M. Lipton and W. Still, *Journal of Computational Chemistry*, 9, pp. 343.
28. A. Smellie, S. D. Kahn, and S. L. Teig, *Journal of Chemical Information and Computer Science*, 35, 285 (1995).
29. M. Zhang, L. Kaviraki, *Journal of Chemical Information and Computer Science*, 42, 64 (2002).
30. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature*, 289, 377 (1981).
31. J. J. Skehel, P. M. Bayley, E. B. Brown, S. R. Martin, M. D. Waterfield, J. M. White, I. A. Wilson, D. C. Wiley, *PNAS USA*, 79, 968 (1982).
32. N. K. Sauter, M. D. Bednarski, B. A. Wurzburg, J. E. Hanson, G. M. Whitesides, J. J. Skehel, and D. C. Wiley, *Biochemistry*, 28, 8388 (1989).
33. C. M. Carr and P. Kim, *Cell*, 73, 823 (1993).
34. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *Journal of Computational Chemistry* 4, 187 (1983).
35. J. Schlitter, *Journal of Molecular Graphics*, 12, 84 (1994).
36. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acid Research*, 28, 235 (2000).
37. T. Lazaridis and M. Karplus, *Proteins: Structure, Function and Genetics*, 35, 133 (1999).
38. J. Denavit and S. Hartenberg, *ASME Journal of Applied Mechanics*, pp. 215 (1955).
39. J. McCarthy, *Geometric Design of Linkages*, (2000).
40. L. C. T. Wang, and C. C. Chen, *IEEE Transactions on Robotics and Automation*, 7(4), 489, (1991).

Received: 11 December 2003. Revised/Accepted: 19 December 2003