Viral Protein Linear (VPL) Nano-Actuators

A. Dubey[1], G. Sharma[1], C. Mavroidis[1,*], S. M. Tomassone[2], K. Nikitczuk[3], M.L. Yarmush[3]
[1]Department of Mechanical and Aerospace Engineering, Rutgers University, 98 Brett Road, NJ 08854
[2]Department Chemical and Biochemical Engineering, Rutgers University, 98 Brett Road, NJ 08854
[3]Department of Biomedical Engineering, Rutgers University, 98 Brett Road, NJ 08854
* Author for Correspondence, Tel 732-445-0732, Fax 732-445-3124, Email: mavro@jove.rutgers.edu

ABSTRACT

Computational studies are performed to predict and verify 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 enivronment. 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.

1 INTRODUCTION

The recent explosion of research in nano-technology, combined with important discoveries in molecular biology have 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.

In this project, we are studying the development of Viral Protein Linear (VPL) nano-motors and their integration as actuators in bio-nano-robotic systems. The project consists of three research phases: 1) Development of concepts for novel bio-nano-motors and devices; 2) Performance of computational studies to develop models and design procedures that will predict and optimize the performance of the proposed bio-nano motors 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 will 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 THE VPL MOTOR AND ITS TYPES

The basis of VPL motors is a pH dependant conformational change [1] observed in surface proteins (envelope glycoproteins) of certain retroviruses, such as the HIV1 virus. The change in the 3D structure and mechanical properties produces a linear-like motion. Envelope glycoproteins of various retroviruses play an important role in 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.

Peptides from different viruses can result in different VPL motors that can have different properties such as different 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 and the mode of infection. The envelope glycoproteins of these viruses can be divided into two subunits, the transmembrane (TM) and the surface subunit, which are a result of proteolytic cleavage of a common precursor protein. The surface subunit serves to recognize the cell to be infected when it comes in the vicinity of the virus with the help of receptors located on the cell surface. The TM subunits 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 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 the fusogenic state.

We have performed computational and experimental studies 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 [2]. BHA and pure HA were shown to undergo similar pH-dependant conformational changes which lead to membrane fusion [3]. 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. There is a specific region (sequence) in HA2, which tends to form a coiled coil. In the original X-ray structure of native HA, this region is simply a random loop. A 36 amino-acid residue region, upon activation, 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 precede it. This process relocates the hydrophobic fusion peptide (and the N-terminal of the peptide) by about 10 nm. In a sense of bio mimicking, 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.

3 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 that are based on the calculation of the energy released or absorbed during the transition from native to fusogenic state. We use the MD software called CHARMM (Chemistry at Harvard Molecular Mechanics) [4]. 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. With CHARMM, we can 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) [5]. TMD is used for approximate modeling of processes spanning long time-scales and relatively large displacements. Because the distance to be traveled by the N-terminal of the viral protein is relatively very 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' towards a final configuration from an initial configuration by applying constraints. The constraint is in the form of a bias in the force field.

In this project, the two known states are the native and the fusogenic states of the loop36. The structural data on these two states was obtained from Protein Data Bank (PDB) [6]. 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 initial and final states of the peptide.

By using TMD we have been able to prove that the final state of the VPL shown in Figure 2 is feasible, and it is the result of the action of the applied potentials. However, from the energy graph in Figure 3 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. The opening of the helical region followed by the adjustment of the remaining loop into an α-helical form is what requires more energy. This process occurs solely due to the pH drop in the natural setting of VPL since it is the rest of the large protein attached to the ends of this loop-36 that affects its behavior.

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/ $Å^2$ applied only to the non-hydrogen atoms. The transformation is achieved by using the artificial RMSD constraint such that a conformation close to the final state is approached successfully. The energy plot using solvation function EEF1 [7] is shown in Figure 3 and the simulation snapshots for the intial and final states are shown in Figures 1 and 2.

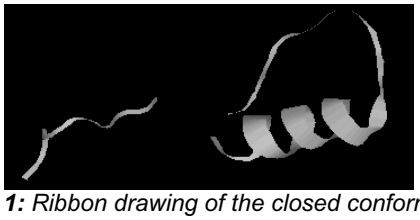

FIGURE 1: *Ribbon drawing of the closed conformation of 36-residue peptide as obtained from PDB entry 1HGF*

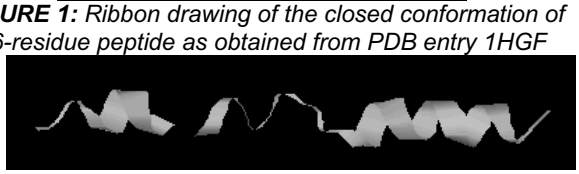

FIGURE 2: Ribbon drawing of the open conformation as obtained by TMD simulations. There is a noticeable increase in alpha-helical content and the peptide opens.

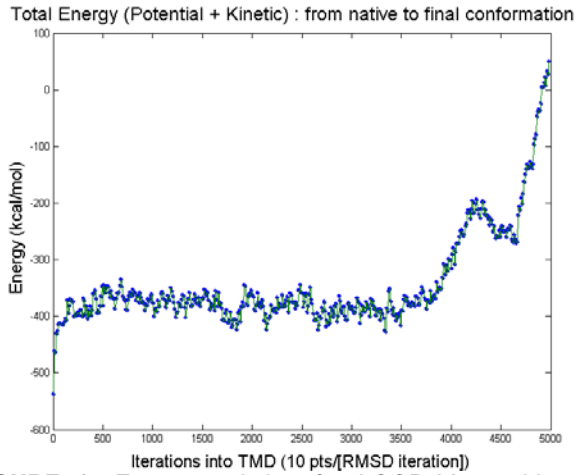

FIGURE 3: Energy variation for LOOP-36 peptide with solvation model EEF1.

4 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 towards the workpspace 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 acid residues. For kinematics 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-Cα-C-) sequence. To the Cα atom is also attached a side-chain (R) which is different for each residue. 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 is thus non- rotatable. There are however two bonds which are free to rotate. These are the N-Cα and Cα-C bonds and the rotation angles around them are known as *phi* (φ) and *psi* (ψ) respectively. These angles determine the 3-D structure of the protein. Therefore a protein can be considered to be a serial linkage with K+1 solid links connected by K revolute joints. 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 [8].

4.1 Direct Kinematics

The direct kinematics 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 4). 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 $\mathbf{z}_{i-1}$. Similarly, a local frame $F_i = \{Q_i; \mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i\}$ is attached to bond b_i [9]. The Protein Denavit-Hartenberg (PDH) parameters are defined to facilitate the geometric representation of one frame to another [10] as follows: a_i is the distance from $\mathbf{z}_{i-1}$ to $\mathbf{z}_i$ measured along $\mathbf{x}_{i-1}$; α_i is the angle between $\mathbf{z}_{i-1}$ and $\mathbf{z}_i$ measured about $\mathbf{x}_{i-1}$; b_i is the distance from $\mathbf{x}_{i-1}$ to $\mathbf{x}_i$ measured along $\mathbf{z}_i$; and θ_i is the angle between $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$ measured about $\mathbf{z}_i$. 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 4x4 homogeneous transformation matrix R_i [11] where $c\theta_i$ is $\cos(\theta_i)$, $s\theta_i$ is $\sin(\theta_i)$, and so on, l_i is the length of the bond b_i, θ_i is the torsional angle of b_i, and α_{i-1} is the bond angle between b_{i-1} and b_i:

$$R_i = \begin{bmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i c\alpha_{i-1} & c\theta_i - c\alpha_{i-1} & -s\alpha_{i-1} & l_i s\alpha_{i-1} \\ s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & -c\alpha_{i-1} & l_i c\alpha_{i-1} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

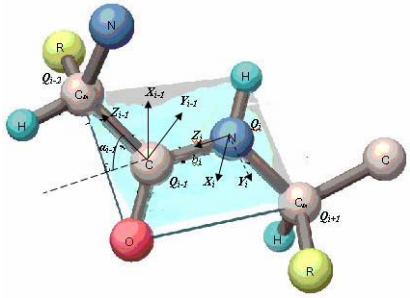

FIGURE 4: Frames F_{i-1} and F_i are attached to parent atom Q_{i-1} and Q_i and bond rotation angle is α_{i-1}.

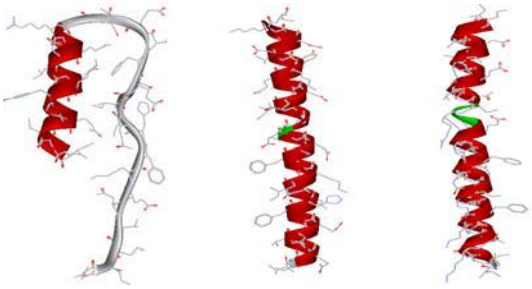

FIGURE 5: (a) loop36 protein in the native state, (b) open state generated by NMR experiments which is similar to that generated by MD, computation time for MD is about 2 hours, (c) open state generated by molecular kinematics, computation time is less than 40 seconds.

A representative result of the direct kinematics module of the BioKineLab Toolbox is shown in Figure 5**.** The results are obtained by running direct kinematics simulations on the native state of loop36 as shown in Fig. 5a. The final state of the same protein obtained from PDB is shown in Fig. 5b. 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 kinematics

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 kinematics module of BioKineLab. Figure 5c shows the final structure generated by BioKineLab which gives a very good approximation of the actual output and clearly shows the relevance of using molecular kinematics for predicting and generating protein conformations.

4.2 Inverse Kinematics

The inverse kinematics problem calculates the VPL motor's torsional angles given an 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. The CCD algorithm was initially developed for the inverse kinematics applications in robotics [12]. 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 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. 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. Figure 7 is the ball and stick model of a segment of the protein before and after the inverse kinematics simulation. Side chains are not shown for clarity.

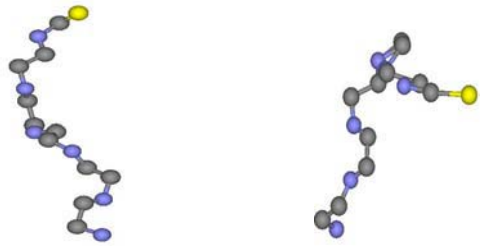

FIGURE 7: Initial conformation of the protein (left) and one of the solutions found by CCD simulations (right).

5 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. It however assures of the stability of the two end states of the system predicted by the kinematic analysis and experimental observations. Kinematics 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-nano-robotics.

6 ACKNOWLEDGEMENTS

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.

7 REFERENCES

[1] Carr C. M. and Kim P., 1993, "A Spring-Loaded Mechanism for the Conformational Change of Influenza Hemagglutinin," *Cell*, **73**, 823-832.

[2] Wilson I.A., Skehel J.J., Wiley D.C., 1981, "Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution", *Nature*, **289**, 377-373.

[3] Skehel J.J., Bayley P.M., Brown E.B., Martin S.R., Waterfield M.D., White J.M., Wilson I.A., Wiley D.C., 1982, "Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion", *PNAS USA*, **79**, 968-972.

[4] Brooks R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S., and Karplus M., 1983, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations", *Journal of Computational Chemistry* **4**, 187-217.

[5] Schlitter J., 1994, "Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions", *Journal of Molecular Graphics*, **12**,84-89.

[6] Berman H.M, Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., and Bourne P.E., 2000, "The Protein Data Bank," *Nucleic Acid Research,* **28**, 235-242.

[7] Lazaridis T., and Karplus M., 1999, "Effective Energy Function for Proteins in Solution," *Proteins: Structure, Function and Genetics*, **35**,133-152.

[8] Gardiner E. J., Willett P. and Artymiuk P. J., "Graph-theoretic techniques for macromolecular docking", *Journal of Chemical Information and Computer Sciences*, 40, 273-279, 2000.

[9] Zhang M., Kavaraki L., 2002, "A new method for fast and accurate derivation of molecular conformations", *Journal of Chemical Information and Computer Science*, 42, 64-70.

[10] Denavit J., and Hartenberg S., 1955, "A Kinematic Notation for Lower Pair Mechanisms Based on Matrices", *ASME Journal of Applied Mechanics*, pp 215-221.

[11] McCarthy J., "Geometric Design of Linkages", *Springer Publications NY*, 2000.

[12] Wang L.C.T, Chen C.C, "A Combined Optimization Method for Solving the Inverse Kinematics Problem of Mechanical Manipulators", *IEEE Transactions on Robotics and Automation*, 7(4):489-499, 1991.