Journal of Mechanical Design

Introduction

Analysis and Design of Protein Based Nanodevices: Challenges and Opportunities in Mechanical Design

Gregory S. Chirikjian
Professor and Chair
e-mail: gregc@jhu.edu
Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218

Kazem Kazerounian
Professor and Fellow of ASME
e-mail: kazem@engr.uconn.edu
Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269-3139

Constantinos Mavroidis
Associate Professor
e-mail: mavro@coe.neu.edu
Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115

Overview

Nanomachines are devices that are in the size range of billionths of meters (10^{-9} m) and therefore are built necessarily from individual atoms. These devices will have intrinsic mobilities that result in their geometry change and hence enable them to perform specific functions. Futuristic scholars and researchers believe that nanodevices will one day be used as "assemblers" in the construction of new materials and objects from inside out [1]; They will be able to "self replicate;" They will be able to enter biological cells to cure disease; They will be able to facilitate space travel; They will be used to clean up the environment; They will be the building blocks of the electronic circuitry and computers [2]. While these claims may prompt profound philosophical and scientific debates for many years to come, they offer humanity with the potential to eliminate poverty, pollution and disease.

In the foreseeable future, nanodevices and molecular machines are expected to lead to breakthroughs in technology and life sciences. Impacted industries include Semiconductors (molecular electronics, nanophotonics, memory), Microsystems (micro electro-mechanical systems—MEMS, communications, optical, fluidics, rf), energy (fuel cells, biomimetics, membranes, carbon systems), materials (compounds, powders, polymers, nanostructures), biotechnology (delivery, lab-on-chip, proteomics, genomics) and medical (drug design, molecular medicine).

Contributed by the Mechanisms and Design Committee for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received: February 28, 2005. revised: March 2, 2005. Associate Editor: J. M. McCarthy.

Given the tremendous potential impact of nanodevices, rational design approaches to these systems based on first principles are conspicuously absent. In part, this is due to an incomplete understanding of the forces that govern the behavior of systems at the nanoscale and in part it is due to the huge numbers of degrees of freedom in these systems and the corresponding computational burden associated with their simulation.

The practice of building nanomachines using large scale equipment faces insurmountable challenges. Given that proteins are the nanodevices of choice for evolution, it is increasingly believed by scientists that the practical and viable approach to the design and fabrication of artificial nanodevices and machines is to use polypeptide chains, protein building blocks found in nature [3–5]. Mother Nature has her own set of molecular machines that have been working for millions of years, and have been optimized for performance and design over the ages. As our knowledge and understanding of these numerous machines continues to increase, we now see a possibility of using these natural machines, or creating synthetic ones from scratch, using nature's components. The main goal in the field of molecular machines is to use various biological elements—whose function at the cellular level creates motion, force or a signal—as machine components. These components perform their preprogrammed biological function in response to the specific physiochemical stimuli but in an artificial setting. In this way proteins could act as motors, mechanical joints, transmission elements or sensors. If all these different components were assembled together in the proper proportion and orientation they would form nanodevices with multiple degrees of freedom, able to apply forces and manipulate objects in the nanoscale world. Developing nanomachines and devices out of protein elements requires the merging of two different research approaches as it is shown in Fig. 1: the inspiration by nature and biology (biomimetics) and the inspiration by large scale machines and the traditional mechanisms and machine theory ("machine nanomimetics").

Proteins are undoubtedly the most important components of biological machines, performing all types of functions in an organism. Indeed, the word protein derives from the Greek "proteios," which literally means "of prime importance." There are three main categories of proteins [6,7]—fibrous, membrane, and globular—and within each of these categories, many types of proteins are classified by their various biological functions: for example, enzymes (which are responsible for catalyzing tens of thousands of chemical reactions in living cells); structural or support proteins (such as bones, muscles and tissue); nutrient and gas transport proteins; proteins of the immune system; and proteins that perform mechanical work.

Proteins are large molecules synthesized from 20 different types of amino acids. Each of these amino acid building blocks (monomers) is referred to as a residue. Tens to hundreds of these residues (amino acid monomers) connect together in a serial manner to create a long chain, known as a polypeptide chain. However, from a kinematics point of view, these polypeptide molecules can be considered to be a chain of miniature rigid bodies connected by revolute (hinge) joints [8]. More specifically, we can

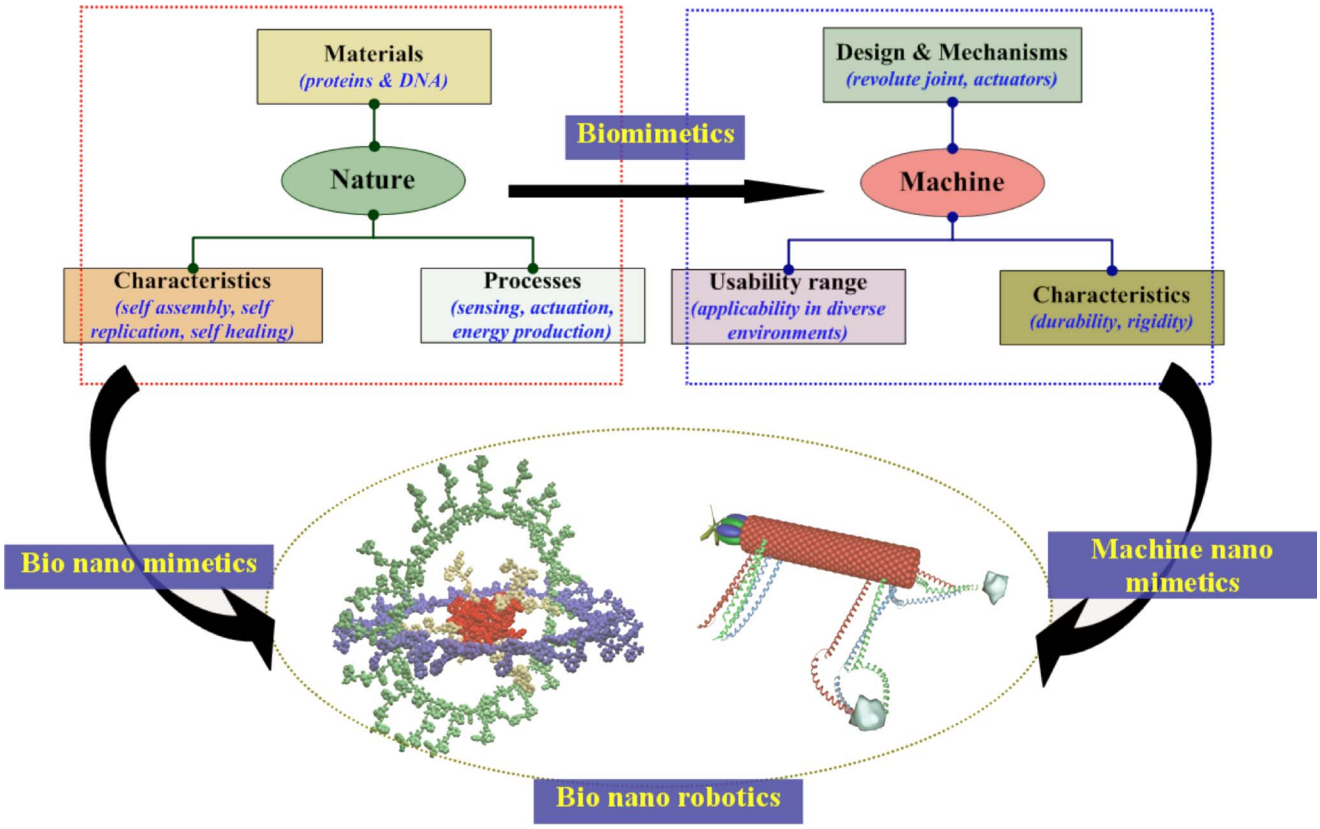

Fig. 1 Bionanorobotics: A field inspired by nature and mechanism and machine theory

denote a protein in its denatured state as a serial linkage with $N+1$ solid links connected by N revolute joint (values for N could be as great as several hundred). The side chains are shorter serial linkages with zero to a few revolute joints, connected to the main links of the backbone. The values of the revolute joints identify the three-dimensional structure (fold) of the protein. The folding occurs under the effect of nuclear forces (among protein atoms as well as between protein atoms and the solvent's atoms [9]. Changes in the environmental parameters, causes corresponding changes in the 3-D geometry of the protein. The final conformation is a relatively stable configuration for which the total potential energy is globally minimized (or in other words the system is in stable static equilibrium).

Being able to accurately predict the three-dimensional structure of a protein based on the known sequences of amino acids in its chain is key to fully understanding a protein's biological functions and thus to manipulating or controlling these functions as a part of disease treatments. Despite more than 50 years of intense work on this challenge, the puzzle of predicting protein folding remains largely unsolved. Indeed, experts refer to protein prediction as the "grand challenge" or the most challenging computational and scientific problem of the century. While the computation of potential energy as the cost function in optimization-based *ab initio* methods to predict the final conformation is rather straightforward, the computational complexity is mind boggling. To put things in perspective, let us examine the computational requirements of a rough exhaustive search for a protein molecule that contains 100 residues (200 revolute joints). If one takes samples on the joint angles at every 36 degrees (10 samples per joint, which is so rough that the results are meaningless), and if energy calculations for each state of the protein take 10–20 seconds (using many parallel supercomputers), then the total computational time required is 3.16×10^{172} years.

Biological polymers are extremely complex. Our ability to predict the structural and dynamic properties of proteins is far from complete. Consequently, the challenges of de novo protein-based molecular design are exceedingly difficult. A viable design strategy for nanomachines requires several enabling core technologies. These core technologies must include quantitative models of forces that are dominant in the nanodimensions, and the motions that these forces induce in the polypeptide chains. Methodologies are also needed for effective mobility (flexibility) analysis and synthesis of the polypeptide chain molecules. Given the vast body of knowledge in theoretical, applied, and analytical kinematics and robotics, the kinematics community is uniquely positioned to make significant contributions to the prediction of protein folding, protein docking, protein engineering. Kinematics (in particular, robot kinematics) can significantly contribute to our understanding of biological systems and their functions at the microscopic level and to the engineering of nanodevices, new materials, diagnostic tools, as well as new treatments and drugs for a variety of diseases.

Several researchers have drawn analogies between protein folding methodologies and robotics methodologies [10–13]. However, most of the robotics-based works actually implemented have been in the fields of motion and path planning and have been undertaken by researchers in the computer science, optimization, and artificial intelligence communities [14,15]. One exception to this is Manocha et al. [16] who used Ragavan and Roth's solution [17] to the general 6-R serial manipulator inverse kinematics problem to facilitate solving the receptor–ligand docking problem for short side chains. In addition, other researchers have sporadically applied some kinematics notations and procedures to problems relating to molecular structures.

In the sections that follow, many of the important kinematics and mechanical issues associated with modeling proteins machines are reviewed. We begin with a discussion of animating protein machines.

Animating Protein Machines

While the protein folding problem represents a "grand challenge" due to uncertainty about the forces that govern folding and the large number of degrees of freedom in polypeptide chains, there are other problems associated with the analysis and design of protein machines that are much more tractable in the near term. Whereas *ab initio* folding algorithms will be critical for the de novo design of proteins for use in nanoscale machines, even in cases where the tertiary structure of an existing protein is already known, the connection between structure and function is not always easy to establish. Two problems of interest in the kinematic analysis of fully folded proteins are: as follows (1) determining the normal modes associated with small motions about the equilibrium conformations of a given protein molecule; and (2) determining pathways for transitions between two known conformations of the same protein molecule. Both goals serve as a way to better understand how proteins function *after* they are folded. This is a critical (though much more tractable) problem than the folding problem itself.

Recently, work based on theoretical and analytical kinematics has started to appear among the vast amount of literature on the simulation of protein motions. In conformational transitions, most of the degrees of freedom are "frozen" and proteins are then viewed as collections of rigid substructures (e.g., alpha helices or large domains) that are connected with kinematic joints such as hinges. This makes the animations easier to simulate during conformational changes between tertiary states. But it is also justified on physical grounds, because it is believed that proteins do not "melt" and reform as they undergo transitions between states as they perform their normal functions.

Chirikjian and co-workers [14] have developed a method for morphing one conformation of a protein into another. This is a more limited problem than protein folding, but it is directly relevant to modeling the behavior of protein machines. Their method, called the elastic network interpolation (ENI), treats a protein as a coarse-grained collection of point masses and/or rigid bodies that are connected with Hookean-spring constraints. This provides a way to "morph" between two different conformations of a protein by using the elastic energy of deformation to serve as a cost function that preserves features throughout an anharmonic motion. It also allows one to efficiently compute the lowest frequency normal modes of a protein without having to be concerned with the full chemical details of the structure.

Statistical Kinematics

As stated earlier, if one were to try to enumerate the states of an unfolded polypeptide chain by sampling each degree of freedom, the number of sampled states would grow exponentially in the length of the chain. Therefore, having a way to generate statistical information about the huge ensemble of polypeptide conformations while simultaneously circumventing the computational cost could represent a breakthrough in the way the unfolded state is characterized. But what statistical information is important to know? In fact, it is desirable to know how probable it is that any amino acid residue in the unfolded polypeptide will come into proximity with any other residue in the same chain as all possible conformations are visited.

Chirikjian and co-workers developed methods for generating workspace probability densities for macroscale "binary manipulators" that can attain huge numbers of discrete states [15, and references therein]. The workspace density is the probability density that the end effector of a discrete-state chain will reach any particular position and orientation in the workspace.

Chirikjian's approach involves partitioning kinematic chains with discrete states into smaller segments, and generating probability densities for each small segment by brute-force enumeration. Probability densities for segments of any length up to the full length of the chain are then obtained by successively convolving the probabilities of adjacent segments [15]. This convolution is not the usual convolution known in engineering, but rather the convolution of functions on the group of rigid-body motions.

Since polymers and unfolded polypeptides can also be viewed as discrete-state chains (where the discrete states are defined by sampling at minimal energy states), the same methodology can be (and has been) applied.

In practice, the associated convolutions are not performed directly, but are computed by using harmonic analysis on motion groups in analogy with the way the FFT is used to compute traditional convolutions. This has been shown to be a useful tool in the statistical mechanics of polymer molecules [15].

In the remainder of this editorial, we will discuss some of the challenges in protein analysis that can benefit from the field of kinematics.

Kinematics Notation

The commonly used convention for describing the geometry of a protein is to represent the Cartesian coordinates of its atoms and their bonds (connectivity). Proteins in the publicly accessible Protein Data Bank are described in this way. To a large extent, kinematics methods of analysis and geometric notation are closely intertwined. Several kinematics notations could potentially be used to describe protein analysis, including Denavit–Hartenberg notation [18], vector notation and analysis [19], tensor methods [20,21], screw coordinates [22], dual numbers [23], quaternion operators [24], constant distance equation method [21], Zero Position Method [25], and train components method [26]. An in-depth study of these notations and their applicability to protein analysis would shed some light on factors such as:

- computational efficiency,
- geometric stability—small changes in the geometry of a protein should not lead to large changes in the kinematic parameters [27].

Hydrogen Bonds and Serial Versus Parallel Robotic Manipulators

In the process of folding, as various portions of the long snake-type chain get closer together, a hydrogen bond (a powerful bond with a relatively fixed length) is created between the hydrogen of an amino group and the oxygen of a carboxyl group. This bond acts as a solid connection that creates a closed loop within an otherwise serial linkage. Some of these developed loops lack internal mobility, while others have a mobility value of one or higher. These internal mobilities act in a similar manner to parallel manipulators and are decisive factors in the continuation of the folding, and, in turn, the function of the protein.

Direct Kinematics

Using the torsion angles as generalized coordinates would reduce the computational complexity of the analysis tremendously. In order to do this however, we would need kinematics-based robot methodologies able to produce the position of all atoms (which are actually body points of various links of the chain) working from the values of the torsion angles. However, this process presents several challenges:

- accommodating and accounting for structural and linkage changes caused by hydrogen bonds,
- checking for and avoiding collisions among various links (atoms), and
- designing an algorithm that will maximize the advantages of parallel computing (by decoupling the direct kinematics of various sections in the long chain).

Inverse Kinematics and Drug Design

The inverse kinematics problem in the prediction of protein folding appears in several forms. The first type is the calculation of the torsion angles when all the coordinates of all the atoms are specified (such as in a PDB file). The observation inaccuracies as well as bond length and bond angle variations throughout the chain make this seemingly straightforward and simple task an important challenge.

The second type comes into play when simulating the behavior of a long snake-type chain under the influence of field forces (atomic attractions and repulsions), a process that is, in principal, similar to that of redundancy resolution of robotic manipulators under energy-favored criteria [28,29]. The primary difference involves the end link trajectory, which is not specified and results in an optimization process that is not constrained by the velocities and accelerations of the end link.

The third type of inverse kinematics problem that arises in protein structure analysis involves protein docking [16,30] (ligand–receptor docking), where the geometric characteristics of a docking location on a receptor (usually in the form of a cavity on a protein molecule) are known and defined by positional information about certain atoms in the mating protein chain (ligand). The challenge is finding a protein (either by searching in a databank or by designing a new protein) that fits the specified position.

Kinematics Condition Analysis of the Chain

Analogous to the study of robotic manipulators, there are numerous issues that will benefit from advanced kinematics analysis of the peptide chains. These include:

- Analysis and determination of the workspace
- Analysis of the mobility and dexterity
- Compliance analysis

Dynamic Analysis and Simulation of the Chain

Methods for dynamic analysis in both robotic manipulators and closed- and open-loop linkages have already been developed and are now mature. There are few differences between the dynamic simulation of a protein chain and the simulation of a hyper-redundant robotic manipulator, although time steps in the integration of the equations of motion must, of course, be much smaller in nanomechanical systems than in macromechanical systems. However, other computational issues can arise when working with molecular dynamics simulations and must be addressed [31,32].

In this special issue, we have included four papers representing significant contributions from the field of mechanical analysis and design in understanding, modeling and design of protein based nanodevices. The first two papers introduce a comprehensive methodology for notation, kinematics, dynamics and motion analysis of the peptide chains. The third paper entails the innovative design of a new linear nanomotors based on protein molecules, and the study of its geometric workspace. The fourth paper offers a two level optimization procedure for the design of proteins.

References

[1] Merkle, R. C., 1996, "Design Considerations For an Assembler, "Nanotechnology, **7**, pp. 210–215.

[2] Merkle, R. C., 2000, "Molecular Building Blocks and Development Strategies For Molecular Nanotechnology," Nanotechnology, **11**, pp. 89–99.

[3] Mavroidis, C., Dubey, A., and Yarmush, M., 2004, "Molecular Machines," Annu. Rev. Biomed. Eng., **6**, pp. 363–395.

[4] Mavroidis, C., and Dubey, A., 2003, "From Pulses to Motors," Nat. Mater., **2**, pp. 573–574.

[5] Ummat, A., Dubey, A., Sharma, G., and Mavroidis, C., 2004, "Bio-Nano-Robotics: State of the Art and Future Challenges," invited chapter in *The Biomedical Engineering Handbook*, 3rd Ed., edited by M. L. Yarmush, CRC Press, City.

[6] Branden, C., and Surname, T. J., 1991, *Introduction to Protein Structure*, 2nd ed. Garland Publishing, City.

[7] Cantor, C. R., and Schimmel, P. R., 1997, *Biophysical Chemistry: The Conformation of Biological Macromolecules*, W.H. Freeman & Co., San Fransisco.

[8] Kazerounian,K., 2002, "Is Design of New Drugs a Challenge for Kinematics?," *Proceedings of the 8th International Conference on Advance Robot Kinematics (ARK)*.

[9] Burkert, U., and Allinger, N., 1982, *Molecular Mechanics*, Oxford University Press, Oxford.

[10] Kavraki, L., 1996, "Geometry and the Discovery of New Ligands," Publisher, City, pp. 435–448.

[11] Parsons, D., and Canny, J. F., 1994, "Geometric Problems in Molecular Biology and Robotics," presented at the *2nd International Conference on Intelligent Systems for Molecular Biology*, Palo Alto, CA.

[12] Emiris, I. Z., and Mourrain, B., 1999, "Computer Algebra Methods for Studying and Computing Molecular Conformations," Algorithmica, **25**, pp. 372–402.

[13] Zhang, M., and Kavraki, L., 2002, "A New Method for Fast and Accurate Derivation of Molecular Conformations," J. Chem. Inf. Comput. Sci., **42**, pp. 64–70.

[14] Kim, M. K., Jernigan, R. L., and Chirikjian, G. S., 2002, "Efficient generation of Feasible Pathways For Protein Conformational Transitions," Biophys. J., **83**, pp. 1620–1630.

[15] Chirikjian, G. S., and Kyatkin, A. B., 2001, *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, Boca Raton FL.

[16] Manocha, D., Zhu, Y., and Wright, W., 1995, "Conformational Analysis of Molecular Chains Using Nano-Kinematics," Computer Applications of Biological Sciences, **11**, pp. 71–86.

[17] Raghavan, M., and Roth, B., 1995, "Solving Polynomial Systems for the Kinematic Analysis and Synthesis of Mechanisms and Robot Manipulators," ASME J. Mech. Des., **117**, pp. 298–306.

[18] Denavit, J., and Hartenberg, R. S., 1955, "A Kinematic Notation For Lower-Pair Mechanisms Based on Matrices," ASME J. Appl. Mech., pp. 215–221.

[19] Chase, M., Year, "Vector Analysis of Linkages," ASME J. Eng. Ind., **85**, pp. 300–308.

[20] Kislitsin, A. P., 1954, "Tensor Methods in the Theory of Spatial Mechanisms," Trudi Seminar Po Teroii Mashin I Mekhanizmov, Akedemia Nauk, USSR, **14**, pp. 51–57.

[21] Osman, M. O., and Segaev, D. N., 1972, "Kinematic Analysis of Spatial Mechanisms by Means of Constant Distance Equations," Trans. Can. Soc. Mech. Eng., **1**, pp. 129–134.

[22] Yuan, M. S., and Freudenstien, F., 1971, "Kinematic Analysis of Spatial Mechanisms by Means of Screw Coordinates," Transactions of the ASME Journal of Engineering for Industry, **93**, pp. 61–73.

[23] Yuan, M. S., and Freudenstien, F., 1964, "Application of Dual-Number Quaternions Algebra to the Analysis of Spatial Mechanisms," Trans. ASME, J. Appl. Mech., **86**, pp. 300–308.

[24] Sandor, G. N., "Principles of General Quaternion-Operator Method of Spatial Kinematic Synthesis," presented at ASME Paper No. 68-APM, 1968.

[25] Gupta, K. C., 1986, "Kinematic Analysis of Manipulators Using the Zero Reference Position Description," Int. J. Robot. Res., **5**, pp. 5–13.

[26] Osman, M. O., Bahgat, B. M., and Dukkipati, R. V., 1981, "Kinematic Analysis of Spatial Mechanisms Using Train Components," ASME J. Mech. Des., **103**, pp. 823–830.

[27] Kazerounian, K., and Qian, Z., 1989, "Kinematics Calibration of Robotic Manipulators," ASME J. Mech., Transm., Autom. Des., **111**, pp. 32–39.

[28] Kazerounian, K., and Nedungadi, A., 1988, "Redundancy Resolution of Serial Manipulators Based on Robot Dynamics," Mech. Mach. Theory, **22**, pp. 631–654.

[29] Kazerounian, K., and Wang, Z., 1988, "Global Versus Local Optimization in Redundancy Resolution of Robotic Manipulat," Int. J. Robot. Res., **7**, pp. 3–12.

[30] Canutescu, A. A., and Dunbrack, R. L. J., 2003, "Cyclic Coordinate Descent: A Robotics Algorithm for Protein Loop Closure," Protein Sci., **12**, pp. 963–972.

[31] Rappé, C. J., 1991, "Molecular Mechanics Across Chemistry," *University Science Books*, Publisher, City, pp. 101–103.

[32] Dubey, A., Sharma, G., Mavroidis, C., Tomassone, S. M., Nikitczuk, K. P., and Yarmush, M. L., 2004, "Computational Studies of Viral Protein Nano-Actuators," J. Comput. Theor. Nanosci., **1**, pp. 18–28.