Biocomplexity Faculty Search Committee,
c/o Prof. Rob de Ruyter van Steveninck,
Department of Physics,
Indiana University,
Swain Hall West 117,
Bloomington IN,
47405-7105

Dear Committee Members,

Enclosed is my application for the position of Assistant Professor in the area of mathematical biology. My research focuses on the applications of differential geometry to biological structure and dynamics. Over the past year, I have developed these methods in collaboration with Alain Goriely, Professor of Mathematics, here at the University of Arizona. The methods utilize an untapped approach to representation of biological structure, and enable a wide variety of novel capabilities complementary to methods for molecular simulation such as molecular dynamics. As described in my statement of research interests in more detail, I have focused on three primary areas: fold space exploration, protein design, and structure determination methodology, each of which apply the basic capabilities we have developed to different ends.

More recently I have made efforts to build a bridge between the geometric description of structure, and physical theory. In particular, the geometric parameters describing structures, or distributions of structures, are a natural set of variables over which to calculate ensemble averages of geometric properties. The fusion of statistical mechanics with differential geometry will enable construction of new classes of models which make quantitative predictions of any experimental observable involving distances or areas, eg. the average distance between two amino acids in a folding protein as measured by FRET experiments, or the distribution of pairwise atomic distances as measured by SAXS. Currently I am focusing on equilibrium problems but am very interested in developing applications to time-dependent processes.

These methods offer new ways to interpret structural data from synchrotron radiation using experimental techniques such as fiber diffraction and small-angle scattering. In the longer term, by incorporating geometry into physical models, they may also provide a route to extraction of kinetic and energetic information from such data.

This work uses a combination of analytical and computational approaches to construct and manipulate models of biomolecules. In the process of development of these methods, a variety of interesting mathematical questions have been raised which are external to the biological applications, and so from the outset this project has proved to be a particularly engaging interdisciplinary undertaking. Because of the interest to both the mathematical and biological communities, Alain Goriely and I have applied for funding for this project through a combined NSF-NIGMS Biology/Math interface award.

I anticipate the project will continue to bring together workers from different fields and with different backgrounds as it grows in scope. As an example, one goal of this work is to create tools for protein design. In collaboration with John Osterhout, a protein chemist with experience in *de novo* protein design, Alain Goriely and I were

invited by DARPA to submit a proposal to test these methods by design of antiviral proteins.

My ability to interact with people from different scientific backgrounds is due to my unusual breadth of training. I have bachelor's degrees in mathematics and physics, and a doctorate in chemistry. My graduate work was in experimental protein crystallography, where I determined the structure of the bacterial F1 ATPase rotary motor. During this time I learned protein biochemistry, molecular biology, and X-ray crystallography. I moved to Tucson, Arizona when my wife, Megan McEvoy, accepted a position as Assistant Professor of Biochemistry. I began my work here as a postdoc with Elizabeth Vierling in this department on biophysical studies of chaperone-substrate interactions. However my interests led me to conclude that work so that I would be able to develop my own research program, the results of which effort are described in the statement of research interests.

Consideration of the employment possibilities for my wife would be appreciated, should my work be of interest to your department.

I have considerable teaching experience in a variety of fields. As an undergraduate I was involved as a reader in various mathematics and physics classes and also worked as a teaching assistant for the physics laboratory. As a graduate student I was a teaching assistant for the general chemistry laboratory and gave occasional lectures on protein crystallography for the graduate biophysical chemistry class. Here at the University of Arizona I have led discussion sections for Biochemistry 462 (the majors' biochemistry course) and this term gave the lectures for the first half of Biochemistry 460 (the nonmajors' biochemistry course). Having participated in teaching at different levels in a variety of subject areas I have experience with a range of curricula.

During this time I have also been guiding student projects. Katie Maish, a first-year Biochemistry graduate student with a strong mathematical background, pursued a rotation project on the differential geometric representations of beta-sheet proteins, and Pick-wei Lau, an undergraduate biochemistry major and computer science minor, has recently chosen to work on applications to protein design. Note that since I am technically an adviserless postdoc, there is no formal mechanism for students to participate in my projects and I have not tried to recruit them. Their enthusiasm suggests to me that my work will be able to bring together quantitatively inclined students from different backgrounds.

In my present position I have independently developed a unique and promising interdisciplinary research program and applied for funding to pursue it further. I have also participated in both the laboratory and classroom training of undergraduate and graduate students. With this experience, I believe I have demonstrated the ability to carry out the responsibilities of a tenure-track faculty position. I look forward to the opportunity to do so.

Sincerely,

Andrew Hausrath
Home Phone: (520)-529-4102
Office Phone: (520)-626-1925
Email: hausrath@email.arizona.edu

Statement of Research Interests

Summary: My research lies at the interface of mathematics and structural biology. I have applied the formalism of differential geometry, which enables construction of abstract shapes and forms, to the description of protein structure. The central idea is to represent a protein fold as a continuous curve which can be smoothly and continuously varied as a way to create and manipulate protein models. Novel capabilities are conferred by this continuous representation. The ability to smoothly deform the curve representing a protein fold makes it possible to systematically explore relationships between protein folds and may enable the identification of favorable regions of fold space not utilized in nature, and this represents the first area on which I will focus. A second area is in development of methodology to create novel proteins. Complete atomic models can be constructed from curves, bringing the powerful analytical methods of the continuum to bear on the problem of protein design. Protein folds can be represented with small numbers of geometric parameters, and a third area of inquiry will be to devise methods for extraction of these geometric parameters from experimental data for the purpose of structure determination, both for proteins and for other biological structures. At present my efforts are focused on development of the theoretical and computational tools for this work. However in the longer term I intend to develop an experimental program which tests and improves these theoretical tools and applies them to the investigation of specific biological problems. I am also particularly interested in pursuing collaborations with researchers whose projects may provide new applications for this formalism.

The remainder of this statement briefly describes the underlying method and then illustrates some examples of how it is used in the three areas described above.

Method: The *path* a protein backbone follows in space, as distinct from the backbone atoms themselves, can be considered as a curve. Curves can be completely specified in terms of the local properties curvature and torsion (in general, curvatures), which describe the bending and twisting of the curve at each point. From curvatures, any curve can be constructed using differential geometry. Likewise, from any curve, the curvatures can be obtained. The manipulation of curvatures to control the shape of curves, and the derived protein models, has many applications.

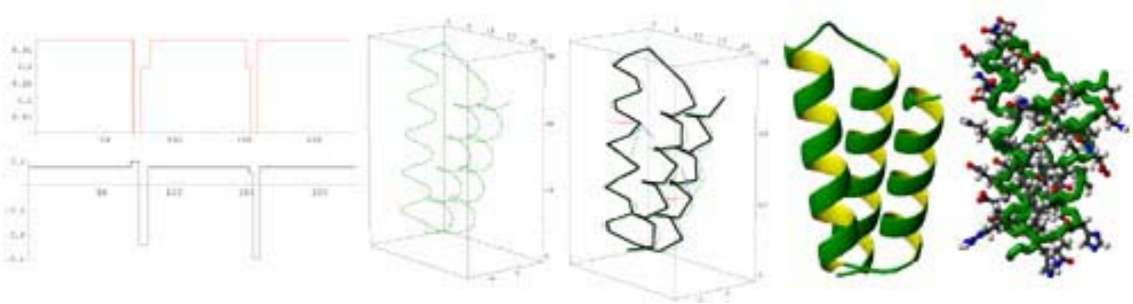

Figure 1: Specification of a 3-helix bundle. A) Curvature (red) and torsion (black) profiles for the curve in B) The three plateaus correspond to the 3 helices in this curve. C) "Cα" positions obtained by selecting points appropriately spaced along the curve. Coordinate systems located at these points on the curve allow construction of D) backbone models E) and complete atomic models.

Exploration of Fold Space: Curvature profiles are continuous functions, but in many cases simplifying assumptions can be made which allow them to be specified by a discrete parameter set. For example, in Fig. 1, the 3-helix bundle is described by

piecewise constant segments. The values and lengths of these segments constitute the parameters needed to describe the curve. The collection of parameters describing the curve can be considered as a vector in the *curvature space*. Each point in the curvature space corresponds to a 3-dimensional space curve, and by systematically searching the curvature space, the realm of possibilities for a given protein architecture can be systematically explored.

For example, the various families of helical repeat proteins, which consist of repeats with the general structure $(\text{helix}_1\text{-turn}_1\text{-helix}_2\text{-turn}_2)^N$ can be accurately modeled by curves described with a periodic curvature profile. The curvature space for this protein architecture has 14 dimensions. *Protein quality functions*, which quantify the adherence of a curve to protein-like geometry, allows exploration of this space. A contour plot of a protein quality function over a curvature space will have "islands" in regions of the space corresponding to protein-like curves.

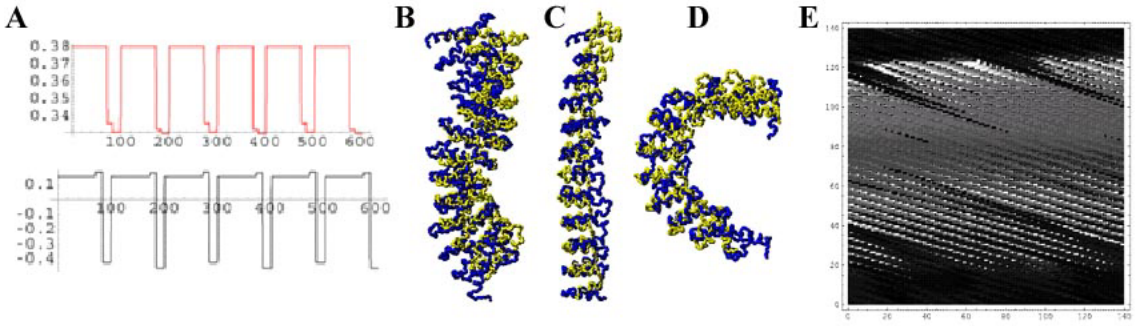

Figure 2: A: Curvature and torsion of a repetitive curve resembling beta-catenin. B) –D) Overlaid ribbon diagrams (blue) and curve-derived models (yellow) for B) β-catenin C) clathrin D) bacterial muramidase E) Contour plot of a protein quality function on a plane containing the points corresponding to the curves in B), C), and D). The presence of many discrete light areas suggests that many protein folds may exist that do not correspond to naturally occurring proteins.

Systematic searches of curvature spaces are computationally intensive, but have the potential to create explicit models of protein folds before they are experimentally observed, and also to address questions about the density and connectivity of fold space. In a sense this can be considered an area of bioinformatics, but which explores the possibilities of form rather than the possibilities of sequence. In the longer term a bridge between the two could be forged.

Algorithms for Protein Design: A theory to postulate plausible novel folds can be tested by applying protein design methods to such folds. Redesigning the sequences of naturally occurring scaffolds is currently feasible, but the ability to postulate and design proteins (particularly functional proteins) not resembling natural proteins is in its infancy. These geometric methods offer new capabilities to this field.

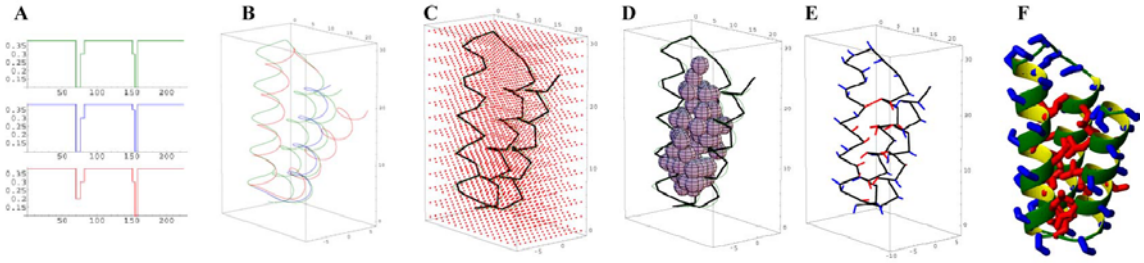

Figure 3: Design tools. A) Subtle changes in curvature profiles result in changes in relative orientation of helices (B). C) Determination of the binary pattern can be accomplished by creating a grid of points. Spheres centered at those points near many Cα positions define an interior volume (D) and E) Cβ atoms located within that volume are assigned hydrophobic character. F) A model with the hydrophobic residues constructed as Leu and the hydrophilic residues constructed as Ser.

Curvature profiles can be manipulated to optimize a design for a desired purpose (eg. surface complementarity to a ligand.) Perhaps the most important single determinant of a protein fold is the binary pattern of hydrophobic and hydrophilic residues. The geometric description enables this determination in terms of curvatures, enabling selection of an optimal fold for a given task rather than modifying a natural protein which may not be well suited to that task.

Algorithms for Structure Determination: While structural analysis by X-ray crystallography or solution NMR is routine, many biological problems are not amenable to these techniques. Other methods of wider applicability but lower information content such as fiber diffraction, cryo-EM, solid-state NMR, and solution X-ray scattering are useful in these cases. The geometric description of protein structure will be most useful in such cases because the small number of curvatures required to specify a model might be obtainable from the limited information content of such data.

Fiber diffraction experiments are interpreted using the theory of diffraction by helices, or in some cases its generalization to the theory of diffraction by coiled-coils. Such structures are spatially repetitive and so can be described with a periodic curvature profile. Being periodic, curvatures describing fiber structures can be represented with a Fourier series. We have found that using only the first term in this series, we recover the theory of diffraction by helices, and that by including 2 terms, we recover the theory of diffraction by coiled-coils. By including more terms, a general theory of fiber diffraction can be obtained. In this way a relationship between the curvatures and the intensities in the fiber diffraction pattern can be obtained.

I am interested in the longer term in applying similar geometric methods to other techniques for which there exists a way of calculating the experimental data from coordinates, such as solid-state NMR or small-angle X-ray scattering. Coordinate models can be built from curves and the expected spectra or scattering profile can be calculated. Curvatures specifying the structure can be obtained by minimization of a residual quantifying the agreement of the observed and calculated data.

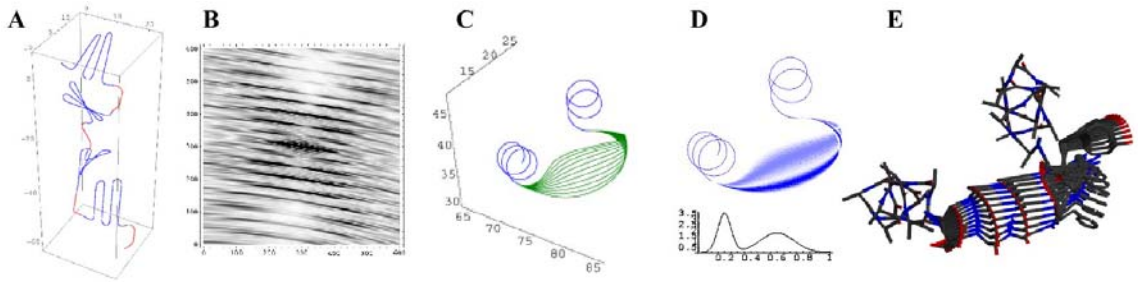

Figure 4: A) A repetitive curve schematically representing a 4 stranded antiparallel beta sheet repeat and B) the calculated fiber diffraction pattern of this structure. C) A disorder surface representing a floppy loop between 2 helices. D) Occupancy density (in blue) on this surface specified by the probability function (below) E) Distribution of structures constructed from the curves in C)

A second emphasis will be on the modeling of static disorder. Proteins are dynamic, flexible molecules and their fluctuations on different timescales may contribute significantly to biological function. The rigid parts of proteins are well-described by atomic coordinate models, but in some cases loops or other local regions undergo motions that are not well-described by harmonic motion about mean atomic positions. The geometric description of protein structure allows specification of curve distributions from which distributions of structures can be created. The parameters describing the breadth of these distributions could then be constrained by X-ray or NMR data, resulting in an estimate of the spatial extent of static disorder.

More generally, a formalism for description of the disordered regions of proteins may prove useful in other contexts. For example, in conjunction with the formalism for fiber diffraction, methods for description of disorder may prove useful in analysis of the diffraction patterns of amyloid fibers, which have approximate but imperfect long-range order. Biological structures often display spatial variability, and a general scheme for description of such structural ensembles is possible in terms of distributions of curvatures. This approach models the range of structures rather than particular instances of a structure.

Analysis of Higher-Order Biological Structure: The geometric formalism is applicable not only to single-chain proteins but to nucleic acids, macromolecular assemblages, organelles, or even larger-scale structures. Structural models have been very useful for protein chemists seeking to understand functional roles of proteins. It may be that explicit, manipulable models for the spatial organization of other structures on different length scales will also be useful to cell and developmental biologists.

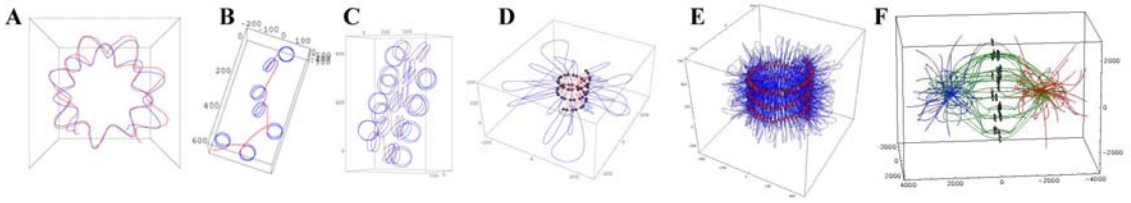

Figure 5: Models of chromatin structure. A) Curve superimposed on trace of phosphate backbone of the nucleosome DNA. B) Beads-on-string model curve C) 30 nm fiber model curve D) Looped fiber model curve E) 700 nm fiber model curve F) Mitotic spindle apparatus model constructed from curves. In each case all parts of the models can be specified in terms of explicit coordinates from the geometric parameters describing the curves. The loops in D) and E), and the curves representing the spindle and asters in F) utilize the method for modeling spatial ensembles.

These types of models could be constrained using data from imaging techniques such as from electron tomography or confocal fluorescence, by fitting curvature parameters to a residual which quantifies agreement with the image or map. As such the method can also serve as a spatial simulation tool. But the real advantage will lie in the ability to integrate information between different length scales. Taking the example of chromatin, each level of structure has one or more natural local coordinate systems. By expressing the models obtained from ultrastructural studies in terms of the local coordinate systems of the higher levels of structure, multi-level descriptions are created in a natural way. In the longer term I will work to integrate the spatial description available through these methods with the quantitative temporal descriptions of biological processes under development by systems biologists.

Statement of Teaching Interests

Although I have some experience with teaching undergraduate biochemistry classes, I would be especially interested in and most effective at teaching courses of a more quantitative nature. Some course areas which would be of interest to me would be physical biochemistry, thermodynamics or statistical mechanics from a biological perspective, and protein crystallography. I would also be interested in developing classes on the theoretical aspects of protein folding, and another which centers on biological structural principles.

In particular, I have long felt that the traditional physical biochemistry class which presents a survey of various techniques would be more useful to research students were it to use "live" examples, namely made use of use data in the format produced by real instruments in the department that the students will be using in their research. I use the general-purpose mathematics program Mathematica for all types of data visualization, model development, and data fitting purposes, and find it very convenient and useful to have a common basis for all such work. I have developed Mathematica scripts for fitting a variety of types of experimental data. I would enjoy developing a course using these as practical examples which would allow students to work on real examples and afterwards they would have working scripts which they could use in their own research.