Interactive Visualization of Protein Dynamics

Henk Huitema and Robert van Liere*

Center for Mathematics and Computer Science, Amsterdam

Abstract

The study of time dependent characteristics of proteins is important for gaining insight into many biological processes. However, visualizing protein dynamics by animating atom trajectories does not provide satisfactory results. When the trajectory is sampled with large times steps, the impression of smooth motion will be destroyed due to the effects of temporal aliasing. Sampling with small time steps will result in the camouflage of interesting motions.

In this case study, we discuss techniques for the interactive 3D visualization of the dynamics of the Photoactive Yellow Protein. We use *essential dynamics* methods to filter out uninteresting atom motions from the larger concerted motions. In this way, clear and concise 3D animations of protein motions can be produced. In addition, we discuss various interactive techniques that allow exploration of the essential subspace of the protein. We discuss the merits of these techniques when applied to the analysis of the yellow protein.

CR Categories and Subject Descriptors: I.3.3 [Computer Graphics]: Picture/Image Generation; I.3.6 [Computer Graphics]: Methodology and Techniques.

Additional Keywords: molecular graphics, essential dynamics, animation, interactive exploration.

1 Introduction

Understanding the function of a protein is fundamental for gaining insight into many biological processes. Proteins are stable mechanical constructs that allow certain internal motion to enable their biological function. These internal motions may be subtle and involve complex correlations between atomic motions, but their nature is inherent in the structure and interaction within the molecule. It is the molecular biologists' challenge to derive internal motions from the molecular structures and interactions in order to identify their functional role. Structural properties of a protein can be obtained with X-ray crystallography or NMR acquisition techniques. Molecular dynamics (MD) simulations are used to simulate the dynamics of a protein at pico/nano-second time scales. MD simulations output one or more trajectory files which describe the coordinates of each individual atom over time. Common practice is to animate these trajectories.

Unfortunately, animating coordinate trajectories with standard techniques do not reveal the most interesting properties of the dynamics. The main problem is that of temporal scale. Animating atom trajectories with large time steps will destroy the impression of smooth motion, while using small time steps will result in the camouflage of interesting motions. To overcome this problem, our approach is to pre-process the trajectories with a motion analysis technique. The technique we use is called *essential dynamics*, [1], which is able to filter out concerted motions from the uninteresting local motions. In this way, clear and concise 3D animations of protein motions can be produced.

*CWI, Department of Software Engineering, P.O. Box 94097, 1090 GB Amsterdam, Netherlands. E-mail {huitema|robertl}@cwi.nl

In this case study, we apply our approach of visualization and interaction of protein dynamics to the Photoactive Yellow Protein (PYP), [2]. This is a light-sensitive and water-soluble protein which was first isolated from the *Ectothiorhodospira halophila* bacterium. PYP functions as a photoreceptor in the negative phototaxis of the cell. The 3D structure of PYP is known; it binds a chromophore via a thio-ester linkage to a unique cystine. It is known that when the chromophore receives blue light its conformation changes. The goal of this work is to gain insight in how the changes of the chromophore effect the overall structure and dynamics of the protein. The visualization and interaction techniques discussed in this paper can be applied to various protein engineering problems.

1.1 Problem Description

In this section we discuss the problem of animating internal protein motions. Consider the line drawing representation of a small part of a protein, as shown in figure 1. Each line represents a bond between two atoms. Arrows are shown to indicate the rotational degrees of freedom a bond may have. These rotations induce displacements of atoms. The total displacement atoms depend on the cumulative rotational degrees of freedom of the atom bonds.

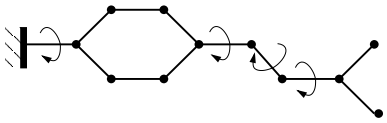

Figure 1: Line drawing representation of a protein. Arrows indicate the degrees of freedom of individual bonds.

Most internal protein motions (usually more than 90 %) are caused by small, independent fluctuations of atoms. These motions have only a local effect and are considered to be uninteresting for the analyses of the protein's function. The motions of interest are called concerted motions: motions that extend over a large number of atoms that induce global structural changes in the protein.

The problem of animating molecular dynamics trajectories is related to temporal scale. If the trajectory is sampled with small time steps the concerted motions will be very slow and difficult to discern. When the trajectory is sampled with large time steps, then the impression of smooth motion will be destroyed due to the effects of temporal aliasing of the local motions.

To overcome these problems a filtering scheme is desired. The goal of such a scheme is to filter out less important motions, resulting in only a set of fluctuations that are caused by concerted motions. Animating only the filtered fluctuations will result in clearer and simpler visualizations of protein dynamics.

2 Related work

There are many techniques to represent protein structures. These include wire frame depictions of bonds, ball and stick representations showing atoms and bonds, ribbon representations for back-

bone structures, space-filling models showing the van der Waal's radius of each atom, etc., [3, 4].

Two common techniques used to visualize motion are based on ribbons and motion blur. A ribbon (or line) is attached to an atom in order to visualize the path. Motion blur of an atom over time can provide the same effect. These technique may suffice in cases when only a small number of motions are to be analyzed. However, applying them to a number of atoms that have large fluctuations will result in excessive cluttering. Usually these techniques work best for a single still image.

Rheingans and Joshi describe three visualization techniques for showing atom positions as probability distributions, [5]. Their techniques have the goal of showing the position and uncertainty of an atom and use rendering methods based on opacity, volume splatting or direct rendering of likelihood volumes. Similar techniques can be used to visualize small atom motions. However, it is doubtful whether the described techniques are suitable for molecular dynamics simulation data, since MD atom fluctuations are much larger than those acquired through crystallography.

Bulatov and Grimes report on several rendering approaches to visualize the evolution of nanoclusters, consisting of several thousands of ions, [6]. They observe that different methods are needed to visualize *slow* and *fast* time scales. Also, they discuss one approach that computes the moments of inertia to the eigenvalues of the inertia tensor of the system. The evolution of an inertia tensor is visualized by plotting its eigenvalues as a function over time. These approaches differ from ours in that clusters of ions have different structural properties than proteins, which have more 3D structure. Inertia tensors do not capture concerted motions.

Various motion analysis techniques for protein dynamics are known; in particular normal mode analysis, [7]. Although essential dynamics differs substantially, the visualization techniques used in this case study can be equally applied to the results of normal mode analysis.

3 Methods

In this section we briefly describe the essential dynamics method and the developed visualization techniques. Essential dynamics has been widely discussed in the literature. See [1] for the original article describing the method. We have developed the interactive visualization techniques for this case study. The techniques have been implemented in a virtual reality environment.

3.1 Essential Dynamics

The basic idea of essential dynamics is to separate the configurational space into two subspaces: an *essential* subspace containing only a few degrees of freedom and a remaining space in which the motions can be considered as *physically constrained*. The essential degrees of freedom describe motions which are relevant for the function of the protein, while the physically constrained subspace describes irrelevant local fluctuations.

The internal motion of a protein is described by a trajectory $\mathbf{x}(t)$, a N-dimensional vector of all atomic coordinates. The essential dynamics method can be seen as a two step procedure:

1. fitting atom trajectories to a reference frame. This step removes overall translational and rotational motions, since only internal motions of the protein are interesting to analyze.

2. constructing a covariance matrix and determining essential motions.

   The symmetric correlation between atomic motions can be expressed in a $N \times N$ covariance matrix C of the positional deviations:

$$C_{ij} \;=\; <(x_i - <x_i>)(x_j - <x_j>)> \qquad (1)$$

   x_i is a position of an atom, and $<>$ denotes the average position of the positions of x_i calculated over the whole trajectory. Atoms moving in a similar direction will result in a positive correlation. Atoms moving in an opposite direction will result in a negative correlation. Atoms moving randomly will result in no correlation.

   The covariance matrix C can be diagonalized by applying an orthonormal transformation matrix T such that:

$$T^T C T = diag(\lambda_1, \lambda_2, ..., \lambda_N); \;\; \lambda_1 > .... > \lambda_N \qquad (2)$$

   The i-column of T are the eigenvectors belonging to eigenvalue λ_i. Eigenvectors indicate directions in the N configurational space, representing correlated displacements of groups of atoms in the protein. Eigenvalues can be interpreted describing the amplitude of these displacements.

   Eigenvectors are sorted by the size of its displacement (i.e. the 'first' eigenvector is the eigenvector with the largest eigenvalue).

3.2 Animation

Filtered atom positions can also be interpreted geometrically (see figure 2). Denote the position of atom k at time step t as $\mathbf{x}_k(t)$ and the essential mode eigenvector as $\mathbf{ev}_i$. The length of the eigenvector is bounded by its corresponding eigenvalues $-\lambda_i, \lambda_i$. The filtered position $\mathbf{p}_k(t)$ is obtained by projecting $\mathbf{x}_k(t)$ onto $\mathbf{ev}_i$.

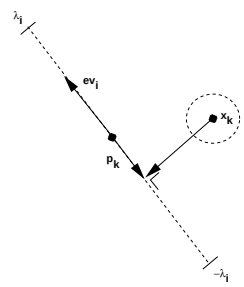

Figure 2: Projecting an atom at position $\mathbf{x}$ onto an essential mode, with a principle eigenvector $\mathbf{ev}_i$. The length of the eigenvector is determined by its corresponding eigenvalue.

The magnitude of all filtered positions is a metric which indicates the amount of displacement with respect to the mean square fluctuation of the eigenvector; i.e. $p(t) = \sum_k \frac{\|\mathbf{p}_k(t)\|}{N}$. Since all correlated motions induce movements in a similar direction, it can be shown that $\mathbf{p}_k(t) = p(t) * \lambda_i * \mathbf{ev}_i$.

Animation of the trajectory is performed by applying the following steps:

1. compute the covariance matrix and the essential eigenvectors.

2. select one or more eigenvectors. If more than one eigenvector is selected, then the sum of the selected eigenvectors is used. Denote the selected eigenvector as $\mathbf{ev}_i$.

3. for each time step t

   - read trajectory data

- fit trajectory into the reference frame by removing global translations and rotations
- compute the mean square fluctuation of the selected eigenvector, $p(t)$
- for each atom A_k :
  - draw A_k at $p(t) * \lambda_i * \mathbf{ev}_i$.

Interpolation between two trajectory time steps, allows the animation to be played with very small time steps. Interpolation is realized by interpolating between the mean square fluctuations of the selected eigenvector $p(t)$.

The given algorithm addresses the temporal scale problem discussed in section 1.1. The covariance analysis has filtered out the small uncorrelated motion and only the concerted motions are shown.

3.3 Interaction

In addition to animation of trajectories, we have implemented a number of interactive techniques which are tailored to explore the configurational space spanned by the essential eigenvectors. These interactive techniques will be illustrated in the next section when applied to PYP.

Here, we briefly mention three of the implemented techniques:

- the user can drag an atom (or group of atoms) along the eigenvector. Dragging an atom to a new position will result in a new displacement p. This displacement is used to position other atoms in the protein.

  Directly manipulating atoms in this way allows the user to interactively explore concerted fluctuations with respect to a selected eigenvector.

- an *angular widget* can be pegged onto any part of the backbone. This widget is useful to monitor the hinge bending properties.

- a *measuring stick widget* can be pegged between two atoms. This widget is useful to monitor the distance between atoms during exploration.

4 PYP Dynamics

As mentioned briefly in the introduction, PYPs 3D structure binds a chromophore via a thio-ester linkage to a unique cystine. It is known that when the chromophore receives blue light, it induces various conformational changes. A hypothesis is that a conformational change on the surface of PYP functions as a signaling device for another protein. The goal of our work is to gain insight into which part of PYP is responsible for this signalling. In particular, we use our visualization and interaction techniques to answer the following questions:

- which domains in PYP are effected when the chromophore changes state? The answer to this will help in finding the signal receiving protein.

- which residues are involved in these conformational changes? By investigating mutants, we can experimentally test our hypotheses that these residues are involved in the signaling.

A three step procedure is followed to address these questions. First, animation of the data provides global overview of the internal protein motions and an indication of concerted fluctuations. Second, exploration techniques are used to gain a deeper understanding of particular concerted motions. Finally, exploration combined with measuring widgets provide quantitative insight into the specific internal protein motions and fluctuations.

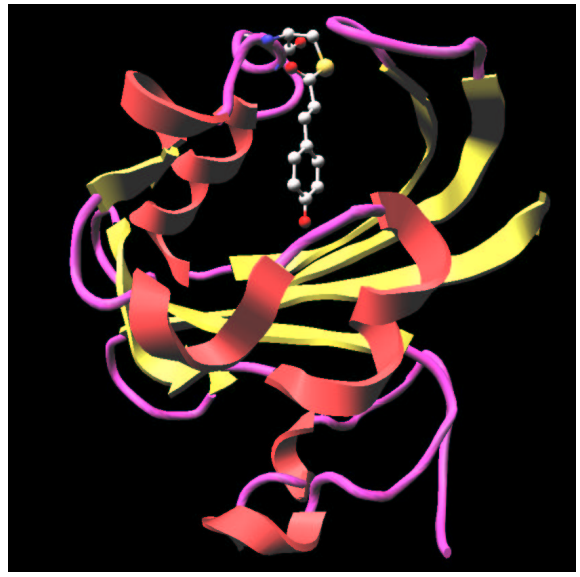

Figure 3: An overview of PYP.

Figure 3 shows an overview of PYP. A cartoon representation is used for the backbone. The side groups are not shown in this figure. The chromophore is shown as a colored ball and stick representation. Animation of the MD trajectory reveals that the chromophore fluctuates considerably along the most relevant essential dynamics eigenvectors. The animation also reveals that these fluctuations are correlated with large motions on the backbone and side chains.

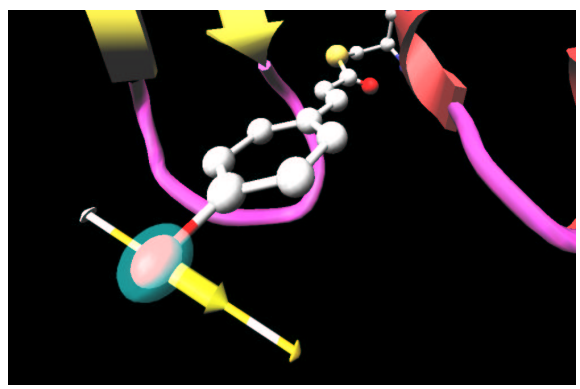

Figure 4: Exploration of the essential eigenspace.

Figure 4 illustrates the exploration of the eigenspace spanned by the relevant essential dynamics eigenvectors. The eigenvector is drawn on a selected atom of the chromophore. The selected atom is drawn as a semi-transparent sphere. The eigenvector is displayed as a yellow vector, the minimum and maximum bounds are displayed on a yellow-white cylinder.

By dragging atoms along their essential eigenvectors, insight into concerted motions can be obtained. In particular, the exploration shows that the motion of the chromophore along the eigenvector induces a contraction of the 50-helix. In addition, as a result of this contraction, the N-terminus region will pivot around the 25-

turn. (The 50-helix, N-terminus region and 25-turn are not shown in the figure 4.)

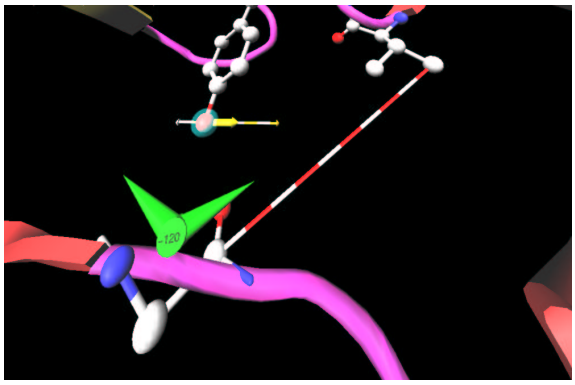

Figure 5: Measuring widgets during exploration.

Finally, figure 5 shows how measuring widgets are used during the exploration of the eigenspace. These measurements can provide quantitative information of the motions. An angular widget (shown as a green compass) is used to measure the dihedral angle that glycine-51 makes with the backbone. During exploration, we can see that the large conformational freedom of this angle is responsible for the contraction of the 50-helix. The distance widget (shown as a red-white cylinder) is used to measure the distance between the 100-loop and the 50-helix.

4.1 Discussion

Animating the raw trajectory data only with standard animation techniques cannot provide the required insight into the dynamics of PYP trajectory data. By using the results of essential dynamics motion analysis, we were able to filter out the interesting motions, resulting in animations which are more easily understood. By selecting only the essential eigenvectors, we were able to focus on interesting concerted motions.

Definite answers for the biochemical questions posed in the previous section cannot be obtained by visualization alone. However, with these visualization techniques we were able to formulate a hypothesis, which can be tested by further biochemical research. A candidate domain for the signaling function was found to be the N-terminal region. Critical residues for the conformational changes are glycine-51 and several residues in the 25-turn.

5 Conclusion

In this case study we have developed a number of techniques for the visualization of protein dynamics. Trajectory data describing atom motions were obtained from molecular dynamics simulations. We have applied essential dynamics methods to filter and display only the correlated internal motions of atoms. We have defined a number of visualization techniques that allow for the animation and interactive exploration of the essential subspace of motions. By applying these techniques to the Photoactive Yellow Protein, we able to explore correlated motions relating to the chromophore function. This resulted in insight to a candidate domain for the signaling function induced by motions of the chromophore. In addition, critical residues for the conformational changes were found. This insight could not be obtained with traditional animation techniques.

The lessons we learned in this case study are:

- Using filtering techniques allows for the study of internal protein motions. Without these filtering techniques these motions are camouflaged by small scale motions.
- Active exploration of the space spanned by the selected eigenvectors yield a more intuitive understanding than that of passive viewing of an animation.
- Exploration in combination with measuring widgets provide the ability to examine very specific properties of a motion.

Although the techniques discussed in this paper are strongly related to protein motions, we believe that similar motion analysis techniques can be found for other temporal data sets. For example, in turbulent flow, some vortices are often strongly correlated, while others are not. Principle component analysis might be used to determine those vortices that contribute to the global flow patterns.

Acknowledgments

We would like to thank Daan van Aalten of the the Department of Biochemistry, University of Dundee and Wim Crielaard of the Department of Microbiology, University of Amsterdam for discussion related to protein motion, essential dynamics and PYP.

References

[1] A. Amadei, A.B.M Linssen, and H.J.C. Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function and Genetics*, 17(4):412–425, 1993.

[2] D. van Aalten, W. Hoff, J. Findlay, W. Crielaard, and K. Hellingwerf. Concerted motions in the photoactive yellow protein. *Protein Engineering*, 11(10):873–879, 1998.

[3] A.J. Olson and D.S. Goodsel. Visualizing biological molecules. *Scientific American*, 267(5):76–81, 1992.

[4] H. Haase, J. Strasser, and F. Dai. VR techniques for the investigation of molecule data. *Computers & Graphics*, 20(2):207–217, 1996.

[5] P. Rheingans and S. Joshi. Visualization of molecules with positional uncertainty. In E. Gröller, H. Löffelmann, and W. Ribarsky, editors, *Proceedings of the Joint EUROGRAPHICS - IEEE TCVG Symposium on Visualization*, pages 299–305. Springer-Verlag, 1999.

[6] V.L. Bulatov and R.W. Grimes. Visualization of molecular dynamics simulations. In *Proceedings of the EUROGRAPHICS UK Chapter*, 1996.

[7] T. Horiuchi and N. Go. Projection of monte carlo and molecular dynamics trajectories onto the normal mode axes. *Proteins: Structure, Function and Genetics*, 10(1):106–116, 1991.