Exhibit D

Part 1 of 2

A Visual Computing Environment for Very Large Scale Biomolecular Modeling

M. Zeller, J.C. Phillips, A. Dalke, W. Humphrey, K. Schulten
R. Sharma†, T.S. Huang, V.I. Pavlović, Y. Zhao, Z. Lo, S. Chu

Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign
405 N. Mathews Avenue, Urbana, IL 61801

†Department of Computer Science and Engineering
Pennsylvania State University
University Park, PA 16802-6106

Abstract

Knowledge of the complex molecular structures of living cells is being accumulated at a tremendous rate. Key technologies enabling this success have been high performance computing and powerful molecular graphics applications, but the technology is beginning to seriously lag behind challenges posed by the size and number of new structures and by the emerging opportunities in drug design and genetic engineering. A visual computing environment is being developed which permits interactive modeling of biopolymers by linking a 3D molecular graphics program with an efficient molecular dynamics simulation program executed on remote high-performance parallel computers. The system will be ideally suited for distributed computing environments, by utilizing both local 3D graphics facilities and the peak capacity of high-performance computers for the purpose of interactive biomolecular modeling. To create an interactive 3D environment three input methods will be explored: (1) a six degree of freedom "mouse" for controlling the space shared by the model and the user; (2) voice commands monitored through a microphone and recognized by a speech recognition interface; (3) hand gestures, detected through cameras and interpreted using computer vision techniques.

Controlling 3D graphics connected to real time simulations and the use of voice with suitable language semantics, as well as hand gestures, promise great benefits for many types of problem solving environments. Our focus on structural biology takes advantage of existing sophisticated software, provides concrete objectives, defines a well-posed domain of tasks and offers a well-developed vocabulary for spoken communication.

1: Introduction

The biomedical sciences are presently undergoing a revolutionary development following the determination of a rapidly increasing number of complex molecular structures of the living cell. Structures encompassing many thousands of atoms, with integral biological functions and of great medical significance, are being discovered at an increasing pace. This development came about only with the advent of high performance computers, powerful molecular graphics software, and a host of program packages regularly used by thousands of researchers. Computational methods play an essential role in structure determination and structure refinement, in graphical interpretations of complex structures, in modeling of biopolymers for drug design and genetic engineering of proteins, and in the physical, mechanistic and dynamic analysis of resolved structures with the goal for understanding the principles underlying biopolymer architecture and function.

1063-6862/97 $10.00 © 1997 IEEE

Computational technology, unfortunately, lags now behind the rapid progress of structural biology and cannot adequately handle the size and number of structures discovered today. Furthermore, current user interfaces suffer from numerous shortcomings, e.g., a poor adaptation to the three-dimensional character of biopolymer structures and a complexity that requires long training periods. Also, gross limitations in locally available computer power restricts the interactive modeling of biopolymers and source codes are, with rare exceptions, inaccessible or poorly documented. Finally, the present paradigm for the rendering of biomolecular structures simply displays positions and types of atoms and bonds, neglecting topological and physical characteristics such as surfaces, voids, pockets, or electrostatic potentials.

We have taken an innovative, collaborative approach to address the above impediments. A visual computing environment, MDScope, has been developed which permits interactive modeling [1, 4, 6]. MDScope connects a powerful molecular graphics program, VMD, with a fast and efficient modeling program, NAMD, specifically designed for parallel computers. MDScope is ideally suited for high performance, distributed computing environments. The program presently links VMD to NAMD, the latter running on a remote high performance parallel computer. Three input methods have been explored in the framework of the VMD program to create an interactive 3D environment: a six (translation and rotation) degrees of freedom, electromagnetically tracked "mouse" has been integrated into VMD for manipulations of 3D objects; voice commands monitored through a microphone and recognized by a speech recognition interface have been introduced to replace cumbersome keyboard commands; hand gestures to manipulate the displayed model are detected through stereo cameras and interpreted using computer vision techniques. A large screen stereo graphics facility, expected to become the standard for biomolecular graphics, has been implemented and is frequently used by researchers.

A long-term goal of VMD is to enable multiple users to interact with the model simultaneously. This interaction and collaboration is expected to significantly shorten the problem-solving cycle in biomolecular modeling since users can guide searches, e.g., for optimal designs, without resorting to a time consuming, non-intuitive cycle of batch jobs carrying out automated searches. Incorporating voice commands will allow the user to be free of the keyboard, and hand gestures will permit the user to easily manipulate the displayed model and to explore different molecular configurations. The combination of both speech and hand gestures will be far more powerful than either individually, and its success will apply to interactive environments for structural biologists as well as to a wide range of other scientific and engineering applications. To accomplish this interaction, highly robust automatic speech recognition (ASR) and automatic gesture recognition (AGR) techniques are necessary. These techniques will be required for such problems as differentiation between commands and casual speech, and distinction between meaningful and meaningless gestures and hand movements.

The primary infrastructure for our project is a large-screen stereographic projection facility, developed in the Theoretical Biophysics Group and shared by a large group of biomedical researchers. The facility employs cost effective and space saving display hardware which is added to a high end graphics workstation and can be easily duplicated at other sites. It produces $8' \times 6' \times 6'$ 3D models in a 120 square foot area. The facility consists of a projector which displays alternating left- and right-eye views onto the screen at nearly twice the rate of ordinary projectors. The images, when viewed through special eyewear, produce a stereo display. In addition, a spatial tracker is used as a 3D input device for the development of a three-dimensional user interface. The program VMD has been designed for this environment as well as for standard monitors. Figure 1 attempts to convey an impression of the system through a photomontage of an actual image (nuclear hormone receptor–DNA complex) and of the actual space.

The key goal of our work is to simplify model manipulation and rendering to such a degree that biomolecular modeling assumes a playful character; this will allow the researcher to explore variations of their model and concentrate on biomolecular aspects of their task without undue distraction by computational aspects. The ultimate goal, illustrated in Fig. 1, will focus on wide ranging improvements of the graphical user interface of the existing programs MDScope which will become possible through the combined expertise of researchers in computational structural biology, parallel computing, numerical algorithm, and intelligent human-computer interaction technology.

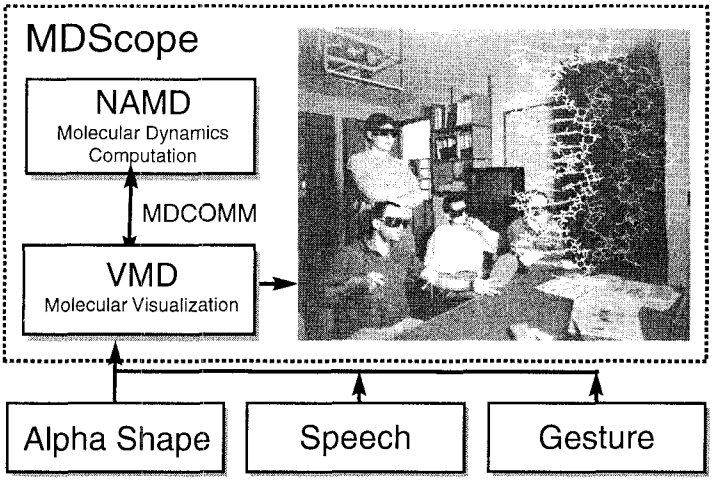


Figure 1. The program VMD, coupled with the program NAMD, and speech and gesture user-interface components. These facilities comprise MDScope, a problem-solving environment for structural biology.

2: MDScope, a Visual Environment for Structural Biology

MDScope is a set of integrated software components which provides an environment for simulation and visualization of biomolecular systems in structural biology. It consists of three separate packages which may be used individually, or together to constitute the MDScope environment [5]. These packages are:

1. The program NAMD, a molecular dynamics program which runs in parallel on a wide variety of architectures and operating systems.
2. The program VMD, a molecular visualization program which displays both static molecular structures and dynamic molecular motion as computed by programs such as NAMD.
3. The MDCOMM software, which provides an efficient means of communication between VMD and NAMD, and allows VMD to act as a graphical user interface to NAMD. Using MDCOMM, VMD provides an interface for interactive setup and display of a molecular dynamics simulation on a remote supercomputer or high-performance workstation, using NAMD as a computational "engine".

Molecular dynamics calculations are computationally very expensive, and require large amounts of memory to store the molecular structure, coordinates, and atom-atom interaction lists. The challenge of efficient calculation of the inter-atomic forces by using high performance computing is addressed in NAMD through the use of parallel computation and incorporation of the Distributed Parallel Multipole Tree Algorithm (DPMTA) [2]. NAMD uses a *spatial decomposition* algorithm to partition the task of computing the force on each atom among several processors. This algorithm subdivides the volume of space occupied by the molecule into uniform cubes (or *patches*), as shown in Figure 2, which are distributed among the processors in a parallel computer. The motions of the atoms in each patch are computed by the processor to which each patch is assigned; as atoms move they are transferred between the patches, and patches are reassigned to different processors in order to maintain a uniform computational load.

The key functions of the program VMD are to visualize biomolecular systems, to allow direct interaction between a user and a molecule being simulated on another computer, and to provide an intuitive user interface for controlling the visual display and remote simulation. VMD uses the

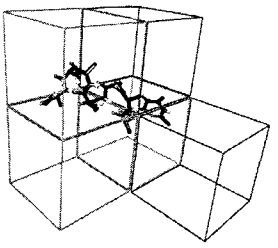


Figure 2. Spatial decomposition of a small polypeptide. Each cube represents a patch in NAMD.

MDCOMM software mentioned above to enable it to initiate, display, and control a simulation using NAMD. As the trajectory of a molecular system is calculated, the coordinates of each atom are sent from NAMD to VMD. Current network technology provides the necessary bandwidth to communicate the atomic coordinate data; the use of a high-performance dynamics program is crucial in order to furnish new data at the speed required for interactive display.

VMD implements many different forms of user interfaces – users may control the program through keyboard commands, a mouse, and a graphical user interface. VMD also implements a mechanism for external programs to serve as user interface components, by allowing them to communicate with VMD through standard network communication channels. This makes it possible for new user interface methods, such as the speech- and gesture-recognition systems discussed in Section 3, to be developed in parallel with VMD.

Development of MDScope is an ongoing project in the Theoretical Biophysics Group at the University of Illinois [1]. All three components of MDScope (NAMD, VMD, and MDCOMM) may be obtained via anonymous ftp (`ftp.ks.uiuc.edu`), or World Wide Web (`http://www.ks.uiuc.edu`). The components may be used individually or in concert, and include the complete source code for the packages as well as extensive documentation describing how to use and modify the programs.

Currently, NAMD is available for a wide variety of architectures and operating systems, including clusters of high-performance Hewlett-Packard, Silicon Graphics, and IBM RS/6000 Unix workstations, as well as the Cray T3D and Convex Exemplar parallel systems. We are in the process of making NAMD available on the IBM SP-2 and SGI Power Challenge architectures. NAMD should be compilable on any system with a C++ compiler and PVM version 3.3.11.

In addition to the full source and SGI binary distributions (IRIX 5.x and IRIX 6.x), the program VMD is already available for Hewlett-Packard (HP-UX 9 and HP-UX 10 using Mesa emulated OpenGL) and Linux workstations. Ports to other platforms, most notably IBM RS/6000 (AIX), will be available soon. VMD displays a graphical rendering of the molecular systems under study, and provides both a text-based console interface and a complete graphical interface for the user. These controls are used to modify the appearance of the molecules and display, to control the display of structural features of the molecules, and to access remote computers running molecular dynamics simulations. Multiple structures may be viewed simultaneously, and a flexible atom selection mechanism enables the user to easily select subsets of atoms for display. VMD includes an extensive text-command processing capability through the use of the Tcl library, a popular and widely available package for script parsing and interpreting. The use of Tcl makes it possible for users to write scripts including such features as variable substitution, control loops, and function calls. The current rendering of a molecule may also be saved in an image file or in a format suitable for use by several image-processing packages. Also, by connecting directly to a remote computer running a molecular dynamics simulation, VMD offers users the capability to interactively participate in an

ongoing simulation, e.g. the option to apply perturbative forces to individual atoms.

Speech and hand gestures are fundamental methods of human communication, and their use for interaction with and control of the display of VMD will greatly improve the utility of the program. Speech- and gesture-recognition user interfaces are being added to VMD to provide a "natural" working environment for researchers.

3: Speech and Gesture Interface

To fully exploit the potential that visual computing environments offer, there is a need for a "natural" interface that allows the manipulation of such displays without cumbersome attachments. In this section we describe the use of visual hand gesture analysis and speech recognition for developing a speech/gesture interface to VMD. The free hand gestures are used for manipulating the 3-D graphical display together with a set of speech commands. We describe the visual gesture analysis and the speech analysis techniques used in developing this interface. The dual modality of speech/gesture is found to greatly aid the interaction capability.

The communication mode that seems most relevant to the manipulation of physical objects is hand motion, also called *hand gestures.* We use it to act on the world, to grasp and explore objects, and to express our ideas. Now virtual objects, unlike physical objects, are under computer control. To manipulate them naturally, humans would prefer to employ hand gestures as well as speech. Psychological experiments, for example, indicate that people prefer to use speech in combination with gestures in a virtual environment, since it allows the user to interact without special training or special apparatus and allows the user to concentrate more on the virtual objects and the tasks at hand [3]. We explore this multimodal nature of HCI involved in manipulating virtual objects using speech and gesture.

To keep the interaction natural, it is desirable to have as few devices attached to the user as possible. Motivated by this, we have been developing techniques that will enable spoken words and simple free-hand gestures to be used while interacting with 3D graphical objects in a virtual environment. The voice commands are monitored through a microphone and recognized using automatic speech recognition (ASR) techniques. The hand gestures are detected through a pair of strategically positioned cameras and interpreted using a set of computer vision techniques that we term automatic gesture recognition (AGR). These computer vision algorithms are able to extract the user hand from the background, extract positions of the fingers, and distinguish a meaningful gesture from unintentional hand movements using the context. We use the context of the VMD environment to place the necessary constraints to make the analysis robust and to develop a command language that attempts to optimally combine speech and gesture inputs.

VMD uses a keyboard and a magnetically tracked pointer as the interface. This is particularly inconvenient since the system is typically used by multiple (6-8) users, and the interface hinders the interactive nature of the visualization system. Thus incorporating voice command control in MDScope would enable the users to be free of keyboards and to interact with the environment in a natural manner. The hand gestures would permit the users to easily manipulate the displayed model and "play" with different spatial combinations of the molecular structures. The integration of speech and hand gestures as a multi-modal interaction mechanism would be more powerful than using either mode alone, motivating the development of the speech/gesture interface. Further, the goal was to minimize the modifications needed to the existing VMD program for incorporating the new interface. The experimental prototypes that we built for both the speech (ASR) and hand gesture analysis (AGR) required the following addition to the VMD environment.

Software. In order to reduce the complexity and increase the flexibility of the program design, a communications layer was added so external programs can be written and maintained independently from the VMD code. These use the VMD text language to query VMD for information or to send new commands. The VMD text language is based on the TCL scripting language. Since all the capabilities of VMD are available at the script level, an external program can control VMD in any way. Both the ASR and AGR programs interact with VMD using this method. For a simple voice command, such as "rotate left 90", the ASR converts the phrase into the VMD text command

"rotate y 90" and sends that to VMD. Similarly, when the AGR is being used as a pointing device, it sends the commands to change the current position and vector of VMD's graphical 3D pointers.

Setup for visual gesture analysis. To facilitate the development of AGR algorithms, we designed an experimental platform shown in Figure 3 that was used for gesture recognition experiments. In addition to the uniformly black background, there is a lighting arrangement that shines red light on the hand without distracting the user from the main 3D display. The setup has the additional advantage that it can be transported easily and is relatively unobtrusive.

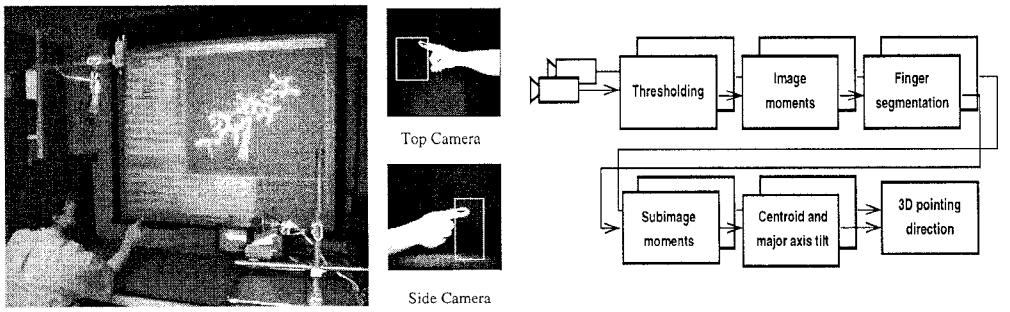


Figure 3. The experimental setup with two cameras used for gesture recognition (left) and overview of the AGR subsystems (right).

Setup for speech analysis. A prototype ASR system has been implemented and integrated into VMD. The system consisted of two blocks: a recorder front-end followed by the recognizer unit. The recorder employed a circularly-buffered memory to implement its recording duties, sending its output to the recognizer unit in blocks. A digital volume meter accompanied this to provide feedback to the user by indicating an acceptable range of loudness. The recognizer that followed was developed by modifying HTK software. This unit performed feature extraction and time-synchronous Viterbi decoding on the input blocks, sending the decoded speech directly via Tcl-dp commands to an SGI Onyx workstation where the VMD process resided.

Speech/gesture command language. In order to effectively utilize the information input from the user in the form of spoken words and simple hand gestures, we have designed a command language for MDScope that combines speech with gesture. This command language uses the basic syntax of $< action >< object >< modifier >$. The $< action >$ component is spoken (e.g., "rotate") while the $< object >$ and $< modifier >$ are specified by a combination of speech and gesture. An example is, speaking "this" while pointing, followed by a modifier to clarify what is being pointed to, such as "molecule", "helix", "atom", etc., followed by speaking "done" after moving the hand according to the desired motion. Another example of the desired speech/gesture capability is the voice command "engage" to query VMD for the molecule that is nearest to the tip of the pointer and to make the molecule blink to indicate that it was selected and to save a reference to that molecule for future use. Once engaged, the voice command "rotate" converts the gesture commands into rotations of the chosen molecule, and the command "translate" converts them into translations. When finished, the command "release" deselects the molecule and allows the user to manipulate another molecule. The ASR and AGR techniques that made the above interaction possible are described next.

3.1: Speech input using ASR

In the integration of speech and gesture within the MDScope environment, a real-time decoding of the user's commands is required in order to keep pace with the hand gestures. Thus there is a need for "word spotting" which is defined as the task of detecting a given vocabulary of words embedded in unconstrained continuous speech. It differs from conventional large-vocabulary continuous speech recognition (CSR or LVCSR) systems in that the latter seeks to determine an optimal

sequence of words from a prescribed vocabulary. A direct mapping between spoken utterances and the recognizer's vocabulary is implied with a CSR, leaving no room for the accommodation of non-vocabulary words in the form of extraneous speech or unintended background noise. The basis for word spotting, also termed keyword spotting (KWS), is dictated by real world applications. Real users of a spoken language system often embellish their commands with supporting phrases and sometimes even issue conversation absent of valid commands. In response to such natural language dialogue and the implications to robust human-computer interaction, standard CSR systems were converted into spotters by simply adding filler or garbage models to their vocabulary. Recognition output stream would then consist of a sequence of keywords and fillers constrained by a simple syntactical network. In other words, recognizers operated in a "spotter" mode. While early techniques emphasized a template-based dynamic time warping (DTW) slant, current approaches are typically armed with the statistical clout of hidden Markov models (HMMs) [8, 9, 12], and recently with the discriminatory abilities of neural networks (NN). These were typically word-based and used an overall network which placed the keyword models in parallel with the garbage models.

Keywords. Table 1 lists the keywords and their phonetic transcriptions chosen for the experiment. These commands allowed the VMD user to manipulate the molecules and polymeric

Keyword	Transcription
translate	t-r-ae-n-s-l-ey-t
rotate	r-ow-t-ey-t
engage	eh-n-g-ey-jh
release	r-ih-l-iy-s
pick	p-ih-k

Table 1. Keywords.

structures selected by hand gestures. In modeling the acoustics of the speech, the HMM system was based on phones rather than words for large vocabulary flexibility in the given biophysical environment. A word-based system, though invariably easier to implement, would be inconvenient to retrain if and when the vocabulary changed.

Fillers. Filler models are more varied. In LVCSR applications, these fillers may be represented explicitly by the non-keyword portion of the vocabulary, as whole words for example. In other tasks, non-keywords are built by a parallel combination of either keyword "pieces" or phonemes whether they be context-independent (CI) monophones or context-dependent (CD) triphones or diphones [9].

Twelve fillers or garbage models were used to model extraneous speech in our experiment. Instead of being monophones or states of keyword models as used in prior experiments in the literature, the models that were used covered broad classes of basic sounds found in American English. There are several things to note. First, the class of "consonants-africates" was not used due to the brevity of occurrence in both the prescribed vocabulary and training data. As observed by [12] and many other researchers, varying or increasing the number of models does not gain much in spotting performance. Second, a model for background silence was included in addition to the twelve garbage models listed. Such a model removed the need for an explicit endpoint detector by modeling the interword pauses in the incoming signal. Note also that the descriptors for the vowel class correspond to the position of the tongue hump in producing the vowel.

Recognition Network. The recognition syntactical network placed the keywords in parallel to a set of garbage models which included a model for silence. These models followed a null grammar, meaning that every model may precede or succeed any other model. A global grammar scale factor (s) and transition probability factor (p) were used to optimize the recognition accuracy and adjust the operating point of the system.

Features and Training. After sampling speech at 16kHz and filtering to prevent aliasing, the speech samples were preemphasized with a first order digital filter using a preemphasis factor of 0.97 and blocked into frames of 25 ms in length with a shift between frames of 10 ms. Each frame of speech was weighted by a Hamming window, and then mel-frequency cepstral coefficients

of sixteenth order were derived and weighted by a liftering factor of 22. Cepstral coefficients were chosen as they have been shown to be more robust and discriminative than linear predictive coding coefficients or log area ratio coefficients. Normalized log energy and first order temporal regression coefficients were also included in the feature vector.

The topology of the HMMs for both keyword-phones and garbage models consisted of five states, the three internal states being emitting states. Following a left-to-right traversal, each state was described by a mixture of five continuous density Gaussians with diagonal covariance matrices. Three iterations of the Baum-Welch reestimation procedure were used in training.

In training the set of fifteen keyword-phones, forty sentences were developed as follows. Each of the five keywords were individually paired with the remaining four. This was then doubled to provide a sufficient number of training tokens. In sum, the sentences were composed of pairs of keywords such as "engage translate" and "rotate pick" which were arranged in such an order as to allow each of the keywords to be spoken sixteen times. Each VMD user proceeded with this short recording session.

In training the garbage models, a much more extensive database of training sentences was required to provide an adequate amount of training data since the twelve broad classes cover nearly the entire spectrum of the standard 48 phones. The TIMIT database was subsequently employed to provide an initial set of bootstrapped models. Retraining was then performed once for one VMD user who had recorded a set of 720 sentences of commonly used VMD commands. These sentences spanned the scope of the VMD diction, including a more detailed set of commands, numbers, and modifiers. This was necessary to provide data normalized to the existing computational environment. Note that the garbage models were trained only once for this experiment. Hence, VMD users only needed to go through the short training procedure detailed above.

Performance. Upon testing the system as a whole with fifty test sentences that embedded the keywords within bodies of non-keywords, wordspotting accuracies ranged to 98% on the trained speaker. The trained speaker refers to each user who trained the keyword-phones regardless of the one who trained the garbage models. This was considered very well by the VMD users for the given biophysics environment, supporting the techniques that were used. In general, false alarms occurred only for those situations where the user embedded a valid keyword within another word. For example, if one says 'translation' instead of 'translate', the spotter will still recognize the command as 'translate'.

3.2: Hand gesture input using AGR

The general AGR problem is hard, because it involves analyzing the human hand which has a very high degree of freedom and because the use of the hand gesture is not so well understood (See [7] for a survey on vision-based AGR). However, we use the context of the particular virtual environment to develop an appropriate set of gestural "commands". The gesture recognition is done by analyzing the sequence of images from a pair of cameras positioned such that they facilitate robust analysis of the hand images. The background is set to be uniformly black to further help with the real-time analysis without using any specialized image-processing hardware.

Finger as a 3D pointer. The AGR system consists of two levels of subsystems (See Figure 3). First level subsystems are used to extract a 2D pointing direction from single camera images. The second level subsystem combines the information obtained from the outputs of the first level subsystems into a 3D pointing direction. To obtain the 2D pointing direction, the first level subsystems perform a sequence of operations on the input image data. The gray-level image is first thresholded in order to extract a silhouette of the user's lower arm from the background. Next, first and second image moments are calculated and then used to form a bounding box for extraction of the index finger. Once the finger is segmented from the hand, another set of image moments is calculated, this time for the finger itself. Finally, based on these moments, 2D finger centroid and finger direction are found. 3D pointing direction is finally determined in the second level subsystem using the knowledge of the setup geometry and 2D centroids and pointing directions. This information is then forwarded to the central display manager which displays a cursor at an appropriate screen position. Our implementation produced a tracking rate of about 4 frames per second, mainly limited by the

inability of the digitization hardware to properly handle multiple video signals. Special purpose hardware can easily improve the performance. However, even with the low sampling rate, the users can achieve a reasonable control of the display.

Gestures for manipulating 3-D display. In addition to recognizing a pointing finger, we have developed a hidden Markov model based AGR system for recognizing a basic set of manipulative hand gestures. Figure 4 gives examples of some of the gestures that were used. We have also developed a gesture command language for MDScope that is mainly concerned with manipulating and controlling the display of the molecular structures. The gesture commands are categorized as being either dynamic (e.g., move back, move forward) or static (e.g., grab, release, stop, up, down).

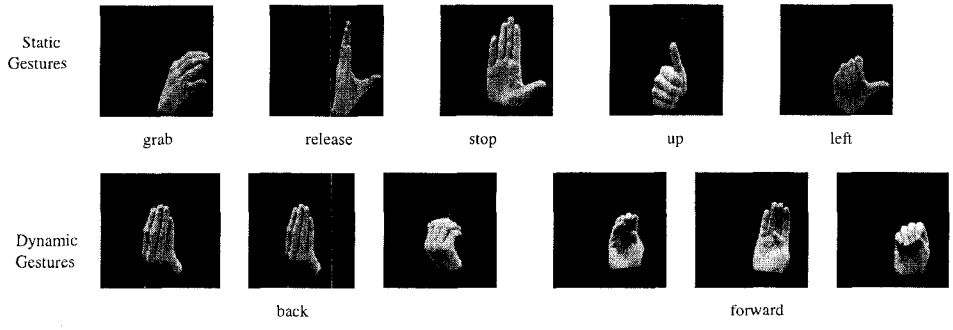


Figure 4. Examples hand gestures used in manipulating a virtual object and interpreted using AGR.

The system uses image geometry parameters as the features that describe any particular hand posture (static hand image). We use the *radon transform* of an image to extract these features [11]. The set of gestures that were used consisted of both static and dynamic gestures (see Figure 4). The recognition system was built by training hidden Markov models for the specific gestures on sample runs. Each gesture in the vocabulary was modeled as a single four-state-HMM. The observations were modeled by a Gaussian mixture of two different sizes (one and three) with a diagonal covariance matrix. Although only 35 training sequences were used the performance of the recognition system was quite good (80% correct recognition rate). The performance is expected to improve with a better model of the hand and by exploiting multimodality.

The experimental speech/gesture interface reported so far could be part of a more general multimodal framework, where other modalities can also be exploited to make the interface more natural and efficient. For example, consider the problem of visual gesture analysis, the following interactions can be exploited to improve the gesture recognition process in the multimodal framework: (a) interaction of gesture and speech, (b) interaction of gesture and the virtual scene, (c) interaction of gesture and gaze direction, and (d) interaction of gesture and graphical display. These interactions are discussed in more detail in Ref. [10].

4: Discussion

Combining high performance parallel computing and high end graphics for research in structural biology will open new avenues for very large scale biomolecular modeling. Efficient parallel algorithms and powerful molecular graphics software provide crucial tools to address the challenges ahead in computational biology. MDScope, as a scalable, parallel and object oriented software package, will contribute to the ongoing effort of biomedical researchers to expand system size and time scales of the simulations. Furtheremore, MDScope provides an excellent testbed where computer vision and speech recognition techniques are used for building a natural human-computer interface, using spoken words and free hand gestures. The prototype speech/gesture interface presented lets

the scientist easily and naturally explore the displayed information. The speech/gesture interface offers a level of interactive visualization that was not possible before. Incorporating voice command control in MDScope enables the users to be free of keyboards and to interact with the environment in a natural manner. The hand gestures permit the users to easily manipulate the displayed model and "play" with different spatial combinations of the molecular structures. The integration of speech and hand gestures as a multi-modal interaction mechanism proves to be more powerful than using either mode alone and forms a basis of future work

Acknowledgments

This work was supported by the Roy J. Carver Charitable Trust, by the U.S. Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0003 and in part by the National Science Foundation Grant IRI-96-34618. Simulations were carried out at the Resource for Concurrent Biological Computing at the University of Illinois, funded by the National Institute of Health (Grant P41RR05969) and by the National Science Foundation (Grants BIR-9318159 and ASC-8902829). J.C.P. was supported by a Computational Science Fellowship from the United States Department of Energy.

References

[1] Source code and documentation for the MDScope applications are available via anonymous ftp (`ftp.ks.uiuc.edu`), or via the World Wide Web (`http://www.ks.uiuc.edu`).

[2] J. A. Board, Jr., R. R. Batchelor, and J. F. Leathrum, Jr. In *Proceedings of the 5th AIAA/ASME Thermophysics and Heat Transfer Conference*, pages 27–34, 1990.

[3] A. G. Hauptmann and P. McAvinney. Gesture with speech for graphics manipulation. *International Journal of Man-Machine Studies*, 38(2):231–249, Feb. 1993.

[4] W. F. Humphrey, A. Dalke, and K. Schulten. VMD – Visual molecular dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[5] M. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. Skeel, K. Schulten, and R. Kufrin. MDScope – a visual computing environment for structural biology. *Comput. Phys. Commun.*, 91(1, 2 and 3):111–134, 1995.

[6] M. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. D. Skeel, and K. Schulten. NAMD— A parallel, object-oriented molecular dynamics program. *J. Supercomputing App.*, 10:251–268, 1996.

[7] V. I. Pavlović, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), July 1997.

[8] J. Rohlicek, W. Russell, S. Roukos, and H. Gish. Continuous hidden markov modeling for speaker-independent word spotting. In *Proc. ICASSP*, pages 627–630, 1989.

[9] R. Rose and D. Paul. A hidden markov model based keyword recognition system. In *Proc. ICASSP*, pages 129–132, 1990.

[10] R. Sharma, T. S. Huang, and V. I. Pavlović. A multimodal framework for interacting with virtual environments. In C. A. Ntuen and E. H. Park, editors, *Human Interaction with Complex Systems*, pages 53–71. Kluwer Academic Publishers, 1996.

[11] M. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70:920–930, 1980.

[12] J. Wilpon, L. Rabiner, C. H. Lee, and E. R. Goldman. Automatic recognition of keywords in unconstrained speech using hidden markov models. *IEEE Trans. on ASSP*, 38(11):1870–1878, 1990.

Speech/Gesture Interface to a Visual-Computing Environment

Rajeev Sharma
Pennsylvania State University

Michael Zeller
University of Southern California

Vladimir I. Pavlovic, Thomas S. Huang, Zion Lo, Stephen Chu, Yunxin Zhao, James C. Phillips, and Klaus Schulten
University of Illinois, Urbana-Champaign

Although recent years have witnessed tremendous progress in 3D immersive display and virtual reality, the corresponding human-computer interaction (HCI) technologies have lagged behind. Current interfaces involve heavy headsets, datagloves, tethers, and other virtual reality devices, which may deter or distract the user. To fully exploit VR's potential for visualizing and interacting with complex information, users must be able to interact with the virtual display in a more natural way (see the "Hand gestures" sidebar on the next page).

In this article, we describe a bimodal speech/gesture interface, which we have integrated in a 3D visual-computing environment used by structural biologists. This interface lets researchers interact with 3D graphical objects in a virtual environment using spoken words and simple, free hand gestures. The reason we used a particular virtual environment context was to set the necessary constraints to make our analysis robust and to develop a command language that optimally combines speech and gesture inputs. Our interface uses

- automatic speech recognition (ASR), aided by a microphone, to recognize voice commands;
- two strategically positioned cameras to detect hand gestures; and
- automatic gesture recognition (AGR), a set of computer-vision techniques, to interpret those hand gestures.

The computer vision algorithms can extract the user's hand from the background, detect different finger positions, and distinguish meaningful gestures from unintentional hand movements.

Our main goal was to simplify model manipulation and rendering to make biomolecular modeling more playful. Researchers can explore variations of their model and concentrate on biomolecular aspects of their task without undue distraction by computational aspects. They can view simulations of molecular dynamics, play with different combinations of molecular structures, and better understand the molecules' important properties. A potential benefit, for example, might be reducing the time to discover new compounds for new drugs.

Virtual environment testbed

The Theoretical Biophysics Group at the University of Illinois, Urbana-Champaign, built the virtual environment we are using: MDScope. MDScope is a set of integrated software components that allows simulation and visualization of biomolecular systems in structural biology. As Figure 1 (next page) shows, three separate packages, which may be used individually or together, constitute the MDScope environment:[1]

We developed a speech/gesture interface that uses visual hand-gesture analysis and speech recognition to control a 3D display in VMD, a virtual environment for structural biology.

- The NAMD (Numerical Analysis Molecular Dynamics) program, a molecular-dynamics program that runs in parallel on various architectures and operating systems.
- The VMD (Visual Molecular Dynamics) program, a molecular-visualization program that displays both static molecular structures and dynamic molecular motion, as computed by programs such as NAMD.
- The MDComm (Molecular Dynamics Communication) software, which provides an efficient means of communication between VMD and NAMD and lets VMD act as a graphical user interface to NAMD. Using NAMD as a computational engine, VMD uses MDComm to provide an interface for interactive setup and display of a molecular-dynamics simulations on a remote supercomputer or high-performance workstation.

The NAMD program

Molecular-dynamics calculations are computationally very expensive and require large amounts of memory to store the molecular structure, coordinates, and atom-to-atom interaction lists. The challenge is to

0272-1716/00/$10.00 © 2000 IEEE

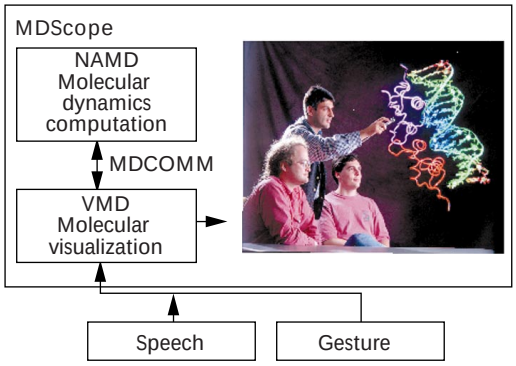
Hand Gestures

The communication mode that seems most relevant for manipulating physical objects is hand motion, also called hand gestures. People use this mode to act on the world, to grasp and explore objects, and to express ideas. Virtual objects, unlike physical objects, are under computer control. Psychological experiments indicate that people prefer to use speech and hand gestures in a virtual environment so that they can interact without special training or special apparatus and concentrate more on the virtual objects and the tasks at hand.[1] Several experimental systems study different aspects of this multimodality. One of the earliest was the "Put-That-There" system,[2] where a frame-based integration architecture fused spoken input and magnetically tracked 3D hand gestures. More recently, the QuickSet system integrated voice input with pen-based gestures to control military simulations.[3] Other systems include VisualMan,[4] Virtual-World, Artificial Life Interactive Video Environment,[5] and Smart Rooms.[6]

To interact naturally in a VR environment, users need as few devices attached to them as possible. However, most demonstrated gesture/speech systems use some hand-held device or instrumented glove, which is often tethered with cables that connect the device to a computer. This hinders user ease in interacting with the computer-controlled environment. A highly specialized application domain, such as simulation of surgery in a VR environment, might justify using such devices, but these cumbersome interface tools deter the everyday user. This problem has spawned active research toward video-based, noncontact gesture analysis that uses video cameras and computer vision to interpret gestures. Despite some progress in vision-based gesture recognition,[7] few systems integrate gestures in a working application.

References

1. A.G. Hauptmann and P. McAvinney, "Gesture with Speech for Graphics Manipulation," *Int'l J. Man-Machine Studies*, Vol. 38, No. 2, Feb. 1993, pp. 231-249.
2. R.A. Bolt, "Put-That-There: Voice and Gesture at the Graphics Interface," *ACM Computer Graphics*, Vol. 14, No. 3, 1980, pp. 262-270.
3. P.R. Cohen et al., "QuickSet: Multimodal Interaction for Distributed Applications," *Proc. Fifth ACM Int'l Multimedia Conf.*, ACM Press, New York, 1997, pp. 31-40.
4. J. Wang, "Integration of Eye-Gaze, Voice and Manual Response in Multimodal User Interface," *Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics*, IEEE Press, Piscataway, N.J., 1995, pp. 3938-3942.
5. P. Maes et al., "ALIVE: Artificial Life Interactive Video Environment," *Intercommunication*, Vol. 7, Winter 1999, pp. 48-49.
6. A. Pentland, "Smart Rooms," *Scientific American*, Apr. 1996, pp. 54-62.
7. V.I. Pavlovic, R. Sharma, and T.S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July 1997, pp. 677-695.



1 The VMD program, coupled with the NAMD program via MDComm. These facilities comprise MDScope, a problem-solving environment for structural biology. External speech/gesture user-interface components can control all aspects of VMD through a communications layer. On the right, three users discuss and manipulate a 3D structure of DNA.

efficiently calculate interatomic forces using high-performance computing. NAMD meets this challenge by using parallel computation and incorporating the Distributed Parallel Multipole Tree Algorithm (DPMTA). NAMD uses a *spatial-decomposition* algorithm to partition the task of computing the force on each atom among several processors. This algorithm subdivides the space occupied by the molecule into uniform cubes (or *patches*), as shown in Figure 2. The algorithm then distributes those patches among a parallel computer's processors. The processor to which each patch is assigned computes the motion of the atoms in that patch. As atoms move, they are transferred between patches, and patches are reassigned to different processors to maintain a uniform computational load.

The VMD program

The key functions of VMD are to visualize biomolecular systems, allow direct interaction between a user and a molecule being simulated on another computer, and provide an intuitive user interface for controlling the visual display and remote simulation. VMD provides various methods for rendering and coloring the structure, such as simple lines, solid bonds, and ribbon diagrams (see Figure 3). VMD uses the MDComm software to initiate, display, and control a simulation using NAMD. When NAMD calculates a molecular system's trajectory, it sends each atom's coordinates to VMD. Current net-

work technology provides the necessary bandwidth to communicate the atomic coordinate data; a high-performance dynamics program is crucial for furnishing new data at the speed required for interactive display.

VMD implements many different forms of user interfaces; users may control the program through keyboard commands, a mouse, or a graphical user interface. These controls let users modify the appearance of the molecules and display, control the display of the molecules' structural features, and access remote computers running molecular-dynamics simulations. Users can view multiple structures simultaneously and, because of a flexible atom-selection mechanism, easily select subsets of atoms for display.

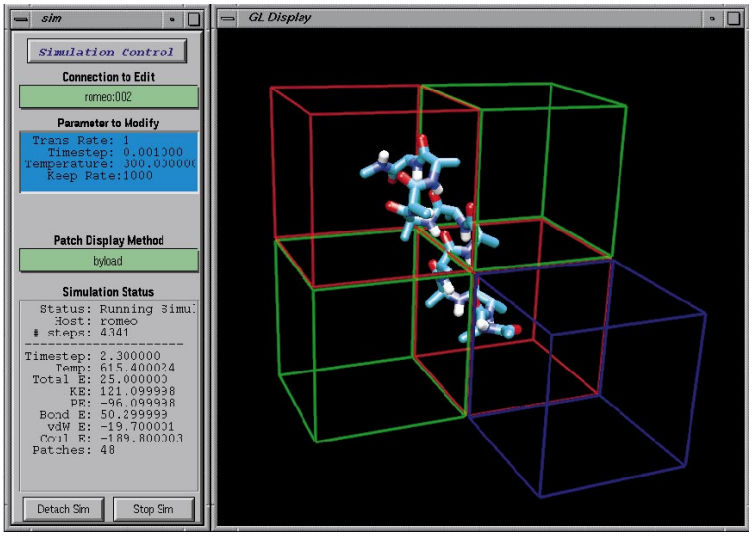
VMD includes an extensive text-command processing capability from the Tcl library, a popular and widely available package for script parsing and interpreting. Tcl lets users write scripts with features such as variable substitution, control loops, and function calls. Users can also save a molecule's current rendering in an image file or in a format suitable for use by several image-processing packages. By connecting directly to a remote computer running a molecular-dynamics simulation, VMD offers users the capability to interactively participate in an ongoing simulation—for example, the option to apply perturbational forces to individual atoms. VMD also implements a mechanism for external programs to serve as user-interface components by letting these programs communicate with VMD through standard network-communication channels. This makes possible new user-interface methods, such as the speech- and gesture-recognition systems discussed later, to be developed in parallel with VMD.

MDScope

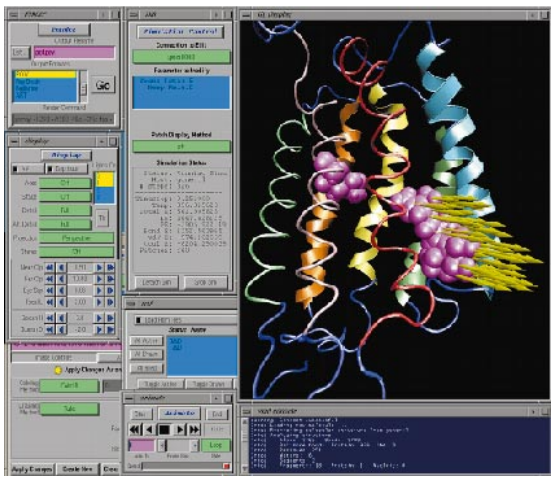
Development of MDScope is an ongoing project. Source code for MDScope applications are available via anonymous ftp at ftp.ks.uiuc.edu or on the Web at http://www.ks.uiuc.edu. This Web site also includes extensive documentation describing how to use and modify the programs. MDScope is available for various architectures, operating systems, and OpenGL-based workstations.

The primary infrastructure for our project is a large-screen stereographic projection facility, developed in the Theoretical Biophysics Group at the University of Illinois and shared by many biomedical researchers. The facility, designed for groups of up to 10 people, uses cost-effective, space-saving display hardware, which is added to a high-end graphics workstation and can be easily duplicated at other sites. It produces 8×6×6-foot 3D models in a 120-square-foot area. The facility includes a projector that displays alternating left- and right-eye views onto the screen at nearly twice the rate of ordinary projectors. The images, when viewed through special eyewear, produce a stereo display.

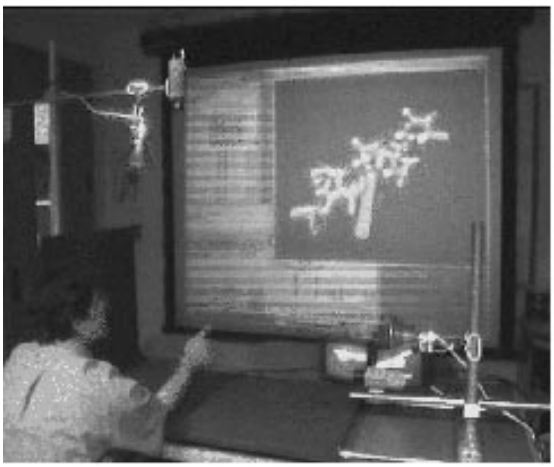
The primary VMD interfaces used by researchers are a keyboard and a magnetically tracked pointer. This is inconvenient because the system is typically used by multiple (six to eight) users, and the interface hinders the visualization system's interactive nature. Speech and hand gestures, on the other hand, are fundamental, natural methods of human communication; their use for interaction with and control of the display of VMD would greatly improve the program's utility. Thus, incorporating voice-command control in MDScope would free users of keyboards so that they could interact with the environment in a natural manner. The hand gestures would permit the users to easily manipulate the displayed model and play with different spatial combina-



2 VMD visualizing the spatial decomposition in NAMD. VMD displays a molecular-dynamics simulation of a small polypeptide computed by the program NAMD on a remote workstation and shown on a local graphics workstation. The form on the left gives information about the simulation state and controls for modifying the simulation parameters. The colored boxes surrounding the protein indicate NAMD's spatial decomposition. Each box's color indicates the relative amount of CPU time required to compute the motion of the atoms in that region (the red end of the color spectrum denotes a greater CPU time), thus providing a way of visualizing the distribution of computational load.



3 A VMD session demonstrating the graphical user interface and the text console. Molecules appear in the display window shown in the upper right corner; GUI components are on the left. Below the graphics library display window is the text console interface for VMD.



4 The experimental setup with two cameras used for gesture recognition.

(a) (b)

5 A sample pair of images from the two cameras used for automatic speech recognition: (a) top camera; (b) side camera.

tions of the molecular structures. Integrating speech and hand gestures as a multimodal interaction mechanism would be more powerful than using either mode alone, thus motivating the development of the speech/gesture interface.

Software

Our goal was to minimize the modifications needed to the existing VMD program for incorporating the new interface. The experimental prototypes that we built for both the speech and hand-gesture analysis required an addition to the VMD environment. To reduce the complexity and increase the flexibility of the program design, we added a communications layer so that external programs could be written and maintained independently of the VMD code. These programs use the VMD text language to query VMD for information or to send new commands. The VMD text language is based on the Tcl scripting language. Because all VMD capabilities are available at the script level, an external program can control VMD in any way. Both the ASR and AGR programs interact with VMD using this method. For a simple voice command, such as "rotate left 90," the ASR converts the phrase into the VMD text command "rotate y 90" and sends it to VMD. Similarly, when the AGR is used as a pointing device, it sends the commands to change the current position and vector of VMD's graphical 3D pointers.

Setup for visual gesture analysis

To facilitate the development of AGR algorithms, we designed an experimental platform (see Figure 4) for gesture-recognition experiments. Within the uniformly black background, a lighting arrangement shines red light on the hand without distracting the user from the main 3D display. The red light helps to localize the hand in video and track it robustly. An alternative would be to track the hand using skin-color segmentation or motion and shape information.[2] However, for the visualization setup, the ambient light is quite low, which makes tracking more difficult without the help of the additional lighting.

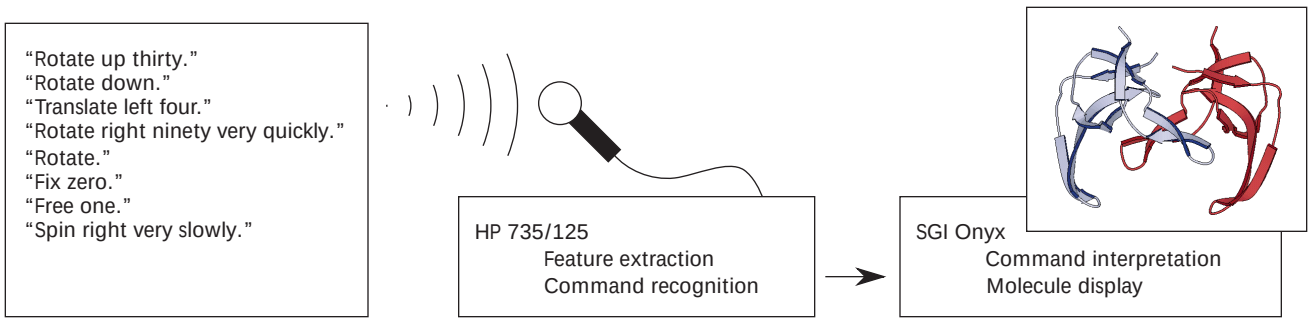
Figure 5 shows a sample pair of images from the two cameras. This setup lets a user sit at the table and use hand gestures to control the graphics display. Moreover, the setup can be transported easily and is relatively unobtrusive.

Setup for speech analysis

We have implemented a prototype ASR system and integrated it into VMD. The system has two blocks: the recorder front-end and the recognizer unit (see Figure 6). The recorder uses circularly buffered memory to implement its recording duties, sending its output to the recognizer unit in blocks. A digital volume meter accompanies this to give the user feedback by indicating an acceptable loudness range. We developed the recognizer using Hidden Markov Models. This unit performed feature extraction and time-synchronous Viterbi decoding on the input blocks, sending the decoded speech directly to the SGI Onyx workstation where the VMD process resides.

Speech/gesture command language

To effectively use the information the user inputs through spoken words and simple hand gestures, we designed a command language for MDScope that combines speech with gesture. This command language uses the basic syntax of `<action> <object> <modifier>`. The `<action>` component is spoken (for example,



6 Setup for the experiments on speech recognition for controlling the virtual display.

"rotate"); a combination of speech and gesture specify `<object>` and `<modifier>`. An example of this basic syntax would be speaking "this" while pointing, followed by a modifier to clarify what is being pointed to, such as "molecule," "helix," or "atom," and then speaking "done" after moving the hand according to the desired motion.

Another example of the desired speech/gesture capability is the voice command "engage," which would query VMD for the molecule nearest the pointer tip, make the molecule blink to indicate that it was selected, and save a reference to that molecule for future use. Once engaged, the voice command "rotate" would convert the gesture commands into rotations of the chosen molecule, and the command "translate" would convert them into translations. When finished, the command "release" would deselect the molecule and let the user manipulate another molecule.

Speech input using ASR

In integrating speech and gesture in the MDScope environment, we needed a real-time decoding of the user's commands to keep pace with the hand gestures. Thus, we needed *word spotting*, which means detecting a given vocabulary of words embedded in unconstrained continuous speech. Word spotting differs from conventional large-vocabulary continuous speech recognition systems, which seek to determine an optimal sequence of words from a prescribed vocabulary. A direct mapping between spoken utterances and the recognizer's vocabulary is implied with continuous speech recognition, leaving no room for the accommodation of nonvocabulary words in the form of extraneous speech or unintended background noise. Real-world applications dictate the basis for word spotting (also termed keyword spotting). Users of a spoken-language system often embellish their commands with supporting phrases and sometimes even issue conversation absent of valid commands. In response to such natural-language dialogue and the implications for robust human-computer interaction, we converted standard continuous-speech recognition systems into spotters by simply adding filler or garbage models to their vocabulary. The recognition output stream then contains a sequence of keywords and fillers constrained by a simple syntactical network. In other words, recognizers operated in a spotter mode.

Although early techniques emphasized a template-based dynamic time-warping slant, current approaches typically wield the statistical clout of hidden Markov models (HMMs), and recent ones come with the discriminatory abilities of neural networks. Typically word-based, they use an overall network that places the keyword models in parallel with the garbage models.

Keywords

Table 1 lists the keywords and their phonetic transcriptions chosen for our system. These commands let the VMD user manipulate the molecules and polymeric structures selected by hand gestures. In modeling the speech acoustics, we based the HMM system on phonemes rather than words, for large vocabulary flexibility in the given biophysical environment. Although implementing a word-based system would invariably be easier, retraining it would be inconvenient if the vocabulary changed.

Table 1. Keywords and phonetic descriptions for our system.

Keyword	Transcription
translate	t-r-ae-n-s-l-ey-t
rotate	r-ow-t-ey-t
engage	eh-n-g-ey-jh
release	r-ih-l-iy-s
pick	p-ih-k

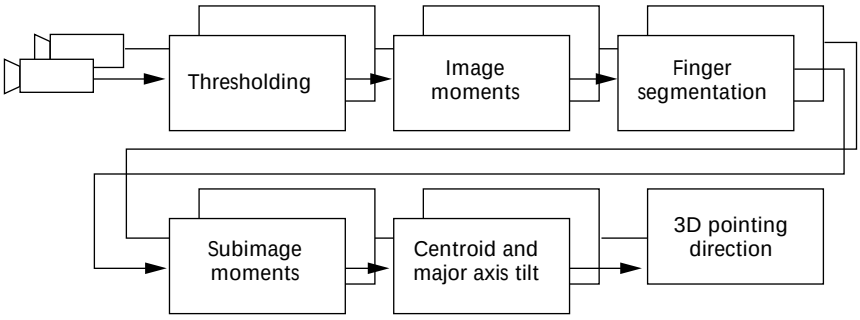
Table 2. Broad sound classes used as garbage models.

Sound Class	Symbol
Vowels-front	vf
Vowels-mid	vm
Vowels-back	vb
Dipthongs	dipth
Semivowels-liquids	svl
Semivowels-glides	svg
Consonants-nasals	cn
Consonants-stops-voiced	csv
Consonants-stops-unvoiced	csu
Consonants-fricatives-voiced	cfv
Consonants-fricatives-unvoiced	cfu
Consonants-whispers	cw

Fillers

Filler models are more varied. In large-vocabulary continuous-speech recognition applications, the non-keyword portion of the vocabulary can represent these fillers explicitly—for example, as whole words. In other tasks, non-keywords are built by a parallel combination of either keyword pieces or phonemes, whether they be context-independent monophones or context-dependent triphones or diphones.

We used 12 fillers (garbage models) to model extraneous speech in our experiment. Rather than using monophones or states of keyword models (as researchers have used in prior experiments), we used models that cover broad classes of basic sounds found in American English (listed in Table 2). Such models adequately cover the English language and are amenable to training. However, we modified these models in two ways. First, we did not use the class of "consonants-africates," because they don't occur frequently in the prescribed vocabulary or the training data. As observed by many researchers, varying or increasing the number of models does not gain much in spotting performance.[3] Second, we included a model for background silence in addition to the 12 garbage models listed. Such a model removes the need for an explicit endpoint detector by modeling the interword pauses in the incoming signal. Also, the descriptors for the vowel class correspond to the tongue hump's position in producing the vowel.



7 An overview of the main steps involved in hand-gesture recognition.

Recognition network

The recognition syntactical network placed the keywords in parallel to garbage models that included a model for silence. These models followed a null grammar, meaning every model may precede or succeed any other model.

Features and training

After sampling speech at 16 kHz and filtering to prevent aliasing, we preemphasized the speech samples with a first-order digital filter using a preemphasis factor of 0.97. We blocked these samples into frames of 25 ms with a shift between frames of 10 ms. We used a Hamming window to weight each speech frame; then we derived 16th-order mel-frequency cepstral coefficients and weighted them with a liftering factor of 22. We chose cepstral coefficients because they are more robust and discriminative than linear-predictive-coding coefficients or log-area-ratio coefficients. We also included normalized log-energy and first-order temporal-regression coefficients in the feature vector.

The HMMs' topology for both keyword phonemes and garbage models had five states, the three internal states being emitting states. Following a left-to-right traversal, each state was described by a mixture of five continuous-density Gaussians with diagonal covariance matrices. We used three iterations of the Baum-Welch reestimation procedure for training.

In training the 15 keywords, we developed 40 sentences. We individually paired each of the five keywords with the remaining four. We then doubled this pairing to provide a sufficient number of training tokens. Thus, the sentences contained keyword pairs such as "engage translate" and "rotate pick," which were arranged so that each keyword could be spoken 16 times. Each VMD user proceeded with this short recording session.

Training the garbage models required a far more extensive database of training sentences to provide an adequate amount of training data. The reason is that the 12 broad classes cover nearly the entire spectrum of the standard 48 phonemes. Subsequently, we used the Timit database to provide an initial set of bootstrapped models. We then conducted retraining for a VMD user who had recorded a set of 720 sentences of commonly used VMD commands. These sentences spanned the scope of the VMD diction, including a more detailed set of commands, numbers, and modifiers. This was necessary to normalize the data to the existing computational environment.

We trained the garbage models only once for this experiment. Hence, VMD users only needed to go through the short training procedure just explained.

Hand-gesture input using AGR

The general AGR problem is hard because it involves analyzing the human hand, which has a very high degree of freedom, and because the use of hand gestures is not well understood. However, we used the context of the particular virtual environment to develop an appropriate set of gestural commands. Our interface recognizes gestures by analyzing the sequence of images from a pair of cameras positioned so that they facilitate robust analysis of the hand images. The background is uniformly black to further help with the real-time analysis without using any specialized image-processing hardware.

Finger as a 3D pointer

The AGR system has two subsystem levels (see Figure 7). The first-level subsystem extracts a 2D pointing direction from single-camera images. The second-level subsystem combines the information obtained from the outputs of the first-level subsystems into a 3D pointing direction. To obtain the 2D pointing direction, the first-level subsystem performs a sequence of operations on the input image data. It first gives the gray-level image a threshold to extract a silhouette of the user's lower arm from the background. Next, the system calculates first- and second-image moments and uses them to form a bounding box to extract the index finger. Once the finger is segmented from the hand, another set of image moments is calculated, this time for the finger itself. Finally, based on these moments, the system determines the 2D finger centroid and finger direction. Finally, the system determines the 3D pointing direction of the hand, using the knowledge of the setup geometry and the 2D centroids. The system then forwards this information to the central-display manager, which displays a cursor at an appropriate screen position.

Gestures for manipulating 3D display

We also developed an AGR system based on HMMs to recognize basic manipulative hand gestures. We categorize these gesture commands as either dynamic (move back or move forward) or static (grab, release, stop, up, or down). The gestures used were mainly for manipulating and controlling the molecular structures' display (see Figure 8). We derived this gesture set by observing the MDScope system's functionality and the

corresponding command set used for interacting with it. In this way, we incorporated the gestures into the system without changing the existing command set.

The system uses image-geometry parameters as the features that describe any particular hand posture (static hand image). We use an image's *radon transform* to extract these features. The radon transform of the image $I(x, y)$ is defined as

$$R(\theta, t) = \int_{(x,y)} I(t\cos\theta - s\sin\theta,\ t\sin\theta + s\cos\theta)ds$$

where $0 \le \theta \le \pi/2$. The image-geometry moment of order k is then given by

$$m^{(k)}(\theta) = \int_t t^k R_0(\theta,\ t)dt$$

where R_0 denotes the radon transform, normalized with respect to the image mass:

$$R_0(\theta, t) = \frac{R(\theta, t)}{\int_v R(\theta, v)dv}$$

The first-order moments constitute the image's center of mass. The higher-order moments provide additional information on image shape. We built the recognition system by training HMMs for the specific gestures on example runs. We modeled each gesture in the vocabulary as a single four-state HMM. We modeled the observations using a Gaussian mixture of two different sizes (one and three) with a diagonal covariance matrix.

Conclusions

Several researchers tested and evaluated the speech/gesture interface. Overall, new users needed only a few minutes to get acquainted with the setup. We asked the researchers to perform a generic task that involved selecting, moving, rotating, and releasing a certain part of a given molecular structure. This task would otherwise involve several keyboard commands or a combination of keyboard and magnetic-tracker input. Every user reported that working with the speech/gesture interface was more convenient and, in most cases, more efficient than the traditional interaction with VMD.

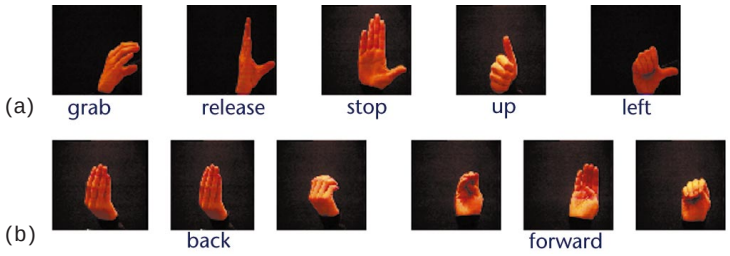
Upon testing the speech system as a whole with 50 test sentences that embedded the keywords within bodies of nonkeywords, we found a word-spotting accuracy of 98 percent on the trained speaker. The trained speaker refers to each user who trained the keywords, regardless of the one who trained the garbage models. The VMD users considered this very good for the given biophysics environment, thus supporting the techniques used. In general, false alarms occurred only for those situations where the user embedded a valid keyword within another word. For example, if someone says "translation" instead of "translate," the spotter still recognizes the command as "translate."

Because of the strict grammar imposed on the allowed speech/gesture commands (which essentially makes the multimodal interaction sequential), the relative timing between the spoken words and hand gestures does not create problems in the interpretation.

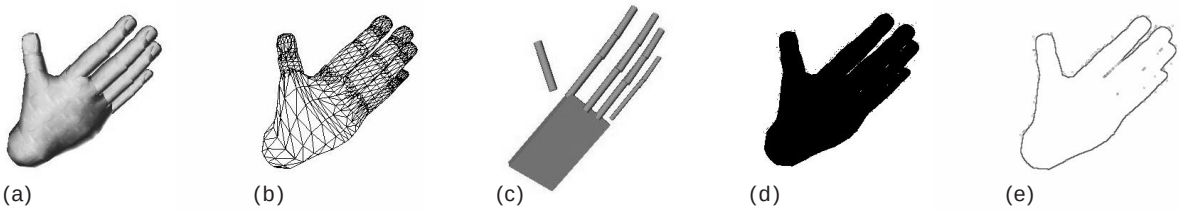
The speech/gesture interface reported could be part of a more general multimodal framework, where other modalities such as gaze could also be exploited to make the interface more natural and efficient.

Using the simple setup with a uniform background, we segmented the hand image from the background in real time. This didn't require any restriction on the user other than being seated at the table, which is quite appropriate for the MDScope system. However, in a more general VR setting, we might need better segmentation techniques. The hand segmentation and corresponding motion analysis can benefit from the other modalities mentioned.

The experimental results for gesture recognition show that, even with simple image moments, the HMM-based approach yields a useful working system. However, a model-based approach can significantly affect recognition performance. For example, there is a trade-off between the reliability and speed of gesture recognition for different levels of the hand model used (see Figure 9).[2] One approach for AGR hand-motion analysis is to



8 Examples of images from hand gestures used to manipulate a virtual object, and interpreted using AGR: (a) static gestures; (b) dynamic gestures.



9 Hand models of varying complexity: (a) 3D textured volumetric model; (b) 3D wire-frame volumetric model; (c) 3D skeletal model; (d) binary silhouette; (e) contour.

consider the motion class called articulated motion for analysis and tracking of the hand. Using the prediction based on articulated motion analysis, we can reliably derive a minimal description of the hand image in real time. The more detailed the hand mode, the better the prediction that can be made of the hand positions under different gestures. Such models can be the basis for developing a suitable feature vector for gesture classification and recognition.[4] The aim would be to replace the simple image moments in our current implementation with a feature vector that can define the more complicated hand gestures needed for manipulating a virtual environment.

Our focus on structural biology takes advantage of existing sophisticated software, provides concrete objectives, defines a well-posed task domain, and offers a well-developed vocabulary for spoken communication. Therefore, the VR setup we considered as a testbed for developing the multimodal interface facilitates the definition of gesture- and speech-recognition problems. Our prototype speech/gesture interface lets scientists easily and naturally explore the displayed information. Integrating speech and hand gestures as a multimodal interaction mechanism has proven to be more powerful than using either mode alone.

From a structural biologist's view, a robust gesture/speech interface for a molecular-graphics framework is useful for several reasons. First, it eliminates typing commands, which would require knowledge of the correct spelling and syntax. Pointing to a structure and saying "rotate this helix 75 left" is easier than entering the command using a keyboard, a mouse, menus, or a 3D tracker. Second, this novel interface simplifies accessing complex molecular-graphics programs for the novice or casual user. It also lets experienced users achieve many tasks in less time while focusing on the scientific merit of the modeling. Third, the gesture/speech interface is particularly useful for teaching and lecturing because it provides a far more natural way for presenting information than typing commands. Currently, because of the tedious task of controlling the VR display, a second person often accompanies the lecturer to operate the molecular-graphics program. Finally, the interface could help in the preparation of figures for publication. It can potentially provide a quicker way to explore different aspects of the model, select the most informative orientation of a biological structure, or simply add color and shading to the model to highlight specific features.

The combination of high-performance computing and high-end graphics for research in structural biology will open new avenues for very large-scale biomolecular modeling. A robust speech/gesture interface will offer a new level of interactive visualization not possible before. ■

Acknowledgments

We gratefully acknowledge the financial support of the National Science Foundation (grants IIS-97-33644, IRI-89-08255, BIR-9318159, IRI-95-02074, IRI-96-34618); the US Army Research Laboratory (Cooperative Agreement No. DAAL01-96-2-0003); Sumitomo Electric Industries; the National Institutes of Health (PHS 5 P41 RR05969-04); and the Roy J. Carver Charitable Trust.

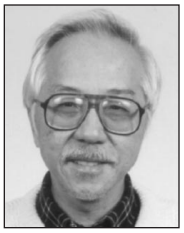
References

1. M. Nelson et al., "MDScope—A Visual Computing Environment for Structural Biology," *Computational Physics Communication*, Vol. 91, Nos. 1–3, Jan. 1995, pp. 111-134.
2. V.I. Pavlovic, R. Sharma, and T.S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July 1997, pp. 677-695.
3. J.G. Wilpon et al., "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models," *IEEE Trans. ASSP*, Vol. 38, No. 11, Nov. 1990, pp.1870-1878.
4. J.M. Rehg and T. Kanade, *Digiteyes: Vision-Based Human Hand Tracking,*" Tech. Report CMU-CS-93-220, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pa.,1993.

Rajeev Sharma *is an assistant professor in the Department of Computer Science and Engineering at Pennsylvania State University, University Park. He received a PhD in computer science from the University of Maryland, College Park.*

Michael Zeller *is Director of Systems Engineering at H&F Aeronautical Technologies, in San Diego. He received his Diplom degree and PhD in physics from the Johann Wolfgang Goethe University in Frankfurt, Germany.*

Vladimir Pavlovic *is a doctoral student in electrical engineering at the Department of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign.*



Thomas S. Huang *is the William L. Everett Distinguished Professor of Electrical and Computer Engineering at the University of Illinois, Urbana-Champaign, and a research professor at the Coordinated Science Laboratory. He also heads the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology. He received his ScD in electrical engineering from the Massachusetts Institute of Technology.*

Zion Lo *received an MS in electrical and computer engineering from the University of Illinois, Urbana-Champaign.*

Stephen Chu *is a PhD candidate in the Electrical and Computer Engineering Department at the University of Illinois, Urbana-Champaign, where he also received an MS in electrical engineering.*

Yunxin Zhao *is an associate professor in the Department of Computer Engineering and Computer Science at the University of Missouri-Columbia. She received her PhD from the University of Washington, Seattle.*

James C. Phillips *is studying for his doctorate in physics at the University of Illinois, Urbana-Champaign, where he also received an MS in physics.*

Klaus J. Schulten *is a professor in the Physics Department at the University of Illinois, Urbana-Champaign. He received his Diplom degree in physics from the University of Munster, Germany, and his PhD in chemical physics from Harvard University.*

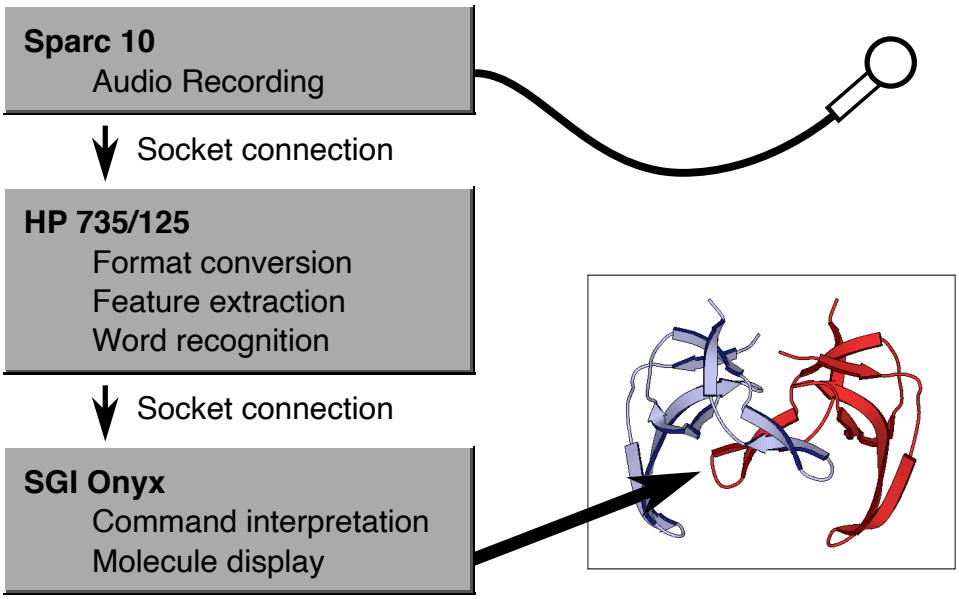
Readers may reach the authors in care of Klaus Schulten, Beckman Inst., 405 N. Matthews Ave., Univ. of Illinois, Urbana, IL 16801; kschulte@ks.uiuc.edu; or Rajeev Sharma, Dept. of Computer Science and Engineering, 220 Pond Lab, Pennsylvania State Univ., University Park, PA 16802; rsharma@cse.psu.edu.

19Jan2000

PROTOTYPE SPEECH RECOGNITION INTERFACE FOR VMD

by Jim Phillips and Zion Lo

The prototype speech recognition interface for VMD was partially and temporarily implemented using the speech recognition software (HTK) from Entropic. Recording of discrete command statements was done on a Sparc 10 in the 3D facility using a noise-cancelling headset microphone. This data was then transferred via a socket connection to a remote computer (either a Sparc 20 or an HP 735) for processing. There audio format conversion, feature extraction, and continuous speech recognition via a hidden Markov model were performed and the resulting sentence was transferred again via a socket connection to the graphics workstation (an SGI Onyx) in the 3D facility. The spoken sentence was then interpreted into commands suitable for the VMD text interface and communicated to VMD via a pvm (portable virtual machine, a message-passing interface for parallel programming) message.



The command set for the prototype speech interface was designed to require purely one-way communication between programs. The interpreter did not employ information on the internal state of VMD. Although the interpreter was capable of providing feedback to the recognition system on a word-by-word basis, this information was not employed. The recognition system contained its own information on the grammatical structure of the command set and thus only gramatically correct commands were passed to the interpreter. This one-way communication placed great restraints on the capabilities of the system. Two types of commands were avaiable. Global manipulation commands (spin, rotate, translate, scale, stop, reset) affected the entire scene and had optional modifiers for direction (left, up, ...), amount (ten, fifty, ...), and rate (slowly, very quickly, ...) where appropriate. Defaults for these modifiers were in most cases dynamic, being based on their value in the last command. For example, after instructing the computer to "rotate left ninety very slowly" the simple command "rotate" would repeat the procedure and "rotate right" would reverse it. The remainder of the commands (hide, show, fix, unfix, activate, inactivate) applied to specific molecules, identified by number, and defaulted to the "top" molecule, set by the "top" command. This allowed the user to manipulate some molecules independently of others.

The accuracy of the speech recognition system was excellent and remarkably speaker independent, even when trained on only one person, although problems with electrical noise persist. Its performance was respectable, with a lag time of two seconds on a Sparc 20 or on an HP 735. In the future this time could be reduced even further for several reasons. First, the hidden Markov model employed searches for a best match for a sentence from the entire grammar. This is obviously a lengthly process and will only grow as commands become more complex. In the future, less ambitious systems based on words rather than sentences will be needed. The set of acceptable words at some point in a sentence could be further reduced by information from the interpreter, but would only be based on prior words rather than later ones. Second, the current implementation is based on six separate programs. Two of these programs were called repetitively from inside a single instance of the recognition program, eliminating the most expensive part of the initialization. In the future, these programs will be combined into a single, non-commercial program which will run continually. Communication with the recorder and interpreter will be via pvm messages rather than socket connections. Finally, processing of data does not begin until the user finishes speaking, wasting several seconds. To save this time and to facilitate interaction between voice commands and gestural input (i.e., "rotate this (user points) molecule (molecule highlights, user turns hand to indicate desired rotation) done") speech recognition will have to be done "on the fly" as the user is speaking, and lag by at most a single word.

Many changes in the "natural language" interpreter system, in VMD, and in the spoken commands will be necessary to incorporate gestural input. Feedback on objects in the scene will be needed. VMD will need to respond to information such as "the user should be trying to point to something...", "...which he says in an alpha helix" by providing visual feedback during the pointing process (highlighting pointed-to objects) and by reporting the final object's identifier back to the interpreter. VMD will then need to handle the manipulation of the object by the user based on additional information from the interpreter and gestural input. The use of gestural input will reduce the complexity of spoken commands since more information will be communicated visually. The command "grab this (point) molecule (move and rotate hand appropriately) done" for a molecular docking operation is much simpler than attempting to construct the desired motion from "rotate" and "translate" commands. On the other hand (no pun intended), detailed spoken commands should not be disposed of. Instructing the computer to "rotate this (point) molecule and this (point) molecule (wave hands) done" is much less efficient than "rotate the red molecule and the blue molecule up fifty degrees", especially if the user is holding a cup of coffee and simply wants to look at the bottoms of the molecules.

I will now outline the basic elements of a WHAT (Waving Hands And Talking) interface to VMD. Commands will follow a basic syntax of <action> <object> <modifiers> which will be communicated to VMD. The <action> component of the command will be completely spoken (with the possible exception of virtual push-buttons if you want to get fancy). These will be a superset of the current commands. Some will be very similar, such as "spin", "rotate", "translate", and "grab", but more powerful because of object selection and gestural input. Others will be higher level, controling simulations of molecules, such as "equilibrate", "minimize", "monitor", "plot", "run", "pause", and even "merge" or allowing interaction with the simulation, such as "pull", "twist", "heat", "cool", and "restrain".

The <object> component of a command may be specified by a combination of speech and gesture. The word "this" would be a cue for gestural (or other pointer) input and could be followed by specific modifiers to clarify what was being pointed to, such as "molecule", "helix", "residue", "atom", "alanine", "oxygen", etc. When the correct item was highlighted by VMD, based on this information, the user would switch from pointing to a affirmation (thumbs up?) to confirm the selection. Multiple objects could be selected with the words "and" or by using "these" and pointing to several objects in turn. Speech could be incorporated as well, giving a selection along the lines of "...all residues in this (point) molecule except for these (point) hydrophobic ones and this (point)

guanine..." or "...all residues within ten angstroms of here (point)...". Object or sets of objects could even be given (preselected) names to make spoken references to them easier than remembering artificial contructs like molecule numbers. For example, "name this (point) molecule bill, spin bill". The exact synax will have to be well thought out to avoid ambiguities.

The <modifiers> portion of the command would also incorporate a mix of speech and gesture, highly dependent on the specific action desired. The translate command could have spoken modifiers similar to the present implementation ("left twenty degrees very quickly"), pure gestural input (moving hand to indicate desired motion, followed by "done"), or a combination of the two ("to here (point) slowly", "in this direction (move hand) five angstroms").

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

MOLECULAR DYNAMICS STUDIES OF THE PROTEIN BACTERIORHODOPSIN

BY

WILLIAM FOWLER HUMPHREY

B.S., Drake University, 1990
M.S., University of Illinois, 1992

THESIS

Submitted in partial fullfilment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1996

Urbana, Illinois

UMI Number: 9712317

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

SEPTEMBER 1996

WE HEREBY RECOMMEND THAT THE THESIS BY

WILLIAM FOWLER HUMPHREY

ENTITLED MOLECULAR DYNAMICS STUDIES OF THE PROTEIN BACTERIORHODOPSIN

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Director of Thesis Research

Head of Department

Committee on Final Examination†

Chairperson

† Required for doctor's degree but not for master's.

O-517

MOLECULAR DYNAMICS STUDIES OF THE PROTEIN BACTERIORHODOPSIN

William Fowler Humphrey, Ph.D.
Department of Physics
University of Illinois at Urbana-Champaign, 1996
Dr. Klaus Schulten, Advisor

Molecular dynamics (MD) simulations are employed to study the structure and function of the protein bacteriorhodopsin (bR), a 26 kD protein which residues in the purple membrane of the bacterium *Halobacterium halobium*. Bacteriorhodopsin undergoes a light-driven cyclic process, which pumps protons across the membrane, in order to maintain a proton gradient necessary for ATP synthesis. The cycle is initiated through a *trans* $\rightarrow$ *cis* isomerization of the chromophore retinal, which is bound to a lysine residue via a protonated Schiff base linkage. The study of bR is facilitated through development of the program VMD for visualization of the simulation results, and the program NAMD for MD calculations on parallel computers. Initially, MD simulations are used to develop a refined three-dimensional structure of the protein, using the experimentally determined electron-microscopy structure of bR as a basis, and to determine equilibrium positions for several water molecules within the protein interior. MD simulations are then used to model the early isomerization reaction events in the bR photocycle, for both the native (wild-type) system and several bR mutants. The simulations reveal the possibility for bR to form two or three unique photoproducts, distinguished by the retinal isomeric state and the orientation of the Schiff base proton relative to nearby water molecules and negatively charged aspartic acids. One particular photoproduct is suggested to lead to successful proton pump activity, while the remaining structures return back to the initial state; this result is supported by simulations of non-functional bR mutants, which do not exhibit formation of the suggested functional photoproduct. The very fast initial retinal photoexcitation and subsequent isomerization reaction are also examined in detail using a combined quantum/classical simulation technique, in which the evolution of the density matrix for the retinal isomerization degree of freedom is computed using the Liouville-von Neumann equation. The simulations result in wild-type bR exhibiting a non-adiabatic crossing between excited states

after 500 fs, while the computed excited-state lifetimes for mutants D85N and D212N are an order of magnitude longer. The results compare well with recent femtosecond spectroscopy data for these systems and demonstrate that the lifetime of the excited state is controlled by the position and slope of the first potential energy surface crossing point.

To my family: Mom, Dad, and Paul

and Heidi

Acknowledgements

There are many people who have aided me tremendously throughout my graduate school career, and I would like to take this opportunity to acknowledge their help. My advisor, Klaus Schulten, has provided guidance and advice during my time in the Theoretical Biophysics Group, and made it possible for me to pursue many interesting projects. I wish to thank Mudi Sheves, who became a second advisor for me during his tenure in our group, for his tremendous patience and suggestions in our studies of bacteriorhodopsin. In addition, Sanjay Kale and Bob Skeel were very helpful in their organization of the development of NAMD. I thank also for their continuous help the people who actually keep the Theoretical Biophysics Group running: Gila Budescu, Marilyn Lynch, and Jo Miller, as well as the very helpful support staff of the Beckman Institute, who have made this a tremendous place to work. During my tenure in the Theoretical Biophysics group, my support was provided by grants from the National Institutes of Health (PHS 5 P41 RR05969-04), the National Science Foundation (BIR-9318159), and the Roy J. Carver Charitable Trust. In addition, I gratefully acknowledge the support of a DOE Computational Science Graduate Fellowship during the period 1993 – 1995.

The development of the program VMD originated with Mike Krogh and Rick Kufrin (when it was still called VRChem); I wish to thank Mike for urging us to use C++ for this project, and I wish to thank Rick for his continual collaboration and advice over the years. The development of the current version of VMD would not have been possible without the hard work and insightful suggestions from Andrew Dalke, my co-author and now project leader. I am grateful for having had the opportunity to work with Andrew and also Mark Nelson in the design and implemention of NAMD, for which Mark, in particular, deserves great praise for his leadership and dedication.

I also thank the many other people who helped make NAMD and VMD reality: Robert Brunner, Attila Gursoy (the older Turkish brother I never had), Jon Leech, and Ken Hamer.

I would like to thank the many members of the Theoretical Biophysics Group with whom I have worked over the past four years. First and foremost, my office mate Tom Bishop has been a tremendous person to work with; Tom was the original (and for some time only) user of VMD, and continually helped with suggestions and bug reports. Feng Zhou, Dong Xu, Helmut Heller, and Andreas Windemuth all helped me learn the ropes of running MD simulations. I wish to thank the many current group members with whom I have worked and who have helped in running, testing, and finding bugs in VMD; these members include Alexander Balaeff, Daniel Barsky, Jeff Canfield, Ivo Hofacker, Xiche Hu, Barry Israelewitz, Dorina Kosztin Ilya Logunov, Hui Lu, Jim Phillips, Benno Pütz, Rama Ratnam, Serguei Stepaniants, Ken Wallace, Willy Wriggers, and Michael Zeller. Ari Shinozaki has my gratitude for skillfully taking over the administration of our group's computers and allowing me more time for actual research. In addition, I would like to thank John Reynders and the many other people in the Advanced Computing Lab at Los Alamos National Laboratory for the opportunity to work with them and for their great friendship.

Finally, I wish to thank all my friends and family members for their support and kindness over the years. In particular, I thank my parents for their help and understanding through a decade of college education, and Heidi Hardner, my companion for many years who I appreciate more than I'll ever be able to express in words.

Contents

Chapter

Chapter 1

Introduction

Since the discovery of the protein bacteriorhodopsin (bR) in the purple membrane of *Halobacterium halobium* by Oesterhelt and Stoeckenius in 1971 [117], a considerable research effort has been carried out in order to uncover the functional role and molecular mechanism of this protein in the life cycle of the host bacterium. Bacteriorhodopsin, in the past 25 years, has been revealed in experimental and theoretical studies as an ideal example of an ion-conducting membrane protein, and, further, as a key protein in one of the simplest photosynthetic organisms in nature [83, 14, 100, 88, 119, 43, 147].

Bacteriorhodopsin has been shown to function as a light-driven proton pump for the purpose of ATP synthesis in *H. halobium*. Under normal aerobic conditions, the bacterium exhibits a reddish color and employs oxidative phosphorylation to synthesize ATP molecules. However, when *H. halobium* is placed in an oxygen-free environment, the bacterium expresses bR molecules for the purpose of utilizing energy from visible light photons for ATP synthesis, i.e., for photophosphorylation [118]. To accomplish this function, bR spans the bacterial cell membrane and undergoes a cyclic reaction driven by absorption of a 568-nm photon to transfer protons from the cell cytoplasm to the cell exterior. A 100 mV transmembrane electric field provides an inwardly-directed current of protons which are utilized by ATPase molecules within the cell to convert ADP to ATP; bR pumps protons outwardly against this electric field, and helps maintain the cell membrane proton gradient (see Fig. 1.1). Membrane patches containing bR exhibit a characteristic purple color, which, along with the name bacteriorhodopsin, arises from the use of a retinal chromophore covalently bound to a lysine residue for the photoreactivity of bR. Retinal, shown in Fig. 1.2, is also the chromophore

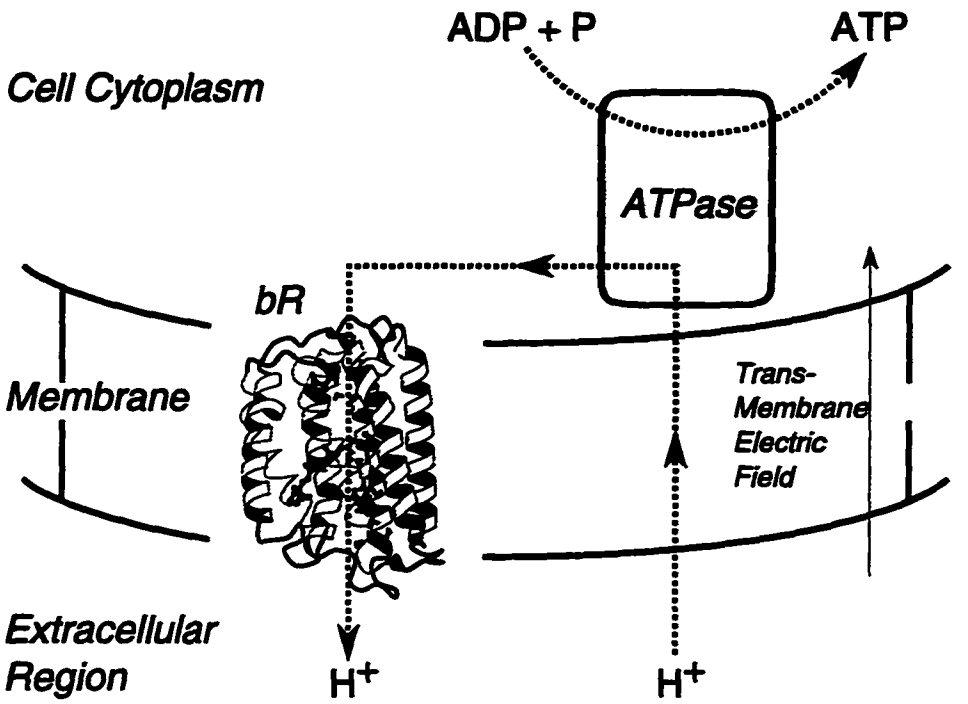


Figure 1.1: Schematic view of proton transport in *H. halobium*. Protons diffuse through the cell membrane and are used by ATPase for ATP production; bR pumps protons out of the cell against an electric field in order to maintain a potential energy gradient.

found in rhodopsin, a protein similar to bR found in photoreceptor cells of the vertebrate eye [117].

Despite its relatively small size (about 4,000 atoms, with a molecular weight of 26 kilodaltons), bR continues to be a subject of extensive study, for both experimental and theoretical research groups interested in discovering the key to the bR proton pumping mechanism. Several factors have contributed to the continuing interest in bR:

- The biological function of bR, namely the pumping of protons across a cell membrane, is an excellent example of an ion transport mechanism in nature; trans-membrane ion transport is of fundamental importance in cell physiology, where it is used to regulate the concentrations of Na, Cl, Mg, K, Ca, and other ions.

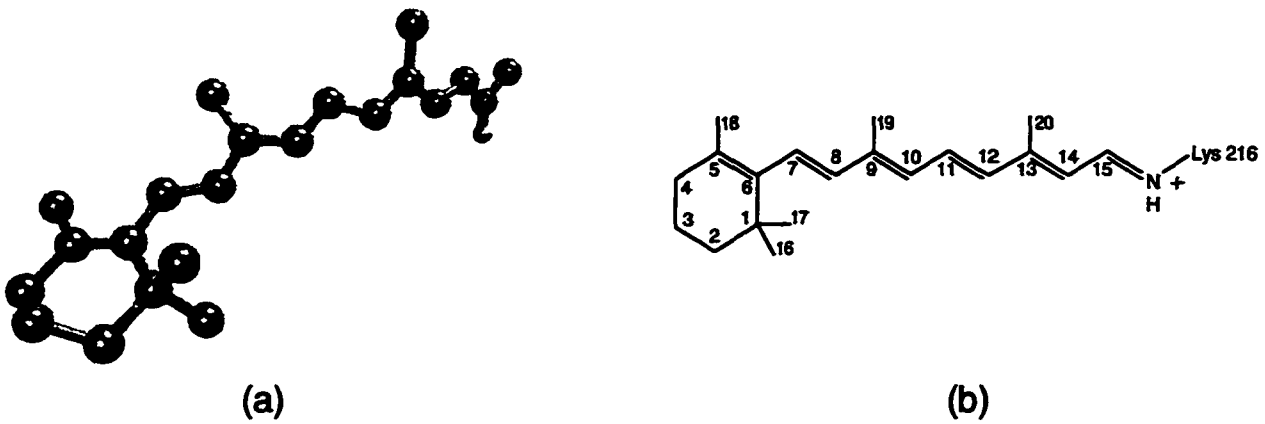


Figure 1.2: Retinal chromophore from bacteriorhodopsin. (a) CPK representation. (b) Atom numbering scheme.

- bR is one of the earliest membrane proteins for which the amino acid sequence has been determined, and one of the very few such proteins for which the three-dimensional structure has been observed to some accuracy [67, 65, 66]. Due to the difficulties in crystallizing membrane-bound systems, this has made bR of particular interest in the study of proteins located in cell membranes.

- Techniques for efficiently producing large quantities of bR and of generating site-specific bR mutants (in which specific amino acids are substituted with selected replacements) have been developed [114]. The availability of these techniques has facilitated many experiments otherwise difficult with only small quantities of the protein.

- bR is one member of a much larger family of retinal proteins, all of which employ a covalently bound retinal for their function. This family includes halorhodopsin (hR), which pumps Cl^- ions in a fashion similar to bR [82]; sensory rhodopsin (sR), which photoactively generates signals for an associated transducer protein [158]; and the visual pigment rhodopsin.

One of the key breakthroughs in the study of bacteriorhodopsin has been the determination of the three-dimensional structure of bR by Henderson and coworkers using electron microscopy (EM) at low temperature [67, 66]. This work demonstrated that bR contains seven transmembrane helical segments, oriented roughly parallel to each other and perpendicular to the plane of the membrane.

The helical regions form a roughly cylindrical channel through which protons are transferred, with retinal positioned near the center of the channel. The knowledge of this structure has made it possible to develop atomic-level models for the proton pump mechanism and to highlight those specific residues which may be directly involved in proton translocation. In addition, this structure may be used as a starting point for computer modeling of the bR at an atomic level of detail, using the technique of molecular dynamics (MD) simulations. Figure 1.3 shows the structure of bR based on molecular dynamics refinement of the EM structure.

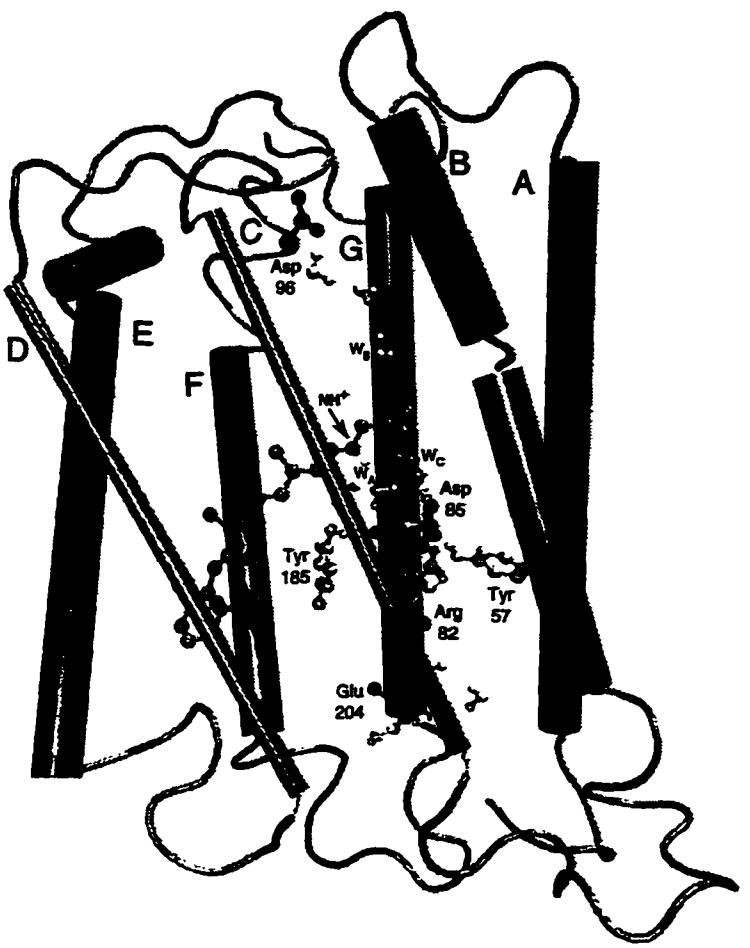


Figure 1.3: Refined bacteriorhodopsin structure used as a starting point in the simulations. The refinement procedure is described in chapter 3. Shown are the seven trans-membrane helices of bR along with internal water molecules and key residues participating in the proton pump cycle. Helices C and D are presented as thin cylinders to reveal the protein interior.

In molecular dynamics simulations, the motion of the atoms in a molecule or molecular aggregate

is computed using Newtonian equations of motion, where each atom i is treated as a point mass m_i with some partial charge q_i. The force on each atom is derived from a model potential energy function which includes terms describing the electrostatic interaction between each atom pair, as well as empirical energy terms describing the coupling of covalently bonded atoms. Within this approximation, in which the charge distribution and molecular bond structure are fixed for a given simulation, molecular dynamics makes it possible to model the dynamical motion of a system at the level of atomic detail. MD simulations may be used to calculate such quantities as transport coefficients or free energies, or may be applied to determine the specific atomic motions responsible for a particular molecular function. It is this latter purpose for which molecular dynamics has been applied here in the study of the proton pump mechanism of bacteriorhodopsin.

The advantage of molecular dynamics over more accurate quantum chemical methods (such as *ab initio* and semi-empirical calculations) is the ability to simulate the motion of many more atoms, e.g., up to 10,000 atoms in a typical MD simulation *vs.* 10 – 100 atoms in an *ab initio* evaluation. It is important to realize, however, that MD simulations are limited in the types of problems they may address, and in the time scales which the simulations may be used to explore. In this regard, bacteriorhodopsin is an ideal candidate for the application of MD to study a specific biological mechanism at the atomic level, for the following reasons:

- It is relatively small, yet too large for direct quantum chemical techniques.
- The proton pump cycle of bR proceeds in several intermediate steps, with the initial reactions occuring very quickly (within 5 picoseconds), well within the accessible time scale for MD simulations.
- The experimentally determined three dimensional structure of bR is available for use as the initial condition in the simulations.

The following chapters report the results of molecular dynamics simulations of bacteriorhodopsin which model the structure and early intermediates in the bR proton pump cycle. Chapter 2 presents a quick overview of the characteristic features of proteins, and a summary of the molecular dynamics

technique including the empirical energy function used to model the protein interatomic forces. Following this introduction, chapters 3 – 6 describe efforts to use molecular dynamics simulations to study the structure and photocycle of bacteriorhodopsin. Chapter 3 presents an overview of the bacteriorhodopsin photocycle, and describes the initial refinement of the equilibrium structure of bR starting from the experimentally determined bR coordinates. After refinement, simulations of the early (picosecond) events in the bR photocycle were carried out, and are described in chapter 4. Chapter 5 proceeds with simulations of the early photointermediates of bR mutant structures (in which selected amino acids have been replaced by alternative residues with distinct structural or chemical properties), and compares the results of these simulations to those in Chapter 4. Chapter 6 concludes with a discussion of a more accurate method for modeling the initial photoexcitation and subsequent motion of retinal using a combined quantum/classical technique.

Appendices A – B describe the development of software tools for molecular dynamics simulations (the program NAMD, in appendix A), and for visualization and analysis of the computed molecular trajectories (the program VMD, in appendix B). Appendix C derives the differential equations employed in the combined quantum/classical simulations described in chapter 6.

Chapter 2

Computer Simulations of Biopolymers

2.1 Introduction

Molecular dynamics simulations have been used for the study of a wide range of problems, starting from monoatomic systems such as elastic hard spheres [2] and models of liquid argon [132, 179], to more complicated molecular assemblies such as water and polyatomic gases [133]. Starting in the mid 1970's, models were developed to study the motion of small protein biopolymers such as the bovine pancreatic trypsin inhibitor [102]. Biological macromolecules, e.g., proteins and nucleic acid polymers, typically consist of $10^2 - 10^6$ atoms or more, a size which prohibits the direct use of quantum mechanical calculations to model their dynamics. Molecular dynamics makes simulation of the motion of atoms in molecules of this size possible by ignoring quantum effects and treating each atom as a point mass subject to Newton's second law. This chapter starts with a brief overview of the structural features of proteins which facilitate their study by molecular dynamics, and concludes with a description of the molecular dynamics technique.

2.2 Protein Structure

As illustrated in Fig. 2.1, protein molecules consist of an inhomogeneous linear chain of repeating subunits; each subunit includes one of the 20 naturally occuring amino acid sidechains, which link planar *peptide bonds* (shown in grey). The amino acid sidechains (group R in Fig. 2.1) are linked to each C_α carbon atom of the protein backbone. Unlike many homogeneous polymers which assume a distribution of lengths and conformations, each protein has a specific length and amino

acid sequence, and it is this sequence which determines the unique way in which the protein folds together as well as the chemical properties of the molecule.

The atoms in the peptide bond group exhibit a partial double bond character, which maintains a rigid, planar conformation for this O=C–N–H unit. The important degrees of freedom for the protein backbone are torsional motions about the single bonds which link the C_α atoms to the preceding and following peptide bonds. It is the specific angles of rotation about these bonds (the angles ϕ and ψ in Fig. 2.1) which determine the overall shape of the protein.

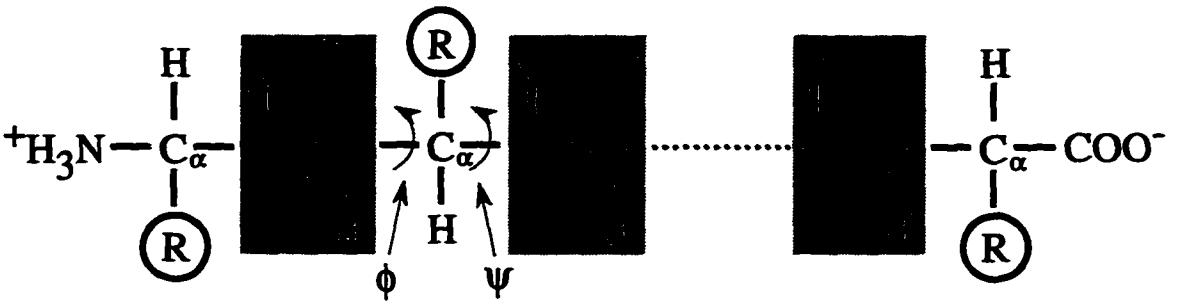


Figure 2.1: Standard protein backbone structure, including the peptide bonds (boxed grey regions), C_α atoms, and amino acid sidechains (R). Proteins conventionally begin at the N-terminus (at left), and end at the C-terminus (at right).

As a result of steric hindrances between the peptide bond units, the C_α atoms, and the amino acid sidechains, not all ϕ-ψ combinations are possible. Pauling and Corey, in 1951, were the first to realize that certain ranges of ϕ-ψ angles lead to low-energy configurations, in which hydrogen-bonds between adjacent amino acids and peptide bond atoms stabilize the local protein conformation and help form distinct secondary structure units [127]. The first two secondary folding motifs discovered were the α-helix, shown in Fig. 2.2a, and the β-sheet, shown in Fig. 2.2b. An α-helix is formed when a single protein backbone segment assumes a helical form such that hydrogen-bonds form between the peptide bond atoms of residue N, and the peptide bond atoms of residues $N+3$ and $N-3$. β-sheet segments form when two separate backbone segments align parallel to each other, and hydrogen-bonds form between aligned peptide bonds. Since the initial discovery of these folding motifs, many other forms have been identified, such as alternate helical structures, parallel

as well as anti-parallel β-sheets, turns, and motifs involving coordinated ions such as Ca, Zn, Mg, or Cu.

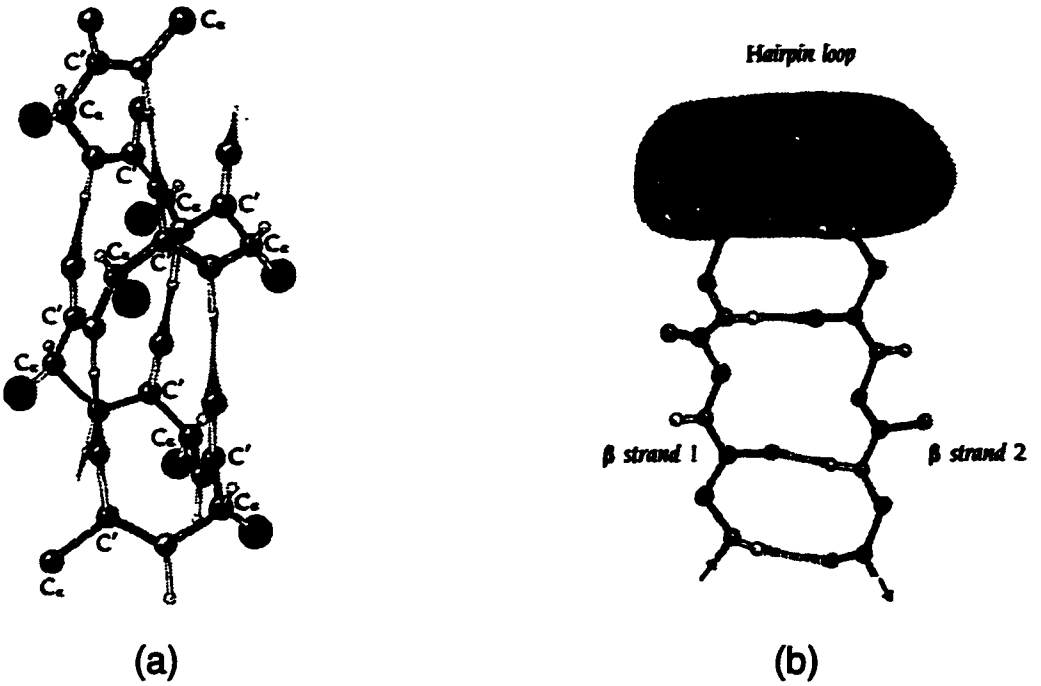


Figure 2.2: (a) An α-helical protein folding motif. (b) A β-sheet protein folding motif. The two β strands in this case are connected via a hairpin loop, a commonly occuring turn in proteins. (After Branden and Tooze, *Introduction to Protein Structure* [20]).

The secondary structure units, in turn, combine to form tertiary, quaternary and higher patterns, such that a hierarchy of levels of protein folding emerges. The structure of bacteriorhodopsin consists of seven α-helices, arranged roughly parallel to each other and connected by turn segments. The helices form a cylinder which spans the bacterial cell membrane, through which protons are conducted during the photocycle.

2.3 Molecular Dynamics

Molecular dynamics simulations calculate the motion of the atoms in a molecular assembly using Newtonian dynamics to calculate the net force and acceleration experienced by each atom. In this model, each atom i is treated as a point mass m_i at position $\vec{r}_i$, with the atoms assigned partial charges q_i which remain fixed throughout the simulation. The covalent bond structure of

the molecule is also assumed not to vary during the simulation. These assumptions limit the range of problems to which MD simulations may be applied; for example, chemical reactions involving bond breakage or charge transfer are not suitable for study by standard molecular dynamics. MD techniques are best suited for structures which involve stable repeating subunits interacting primarily via electrostatic interactions, and which undergo conformational (as opposed to structural) transitions such as ring flips, torsional isomerizations, or hydrogen-bond formation. Many types of motions exhibited by proteins, nucleic acids, membranes, and liquids fit well within this realm of applicability.

2.3.1 Energy Function

An empirical energy function

$$U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R}) \tag{2.1}$$

which depends on the positions (and possibly momenta) of the atoms i is used to determine the force $\vec{F}_i$ on each atom; using Newton's second law, these forces determine the acceleration experienced by each atom:

$$m_i \frac{d^2\vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R}). \tag{2.2}$$

Equation(2.2) represents a set of N second order differential equations, which are solved numerically at discrete time steps $t_0, t_1, \ldots t_M$ to determine the trajectory of each atom $\vec{r}_i(t)$. A standard integration scheme used in molecular dynamics programs is the Verlet method [179], which, given the positions of atoms at time $t - \Delta t$ and at time t, determines the positions at time $t + \Delta t$ using

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i}\vec{F}_i(t). \tag{2.3}$$

The Verlet integration method has the advantage of requiring only one force evaluation per time step, and of not requiring calculation of the atomic velocities.

A number of different energy functions $U(\vec{R})$ have been developed to model different classes of molecules or to model different types of molecular motion. Energy functions currently in use for the study of protein dynamics include the CHARMm [21], Amber [183], and Gromacs [172] force

fields. In these models, the energy function includes terms which describe the interactions between atoms due to the electrostatic (non-bonded) attractions and repulsions between pairs of atoms, and the interactions between covalently bonded atoms which are typically represented by harmonic potentials.

The CHARMm energy function is used in all simulations reported here. This function has the form

$$
\begin{aligned}
U(\vec{R}) \;=\; & \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} + \underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} + \\
& \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}} + \\
& \underbrace{\sum_D \sum_A \left(\frac{A_{DA}}{r_{DA}^k} - \frac{B_{DA}}{r_{DA}^l}\right) \cos^m\theta \cos^n\phi}_{U_{hbond}}
\end{aligned}
\tag{2.4}
$$

which is comprised of two parts: the non-bonded energy terms ($U_{nonbond}$), and the bonded energy terms (U_{bond}, U_{angle}, $U_{dihedral}$, and U_{hbond}). The non-bonded terms include the $1/r_{ij}$ Coulomb interaction as well as a Lennard-Jones 6-12 potential used to model the close-range repulsion between electron clouds and the longer-range attraction due to induced dipoles. These non-bonded quantities involve the interaction of each atom with all other atoms in the system, an $O(N^2)$ calculation; it is the computation of these terms which require the majority of the CPU time in an MD simulation.

The second part of eqn. (2.4), the bonded energy terms, model the motion of atoms due to the covalent bonds in the molecule. These terms include the following:

- U_{bond} (Fig. 2.3a): Oscillations about the equilibrium bond lengths r_0. In biopolymers, bond oscillations are the highest frequency motion in the system, with periods on the order of 10^{-14} s; consequently, the integration time step in MD simulations of proteins is typically taken as one tenth of this, i.e., 1 femtosecond.

- U_{angle} (Fig. 2.3b): Oscillations of three atoms about an equilibrium angle θ_0.

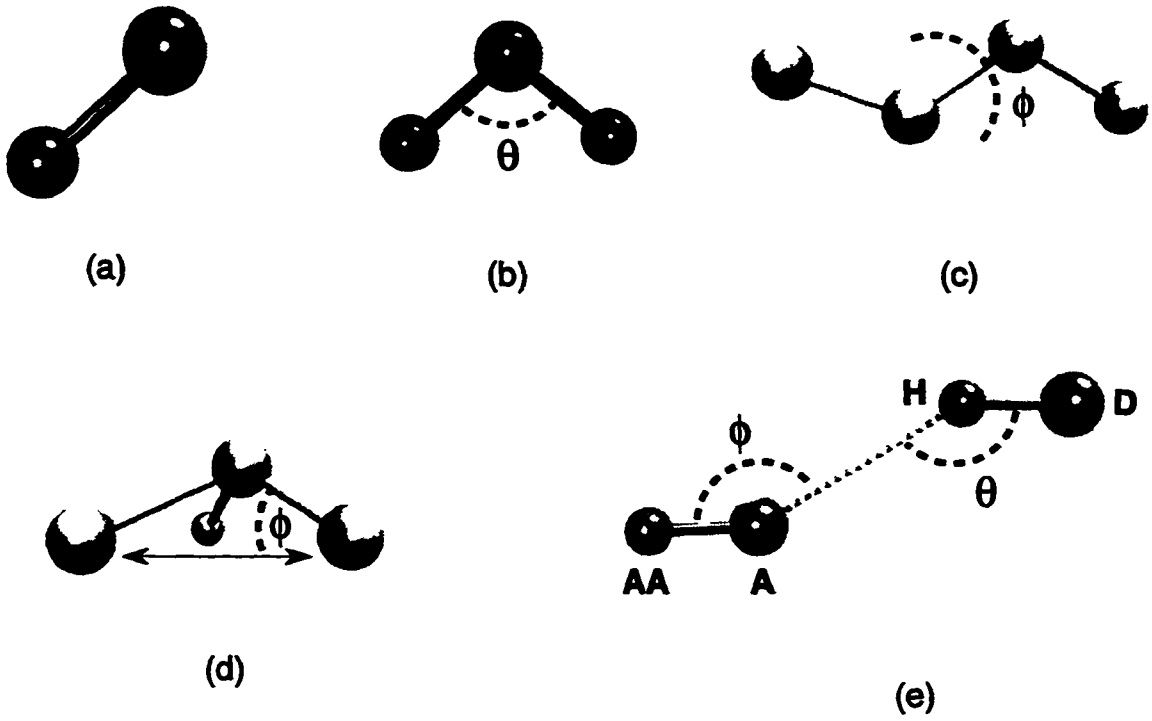


Figure 2.3: Bonded energy terms in the CHARMm molecular dynamics force field. (a) Bond energy. (b) Angle energy. (c) Dihedral angle energy. (d) Improper dihedral angle energy. (e) Hydrogen-bond energy.

- $U_{dihedral}$ (Fig. 2.3c and d): Dihedral angle motions refer to torsional rotation of four atoms about a central bond. The dihedral angle ϕ is defined as the angle between the planes formed by the first three atoms in the dihedral, and the last three atoms in the dihedral. The CHARMm force field also includes *improper* dihedral angles (Fig. 2.3d), which are four-body interactions with the same functional form as dihedral angles but involving a different type of motion. Improper dihedral angles are formed by one central atom moving out of the plane formed by the other three atoms in the improper dihedral, by an angle ϕ.

- U_{hbond} (Fig. 2.3e): MD simulations may optionally include the U_{hbond} term to explicitly model hydrogen-bond interactions between donor (D), hydrogen (H), acceptor (A), and antecedent (AA) atoms. The hydrogen-bond strength is dependent on both the donor-acceptor separation r_{DA} and the orientation of the atoms.

2.3.2 Force Field Parameters

One consequence of the parametrization of the forces acting on each atom as done in eqn. (2.4) is the need to determine a large number of force constants and equilibrium values such as k^{bond}, k^{angle}, θ_0, etc., for each atom. One method employed to reduce the number of parameters in a simulation is to develop a small set of characteristic *atom types*, for which parameters are determined. For proteins, parameters are developed for each amino acid separately, in either the gas or liquid phase, after which the determined values are used for all occurences of that amino acid in simulated systems. Values for force constants in most cases start from best-guess values and are iteratively refined through comparison with experimental data. The comparison data is derived from numerous sources, including: accurate quantum chemical evaluations of the electron densities and ground-state equilibrium configurations; vibrational spectra, compared to calculated normal-mode frequencies; and experimentally measured van der Waal's radii, densities, bulk moduli, and transport coefficients.

2.3.3 Simulation Algorithm

A typical molecular dynamics simulations of a biopolymer begins from an initial structure and initial velocity distribution, and proceeds to solve eqn. (2.2) for successive time steps using eqn. (2.3). The initial coordinates quite often derive from an X-ray, NMR, or electron microscopy structure for the molecule, while the velocities are assigned randomly from a Maxwell-Boltzmann distribution at some temperature T. In the absence of any external coupling or boundary conditions, this scheme is energy conservative and represents simulation in the microcanonical ensemble. Quite often, it is desirable to maintain a constant temperature as well, in order to compare the simulation results to experiments performed at some specific value of T. In this case, a simulation may include an initial period of temperature equilibration, in which the particle velocities are periodically either scaled by a factor

$$\gamma = \sqrt{\frac{3Nk_BT}{\sum m_i v_i^2}} \tag{2.5}$$

or are reassigned from a new Maxwell-Boltzmann distribution. This initial equilibration period continues until partitioning of the total energy among the potential and kinetic degrees of freedom of the molecule reaches an equilibrium.

2.3.4 Parallel Molecular Dynamics

Calculation of the non-bonded energy terms in eqn. (2.4) is a computationally expensive procedure, and several techniques have been developed to reduce the time necessary to evaluate $U_{nonbond}$ or to increase the accessible time scale for simulation. In many cases, a cutoff radius is employed for the non-bonded interactions, where, for each atom, the interaction with other atoms outside this region is taken to be zero. A switching function is used to smoothly vary the nonbond terms from their correct value to zero in an interval near the edge of the cutoff region. To increase the size of the integration time step, which is limited by the high-frequency bond oscillations, the SHAKE algorithm is often used to iteratively constrain bond lengths to their equilibrium values [139]. Recently, parallel computers and algorithms have also been employed to decrease the total time necessary to compute a molecular trajectory.

Parallel computers employ several processors connected by an interprocessor communications network (or shared memory mechanism) to increase the total computational resources available for a given task. Parallelization of molecular dynamics requires the use of an efficient strategy for partitioning tasks among the processors, and several methods have been investigated to parallelize existing molecular dynamics applications. For simulations involving a cutoff of the nonbonded interactions, a spatial decomposition strategy is an effective method to limit the total amount of communication required between processors and to maintain an equal CPU workload. Spatial decomposition partitions the volume of space occupied by the molecule into uniform regions, and assigns the task of computing the motion of the atoms in each region to different processors. We have developed the molecular dynamics program NAMD, which implements the CHARMm energy function and uses spatial decomposition for parallel computation; NAMD was employed for several of the simulation of bacteriorhodopsin reported here, and is described further in appendix A.

2.3.5 Visualization and Analysis of Trajectory Data

Molecular dynamics simulations generate a tremendous amount of data, in the form of position and momentum vectors for each atom at each time step. An important process in MD simulations is visualization of this data. In conjunction with the development of NAMD, we have developed the program VMD to display and analyze the computed molecular trajectories. This program has been employed to study the results of MD simulations in all studies of bacteriorhodopsin reported here, and is described further in appendix B.

At present, MD simulations proceed in two steps: generation of output data files containing the computed molecular coordinates, followed by visualization of these data files. What is often desired, however, is a single step: visualization during calculation, which makes possible immediate feedback on and interactive participation in the simulation while it is in progress. VMD is one attempt to provide this functionality: it includes a graphical user interface to initialize an MD simulation using NAMD on a remote computer, and to display the computed time steps of the molecule under study as they are calculated. Further, it offers interactive controls to the user: using a pointer device, the user may add perturbative forces to selected atoms, amino acids, or protein fragments which are included in the dynamics of the system. Appendix B discusses this capability and the features of VMD in detail.

Chapter 3

Structural Refinement of Bacteriorhodopsin[1]

3.1 Introduction

The publication of the electron microscopy (EM) structure of bacteriorhodopsin by Henderson and coworkers [66] has been a seminal event for research on retinal proteins. It provides a necessary reference for atomic-level models of the bR proton pump mechanism, and acts as a starting point for molecular dynamics simulations of the protein. Unfortunately, the EM structure does not yet resolve sufficient detail to explain fully the mechanism of proton pumping in bR. For example, the EM structure resolves only the membrane spanning residues of the seven α-helices of bR, but not the inter-helical loops. The EM structure also does not resolve any of the water molecules inside bacteriorhodopsin. In this chapter, we describe a refined structure of bR which has been developed through modification and energy minimization of the original EM structure, and which has been tested for stability and accuracy in molecular dynamics simulations.

3.1.1 Proton Pump Cycle

Bacteriorhodopsin transfers protons from the cytoplasmic to the extracellular side of the membrane through a cyclic process initiated by absorption of a photon. This photon triggers an isomerization of retinal, which proceeds then through several intermediate states identified by their absorption spectra. An accepted kinetic scheme for this cycle is an unbranched series of intermediates shown

[1]Portions of this chapter have been taken from the manuscript: W. Humphrey, I. Logunov, K. Schulten, and M. Sheves, Molecular Dynamics Study of Bacteriorhodopsin and Artificial Pigments, *Biochemistry* 33:3668–3678, 1994.

in Fig. 3.1, where the indices on each intermediate (J, K, L, etc.) denote the absorption maxima of the respective species.

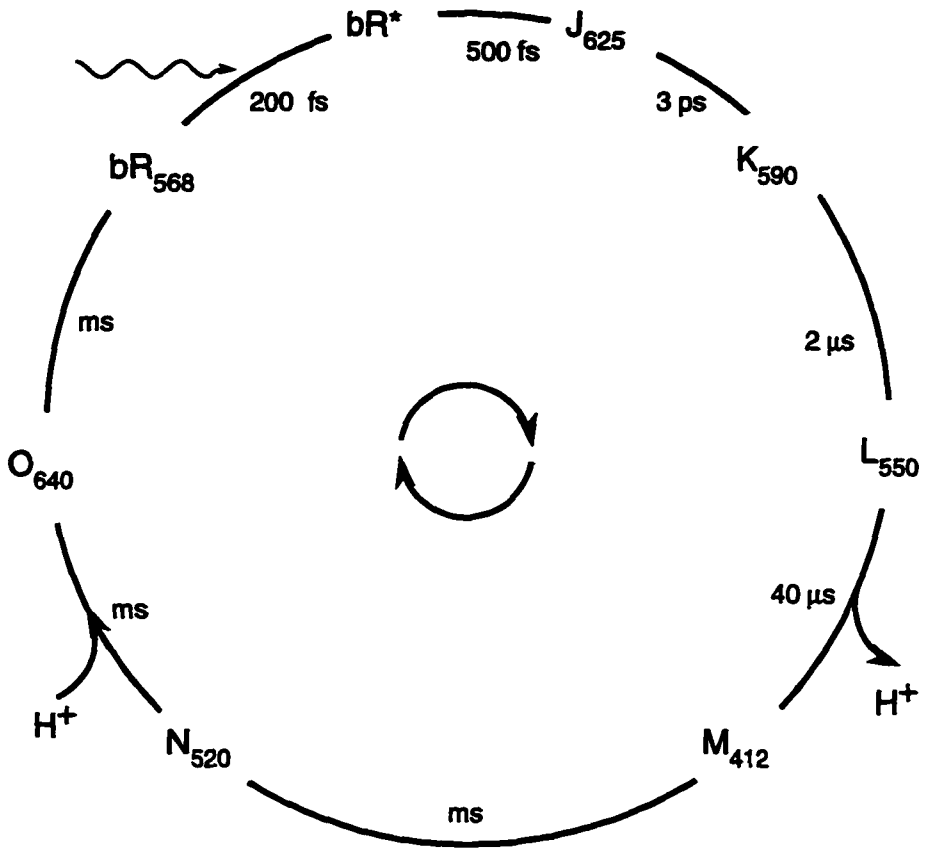


Figure 3.1: Bacteriorhodopsin photocycle.

The photocycle can be considered in two parts: the early intermediates which encompass the reaction [93, 99]

$$bR_{568} \overset{h\nu}{\rightsquigarrow} bR^* \rightarrow J_{625} \rightarrow K_{590} \leftrightarrow L_{550} \tag{3.1}$$

followed by the late intermediates [93, 126, 149, 59, 174]

$$L_{550} \leftrightarrow M_{(I)} \rightarrow M_{(II)} \leftrightarrow N_{520} \leftrightarrow O_{640} \rightarrow bR_{568}. \tag{3.2}$$

Photoexcitation of bR to bR* initiates an isomerization about the C_{13}-C_{14} bond of retinal to form a 13-*cis* isomer; until recently, this reaction has been considered to finish after formation of the J_{625} state in 500 fs [99, 37, 40]. This picture is supported by the fact that J_{625} cannot be trapped

even at low temperatures. However, recent femtosecond spectroscopy experiments have suggested that isomerization does not occur immediately, and thus J_{625} may represent a partially twisted, unstable state (see ref [64], and chapter 6).

By K_{590} and L_{550}, the system is in a fully 13-*cis* state [169, 18]. The $L_{550} \rightarrow M_{412}$ transition is accompanied by transferral of the Schiff base proton to Asp-85, due to lowering of the Schiff base pK_a and concomitant increase in the pK_a of Asp-85 [150, 108, 106, 19, 160, 57]. Subsequently, a proton is given off to the outside of the cell via a proton release group, which has been recently identified as Glu-204 [23, 6, 136]. During the $M_{412} \rightarrow N_{520}$ transition, a proton is transferred to the Schiff base from Asp-96, which then takes up a proton from the cytoplasmic environment [57]. The reaction cycle is completed as the protein returns to bR_{568} in the final transitions.

Several researchers have proposed that the $L_{550} \rightarrow M_{412} \rightarrow N_{520}$ transitions involve actually two M_{412} forms, $M_{(I)}$ and $M_{(II)}$ [33, 86, 94, 110, 149, 58, 134]. Detailed models were suggested by Orlandi & Schulten [121], and Váró & Lanyi [175, 177, 176], who suggested that the crucial change which is responsible for the difference in the exposed site of the Schiff base from outside to inside is attributed to the $M_{(I)} \rightarrow M_{(II)}$ transition.

3.1.2 Evidence for Internal Water Molecules

Several studies have suggested the presence of water molecules within bacteriorhodopsin. The involvement of water in stabilizing the protonated Schiff base was suggested by Dupuis *et al.* [41], by Hildebrandt & Stockburger [71] based on resonance Raman studies of dried membrane, and by de Groot *et al.* [34] based on ^{15}N NMR studies. Neutron diffraction data has indicated at least 11 water molecules within the interior of bR, with four located near the Schiff base binding site [125]. Recently it was suggested [56] that in bR a specific orientation between the Schiff base and carboxyl groups allows water molecules to form a well defined structure which bridges the charged groups. This structure stabilizes the ion pair (protonated Schiff base/counterion) and leads to the unusually high pK_a value of the protonated Schiff base and low pK_a of the aspartic acids. In addition, water molecules have also been suggested to participate in the transfer of the proton

from Asp-96 to the Schiff base [66, 175]. Water is also expected to contribute significantly to the pK_a values of relevant titratable groups in bR [5, 56]. Observed changes in the water structure during the photocycle [96] and changes in the absorption maximum of bR_{568} upon dehydration [71, 125, 135] support an important role for water in bR. The participation of water and amino acid side groups in proton transport in bR has been discussed also in Brünger *et al.* [25].

3.1.3 Outline of Chapter

In this chapter, we extend the work of earlier MD studies of bR and further refine the structure for bR_{568} by means of molecular dynamics simulations, focusing mainly on the retinal binding site. The inter-helical loop regions of bR, missing in the EM structure, had been added and the structure had been refined in a previous molecular dynamics study of bR by Nonella *et al.* [116], and a first attempt to place water into bacteriorhodopsin and to further refine its structure had been made by Zhou *et al.* [189]. These earlier refinements were improved in several respects. First, we incorporated into the simulations new data by Henderson (R. Henderson, private communication) which shifts the D helix 3 Å toward the cytoplasmic side of bR. Second, we changed significantly the refinement protocol of the earlier studies [116, 189]: after addition of the loops and of sixteen water molecules the structure was constrained back to the Henderson coordinates of the transmembrane segment. Third, we employed observations on bacteriorhodopsin reconstituted with retinal analogues (see for a review: [123]) as tests of our structure. In doing so we realized that combined atom simulations, e.g., representations of CH_2 moieties by effective single atoms, are not sufficient to account for sterical interactions of retinal with its binding site. Therefore, we adopted as a fourth improvement an all atom description for the refinement.

3.2 Methods

The original electron microscopy (EM) structure of bR, as determined by Henderson *et al.* [66], resolved the membrane-spanning helical residues; the turns had been added to this structure and equilibrated by molecular dynamics simulations in Nonella *et al.* [116]. The backbone atoms of the resulting structure had an RMSD of 2.8 Å as compared to the EM structure. Our refinement

employed the structure in Nonella *et al.* [116] as a starting point. The program X-PLOR [26], which uses the CHARMm energy function [21] was used for all molecular dynamics simulations. All simulations were carried out at a temperature of 300 K, employed a cut-off distance of 8 Å, and a dielectric constant of $\varepsilon = 1$ for the evaluation of Coulomb interactions.

Both the structure proposed by Henderson *et al.* [66] and refined in Nonella *et al.* [116] omit specification of the hydrogen atoms explicitly, except for the protein backbone and for protonated amino acids. In the present study, all atoms are described explicitly. We also employ an explicit hydrogen-bonding term in the energy function. The all atom representation was obtained by adding hydrogens to the structure in Nonella *et al.* [116] using the `HBUILD` command of X-PLOR. The parameters and charges used in the simulations are, respectively, the `parmallh3x.pro` parameters and `topallh6x.pro` charges [21, 26], except for retinal.

The charges employed for the protonated Schiff base of retinal in bR are those reported in [116] and used in [189] with slight modifications. The original charges were chosen for the united atom model; the explicit hydrogens added to the retinal backbone were given partial charges of 0.03 in units of proton charge, with the corresponding heavy atom charges reduced by 0.03 to preserve net charge. The charges assumed in Nonella *et al.* [116] and Zhou *et al.* [189] for the Schiff base nitrogen and proton were -0.509 and 0.519, respectively; however, quantum chemical calculations of retinal charge distributions [165, 15, 28] have indicated that these charges might be too large. In our simulations, we have reduced the charges on these atoms to -0.309 and 0.319 for the Schiff base nitrogen and proton, respectively. It is not clear yet if this is the best choice. In a recent extensive quantum chemical calculation of a retinal Schiff base fragment, Nina & Smith determined the respective charges as -0.727 and 0.505 (Nina & Smith, in press).

Equilibrium bond lengths, angles, and torsional angles for retinal are those determined by recent X-ray crystallography studies [141]. The retinal geometry in bR_{568} is all-*trans*. Force constants for retinal are derived from molecules with similar chemical structures for which parameters are available in CHARMm and X-PLOR. Retinal backbone single- and double-bond torsional barriers are given in Table 3.1. Barriers for the backbone single bonds are 5 kcal/mol, and for the double

bonds 30 kcal/mol, with the exception of the bonds in the vicinity of the Schiff base. The C_{13}-C_{14} and C_{15}-N_{SB} double-bond barriers are reduced to 20 kcal/mol, and the C_{14}-C_{15} single-bond barrier is increased to 10 kcal/mol, in keeping with quantum chemical calculations in [150, 149].

ϕ_i	k_{ϕ_i} (in kcal/mol)	ϕ_i	k_{ϕ_i} (in kcal/mol)
C_5-C_6-C_7-C_8	5.0	C_6-C_7-C_8-C_9	30.0
C_7-C_8-C_9-C_{10}	5.0	C_8-C_9-C_{10}-C_{11}	30.0
C_9-C_{10}-C_{11}-C_{12}	5.0	C_{10}-C_{11}-C_{12}-C_{13}	30.0
C_{11}-C_{12}-C_{13}-C_{14}	5.0	C_{12}-C_{13}-C_{14}-C_{15}	20.0
C_{13}-C_{14}-C_{15}-N_{SB}	10.0	C_{14}-C_{15}-N_{SB}-C_ϵ	20.0

Table 3.1: Retinal backbone torsional barriers k_{ϕ_i}, in the torsional potentials $E_{dihe} = \frac{1}{2}k_{\phi_i}[1 + \cos(2\phi_i + \delta_i)]$.

The protonation states of titratable groups in bR are as given in the standard X-PLOR amino acid topology files, except for Asp-96 and Asp-115, which are assumed protonated according to the observations in [57, 47]. As suggested previously [10, 189], Arg-82 has been moved up to a position facing toward the cytoplasmic side of the protein. In this position, Arg-82 is part of the Schiff base counterion complex.

In the original structure proposed by Henderson *et al.* [66], helix D was the least resolved segment of the seven α-helix protein, because it does not contain any bulky aromatic side chains. It has since been realized that this helix should be positioned 3 Å toward the cytoplasmic side of bR (R. Henderson, private communication). To achieve this in our structure, we modified the original Henderson coordinates by moving helix D using the molecular graphics program Quanta [109] and harmonically constrained the C_α atoms of all seven helices in the starting structure to the corresponding shifted Henderson coordinates. These constraints were applied for 30 picoseconds, followed by 10 picoseconds of equilibration, and energy minimization.

After these calculations, sixteen water molecules were placed within the interior of bR, in three regions: above the Schiff base toward the cytoplasmic side, around the Schiff base, and below the Schiff base toward the extracellular side. These water molecules are described according to the TIP3P water model [77]. Six water molecules were placed in the region proximate to the Schiff base, four were placed above the Schiff base in the region proximate to Asp-96, and six were placed

below the counterion in the region close to the extracellular side. After the placement of water, the system was equilibrated for 20 picoseconds, and energy minimized. The resulting structure has been further simulated for 30 picoseconds to test the stability of the structure and determine various properties, e.g., RMSD-values.

Retinal modifications were performed on the final bR_{568} structure. The modifications were accomplished with the PATCH facility of X-PLOR; initial coordinates were assigned using the positions of replaced atoms, and using the HBUILD command to generate coordinates for explicitly added hydrogens. For all simulations, including retinal analogues, we carried out energy minimization after the modifications, followed by five picoseconds of equilibration and five picoseconds of dynamics, and further energy minimization. Table 3.2 lists the partial charges used in our simulations of retinal analogues.

4-dimethylamino				4-methyl and 2-methyl	
N_4	-0.060				
C_{21}	0.030	C_{22}	0.030	C_{21}	-0.040
H_{21A}	0.000	H_{22A}	0.000	H_{21A}	0.000
H_{21B}	0.000	H_{22B}	0.000	H_{21B}	0.000
H_{21C}	0.000	H_{22C}	0.000	H_{21C}	0.000

Table 3.2: Partial charges for retinal analogues.

3.3 Results

3.3.1 System Construction

Following the addition of explicit hydrogens to the structure proposed by Nonella *et al.* [116] simulation G1 constrained the C_α atoms of this structure to the corresponding atoms in the structure of Henderson *et al.* [66] with helix D shifted 3 Å toward the cytoplasmic side of the membrane, as described in Methods. The resulting structure had an RMSD of 1.8 Å for the backbone C_α atoms as compared to the shifted Henderson structure, a value much improved compared to the 2.8 Å RMSD for the structure as proposed by Nonella *et al.* For helix D, the RMSD was 2.3 Å, which is larger than the total RMSD, but still reduced, compared to the earlier value of 2.8 Å.

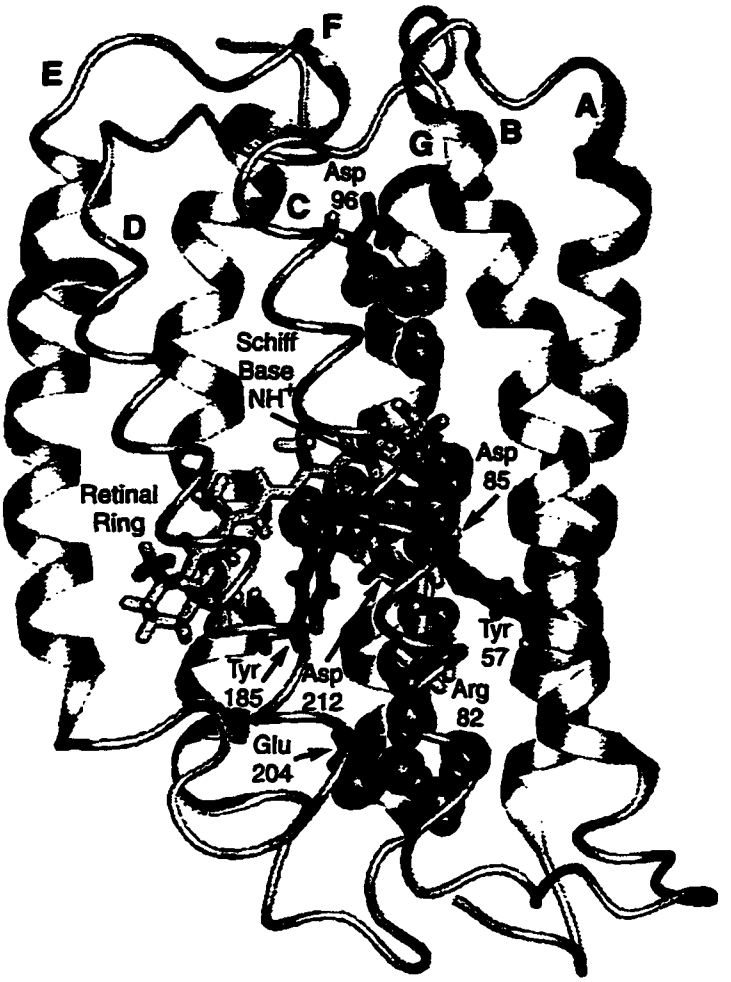


Figure 3.2: Ribbon diagram of bacteriorhodopsin, showing residues which are implicated in the proton pump mechanism. Water molecules placed within the protein interior are represented as solid spheres. Helices C and D are shown as thin ribbons to reveal the retinal binding site.

Following simulation G1, water molecules were added to the structure as described in Methods. In simulation G2, this system was equilibrated for 20 picoseconds, followed by final energy minimization. The resulting structure, bR_{568}, is presented in Fig. 3.2 – 3.6. Figure 3.2 shows the protein backbone, water, and retinal, as well as the added loop regions. The retinal long axis is at an angle of about 30° to the plane of the membrane, somewhat larger but still consistent with measured angles of 20° ± 10°. The methyls at C_9 and C_{13} remain pointing to the cytoplasmic side [70, 42, 89, 66]. In the vicinity of the retinal ring, the average tilt of the retinal long axis is somewhat larger than the average tilt of retinal as a whole, by about 10°, due to a slight curvature

in the retinal backbone.

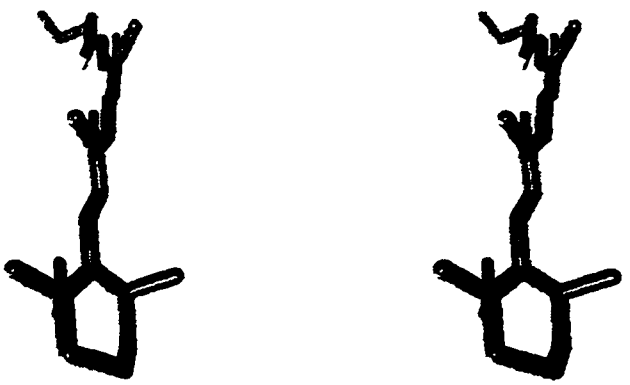


Figure 3.3: Stereo diagram of retinal's twisted configuration. Thin dark cylinders originating from the C_5, C_9, and C_{13} positions indicate axes used to measure the orientation of methyl groups (Table 3.3).

An interesting feature of the bR_{568} structure is a distinctive twist of retinal's backbone, as shown in Fig. 3.3. Obviously, retinal is not planar in this structure; instead it displays a "corkscrew" configuration, as one views retinal along its backbone. This twist is a result of torsions around retinal's single bonds, especially around the C_5-C_6, C_{10}-C_{11}, and C_{12}-C_{13} bonds by 15°, 15°, and 30°, respectively. In Fig. 3.3, thin cylinders originating from the C_5, C_9, and C_{13} atoms indicate axes used to determine orientation angles for the methyl groups at these positions. The vertical cylinders, representing the z axis, are parallel to the membrane normal. The horizontal (out of the page) axes, representing the x axis, are orthogonal to the z axis and in the plane formed by the membrane normal and the average direction of the retinal backbone. Table 3.3 lists the angles which define the orientation of the retinal backbone methyl groups. The angle θ measures the tilt from the z axis, and the angle ϕ measures the orientation of this tilt from the x axis.Also listed in Table 3.3 are measured values for θ determined by solid-state ^{2}H-NMR studies [171].

Methyl	ϕ	θ_{sim}	θ_{exp}
C_{18}	-64.7	48.6	37
C_{19}	-27.8	41.4	40
C_{20}	57.8	34.5	32

Table 3.3: Twist angles for retinal methyl groups. Values for θ_{exp} were determined by solid-state ^{2}H-NMR studies [171].

A detailed picture of the Schiff base region is given in Fig. 3.4, which shows the relative positions of active groups in the binding site, as well as the positions and orientations of water molecules in this region. Possible hydrogen-bonds are shown as dashed lines accompanied by the associated distances from hydrogen to hydrogen-bond acceptor. An interesting feature of this structure is associated with a hydrogen-bonding network within the binding site involving water molecules and protein residues. Within the immediate Schiff base linkage vicinity, there are four water molecules; two additional water molecules are in close proximity. The water molecules most proximate to the Schiff base all have three or four hydrogen-bonding partners. The Schiff base is hydrogen-bonded to only one water molecule; this hydrogen-bond differs from a linear conformation by 31°, in keeping with resonance Raman and FTIR spectra which suggest that weak hydrogen-bonding prevails between the protonated Schiff base linkage and its environment [3, 78, 4].

Both oxygens of the Asp-85 carboxylate are located at a distance of 6 Å from the Schiff base. One oxygen of the Asp-212 carboxylate is located at a distance of 4.6 Å from the Schiff base, the other at a distance of 6.5 Å. Asp-85 has two water molecules hydrogen-bonded to the carboxylate, which is also bound to Arg-82. There is no direct interaction of the Schiff base with Asp-85, the two groups being bridged by two water molecules. While Asp-212 is located closer to the Schiff base than Asp-85, it also is connected to retinal only indirectly through one water molecule. Figure 3.4 shows that there are up to four groups within hydrogen-bonding distance of the carboxylate oxygen of Asp-212 closest to the Schiff base, shielding it from the Schiff base charge. While it is not feasible for all these groups to hydrogen-bond to the same oxygen of Asp-212, their proximity does provide an effective screen for the carboxyl group. A water molecule is hydrogen-bonded to the other oxygen of the Asp-212 carboxylate; this water molecule is also hydrogen-bonded to Tyr-57

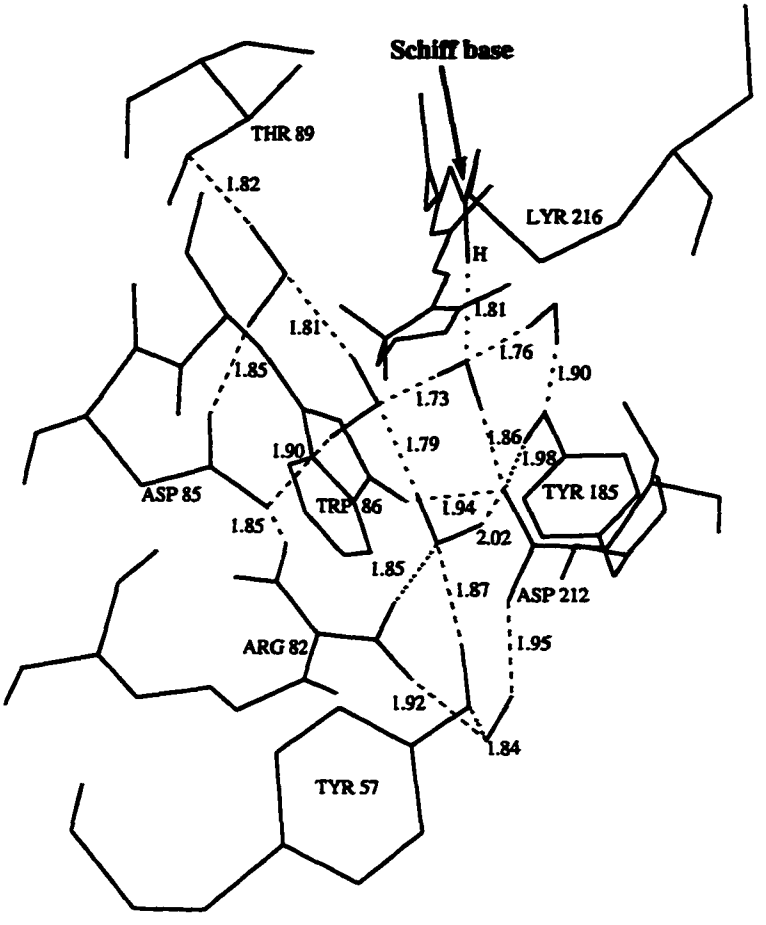


Figure 3.4: Detailed view of the retinal binding site, showing positions of water molecules and important residues, and distances of possible hydrogen-bonds.

and Arg-82.

Tyr-57 is hydrogen-bonded to two water molecules in this structure: one of the waters is hydrogen-bonded to Arg-82, to another water molecule, and possibly to Asp-212 as well; the other water is hydrogen-bonded to Arg-82 and possibly to Asp-212. Tyr-185 is within hydrogen-bonding distance of Asp-212, as well as being hydrogen-bonded to one water molecule which is bound to the water molecule near the Schiff base. Both tyrosine residues have their hydroxyl groups oriented toward the Schiff base environment; however, they are separated enough from the Schiff base to preclude any direct interaction: Tyr-57 is separated by 6.9 Å from the Schiff base, and Tyr-185 by 4.6 Å. Arg-82 has been directed in our bR_{568} structure toward the Schiff base, and remained in

this position throughout all refinement steps, hydrogen-bonding directly to Asp-85 and, indirectly, to Asp-212 and to Tyr-57 via water molecules. Trp-86 seems to participate in the least number of hydrogen-bonds, the only candidate being a possible bond with Asp-212. Trp-86 serves the role of maintaining a boundary for the hydrophilic region around the Schiff base. Thr-89 also participates in the hydrogen-bonded network described, with its hydroxyl group hydrogen-bonded to a water molecule which is also hydrogen-bonded to Asp-85.

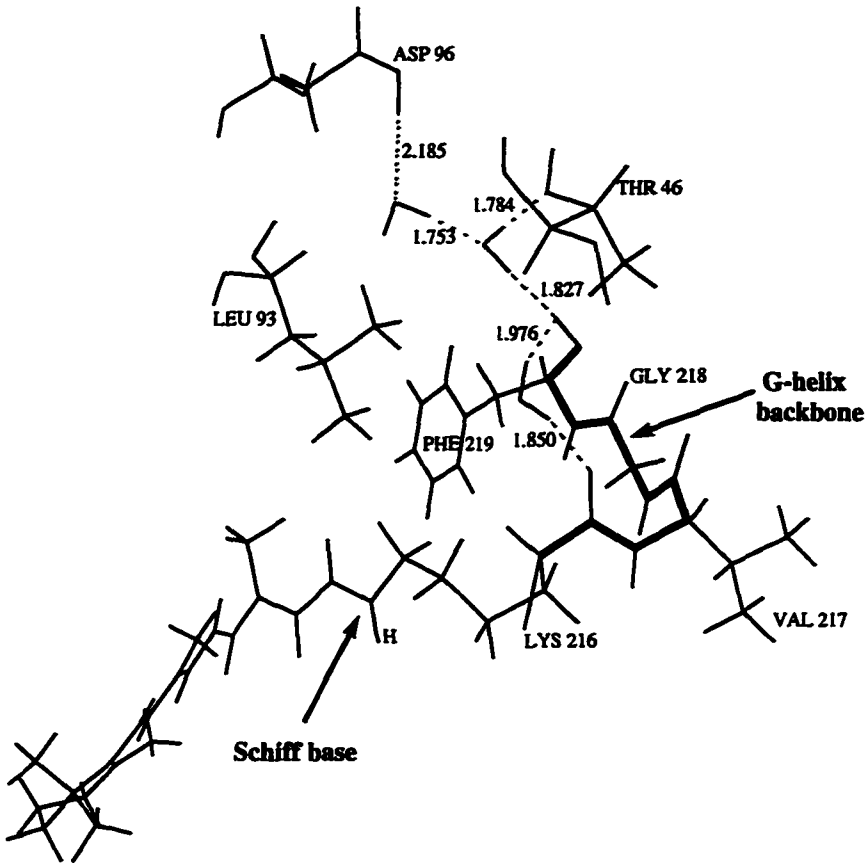


Figure 3.5: View of the region above the Schiff base, toward the cytoplasmic side of the membrane.

Above the Schiff base, toward the cytoplasmic side of the membrane, helix G exhibits an unusual conformation. Figure 3.5 provides a detailed view of this region, indicating possible hydrogen-bonds by dashed lines. There are three water molecules between the protonated Asp-96 and the Lys-216 to which retinal is attached. These water molecules, along with Phe-219, form a hydrogen-bonded

chain between Asp-96 and the carbonyl of Lys-216. The helix G backbone along this chain is seen to be distorted from a standard α-helical configuration, with the carbonyl of Phe-219 oriented perpendicular to the helix instead of parallel. This carbonyl is hydrogen-bonded to two water molecules, which breaks the standard hydrogen-bond between the CO group of Phe-219 and the NH group of Leu-223. One of these water molecules is hydrogen-bonded to the carbonyl of Lys-216, forming a weak hydrogen-bond due to a 60° angle between the hydrogen-donor pair and the acceptor-antecedent pair. The other water molecule is hydrogen-bonded to Thr-46 and also to a third water molecule which hydrogen-bonds to Asp-96. This chain provides an indirect interaction between Thr-46 and Asp-96.

To test if the distortion in helix G is a result arising from the particular placement of water molecules used here, two simulations were done, each starting from the refined bR_{568} structure and running for 25 ps. In the first simulation, the two water molecules hydrogen-bonded to helix G above the Schiff base vicinity were removed. During this time, the helix maintained the distortion from an α-helical form, with the water molecule originally hydrogen-bonded to Asp-96 moving to form a hydrogen-bond with Phe-219, reinstating the hydrogen-bond lost when the water molecules were removed. In the second, all three water molecules placed in the region between the Schiff base and Asp-96 were removed. Within the 25 ps simulated, the helix remained in the distorted form, but it is possible the helix could return to a standard α-helical state given more time.

NMR spectroscopy may be able to reveal the distance from the C_{19} atom of retinal to the closest tryptophan residue (J. Herzfeld, private communication). In our bR_{568} structure, with retinal oriented with its methyl groups at the C_5, C_9, and C_{13} positions toward the cytoplasmic side of the membrane, the closest tryptophan residue is Trp-182, located 3.7 Å from C_{19}. In the EM structure, with retinal oriented in the same manner, this distance is 3.4 Å [66].

3.3.2 Structure Stability

To test the stability of the described bR_{568} structure, we carried out simulation G3 which lasted 30 picoseconds. Figures 3.7 – 3.8 summarize the results of this simulation. In Fig. 3.7, the RMSD

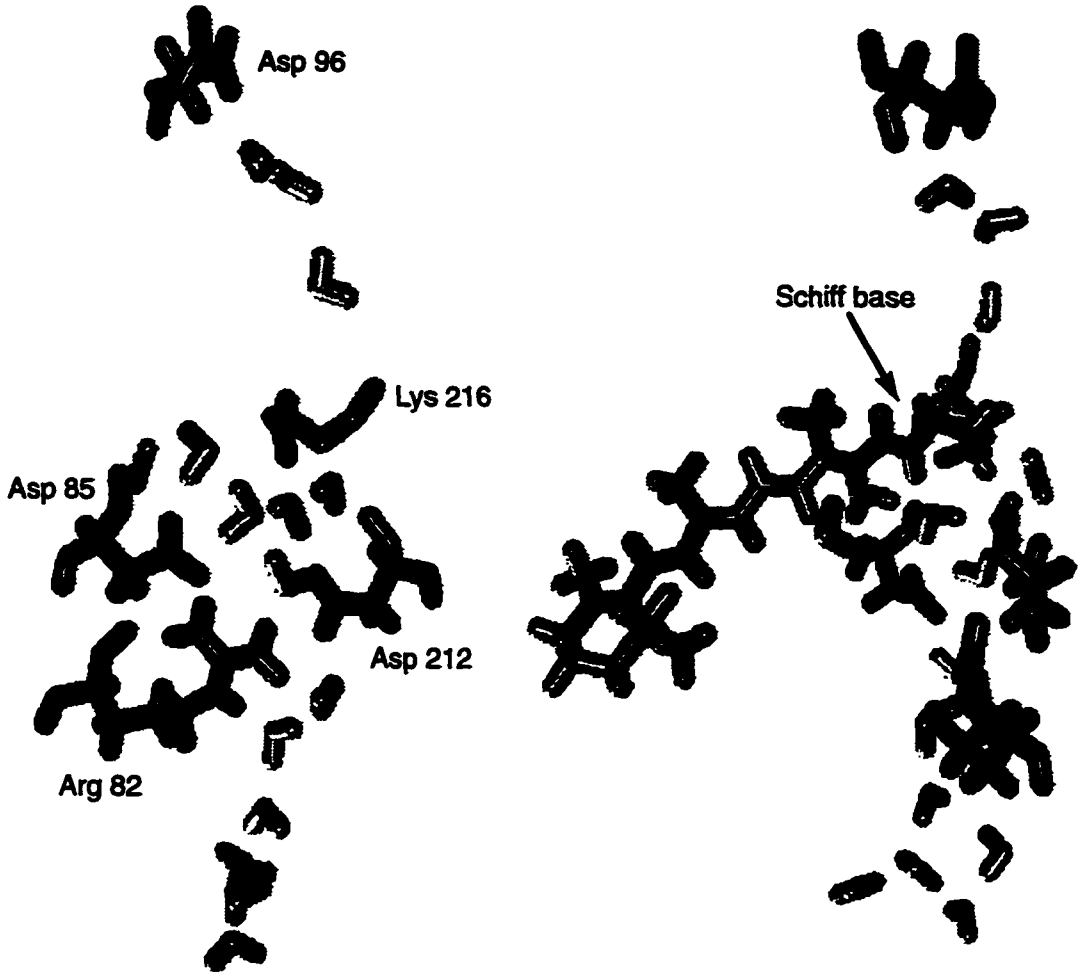


Figure 3.6: Solid model images of retinal and key groups near and above the binding site region, from two different perspectives. The left image shows shows only a portion of retinal for clarity.

of the C_α atoms is shown, with the seven α-helices A through G indicated by boxed regions. The turn regions exhibit the largest deviations, while helices B, C, and G show small deviations, of 0.4 - 0.5 Å. For all the helical portions except helix D, the RMSD is less than 0.8 Å, while for helix D the residues close to the cytoplasmic side show higher RMSDs. Helix D is expected to have a somewhat larger mobility, since it is one of the smallest helices in bR, and does not contain any bulky residues.

Figure 3.8 shows the distances of the carboxylate oxygens of residues Asp-85 and Asp-212 to the Schiff base proton, as well as the RMSD of the C_α atoms relative to bR_{568}, during simulation G3. Asp-212 is seen to remain rather immobile, with distance values of 4.6 Å and 6.5 Å for the

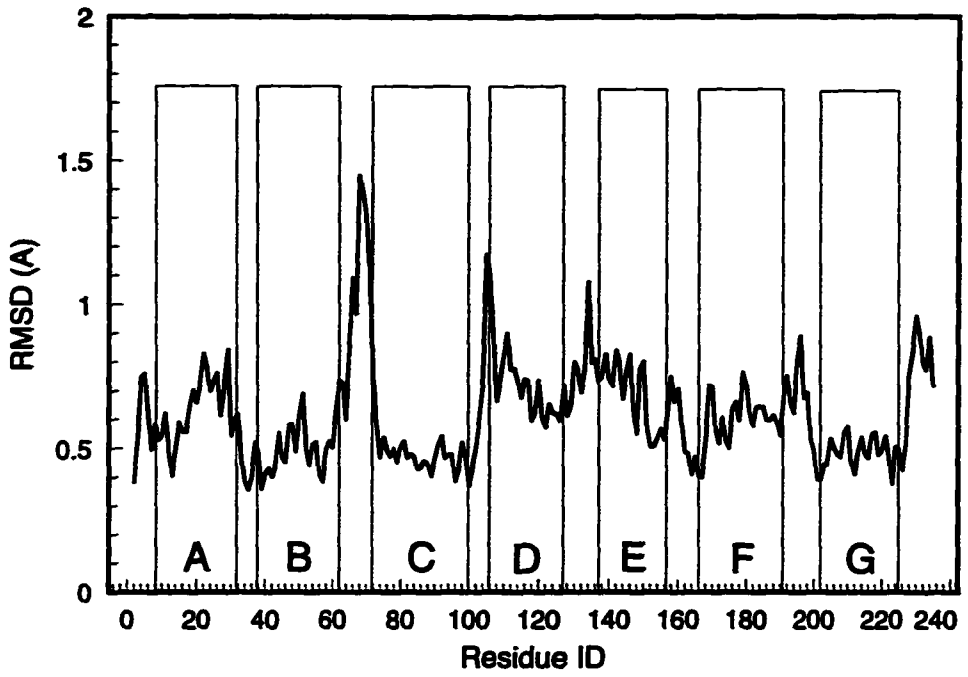


Figure 3.7: RMS deviations for C_α atoms resulting from simulation G3. Boxed regions A through G indicate the α-helical segments of bR.

two oxygens. Asp-85 exhibits larger fluctuations, but its motion is still relatively small, with both oxygens maintaining a distance of about 5.8 Å. The RMSD of the C_α atoms fluctuates slightly during the simulation, mostly staying in the region of 0.8 Å to 1.2 Å. Following an initial relaxation of about 5 ps, where the RMSD increases to around 0.8 Å, the RMSD remains close to an average value of 1.0 Å. The immobility of residues Asp-85 and Asp-212 and the relatively small RMSD of the protein provides a good indication of the stability of retinal's counterion complex, and overall structure.

3.3.3 Retinal Analogue Simulations

Experiments with modified retinal chromophores have provided useful information on the behavior of retinal within the binding site, and on the surrounding environment [123]. In this study we focus on modifications to the β-ionone ring region of retinal, which provide clues about the orientation and location of the retinal ring within the binding site. Figure 3.9 illustrates the analogues considered here. It was observed that introduction of bulky substituents at the ring C_4 position

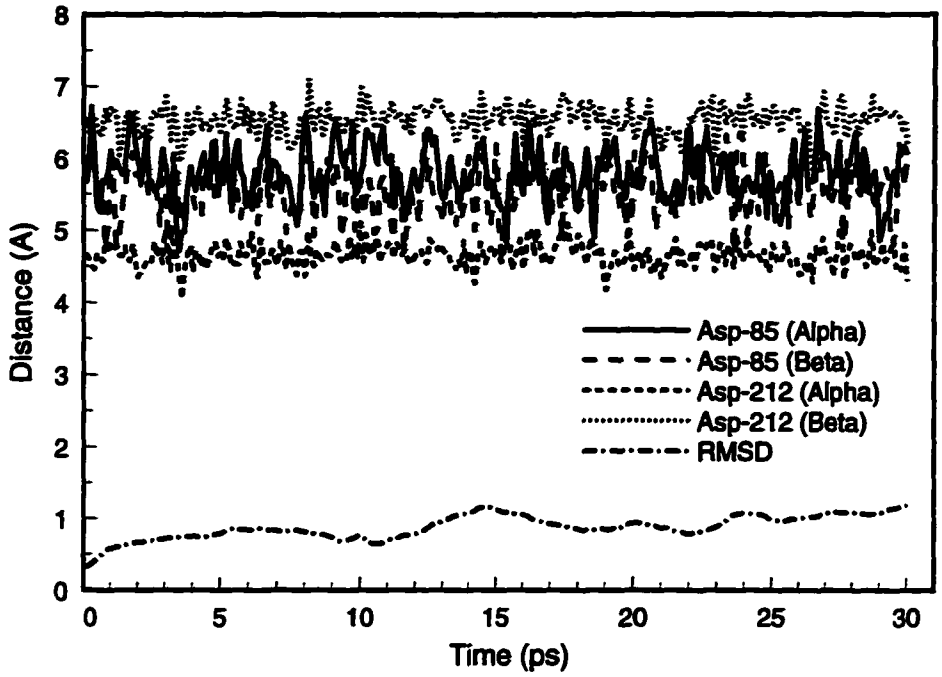


Figure 3.8: Distances of oxygen atoms in the carboxylate of Asp-85 and Asp-212 to the Schiff base proton, and RMS deviations of C_α atoms from bR_{568}, during simulation G3.

blue-shifted the absorption maximum of bR considerably [155]. In contrast, substitution at the C_2 position with a methyl group did not cause any significant change in the absorption maximum (Sheves and Friedman, unpublished results). Simulations of these artificial pigments, related to the spectral shifts observed for the respective pigments, should provide a measure of the consistency and accuracy of the refinement suggested here. Results for these simulations are given in Figs. 3.10 and 3.11.

4-dimethylamino Modification

Simulations R1A and R1B were carried out to determine the effect of the addition of a dimethylamino group to the C_4 position of retinal, and resulted in two structures, bR_{R1A} and bR_{R1B}. The C_4 position has two hydrogen atoms bound to the carbon, thus, there are two possibilities for attachment of the dimethylamino group. We define the plane of retinal facing toward the extracellular side as α, and that facing to the cytoplasmic side as β. Simulation R1A models substitution at the α-position, R1B at the β-position.

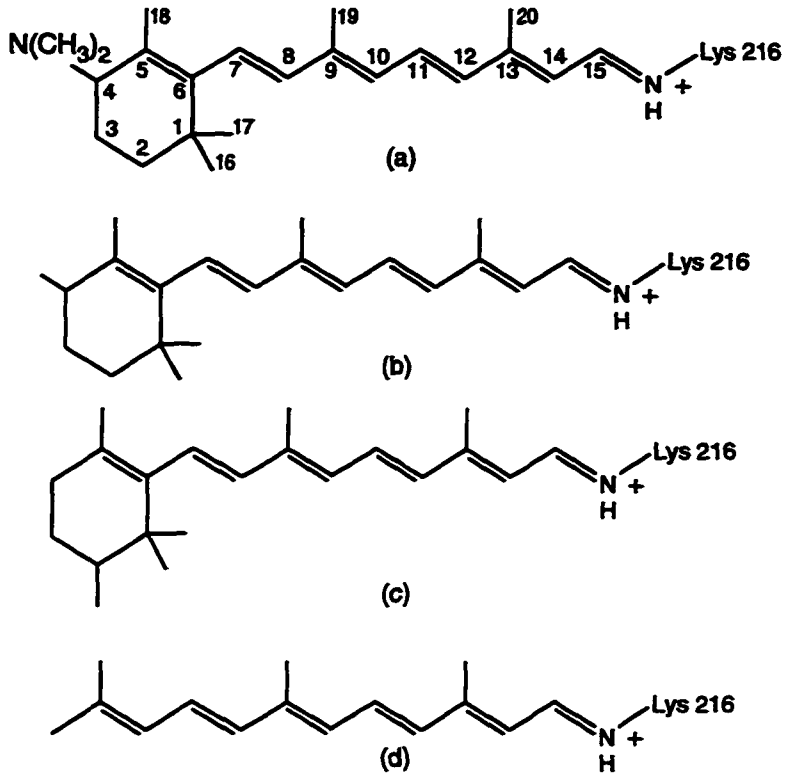


Figure 3.9: Retinal analogues: a) 4-dimethylamino analogue and numbering scheme of retinal atoms. b) 4-methyl analogue. c) 2-methyl analogue. d) Cleaved-ring analogue.

Figure 3.10 summarizes the results of both simulations. Shown in this figure are the torsional angles of bonds along the retinal backbone. For the native pigment, all retinal backbone bonds are in the s-*trans* configuration, with the C_6-C_7, C_{10}-C_{11}, and C_{12}-C_{13} bonds slightly twisted, giving rise to the corkscrew configuration of retinal in bR_{568} discussed above. The changes of retinal's conformation in going from bR_{568} to bR_{R1A} are quite dramatic. The C_6-C_7 and C_8-C_9 bonds rotate from an s-*trans* to an s-*cis* conformation, while the C_{12}-C_{13} bond relaxes completely to the s-*trans* state. Both the C_6-C_7 and the C_8-C_9 bond twist by about 30° from planar s-*cis* conformations. In the case of bR_{R1B}, the resulting conformation shows some change relative to bR_{568}, but all retinal backbone bonds remain in the s-*trans* conformation. The main changes seen are in the C_6-C_7 and C_8-C_9 torsional angles. The rotation about the C_6-C_7 bond measures about 20°, reducing the 30° twist angle of native retinal. The rotation about the C_8-C_9 bond, however, measures about 30°, twisted from the native s-*trans* configuration, a larger overall twist than for the C_6-C_7 bond.

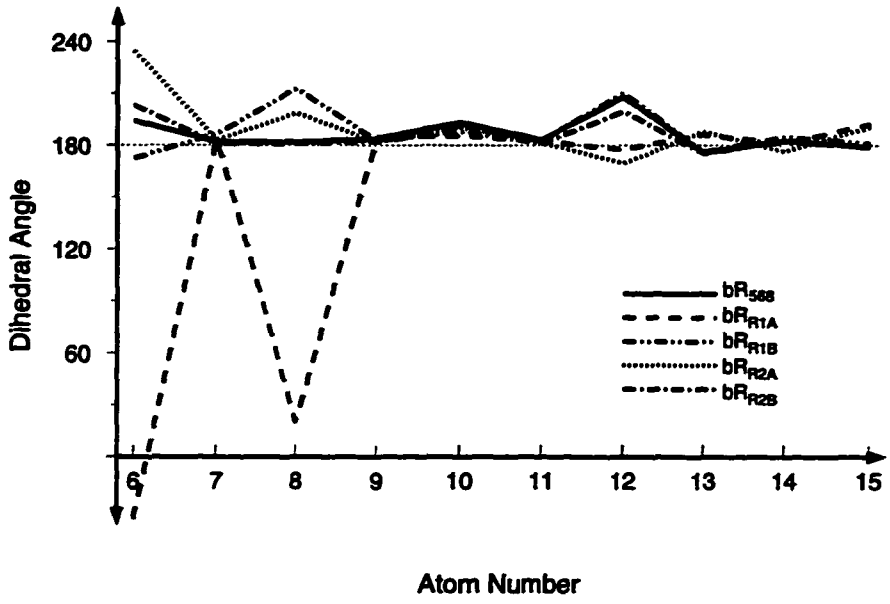


Figure 3.10: Comparison of dihedral angles along retinal backbone, for substitutions at the C_4 position; the numbering scheme is defined in Figure 3.9.

4-methyl Modification

Simulations R2A and R2B modeled the effect of substitutions at the C_4 position of retinal with a methyl group, and resulted in structures bR_{R2A} and bR_{R2B}. Attachment to the α-position was studied in simulation R2A, attachment to the β-position in simulation R2B. The results of these calculations are summarized in Fig. 3.10. In bR_{R2A}, retinal maintains its *s-trans* conformation along the polyene skeleton, but many of the dihedral angles show significant twists. The torsion about the C_8-C_9 bond is about 20°, and the C_{12}-C_{13} single bond rotates by about 30° from the native conformation. Most significantly, retinal twists by over 45° about the C_6-C_7 single bond, such that this bond is twisted by 60° from *s-trans*. Retinal in bR_{R2A} has a very twisted form, with the ring region in particular twisted relative to the plane of the membrane (offset somewhat, by opposite rotations about the C_6-C_7 and C_{12}-C_{13} bonds). In bR_{R2B}, the system stays very close to the bR_{568} conformation, with only slight rotations about the C_6-C_7 and C_{12}-C_{13} bonds.

2-methyl Modification

Simulations R3A and R3B modeled the effect of substitutions at the C_2 position of the retinal ring with a methyl group, and resulted in structures bR_{R3A} and bR_{R3B}. As for the previous analogues, R3A modeled the attachment to the α-position, and R3B to the β-position. Figure 3.11 summarizes the results of these calculations. For both analogues, there is very little change with respect to the native pigment. bR_{R3A}, in particular, has a retinal conformation essentially identical to that of bR_{568}; bR_{R3B} contains only two significant changes, one being a 15° twist about the C_8-C_9 bond, away from the equilibrium angle[1], and the other a 10° relaxation about the C_{10}-C_{11} bond toward the equilibrium angle. In effect, these two counter-rotations in bR_{R3B} leave the retinal ring position and orientation basically unchanged.

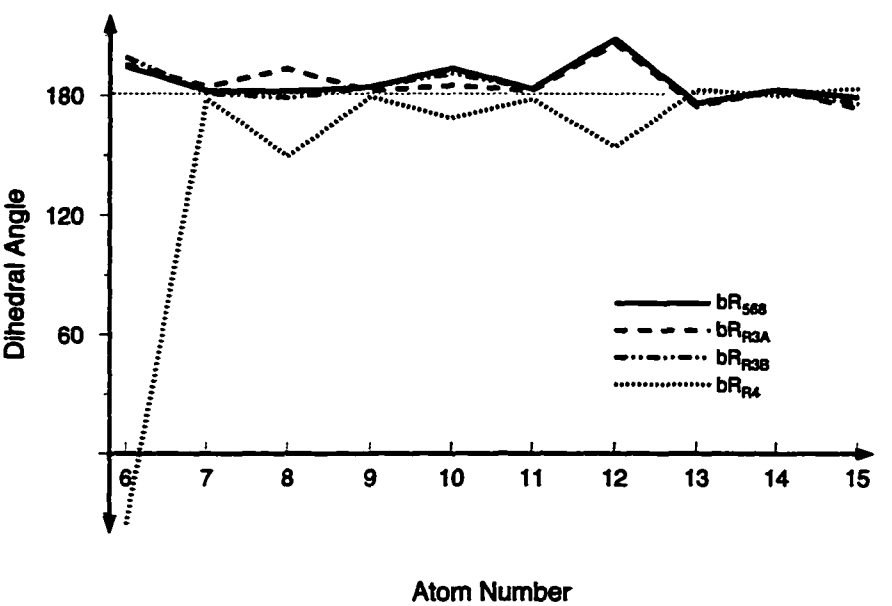


Figure 3.11: Comparison of dihedral angles along retinal backbone, for substitutions at the C_2 position and for the cleaved-ring analogue.

Ring Cleavage Modification

Simulation R4, resulting in structure bR_{R4}, was carried out to simulate the effect of cleaving the ring between the 3- and 4-carbons, and between the 1- and 6-carbons. Experiments with a cleaved-ring retinal analogue indicate that the absorption maximum is blue-shifted [154]. The results for

structure bR_{R4} are summarized in Fig. 3.11. This analogue changes considerably with respect to the native pigment; the chromophore twists about every single bond along the retinal chain except the C_{14}-C_{15} bond which has a higher torsional barrier than the other single bonds (see Table 3.1). The analogue rotated from s-*trans* to s-*cis* about the C_6-C_7 bond, resulting in the terminal ring fragment moving toward the extracellular side. The region into which these atoms move is occupied, in the case of native retinal, by the constituents of the β-ionone ring. The retinal analogue also exhibits a strong overall twist, but in an orientation opposite to that of bR_{568}. The dihedral angles of the native pigment which deviate from their equilibrium values are all larger than 180°, while the dihedral angles of bR_{R4} rotate to angles less than 180°. This results in a retinal with a chirality opposite to that of bR_{568}.

3.4 Discussion

Simulations G1 through G3 provided a refined structure of bR_{568} which is in much closer agreement with the EM structure than the structure in Nonella *et al.* and Zhou *et al.* as judged by the RMSD values. Simulations R1A-B through R4 modeled artificial pigments in an attempt to judge the validity of the bR_{568} structure, and to attempt to interpret the observed spectral shifts of retinal analogues.

3.4.1 Structure Refinement

A distinctive feature of the final structure of bR_{568} determined in simulation G2 is a corkscrew twist in the retinal backbone, due to rotations about the single bonds. Retinal also exhibits a slight curvature along the polyene chain. This curvature did not emerge in the previous refinements of bR [116, 189]. Circular dichroism (CD) experiments on bR have shown that in the trimeric form of bR in the membrane the CD spectrum is characterized by separated negative and positive bands, whereas the monomeric form is characterized by only one positive band. The trimeric spectrum was interpreted as due to exciton coupling [68, 44, 87]. Recently it has been suggested that the CD spectrum originated from rotations around single C-C bonds along retinal's backbone which would break the planar symmetry and provide an angular component to π-electron motion [46].

Solid-state ^{2}H-NMR studies of retinal in bR also indicate a twisted retinal, and suggest the retinal backbone to be curved, bending toward the extracellular side [171]. Table 3.3 compares the experimentally determined tilt angle θ of retinal methyl groups C_{18}, C_{19}, and C_{20} to the angles for bR_{568}. These angles compare quite well, with the largest difference seen for C_{18}, which is the methyl group attached to the retinal ring. The experiments were not able to determine the twist angle ϕ, which determines the orientation of the tilted methyl group; in bR_{568}, the methyls C_{18} and C_{19} are twisted to one side of the membrane normal, with the C_{20} methyl twisted to the other side. It is possible that retinal is capable of twisting in either direction, which has been suggested as an explanation for an observed biphasic band in the CD spectrum of bR [185].

The retinal ring is found at an angle of about 20° relative to the plane of the Schiff base moiety, this orientation being due to the cumulative effect of several rotations about the single C-C backbone bonds, the twist about the C_6-C_7 bond being very small, such that retinal is in a 6-s-*trans* conformation as previously suggested [63]. FTIR linear dichroism and photoselection measurements combined with quantum-chemical calculations have indicated twists around the retinal single bonds [50]. Slight rotations about these single bonds should not affect considerably the absorption maximum and the opsin shift induced by a reduction of the C_6-C_7 torsional angle [1].

The Lys-216 dihedral angles are also seen to be slightly twisted in this structure from their equilibrium positions, by almost 20° in the case of the Schiff base $N - C_\epsilon$ bond, to 5° or 10° for the remaining Lys-216 bonds $C_\epsilon - C_\delta$, etc. Experimental studies have indicated an unusual coupling between the 15-H rock vibrations and the backbone vibrations of the lysine. This unusual coupling might be associated with a twisted lysine conformation [55].

Many recent reports have indicated an important role for water in the structure and function of bR. Vibrational [39, 71] and NMR [34] studies have suggested the presence of bound water within the retinal binding site, and the possibility has been suggested that water molecules participate in proton transfer from Asp-96 to the Schiff base [66, 27]. It was also suggested that water plays an important role in maintaining the high pK_a of the Schiff base and the low pK_a of Asp-85 in the native pigment [5, 56]. Sixteen water molecules, placed within the refined bR_{568} structure,

maintain positions within the protein in three distinct regions: below the Schiff base toward the extracellular side, toward the cytoplasmic side between the Schiff base and Asp-96, and in the region of the Schiff base. The latter water molecules form a complex network of hydrogen-bonds within the binding site. Vibrational spectroscopy studies of threonine mutants of bR have suggested the possibility of a "proton wire" connection from Thr-89 to Asp-212, which is formed in our bR_{568} by three water molecules connecting these two residues [138].

It is currently believed that a major contribution to the opsin shift in bR is a weak Schiff base-counterion electrostatic interaction. Water molecules within the binding site bridge the Schiff base linkage and its negatively charged carboxylate neighbors and, according to our refinement, maintain a relatively large distance between the Schiff base and residues Asp-85 and Asp-212, namely, 6.0 Å and 4.6 Å, respectively. The distances are considerably larger than the respective distances of 4.1 Å and 3.7 Å in the Henderson structure [66]. The large distances are in keeping with a weak electrostatic interaction which contributes to the opsin shift [146]. In our refined structure, there is no direct interaction between the Schiff base and the aspartic acids; simulations without water, however, have all resulted in Asp-85 and Asp-212 remaining very close to the Schiff base, i.e., at distances similar to those of the Henderson structure [66]. Experiments with dehydrated bR have resulted in a blue-shifted absorption maximum [71, 125, 135] which could be the result of removal of at least part of the water molecules in the vicinity of the Schiff base, and of a corresponding attraction of Asp-85 and Asp-212 toward the Schiff base. The Schiff base is also seen to form only one hydrogen-bond, which is relatively weak due to the fact that the hydrogen-bond is not linear, with an angle of about 30°. Accumulated evidence from resonance Raman spectroscopy and FTIR both of bR and model compounds [3, 78, 4] indicated a weak hydrogen-bond between the Schiff base linkage and its environment. The present simulations suggest that this weak hydrogen-bond is not between the Schiff base and the counterions Asp-85 or Asp-212, but rather between the Schiff base proton and water. We note that recent model compound studies [90] have suggested that a weak electrostatic interaction between retinal and its counterion prevails also in bovine rhodopsin; however, in this system the Schiff base proton maintains a strong hydrogen-bond with

its environment. Thus, there is no direct correlation between the Schiff base-counterion electrostatic interactions and the strength of hydrogen-bonding to the Schiff base proton.

The counterion residues Asp-85 and Asp-212 differ in their orientation and environment in the structure presented here. Both oxygens of the Asp-85 carboxylate are equidistant from the Schiff base, while one of the oxygens of the Asp-212 carboxylate is significantly closer to the Schiff base than the other. Asp-212 also participates in more hydrogen-bonds than Asp-85, due to a larger number of closely adjacent groups. These groups provide a very effective screen for Asp-212, which could possibly explain why Asp-212 interacts electrostatically weakly with the retinal positive charge. Model compound studies have indicated that strong hydrogen-bonding to the counterion induces red shifts in the absorption maximum of the retinal protonated Schiff base due to a diminished interaction between the retinal positive charge and the negatively charged counterion [4, 1]. Replacement of Asp-212 by asparagine did not affect significantly the absorption maximum of retinal, in contrast to Asp-85 replacement, which induced a red shift in the absorption maximum [111, 104]. The difference in Asp-212 and Asp-85 environments may also account for their different behavior in M_{412} formation following light absorption, provided that the differences are also reflected in the L and M intermediates. It was found that Asp-85, and not Asp-212, serves as the proton acceptor in bR's pump cycle. The hydrogen-bonding of Asp-212 with the neighboring protein residues may stabilize the negative charge on Asp-212 reducing its pK_a and preventing a proton transfer from the retinal Schiff base to Asp-212. In contrast, the negative charge on Asp-85 is stabilized only by water that forms a bridge betweeen the protonated Schiff base linkage and Asp-85. Following light absorption this structure might be perturbed and the ΔpK_a between Asp-85 and the Schiff base will be decreased to induce the proton transfer observed between these two groups.

The water structure and unusual helix G conformation above the Schiff base (toward the cytoplasmic side) suggest one possible means for establishment of a network for transfer of a proton to the Schiff base from Asp-96. In M, it has been observed that the lysine carbonyl adopts an unusually low amide I frequency, which may be the result of hydrogen-bonding to water (Takei

et al., submitted). In our refined bR_{568} structure, a water molecule is bound to the lysine carbonyl. This unusual structure may allow, following light absorption and a further perturbation in the lysine backbone structure, the formation of a new water arrangement between the lysine carbonyl and possibly Asp-96 to account for the low Amide I lysine frequency observed in *M*. From test simulations with two or three of the water molecules in this region removed, the helix G distortion does not seem to be a result of the particular placement of water molecules used. Two Gly residues in this region of helix G (218 and 220) may provide some extra flexibility for the helix to maintain its unusual configuration.

3.4.2 Retinal Analogue Simulations

Retinal analogues modeled in simulations R1A-B through R4 were used to test our refined bR_{568} structure in the vicinity of the retinal ring, and to explain some of the structural rearrangements which accomodate artificial pigment experiments. It has been observed that the absorption maximum of bR is very sensitive to substitution at the C_4 position of the retinal ring, but relatively insensitive to substitutions at the C_2 position. Substitution of a methyl group at the C_4 position resulted in two photoactive pigments, which absorbed at 550 nm and 460 nm, each with its own photocycle. Substitution of the bulkier dimethylamino group at the C_4 position resulted in a single pigment absorbing at 460 nm [155]. In contrast, substitution of a methyl group at the C_2 position did not significantly affect the absorption maximum.

Both C_2 and C_4 positions of retinal are achiral; however, after incorporation into the protein binding site the two enantiomers are different and chiral discrimination by the protein is possible. It was previously suggested [152] that only one enantiomer at the C_4 position was bound to bovine opsin. In the present simulations we investigate the possibility of pigment formation for both enantiomers. For substitution of a methyl at the C_4 position, one possible explanation for the two observed pigments is that each pigment consists of a different enantiomer. The fact that bR_{R2A} contains very large rotations about retinal's single bonds, particularly about the C_6-C_7 bond, suggest that the respective enantiomer corresponds to the strongly blue-shifted 460 nm peak.

bR_{R2B} shows very little conformational change from the native pigment and suggests that the corresponding enantiomer produces the 550 nm peak. Trp-138, which is in very close proximity to the α-position of the 4-carbon (3.6 Å from the $C_{\delta 1}$ atom of Trp-138 to the α hydrogen of C_4) could account for this effect, in that substitution at the α-position results in a steric hindrance which causes a twist in retinal, while substitution at the β-position does not lead to strong sterical interactions. For the addition of a larger dimethylamino group to the C_4 position, a similar explanation is suggested. bR_{R1B}, with this substitution at the β-position, resulted in a highly twisted retinal which could explain the blue-shifted 460 nm peak. The very dramatic double bond rotation observed for bR_{R1A} could indicate that binding of the dimethylamino group to the α-position requires a very large conformational change, perhaps too large to allow this chromophore to constitute with bacterioopsin. The proximity of Trp-138 prevents attachment of the bulky dimethylamino group to the α-position, but this group can be accomodated at the β-position at the price of some twists in retinal.

Simulations of chromophores with a methyl group at the C_2 position produce no significant change in the bR structure, which compares very well with the unshifted spectra observed for these analogues.

In simulations R1A-B through R3A-B, the structure in the vicinity of the Schiff base has been examined for changes induced by retinal modifications. In the six simulations, no major conformational changes have been observed: Asp-85 and Asp-212 maintain their relative distances from the Schiff base, the network of hydrogen-bonded water molecules is maintained, and the overall structure of the residues surrounding the binding site is the same. Experiments have shown, in fact, that the C=N stretching frequency does not change for substitutions at the C_4 position. Simulations R1A-B through R3A-B also support the direction of the methyl groups of retinal, which are oriented in our simulations toward the cytoplasmic side of the membrane. Orientation of these methyl groups toward the extracellular side would reverse the roles of the C_2 and C_4 position.

In simulation R4 of a linear chromophore (see Fig. 3.9), retinal developed twists about all of its single bonds, except the C_{14}-C_{15} bond and, in particular, isomerized to a twisted 6-s-*cis*

conformation. This conformation resembles the conformation of the protonated Schiff base of retinal in solution, the latter compound absorbing at 440 nm. Our simulation could explain the observed blue-shift (relative to bR_{568}) for this retinal analogue to 532 nm, which is red-shifted relative to the absorption maximum of 440 nm in solution due to the protein maintaining other interactions contributing to the opsin shift, such as the weak electrostatic interactions between the Schiff base linkage and the counterion (in solution the counterion is very close to the protonated Schiff base). This could also shed light on a factor responsible for the chromophore existing in a 6-s-*trans* conformation once bound within the apoprotein – with the β-ionone ring intact, the surrounding residues require a 6-s-*trans* state, however, if this ring is removed, retinal acquires the freedom to adopt the preferred 6-s-*cis* state.

3.5 Conclusions

A cardinal feature of the refined bR_{568} structure is its close agreement with the EM structure of Henderson and coworkers [66]. The RMSD for the C_α backbone atoms comparing the membrane spanning part of the two structures is 1.8 Å, which is significantly less than the RMSD value of 2.8 Å for the structure reported in Nonella *et al.* [116], i.e., refined with a protocol which did not include water, explicit hydrogens and a D helix at the new position. The new RMSD value of 1.8 Å is well within the range of RMSD values arising for structurally well-resolved proteins. Typical RMSD values comparing crystal structures and molecular dynamics simulations for backbone atoms are in the range of 1.5 Å – 3.0 Å, deviations of about 2.5 Å being common. A simulation of the L7/L12 ribosomal protein (CTF) in solution compared to its crystal structure yielded particularly small deviations of 1.05 Å (C_α) and 1.63 Å (all atoms) [32]. While larger than the latter example, our small RMSD value indicates that our structure fits well with the EM structure [66], and that water placed within the protein interior does not cause major structural rearrangements. The small RMSD value encourages further simulation studies since the low value implies that bR, as modeled by molecular dynamics, is structurally stable even though it was here modeled in vacuum rather than in a membrane.

What can we learn from the bR_{568} structure about the protein's function? Obviously, a key element to an eventual answer will be the interactions in retinal's binding site. In this respect the qualitative agreement of our simulations of retinal analogues with respective observations indicate that the present structure reproduces the ring region of the binding site well. We have not described above our many failed retinal analogue simulations which arose with the structures in Nonella *et al.* and Zhou *et al.* [116, 189] or with descriptions which do not include hydrogens explicitly. The latter implies that structural detail at a high resolution in the 1 Å range is properly represented in the ring region for the present structure.

Even more essential than the ring region is the Schiff base region of the binding site. The respective structural features constitute certainly the most conspicuous, but also the most daring, features of the refined bR_{568}. In agreement with the structure in Zhou *et al.* [189] it is predicted that the most proximate groups to the Schiff base of retinal are water molecules. The generally held notion that the interactions of the Schiff base and the protein are dominated by Asp-85 and Asp-212 needs to be revised: the counterion of retinal is a hydrogen-bonded complex which involves several water molecules, Asp-212, Asp-85, Tyr-57, Tyr-185, Arg-82 and Thr-89. The water molecules prevent Asp-85 and Asp-212 from approaching the Schiff base hydrogen and forming direct hydrogen-bonds. Simulations of the primary photoreaction (described in chapters 4 and 5) have revealed that the interactions between the counterion complex and the protonated Schiff base in the native case at $T = 300$ K force the nitrogen-proton bond of the Schiff base to point towards the counterion, leaving retinal in a strained conformation, relieved only after the proton is transferred to the counterion complex. The role of the counterion – retinal interactions is also borne out by the observation that an Asp-85 to Asn-85 [157] mutation apparently extends the life time of the excited state of bR_{568} by a factor of 20.

The counterion structure is functionally significant since it determines the spectrum of the pigment, the ΔpK_a value between the Schiff base and the acceptor group Asp-85, and the proton transfer path between the Schiff base and Asp-85. The spectral shifts of retinal during the pump

cycle indicate that the counterion complex is altered during the pump cycle, e.g., upon photoisomerization, making the transfer path between the Schiff base and Asp-85 more favorable than the path to Asp-212. The preliminary nature of the counterion structure as presented in this chapter is more than offset by the evident fact that only structural details beyond that provided in Henderson *et al.* [66] can explain the mechanism of bR. Obviously, water will affect protein-retinal interactions and proton pathways; such water requires much higher resolution than is presently in sight. In case the water is mobile in the protein, it might not be resolved at all. Models of bR which include water are a necessity for interpreting the properties of bR; the present work has outlined refinement procedures which yield such models.

Another essential region of bR_{568} is the space between the retinal Schiff base and the proton donating side group Asp-96. This region provides the setting for the switching processes subsumed under the so-called *M* intermediate of the pump cycle. *M* formation is linked to proton transfer from retinal to a group connected with the extracellular side; *M* decays when a proton is transfered from a group in contact with the cytoplasmic side to retinal. One expects that retinal, through conformational transitions (e.g., reorienting its proton binding position from pointing in the direction of the extracellular side to the direction of the cytoplasmic side) cooperates with the protein in carrying out the switching processes. In this respect two details of the refined structure are of interest. First, the unusual conformation of the G helix. This helix becomes more ordered in the *M* state, as observed by Subramaniam *et al.* [162]. Second, three water molecules which bridge the 12 Å distance between retinal and Asp-96 in the refined structure could actively participate in the processes called the *M* intermediate; the three water or possibly further water, as suggested in Henderson *et al.* [66], can participate in the transfer of a proton from Asp-96 to retinal. One would expect, therefore, that an interpretation of the key *M* intermediate of bR's proton pump cycle requires a model of bR which includes water.

Obviously, it would be most desirable to have available structures of bR and of key intermediates of its pump cycle at a resolution which reveals all waters and essential conformational transitions, unfolding, thereby, the mechanism of bacteriorhodospin. Unfortunately, bR as a membrane protein

presently defies structure determination at high resolution, while its function as a proton pump may rely on minute structural detail. A resolution of this dilemma demands simulations which start with bR structures refined through addition of water as provided in the present work. There are several avenues for such simulations. First, the refined structure can be employed to evaluate the electronic absorption spectrum of bR, of its mutants, of bR reconstituted with retinal analogues, and of bR's pump cycle intermediates. Second, a renewed attempt of interpreting vibrational spectroscopy of bR can be launched (for earlier attempts see [62, 38, 130]). Third, the structure provided can be employed to evaluate the pK_a values relevant for the pump cycle; such evaluation for an unrefined, water-free structure has been provided in [10], but it is expected that water present in the protein will affect pK_a values strongly [56, 140]. Lastly, one should again embark on the efforts started in Nonella *et al.* and Zhou *et al.* [116, 189] to describe the pump cycle of bR and of its mutants. The following chapters report on simulations of the early photointermediates for wild-type bR (chapter 4) and bR mutants D85N, D85T, D212N, and Y57F (chapter 5).

Chapter 4

Simulations of the Bacteriorhodopsin Early Intermediates[1]

4.1 Introduction

Although bacteriorhodopsin (shown in Fig. 3.2) is a small membrane protein, it conjoins for its bioenergetic function a multitude of properties: it is a pigment; it transports protons; it is extraordinarily stable under intense light, yet undergoes continuously a cyclic reaction process; its consecutive reaction steps extend from extremely fast (500 fs for the initial photoisomerization) to slow (a complete cycle lasts a few ms) [93]. bR's most intriguing attribute may be that it has resisted a two decades-long intense research effort and has not revealed the riddle of its pump mechanism. This is surprising since (i) this mechanism apparently is closely tied to retinal which can be observed well through resonance Raman and Fourier transform infrared (FTIR) spectroscopy, since (ii) the protonation states of key amino acid side groups can be monitored well, and since (iii) the protein exhibits only small conformational transformations during its pump cycle. The solution of bR's structure by Henderson *et al.* [66] has been a major advance since it established a logical arrangement of the essential amino acids involved in the pump cycle, but this structure has also failed to reveal bR's pump mechanism.

We believe that the mechanism of bR's pump cycle has been elusive due to the fact that observations have not revealed a key player in the pump cycle, water. bR is known to contain water

[1]Portions of this chapter have been taken from the manuscript: W. Humphrey, D. Xu, M. Sheves, and K. Schulten, Molecular Dynamics Study of the Early Intermediates in the Bacteriorhodopsin Photocycle, *J. Phys. Chem.* 99:14549-14560, 1995.

molecules [125]; their hydrogen-bond network with each other, with retinal, and with amino acid side groups should play a key role in proton transport. One needs to know where these water molecules are located in bR and how they participate in the pump cycle. The water molecules might rearrange during the pump cycle, open and close proton pathways and, thereby, support the vectorial character of the pump. Another obstacle towards our understanding of bR's pump mechanisms is a lack of knowledge of the exact geometry of retinal in the early intermediates of the pump cycle. This geometry is mechanistically crucial since retinal transfers a proton from its Schiff base nitrogen to Asp-85 and, eventually, to bR's extracellular side (downward in Fig. 3.2) and receives a proton from Asp-96 and, eventually, from bR's cytoplasmic side (upward in Fig. 3.2). However, according to the structure in Henderson *et al.*, a pure all-*trans* $\rightarrow$ 13-*cis* photoisomerization, as suggested by a straightforward interpretation of observations, would render the Schiff base nitrogen with its proton pointing in the wrong, i.e., upward, direction during the initial stage of the pump cycle.

Retinal's Schiff base binding site contains a complex counterion including Arg-82, Tyr-57, Tyr-185, Trp-86, and negatively charged Asp-85 and Asp-212 [34, 35, 66, 137, 111]. It has been suggested that water molecules, within this binding site, participate in a hydrogen-bond network connecting the stated residues and the Schiff base proton (SBP) [41, 71, 34, 125, 56, 98]. FTIR data suggest that a water molecule, weakly hydrogen-bonded to the Schiff base in the wild-type, forms a hydrogen-bond with both the Schiff base and Asp-85 in L_{550} [98, 81]. In addition, water molecules were observed to form hydrogen-bond chains suitable for proton transport, e.g., between Arg-82 and Glu-204. The exact number, location and orientation of water molecules suggested in these studies, however, is by no means definitive.

As described in chapter 3, the bR proton pump cycle proceeds through several intermediate states characterized by their absorption maxima, which fall into two categories: early intermediates, eqn. (3.1), and late intermediates, eqn. (3.2). Our focus in this chapter is on the intermediates in eqn. (3.1). The key determinant for bR's photoreaction $bR_{568} \rightarrow K_{590}$ is the excited state potential energy surface of retinal and its crossings with the ground state surface. Despite the ubiquitous

occurrence of photoisomerization processes in polyene-type compounds, strikingly little is known about the potential surfaces involved, neither the number of states contributing surfaces nor the shape of the surfaces. This situation is compounded by the fact that polyene electron systems pose a formidable challenge to quantum chemistry due to the highly correlated nature of the involved electronic states, which requires extended multi-electron basis sets for suitable descriptions [73]. *Ab initio* quantum chemical methods, at present, cannot determine the relevant excited state potential energy surfaces for electron systems of the size found in retinal. A combined QM/MM technique such as the QCFF/PI method, which has been applied to the study of the photoisomerization event in rhodopsins [181, 182], does allow determination of an excited-state potential surface which takes into account the surrounding chromophore-protein interactions. While computationally feasible for a system the size of bR, it is hard to judge the accuracy of the resulting surfaces and non-Born-Oppenheimer terms; in this respect we need to keep in mind that a definitive calculation would require a multi-dimensional potential surface at an accuracy of about 1 kcal/mol at any point. A rational, straightforward approach, at present, is to use simple model potential surfaces in MD simulations in order to learn how such surfaces, together with sterical effects of retinal's binding site and effects of inertia, determine retinal's photoproducts.

We investigate below two excited state potential surfaces, both of which induce an isomerization around retinal's C_{13}-C_{14} double bond. One surface disfavors a concurrent isomerization around retinal's C_{14}-C_{15} single bond through a large energy barrier, the other surface permits such isomerization. The chosen excited state potential surfaces reflect two models proposed for the photoisomerization process of retinal in bR, an all-*trans* $\rightarrow$ 13-*cis*, 14-*trans* reaction [156, 53], or an all-*trans* $\rightarrow$ 13,14-*dicis* reaction [150, 144, 121, 149, 58].

4.1.1 Outline of Chapter

In this chapter we address then the following questions connected with the intermediates in eqn. (3.1): How do features of the excited state potential surface determine the photoisomerization products? How does the binding site guide retinal's photoisomerization? How does the initial photoprocess

depend on initial conditions? How does water participate in the early intermediates? Can one recognize a J_{625} intermediate? What distinguishes the K_{590} and L_{550} intermediates? How do the initial steps depend on temperature? In section "Methods" we introduce the excited state potential surfaces employed in our simulations as well as molecular dynamics and annealing methods used. We also introduce a notation characterizing the different simulations carried out. In section "Results" the simulations of photoproducts up to the L_{550} stage are presented for the two potentials used. In section "Discussion" the results are analyzed in terms of the known bR structure, the involvement of waters and proton transfer paths. "Conclusions" summarizes the key results of this chapter and suggests future studies.

4.2 Methods

Molecular dynamics simulations reported here are based on the refined and equilibrated structure of bR reported in Humphrey *et al.* [74] derived from the structure reported in Henderson *et al* [66]. Following Humphrey *et al.* the present description of bR involves explicit hydrogens, includes sixteen water molecules placed and equilibrated within the protein interior and assigns standard protonation states to all side groups, except to Asp-96 and Asp-115, which are assumed to be protonated [10, 189]. The water molecules are modeled using TIP3P parameters [77]. The program X-PLOR [26] with the CHARMm force field [21] was used for all simulations. bR was modeled in vacuum at a temperature of 77 K and 300 K. A cut-off distance of 8 Å and a dielectric constant of $\epsilon = 1$ were used for the evaluation of Coulomb forces. All simulations use the standard X-PLOR protein topology file `topallh6x.pro` and parameter file `parmallh3x.pro` to model bR. The retinal topology and parameters for the equilibrium (bR_{568}) configuration are the same as used in Humphrey *et al.* [74], with the exception of the C_{15}-C_N dihedral energy barrier, which was increased to 30 kcal/mol to inhibit rotation about this bond during the isomerization process. Table 4.1 lists the partial charges assigned to retinal atoms; these charges were determined with Gaussian92 [54], using a Mulliken populational analysis at the MP2/6-31G level.

The potential surfaces assumed in the modeling of the initial photoisomerization of retinal

atom	charge	atom	charge	atom	charge
N	-0.47	H_{SB}	0.38	C_{11}	-0.09
H	0.31	C_1	-0.07	C_{12}	-0.18
C	0.51	C_2	-0.26	C_{13}	0.13
O	-0.51	C_3	-0.28	C_{14}	-0.22
C_α	0.07	C_4	-0.33	C_{15}	0.20
H_α	0.09	C_5	0.02	C_{16}	-0.40
C_β	-0.18	C_6	0.12	C_{17}	-0.40
C_γ	-0.18	C_7	-0.17	C_{18}	-0.48
C_δ	-0.18	C_8	-0.14	C_{19}	-0.49
C_ϵ	-0.04	C_9	0.11	C_{20}	-0.49
N_{SB}	-0.63	C_{10}	-0.19		

Table 4.1: Retinal partial charges.

differ in their dependence on the C_{14}-C_{15} dihedral angle. Earlier studies [116, 189] had employed for this purpose a schematic potential which combined the excited state and the ground state into a single surface. In the present study, the three phases of photoisomerization, shown schematically in Fig. 4.1, namely, excited state dynamics (phase I), surface crossing (phase II), and ground state dynamics (phase III), are described through three separate surfaces. The potential for phase I, governing steps (a) and (b) in Fig. 4.1, is modeled through a surface with a maximum at the *trans* and *cis* positions of the C_{13}-C_{14} dihedral angles and a minimum at the 90° twist of this bond. The non-adiabatic crossing from the excited state to the ground state potential surface in phase II, i.e., step (c) in Fig. 4.1, is modeled through a C_{13}-C_{14} bond dihedral angle potential with a single minimum at 13-*cis*. For phase III, i.e., step (d) in Fig. 4.1, the conventional dihedral potential is reinstated.

The surfaces are modeled through simple analytical expressions governing the dependence of the energy on the dihedral angles ϕ_1 and ϕ_2 of retinal's C_{13}-C_{14} and C_{14}-C_{15} bonds, respectively. For this purpose additive contributions $E_1^{dihe} + E_2^{dihe}$ were assumed, with the two terms given by

$$E_i^{dihe} = \frac{1}{2} k_i \left[1 + \cos\left(n_i \phi_i + \delta_i\right)\right] \tag{4.1}$$

Table 4.2 provides the potential energy parameters for the ground state (phase III) and Table 4.3 provides the parameters for phase I and II potentials. Figures 4.2 and 4.3 illustrate the potential energy surfaces. The potentials are chosen such that the transition from the initial ground state

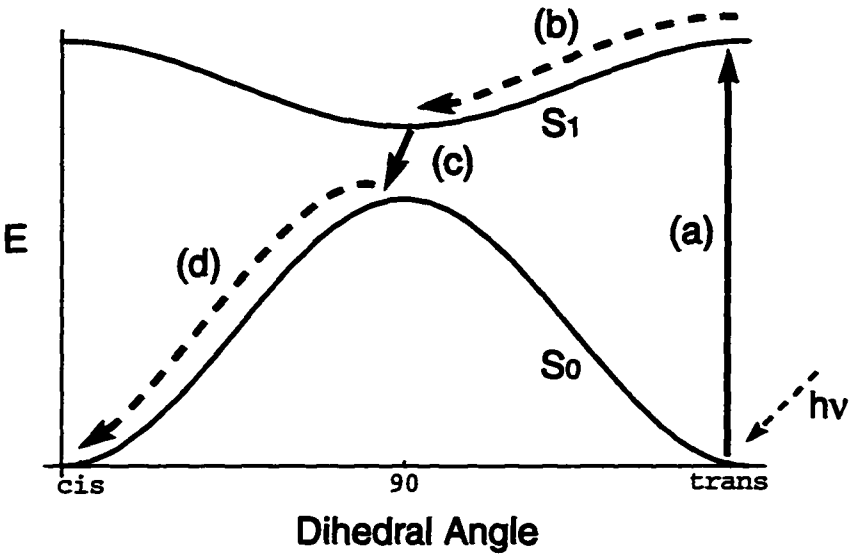


Figure 4.1: Schematic representation of the photoexcitation of retinal and subsequent isomerization. bR absorbs a photon $h\nu$ and is excited from its electronic ground state S_0 to its excited state S_1. (a) S_1 has a potential surface with a minimum near the S_0 maximum. After promotion to the S_1 surface retinal moves to the minimum of this surface within 200-300 fs. (b) Retinal then crosses back to the ground state, i.e., S_0, surface. (c) On the latter surface retinal completes the *trans-cis* isomerization (d).

(phase III) potential to the phase I potential, i.e., step (a) in Fig. 4.1, imparts on retinal (in an all-*trans* geometry) 50.4 kcal/mol which corresponds to the energy of a 568–nm photon. It is important to note that this energy, according to our model, is stored at the beginning of phase I solely in the degree of freedom of torsional rotation about the C_{13}-C_{14} bond.

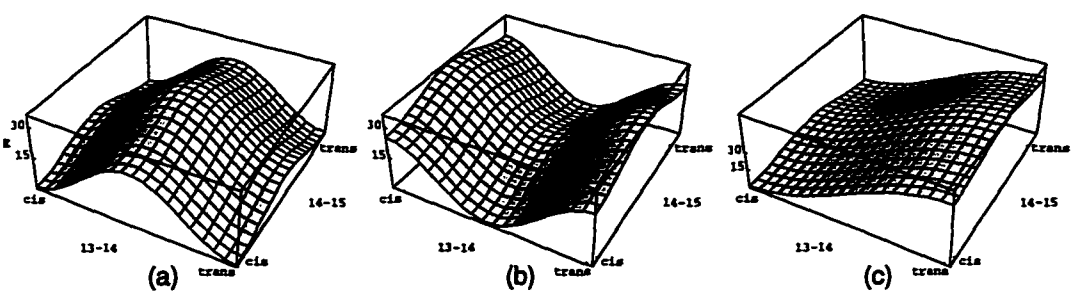


Figure 4.2: 13-*cis* potential energy surface for C_{13}-C_{14} and C_{14}-C_{15} dihedral angles. (a) Equilibrium surface (phase III). (b) Inverted-potential surface (phase I). (c) Single-minimum surface (phase II).

To investigate a possible dependence of the photoisomerization on the initial state, 50 independent trials, characterized through independently chosen random velocities of all protein atoms, were

k_i (in kcal/mol)	n_i	δ_i (in deg)	
C_5-C_6-C_7-C_8	5.0	2	176.6
C_6-C_7-C_8-C_9	30.0	2	176.6
C_7-C_8-C_9-C_{10}	5.0	2	176.6
C_8-C_9-C_{10}-C_{11}	30.0	2	176.6
C_9-C_{10}-C_{11}-C_{12}	5.0	2	176.6
C_{10}-C_{11}-C_{12}-C_{13}	30.0	2	176.6
C_{11}-C_{12}-C_{13}-C_{14}	5.0	2	176.6
C_{12}-C_{13}-C_{14}-C_{15}	25.2	2	176.6
C_{13}-C_{14}-C_{15}-N_{SB}	10.0	2	176.6
C_{14}-C_{15}-N_{SB}-C_ϵ	30.0	2	176.6

Table 4.2: Equilibrium (phase III) parameters used for the torsional potentials of retinal.

	Phase I				Phase II		
	k_i (in kcal/mol)	n_i	δ_i (in deg)		k_i (in kcal/mol)	n_i	δ_i (in deg)
	13-*cis* isomerization				13-*cis* isomerization		
$i=1$:	-25.2	2	180.0	$i=1$:	50.4	1	180.0
$i=2$:	10.0	2	176.6	$i=2$:	10.0	2	176.6
	13,14-*dicis* isomerization				13,14-*dicis* isomerization		
$i=1$:	-25.2	2	180.0	$i=1$:	25.2	1	180.0
$i=2$:	1.0	2	176.6	$i=2$:	25.2	1	180.0

Table 4.3: Retinal dihedral angle parameters.

carried out. Each trial started from the same bR_{568} structure, but with different initial velocities chosen from a Maxwell distribution at 77 K or 300 K. Simulations at 300 K involved an initial 100 fs of equlibration dynamics with a phase III potential as shown in Figs. 4.2a. and 4.3a. Simulations at 77 K extended this equlibration to 5 ps, accounting for the need of a longer equilibration time since the structure of bR_{568} was actually determined at T = 300 K. Following the initial equilibration, for each trial the excited state (phase I) potential, as shown in Figs. 4.2b. and 4.3b., was applied for 250 fs. After this period the potential governing phase II of the photoprocess, as shown in Figs. 4.2c. and 4.3c., was applied for 250 fs. Finally, the ground state (phase III) potential function was restored and 4.5 ps of further dynamics completed. Restart files were used between each switch of potential functions to preserve the dynamical state.

Since the $K_{590} \rightarrow L_{550}$ transition requires a microsecond time period, which cannot be covered by molecular dynamics simulations, we employed simulated annealing [173], using representative K_{590} structures (see below) as starting points. Simulated annealing is often applied at very high

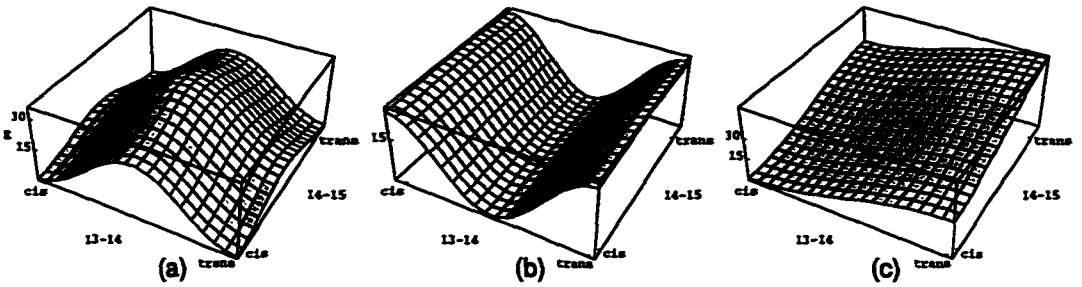


Figure 4.3: 13,14-*dicis* potential energy surface for C_{13}-C_{14} and C_{14}-C_{15} dihedral angles. (a) Equilibrium surface (phase III). (b) Inverted-potential surface (phase I). (c) Single-minimum surface (phase II).

temperatures, typically 1000-4000 K [24]. Such high temperatures were found to adversely affect the structure of bR and, therefore, the maximum annealing temperature was limited to 500 K. The SHAKE algorithm [139] for bond-length constraints was used for simulations at T > 300 K, in order to maintain stable numerical integration with a 1 fs timestep. Since dihedral angles of the single bonds of retinal's backbone become too flexible at high temperatures, during the annealing process torsional barriers of 5 kcal/mol to 10 kcal/mol for the C_6-C_7, C_8-C_9, C_{10}-C_{11}, and C_{12}-C_{13} bonds of retinal were assumed to stabilize its geometry.

The temperature protocol used to determine the L_{550} intermediate is presented in Fig. 4.4. First, torsional potentials for retinal were changed as described, and a 100-step conjugate gradient energy minimization calculation was carried out. Then, starting at $T_0 = 500$ K, the system was simulated for 100 fs using temperature coupling to rescale velocities to the appropriate temperature; for this purpose a frictional force

$$\vec{F}_j = -m_j \vec{v}_j \gamma (T_0/T - 1) \tag{4.2}$$

with $\gamma = 100$ ps^{-1} was applied to each atom j with mass m_j and velocity $\vec{v}_j$. This was followed by a 50-step conjugate gradient energy minimization. The annealing calculations were repeated at 10 K intervals from 500 K to 300 K. The frequent velocity reassignment facilitated a wider search of conformation space and the energy minimization helped avoid instabilities. After a temperature of 300 K was reached, an analogous annealing cycle was started at 400 K, reducing temperatures

again to 300 K. The annealing calculations were then followed by a 15 ps equilibration phase at 300 K.

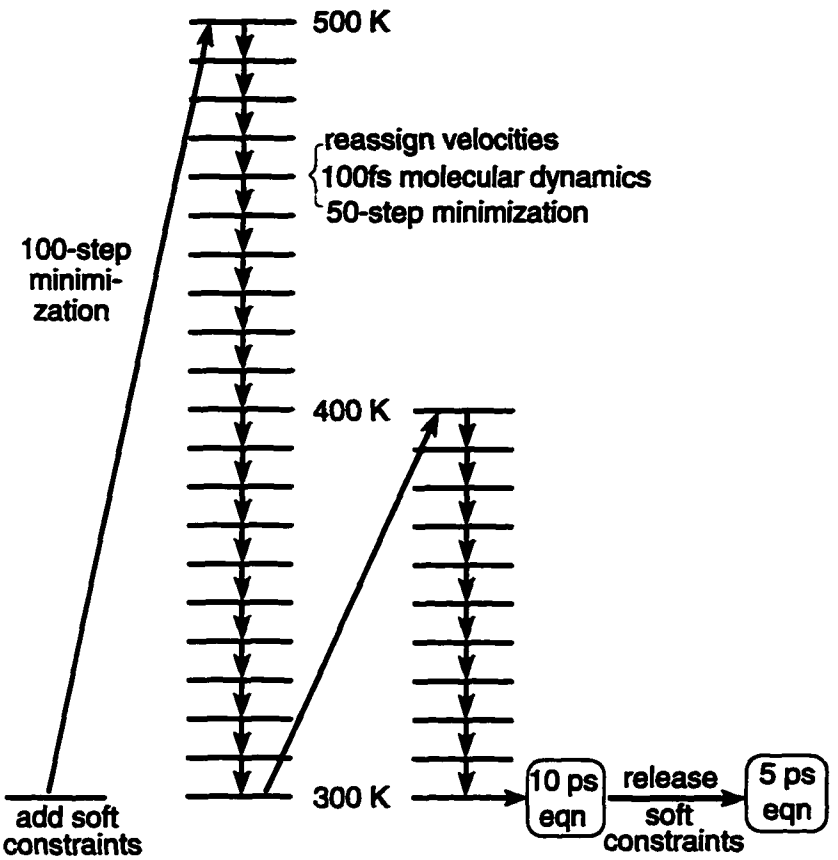


Figure 4.4: The temperature protocol of the simulated annealing process. The label "eqn" refers to a conventional molecular dynamics simulation at 300 K. Soft constraints refer to a change of torsional potentials of retinal which are invoked to keep retinal planar during the annealing process and are released thereafter.

Simulations $A1_{300}$ ... $A50_{300}$ and $A1_{77}$ ... $A50_{77}$ modeled the 13-*cis* photoisomerization and subsequent equilibration at 300 K and 77 K, respectively, while simulations $B1_{300}$ ... $B50_{300}$ and $B1_{77}$ ... $B50_{77}$ modeled the 13,14-*dicis* photoisomerization and equilibration at 300 K and 77 K. The numbers 1–50 refer to the specific trial run, and the subscripts 300 and 77 refer to the modeled temperature. One representative structure, from simulations $A1_{300}$ ... $A50_{300}$ and from $B1_{300}$ ... $B50_{300}$, referred to below as K_1 and K_2, was selected as a starting point for the annealing process outlined above to describe the 2 μs $K_{590} \rightarrow L_{550}$ transition. Simulation C, which resulted in structure L_1 starting from K_1, describes the 13-*cis* model; simulation D, resulting in structure

L_2 starting from K_2 describes the 13,14-*dicis* model.

4.3 Results

Figure 4.5 illustrates the key features of the binding site of bR for the initial structure used in all isomerization simulations. Within this structure, four water molecules (W_A, W_B, W_C, W_D) make direct hydrogen-bonding contact with the SBP, either before or after the isomerization process. W_A is directly hydrogen-bonded to the SBP in the starting structure. W_B is located above the SBP toward the cytoplasmic side of the membrane, and is part of a chain of three water molecules (W_{B1}, W_{B2}, W_{B3}) connected to Asp-96. W_C is in the region between the SBP and Asp-85, and makes a hydrogen-bond with Thr-89 (not shown), as well as with Asp-85 and another water molecule. W_D is near the SBP, outside of hydrogen-bond contact, but still below the Schiff base toward the extracellular side of the membrane. W_D is involved in a hydrogen-bond with W_A and also with Tyr-185.

It was observed in our simulations of excited state and potential crossing dynamics that the orientation of the N-H^+ Schiff base bond, after isomerization and equilibration, assumed basically one of three possible configurations: N-H^+ pointing "up", i.e., toward the cytoplasmic side of bR, N-H^+ pointing "down", i.e., toward the extracellular side of bR, and N-H^+ pointing roughly perpendicular to the membrane normal. To measure this orientation, we define θ_{SB} as the angle between a line formed by N-H^+ and a line connecting the Schiff base nitrogen and the Asp-96 carboxyl: for small θ_{SB}, N-H^+ points toward Asp-96, while for $\theta_{SB} \approx 180°$ the orientation of N-H^+ is toward the extracellular side of the protein. For each set of 50 trials, the photoisomerization products were grouped into one of four cases based on the value of θ_{SB} and the configuration of the C_{13}-C_{14} and C_{14}-C_{15} dihedral angles. The definitions of these cases are given in Table 4.4. This table also lists the percentage of occurrence for simulations $A1_{300}$ ... $A50_{300}$, $B1_{300}$ ... $B50_{300}$, $A1_{77}$ ... $A50_{77}$, and for $B1_{77}$ ... $B50_{77}$ (see Methods for definitions of these simulations). Figure 4.6 provides illustrations of the key features of representative case 1 (Fig. 4.6a.), case 2 (Fig. 4.6b.), and case 3 (Fig. 4.6c.) structures.

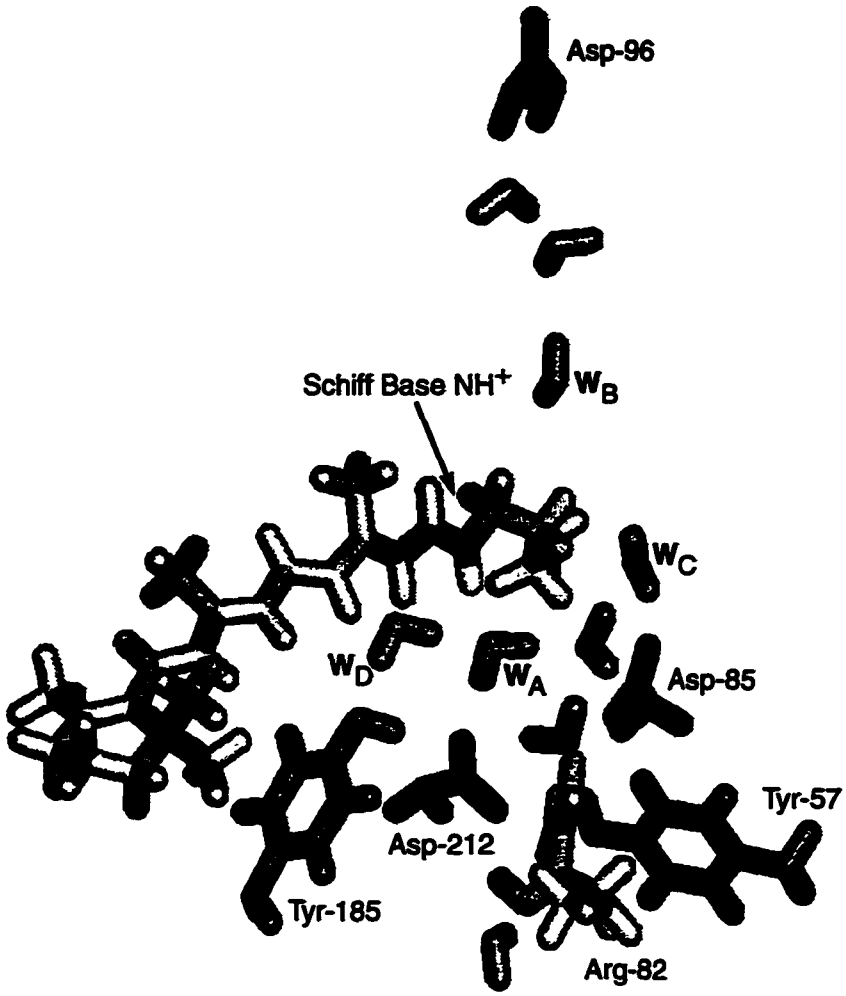


Figure 4.5: Structural features of the binding site for the starting structure used in all isomerization trials.

4.3.1 13-*cis* Isomerization Model

300 K

According to the textbook model of an all-*trans* $\rightarrow$ 13-*cis* photoisomerization, retinal, after initial light excitation, rotates about its C_{13}-C_{14} bond by 180° and leaves the N-H^+ bond oriented in the direction of Asp-96. However, Asp-96 acts as the proton donor rather than acceptor in the photocycle such that this orientation does not appear desirable from a mechanistic point of view. In 58 % of simulations $A1_{300}$ ... $A50_{300}$, this orientation of N-H^+ is observed, and retinal photoisomerizes into a basically planar 13-*cis* conformation. In 36 % of simulations $A1_{300}$ ... $A50_{300}$, however, the Schiff base photoisomerizes such that the N-H^+ bond points roughly perpendicular

Case	Definition	300 K		77 K	
		13-*cis*	13,14-*dicis*	13-*cis*	13,14-*dicis*
1	13-*cis*, $\theta_{SB} \leq 60°$	58	12	72	0
2	13-*cis*, $\theta_{SB} > 60°$	36	2	0	0
3	13,14-*dicis*, $\theta_{SB} > 90°$	6	28	28	76
4	all-*trans*, $\theta_{SB} > 90°$	0	50	0	22

Table 4.4: Definitions of cases used to categorize isomerization trials, and percentage of cases present in each set of simulations.

to the membrane normal, midway between being oriented toward Asp-96, and being oriented in the original all-*trans* direction (toward the Asp-85/Asp-212/water counterion complex). Retinal accommodates this conformation through a series of twists about its single bonds, particularly, through a large twist about its C_6-C_7 bond. The resulting N-H^+ orientation is stabilized by a strong hydrogen-bond formed between N-H^+ and water W_C, such that the Schiff base participates in a direct hydrogen-bond chain to Asp-85, the acceptor for retinal's proton during the $L_{550} \rightarrow M_{412}$ transition [19]. The remaining 6 % of simulations $A1_{300}$ ... $A50_{300}$ result in N-H^+ pointing toward the extracellular side leaving retinal in a 13,14-*dicis* conformation, despite the large (10 kcal/mol) barrier for rotation about the C_{14}-C_{15} bond.

Figure 4.7 shows the time evolution of various angles and distances which measure the configuration of retinal and surrounding residues during the photoisomerization. The quantities shown are averages for all trials in simulations $A1_{300}$ ... $A50_{300}$ which resulted in case 1 structures, at 50 fs intervals. The upper graph in Fig. 4.7 indicates the time evolution of the rotations about the C_6-C_7, C_{13}-C_{14} and C_{14}-C_{15} bonds, and the value of θ_{SB} defined above. The lower graph in Fig. 4.7 shows the average distance between the SBP and the Asp-85 and Asp-212 carboxyl groups. The lower graph also provides the distance from the SBP to waters W_A and W_B. The figure shows that the isomerization of case 1 trials completes within the first 500 fs, with θ_{SB} near 45°. Retinal experiences some twisting about the C_6-C_7 bond during the initial picosecond of simulation, but returns to a mainly *trans* geometry for this bond during the final 4.5 ps. The distance from the SBP to both Asp-85 and Asp-212 increases during the case 1 simulation. Water W_A breaks its hydrogen-bond with the Schiff base very early, and water W_B (in the region between the Schiff

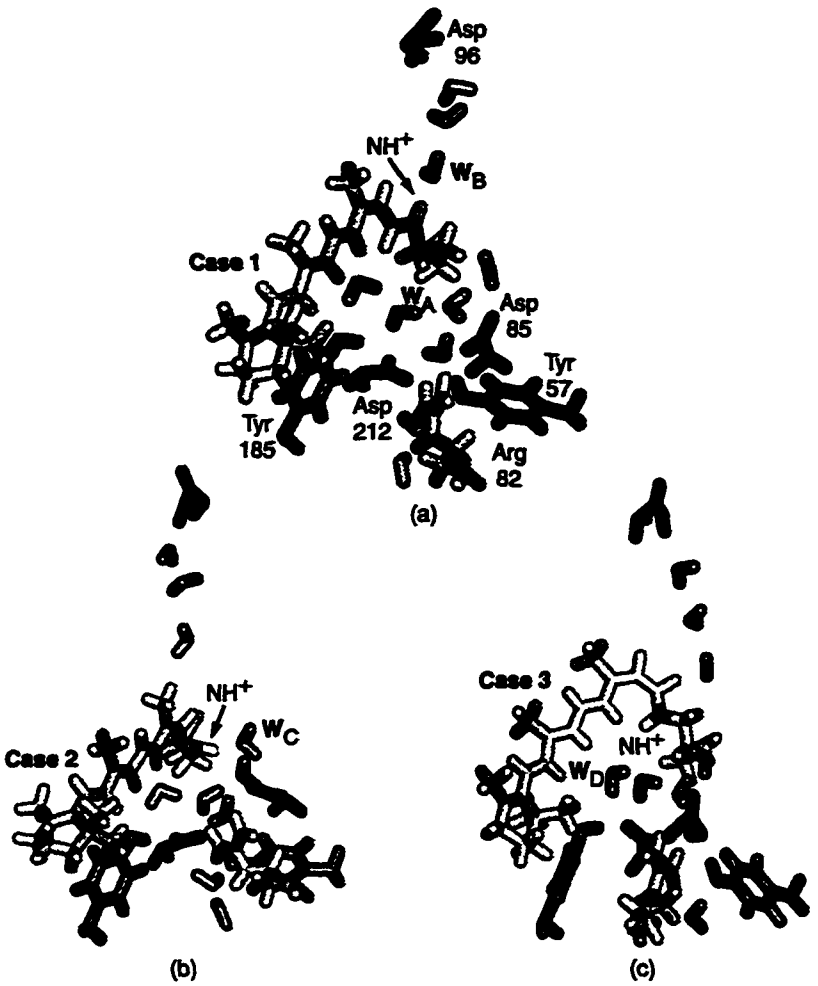


Figure 4.6: Photoisomerization products observed in simulations $A1_{300}$ ... $A50_{300}$ and $B1_{300}$ ... $B50_{300}$: (a) Case 1. (b) Case 2. (c) Case 3.

base and Asp-96) moves closer to the SBP. In some trials, this water molecule moves close enough to hydrogen-bond to the Schiff base. However, on average, this water does not get close enough to make hydrogen-bond contact by the end of 5 ps.

The averaged time evolution results for case 2 outcomes of simulations $A1_{300}$ ... $A50_{300}$ are shown in Fig. 4.8. In this case, the Schiff base forms a strong hydrogen-bond with water W_C, as clearly shown by the SBP–W_C distance. This hydrogen-bond stabilizes the N-H^+ bond in a direction with $\theta_{SB} \approx 90°$. In Fig. 4.8, the isomerization to 13-*cis* is seen to complete within the initial 500 fs, i.e., during phase I and II. However, retinal maintains a 15° to 20° twist about both the C_{13}-C_{14} and C_{14}-C_{15} bonds after isomerization. These twists, along with the observed

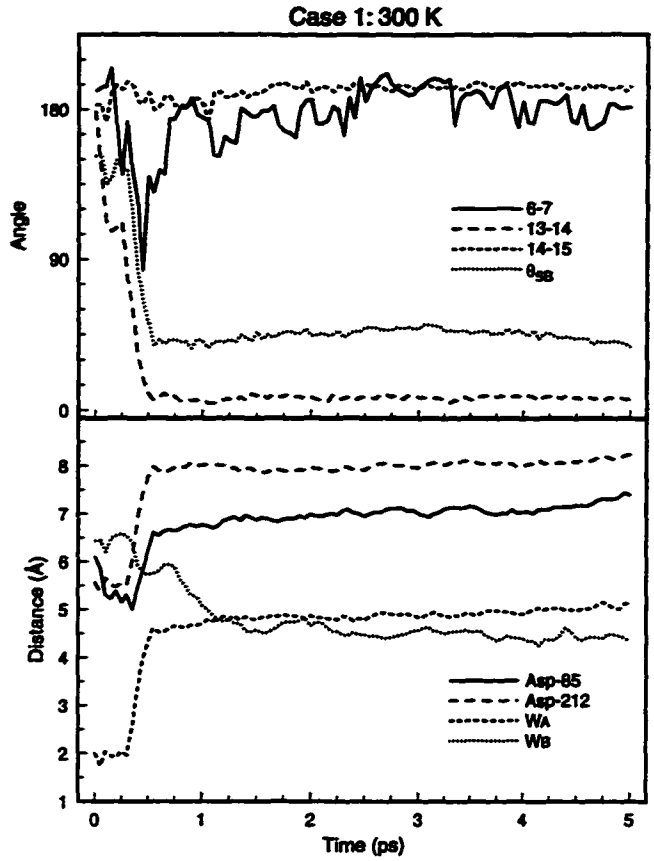


Figure 4.7: Averaged time evolution for case 1 trials in simulations $A1_{300}$ ... $A50_{300}$. The upper graph shows the values for retinal dihedral angles and θ_{SB} at 50 fs intervals, averaged over all trials classified as case 1. The lower graph gives the values for the distance of the aspartic acid carboxyl groups to the SBP, and the distance of water molecules W_A and W_B to the SBP.

large rotations about the C_6-C_7 single bond, allow retinal to maintain the N-H^+ bond oriented at a $\theta_{SB} \approx 90°$ angle. With retinal in this conformation, the distance from the SBP to Asp-85 decreases slightly by the end of the photoisomerization, whereas the distance to Asp-212 increases slightly.

77 K

Simulations $A1_{77}$ ... $A50_{77}$ studied the 13-*cis* isomerization reaction in the same manner as simulations $A1_{300}$ ... $A50_{300}$, but at 77 K. Experiments have shown that the K_{590} state can be trapped at this temperature, with little change in the spectral absorption maximum [11]. Table 4.4 lists the results for the simulations $A1_{77}$ ... $A50_{77}$. Of the 50 trials, 72 % result in case 1 structures, and the rest result in 13,14-*dicis* structures (case 3). In contrast to the simulations at 300 K, no

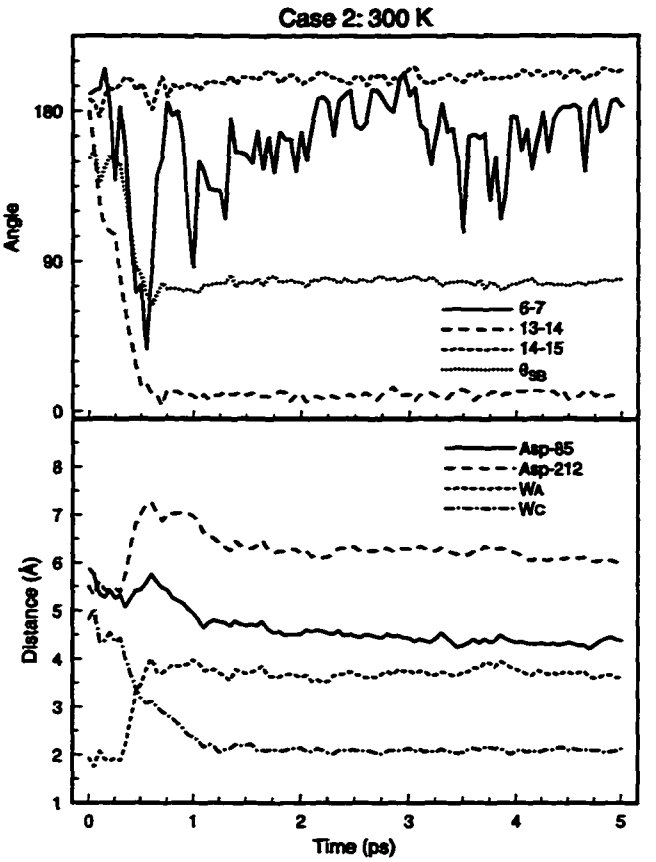


Figure 4.8: Averaged time evolution for case 2 trials in simulations $A1_{300}$ ... $A50_{300}$. The upper graph shows the values for retinal dihedral angles and θ_{SB} at 50 fs intervals, averaged over all trials classified as case 1. The lower graph gives the values for the distance of the aspartic acid carboxyl groups to the SBP, and the distance of water molecules W_A and W_C to the SBP.

case 2 photoproducts arise. Figure 4.9 shows the averaged time evolution for the case 1 structures with the 13-*cis* parameters at 77 K, the presentation corresponding to that in Figs. 4.7 and 4.8 for the 300 K case. Only the angular quantities are shown here, to exemplify the low temperature behavior as compared to the room temperature behavior (c.f. Fig. 4.7, top). The dynamics for case 1 photoproducts is the same at 77 K and 300 K, but with much smaller deviations from the mean for 77 K. The photoisomerization completes also within 500 fs. The resulting θ_{SB} is about 45°, indicating an N-H^+ bond pointing toward Asp-96. Small initial perturbations about the C_6-C_7 bond arise within the first picosecond; retinal assumes eventually a nearly planar conformation. A similar behavior is seen for trials leading to case 3 photoproducts.

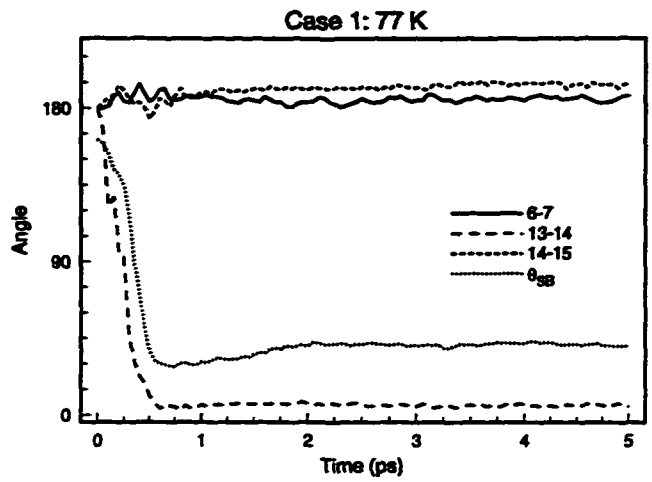


Figure 4.9: Averaged time evolution for case 1 trials in simulations $A1_{77}$ ... $A50_{77}$.

4.3.2 13,14-*dicis* Isomerization Model

300 K

The 13,14-*dicis* photoisomerization is modeled here in simulations $B1_{300}$ ... $B50_{300}$ by applying phase I and phase II potentials as described in Methods, and illustrated in Fig. 4.3. In 50 % of the simulations $B1_{300}$ ... $B50_{300}$, retinal did not complete the all-*trans* $\rightarrow$ 13,14-*dicis* photoisomerization, actually remaining instead in the all-*trans* (case 4) conformation. Of the remaining trials, 28 % resulted in 13,14-*dicis* retinal with N-H^+ pointing towards the extracellular side, i.e., resulted in case 3, and 14 % formed 13-*cis* (case 1 and case 2 structures). The remaining 8 % of simulations $B1_{300}$ ... $B50_{300}$ resulted in a highly distorted 12,14-*dicis* retinal. There is no experimental evidence for the latter photoisomerization product.

Only 50 % of the trials in simulations $B1_{300}$ ... $B50_{300}$ resulted in an isomerized retinal; of the runs which did succeed in converting native bR, over one half formed a 13,14-*dicis* product (case 3). Figure 4.10 shows the averaged time evolution for these case 3 runs, the presentation corresponding to that in Figs. 4.7 and 4.8. For this case, the isomerization completes within the first 500 fs. There arises some twist about the C_6-C_7 bond during the initial stage, but this twist is reduced during the ensuing 4.5 ps of simulation. The orientation of N-H^+, as measured by θ_{SB}, remains toward the extracellular side of bR, i.e., $\theta_{SB} > 90°$. There is little to no twist about the

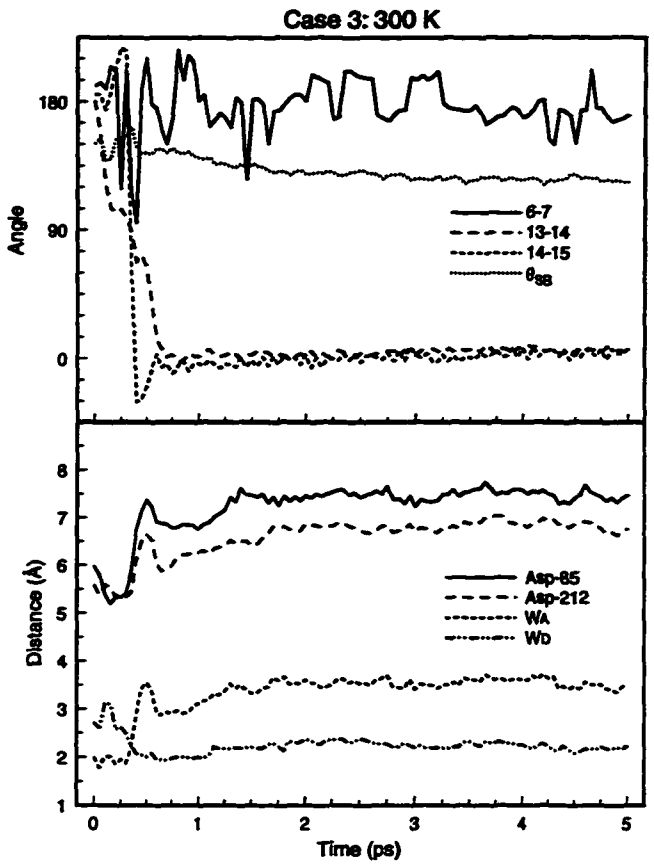


Figure 4.10: Averaged time evolution for case 3 trials in simulations $B1_{300}$ ... $B50_{300}$. The upper graph shows the values for retinal dihedral angles and θ_{SB} at 50 fs intervals, averaged over all trials classified as case 1. The lower graph gives the values for the distance of the aspartic acid carboxyl groups to the SBP, and the distance of water molecules W_A and W_D to the SBP.

C_{13}-C_{14} and C_{14}-C_{15} bonds after the first picosecond. In Fig. 4.10, it is seen that both aspartic acid groups end up further away from the SBP after the 5 ps simulation, each by approximately 1 Å. Water W_D replaces water W_A as the Schiff base hydrogen-bond partner, and W_A moves further away from the Schiff base.

77 K

Molecular dynamics simulations $B1_{77}$ ... $B50_{77}$ were carried out at 77 K employing also the potential surfaces favorable for an all-*trans* → 13,14-*dicis* photoisomerization. The results are summarized in Table 4.4 as well. Of the 50 trials, 76 % result in case 3 structures, with retinal in the expected 13,14-*dicis* isomeric state and the N-H^+ bond pointing to the extracellular side. Of the rest, 22 %

of the simulations did not complete the isomerization and returned to the all-*trans* state. In one case (2 %) a 12,14-*dicis* product also was observed. None of the simulations produced a 13-*cis* product. This is due to the effect of the binding cavity which favors an all-*trans* $\rightarrow$ 13,14-*dicis* photoisomerization. As in simulations $A1_{77}$ ... $A50_{77}$, the motion of retinal for the low temperature simulations $B1_{77}$ ... $B50_{77}$ exhibits less deviation from the mean than the 300 K simulations. Key differences between simulations $B1_{77}$ ... $B50_{77}$ and $B1_{300}$ ... $B50_{300}$ are the much larger number of case 3 structures, and the lack of 13-*cis* (case 1, 2) for 77 K.

4.3.3 Determination of L Structure

As described in Methods, simulated annealing was employed to bridge the 2 μs time scale, the respective simulations still being extremely time consuming. Therefore, only a single case 2 structure, termed K_1, was selected from $A1_{300}$ ... $A50_{300}$, and a single case 3 structure, termed K_2, was selected from $B1_{300}$ ... $B50_{300}$ as a starting point for two annealing calculations. The structures K_1 and K_2, taken from the final state after the 5 ps isomerization/equilibration dynamics, are considered candidates for the K_{590} state and, as starting points in simulations C and D yield candidates for the L_{550} intermediate. The respective structures are denoted as L_1 and L_2. Figure 4.11 presents the configuration of the bR binding site region for L_1 and L_2.

During simulation C, the N-H^+ bond maintained its orientation perpendicular to the membrane normal, as indicated in Table 4.5. This table compares values for the orientation of the N-H^+ bond and for the distances from the SBP to the Asp-85/Asp-212 carboxyls in K_{590} and L_{550}. One can see that the nearby aspartic acids maintained nearly the same relative distance to the SBP during this simulation, while the Asp-212 distance decreased by only 0.4 Å. During simulation D, the N-H^+ bond remained pointing to the extracellular side of bR, indicating that retinal in both the K_{590} and the L_{550} intermediate maintained a 13,14-*dicis* configuration. However, during simulation D the separations between the SBP and the aspartic acid carboxyls of Asp-85 and Asp-212 decreased significantly; for example, the Asp-212 carboxyl moved 2.4 Å closer to the SBP.

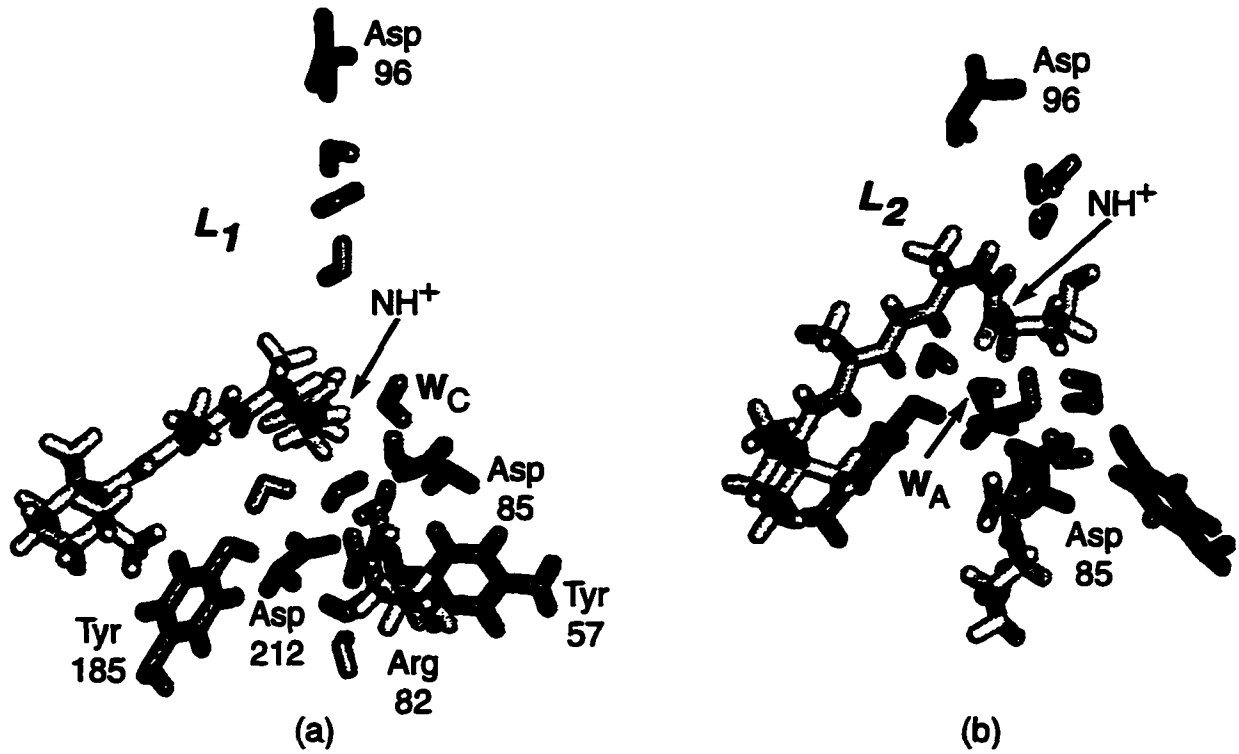


Figure 4.11: Retinal binding site region structural features for the simulated L intermediates: (a) L_1. (b) L_2.

4.4 Discussion

Simulations of the bR photocycle have been carried out previously, using both molecular dynamics methods [116, 189] and combined QM/MM methods [181, 182]. The present study differs in several respects from the earlier MD investigations by Nonella *et al.* and Zhou *et al.*, namely, in that we start from an all-atom bR structure with a new placement of water molecules [74], in that we employ a three phase description of the photoisomerization process, and in that we carry out a series of

	θ_{SB}	SBP – Asp-85	SBP – Asp-212
bR_{568}	158.2°	5.71 Å	5.42 Å
K_{590} (case 2)	79.4°	4.37 Å	6.00 Å
L_1	80.7°	4.17 Å	5.57 Å
K_{590} (case 3)	123.5°	7.47 Å	6.76 Å
L_2	117.6°	6.44 Å	4.36 Å

Table 4.5: Comparison of retinal configuration data between bR_{568}, K_{590}, and L_{550} intermediates from simulations C (13-*cis*) and D (13,14-*dicis*).

fifty simulations for each model or temperature tested. The present study differs also from the earlier QM/MM studies, by using a classical approximation for the photoisomerization process and by using as a starting the structure the refined bR_{568} structure described in Humphrey *et al.* [74]. Recently, molecular dynamics studies of the bR photocycle have been reported which also calculate the pK_a of the simulated intermediate structures [142, 48].

The quantum yield for formation of the K_{590} intermediate in the primary photoprocess of bR has been determined to be 0.64 ± 0.04 [143, 60, 166]. The quantum yield is independent of temperature down to 108 K [7]. This yield implies that a majority, but not all, of the trajectories describing the photoprocess of bR result in an isomerized retinal. To account for this behavior we carried out molecular dynamics simulations on schematic potential surfaces describing the ground and excited state of retinal, rather than enforcing an isomerization, i.e., a quantum yield of unity, as done in previous studies [189, 116]. Simulations were initially conducted with the phase II step (see Methods) eliminated, and the phase I duration extended or reduced to 200 fs, 300 fs, 400 fs, 500 fs, 700 fs, and 1000 fs. At most 5 % of the respective trials completed the isomerization to 13-*cis* or 13,14-*dicis*; most trials returned, instead, to the initial all-*trans* configuration after rotating to near 90° about the C_{13}-C_{14} bond in phase I. We therefore applied a second potential to bias, in phase II (i.e., step (c) in Fig. 4.1), retinal's torsion towards completion of the all-*trans* $\rightarrow$ 13-*cis* isomerization. Such a bias arises in the actual photochemistry if the crossing point of the S_1 and S_0 states is shifted toward the *cis* geometry, or if the actual surface crossing process favors the direction all-*trans* $\rightarrow$ 13-*cis*. Deflection of *trans* $\rightarrow$ *cis* isomerization reactions back toward the *trans* configuration has been found also in previous simulation studies [181, 182]. It is possible that the observed deflections in this study are the result of the particular potential energy surface used, which may not include all relevant torsional mode couplings.

The results of simulations with different initial conditions predict that the bR quantum yield may result from several similar photoisomerization reaction pathways, only a subset of which leads to a cycle which actually pumps protons. It is as yet not clear if the heterogeneity of photoisomerization products emerging in our simulations is an artifact, e.g., due to the schematic potential

surfaces used, or if these products correspond to multiple cycles all occuring at the same time, as has been suggested [36, 185, 45].

4.4.1 J Intermediate

Relatively little data have been acquired for the J_{625} intermediate, in part, due to its short lifetime of 5 ps [128]. Resonance Raman spectroscopy at 3 ps resolution has indicated that photoisomerization to the J_{625} state results in strong hydrogen-out-of-plane (HOOP) motion which decreases within 4 ps as the protein converts to K_{590} [38]. In all simulations which complete the photoisomerization, photoproducts appeared within the first 500 fs, i.e., within the time period over which phase I and II potentials were applied. The fast changes during this time affected hydrogen-bonds between the Schiff base and surrounding water molecules as well as twists of retinal's single and double bonds. Between 500 fs and 5 ps, i.e., the time of formation of the K_{590} intermediate, little structural change was seen.

The actual existence of a structurally distinct J_{625} state is still controversial. Recent work has hinted at the possibility that the J_{625} state intermediate arises due to a dipole moment induced in bR through the charge shift connected with retinal's photoexcitation [186]. This suggestion precludes the need for specific structural changes from J_{625} to K_{590}, and can explain why J_{625} cannot be trapped at low temperatures. Our simulations support this suggestion, due to the small amount of structural change seen in the 500 fs – 5 ps dynamics.

4.4.2 K Intermediate

In contrast to J_{625}, much is known about bR's K_{590} intermediate which is blue-shifted relative to the preceding J_{625} state, but is red-shifted relative to bR_{568}. This state is formed within 4 ps following the isomerization reaction, and has been determined to involve a 13-*cis* retinal. Our simulations were conducted for 5 ps, i.e., over a time period which matches the observed time of K_{590} formation. The structure at the end of the 5 ps simulation period is associated with the K_{590} state.

Another observed change from J_{625} to K_{590} is a decrease of the intensity of HOOP modes [38],

which corresponds to an increase in the planarity of retinal. All three photoproducts, i.e., cases 1–3, experience a reduction in the twist about the C_6-C_7 bond in the final 4 ps of simulation. Case 3 exhibits a definite reduction in the oscillation of this C_6-C_7 twist, while case 1 does not show much change after the initial isomerization. Case 2 displays the largest oscillations and overall twist of the C_6-C_7 bond; this is true for the other dihedral angles along the retinal backbone as well (not shown). In Fig. 4.8, the amplitude of these oscillations and the amount of twist for the C_6-C_7 bond decreases during the time from 500 fs to 3 ps, and then shows a slight increase in the 3 ps to 5 ps period. This might indicate that the K_{590} state has not stabilized for case 2 structures. Regardless, structural changes capable of explaining the observed HOOP modes in K_{590} compared to J_{625} are present in cases 2 and 3, and to a lesser degree in case 1.

An important feature of the K_{590} intermediate simulations is associated with the prediction that different structures are obtained at 300 K and 77 K. This is actually in keeping with experimental data which indicate different absorption maxima for the two species (590 nm and 603–606 nm, respectively) [129, 8], and different resonance Raman spectra [38]. Both resonance Raman and FTIR spectra indicate strong HOOP intensities in the spectra of K_{590} at 77 K, which are attributed to twists around single bonds [31] or to a C=N out-of-plane twist [55]. However, the simulations of K_{590} at 77 K indicate a planar conformation adopted by the retinal chromophore. This discrepancy might indicate that the simulations do not describe accurately the retinal conformation in K_{590} at 77 K, or that the HOOP intensity originates from another reason such as steric interaction with neighboring residues. Calculations of the resonance Raman spectra for rhodopsin [181] have demonstrated the need to take into account the surrounding protein conformation restraints on the chromophore, suggesting that the observed HOOP intensities result in part from chromophore-protein strain.

4.4.3 L Intermediate

While the very short time scales of the $bR_{568} \rightarrow J_{625} \rightarrow K_{590}$ transitions allow us to directly simulate these events, modeling of the significantly longer $K_{590} \rightarrow L_{550}$ transition (of about 2 μs) relied on the simulated annealing method. Use of this method implies a loss of time scale and we cannot

claim that simulations C and D above cover a complete 2 μs transition period. The simulations do, however, provide structures significantly evolved in time from the K_{590} intermediates and may be used to explore further the possible structural changes in the two photocycle models considered here.

In L_1, the orientation of the SBP perpendicular to the membrane normal, and the hydrogen-bond with water W_C, is maintained during the annealing. Most importantly, L_1 allows for proton transfer from the Schiff base to Asp-85: the network of hydrogen-bonds formed by water molecules in the binding site, to which both the SBP and Asp-85 are connected, provides a clear path to transfer the proton. The orientation of the Schiff base with its proton perpendicular to the membrane normal allows for it to connect to the hydrogen-bond chain, which would not be possible with the SBP pointing toward Asp-96. The decrease in the distance between the Asp-85 carboxyl and the SBP from bR_{568} to L_1 fits well with proton transfer to Asp-85, as opposed to Asp-212; the SBP – Asp-212 carboxyl distance does not change much from bR_{568} to L_1.

In L_2, both carboxyl groups from Asp-85 and Asp-212 move closer to the SBP by 0.5 Å and 2.1 Å, respectively (relative to K_{590}). For this 13,14-*dicis* L_{550} state, the SBP is still oriented toward the extracellular side of the membrane, and maintains a connection to the water molecule hydrogen-bond complex in the vicinity of the counterion. This allows the L_2 structure to contain a direct pathway for proton transfer from the Schiff base to Asp-85, as is also available in the L_1 structure. A large difference between L_1 and L_2, besides the isomerization state of retinal, is the distance of the SBP to the counterion residues. In L_1, Asp-85 is the closest group to the SBP, while in L_2, Asp-212 is much closer (by 2 Å).

Recent FTIR data suggest that structural changes occur in the vicinity of Asp-96 during the $bR \rightarrow L_{550}$ transition [19, 57, 97]. Our simulations reveal a weakening of a hydrogen-bond which exists between the Asp-96 carboxyl and Lys-41 both for L_1 and L_2. L_1 shows agreement with FTIR data on changes to the retinal binding site, which suggest stronger hydrogen-bonding to the Schiff base N-H^+ in L_{550} as compared to bR_{568} [95], and indicate that by the formation of L_{550} a water molecule is hydrogen-bonded with both the Schiff base and Asp-85 [98, 81] resulting in a distorted

13-*cis* chromophore. We found that in L_1, the hydrogen-bond interaction between the Schiff base and its proximal water molecule is considerably stronger than the Schiff base – water hydrogen-bond in bR_{568}, while the opposite is true for L_2. As stated above, this strong hydrogen-bond maintains in L_1 the orientation of the Schiff base. In L_2, a strong hydrogen-bond is not necessary to keep the SBP near the counterion region since the all-*trans* $\rightarrow$ *dicis* isomerization provides this orientation naturally.

4.4.4 Retinal Motion During Early Photocycle: Which Model Correlates Better With Experiment?

The key question arises, which photoproduct, case 1, case 2, or case 3, represents most closely the actual structure of bR during the early steps of the photocycle? For the two isomerization models, case 1 or 2 structures primarily result from 13-*cis* simulations, and case 3 or 4 structures primarily result from 13,14-*dicis* simulations. It is useful to consider the question posed for each model separately.

Case 1 photoproducts occur most frequently for the 13-*cis* model, both at 300 K and at 77 K. If one identifies the average structure at 500 fs with the J_{625} intermediate and that at 5 ps with the K_{590} intermediate, little change can be noted between J_{625} and K_{590}. The K_{590} structure also does not explain how the Schiff base can transfer a proton to Asp-85 at a later stage. Case 2 photoproducts, on the other hand, provide a pathway for this proton transfer, via water molecule W_C. The function of water W_C is crucial; our simulations demonstrate that W_C has a profound effect on the photoreaction of retinal, stabilizing the N-H^+ bond in a particular orientation and opening a pathway for proton transfer. The resulting L_{550} structure, i.e., L_1, also favors proton transfer to Asp-85, since Asp-85 draws closer to the SBP in going from K_1 to L_1.

The motion of retinal during the photoisomerization is shown in Fig. 4.12 at 500 fs intervals for the first 1.5 ps of a simulation which resulted in case 2 photoproducts. The figure demonstrates that single bond torsions participate in the isomerization such that retinal assumes eventually a strained configuration with the N-H^+ bond pointing sideways, i.e., neither to the cytoplasmic nor to the extracellular side. The final retinal geometry is stabilized through the interaction of the

SBP with water W_C. This strain may be of functional importance; it might be released only after proton transfer from the Schiff base to Asp-85 (and concomitant weakening of the interaction with W_C) such that in the M_{412} intermediate the N-H^+ bond points towards Asp-96 from where it receives a proton. Figure 4.13 summarizes the suggested binding site motion for the early intermediates $bR_{568} \rightarrow J_{625} \rightarrow K_{590} \rightarrow L_{550}$. The figure points to an important role played by water in our simulations: it stabilizes the strained retinal geometry of the early intermediates, it provides the proton transfer path from retinal to Asp-85, and it is suitably placed in the retinal – Asp-96 interstitial space to provide a transfer route for water in the later part of bR's proton pump cycle. It should be emphasized that waters were not placed in the refinement reported in [74] to accommodate the mechanism described, rather this mechanism emerged after the refinement had been completed.

It is of interest to contrast the early intermediates simulated at room temperature with those simulated at 77 K. While case 1 occurs with higher frequency than case 2 at 300 K, the absolute ratios might not be as significant as the fact that case 2 does not occur at all at 77 K, while case 1 occurs very often. This behavior might provide a clue as to why at low temperatures (180 K) bR irradiation does not produce a high yield of M_{412} following temperature elevation: the path to case 2 photoproducts might be blocked at low temperature such that only case 1 photoproducts develop. However, the case 1 photoproducts cannot form the M_{412} intermediate since they do not exhibit a proton pathway to Asp-85. In this respect, it is interesting to note that irradiation of bR at 180 K detected a significantly lower yield of M_{412} upon warming [80, 79]. It was suggested that at low temperatures there is a branching reaction at the L_{550} stage such that the formation of M_{412} is inhibited and L_{550} reverts thermally to the parent pigment. This might be due to pure temperature effects on the two alternative pathways, or to different K_{590} structures as suggested by our simulations. Alternatively, it is possible that M_{412} is not accumulated since the rates of its formation and decay are similar. The blockage of case 2 photoproducts might be due to a reduced mobility of the water molecules or of the surrounding residues. This suggestion is corroborated by the difference in the absorption maxima of the K_{590} intermediate at 300 K and 77 K, as well as by

the difference in the resonance Raman spectra of these two K_{590} intermediates. This suggestion is in conflict, however, with the observed quantum yield which is temperature-independent to as low as 108 K [7]; again, the quantum yield for the reactions to case 1 and case 2 structures could be similar, with the same fraction of bR returning to all-*trans* from the S_1 excited state.

Several mutants of bR have been studied in recent years, many of which contain reduced proton pumping activity, and some of which can be induced to pump protons in the opposite direction, for example, D85T [168]. It would be of interest to apply the simulations reported here to such mutants and investigate if the occurrence of case 2 structures correlates with the mutants' proton pumping activity. In particular, one would expect that the occurrence of case 1 structures, with N-H^+ pointing toward Asp-96, correlates with proton pumping in the cytoplasmic direction. Respective simulations are currently in progress.

We finally compare the behavior of retinal governed by an excited state potential favoring 13,14-*dicis* photoproducts. For this potential, a large percentage of simulations failed to isomerize at all; of the remaining cases, only half the photoproducts assumed a 13,14-*dicis* geometry. The 13,14-*dicis* structure itself (case 3) certainly contains attributes which fit well with proton transfer to Asp-85, since the SBP remains connected to the counterion region hydrogen-bond network. The simulated low temperature behavior suggests a strong increase of case 3 products which cannot be reconciled with observations. However, this alone would not argue strongly against the 13,14-*dicis* model. Our strongest argument against this model actually stems from a separate study [187] which showed that only an M_{412} intermediate formed from the L_1 intermediate induced the observed tilt of the F helix [162] and the observed shift of C_{20} of retinal [69].

4.5 Conclusions

This chapter is an attempt to delineate from the model of bacteriorhodopsin, which resulted from electron microscopy data [66], the mechanism of this protein. In earlier studies we had complemented the model through refinement efforts, adding loop regions of bR, optimizing side group and

helix placements and, in particular, adding internal water molecules. Molecular dynamics simulations reported here built on the refined model and showed how various properties of retinal and of its immediate protein environment can control the initial photoisomerization process. The simulations revealed that details of the excited state potential surface involving carbons C_{13}, C_{14} and C_{15} are crucial determinants of bR's photoisomerization products, the nature of which determines the mechanism of bR's proton pump cycle. Careful quantum chemical calculations of the potential surface, possibly of several closely lying surfaces, are critical for further progress in our understanding of the mechanism of bacteriorhodopsin; these calculations are presented in chapter 6.

The simulations presented here indicate also that the crossing from retinal's excited state to its ground state potential surface controls the quantum yield of bR's phototransformation. An excited state surface which is symmetric between the all-*trans* and 13-*cis* retinal isomers together with an instantaneous transfer to the ground state after 250 fs of excited state dynamics underestimates grossly the quantum yield of the all-*trans* $\rightarrow$ 13-*cis* photoisomerization. The observation of a very high quantum yield for the photoinduced backreaction $K_{590} \rightarrow bR_{568}$ is also indicative of an asymmetry in the potential function and, hence, in the crossing from the excited state to the ground state surface. In the present study it was necessary to enforce an asymmetric crossing through a phase II potential surface as depicted in Figs. 4.2 and 4.3. It is possible that the phase II potential needed to bring about sufficiently large quantum yields reflects the complex nature of the potential surface crossing in retinal which involves a variety of non Born-Oppenheimer terms and might require a detailed quantum mechanical description. Such a description has been provided, e.g., for the isomerization of *cis*-hexatriene [120].

An important outcome of our simulations has been the stochastic character of bR's photoreaction. Earlier descriptions [116, 189] had enforced a complete photoisomerization through a respective force field such that all simulations lead to a unique K_{590}-like intermediate. The use of a more genuine excited state potential surface in the present simulations allowed us to study systematically the effect of different initial conditions on the photoreaction. Of course, the well-known quantum yield of bR of 0.64 ± 0.04 [143, 60, 166], i.e., a value significantly less than unity, also indicates that

bR's photoisomerization has strong probabilistic attributes. Our simulations indicate, actually, that there do not only exist two outcomes of bR's photoreaction, i.e., all *trans* retinal corresponding to bR and 13-*cis* retinal corresponding to K_{590}, but rather four outcomes (see Table 4.4 and Fig. 4.6). Reports in the literature of inactive photocycles of bR [45] and the observation that bR mutants can produce proton currents directed towards the cytoplasmic side [168] appears to point also to a possible side reaction after the photoisomerization. Nevertheless, the different outcomes of the simulated photoreaction in our study have been a surprise which deserves further experimental and theoretical investigations. Such investigations will also require a better, i.e., accurate quantum chemical, description of the excited state potential surfaces of retinal.

Despite the uncertainties of the excited state potential employed here, a definite candidate for the K_{590} intermediate emerged from our simulations. This candidate, shown in Fig. 4.13, involves an N-H^+ bond pointing in a direction orthogonal to bR's long axis, stabilized by a water molecule which connects the SBP to Asp-85. This intermediate is ideally suited to realize the switch needed to explain the proton pump mechanism of bR: transfer of the SBP to Asp-85 weakens the interaction of the Schiff base nitrogen with the mentioned water such that torsional strains in retinal lead to a reorientation of the Schiff base, leaving the nitrogen pointing towards Asp-96.

If one accepts our suggestion for the identity of the K_{590} intermediate the conclusion can be drawn from our study that neither the J_{625} intermediate nor the L_{550} intermediate is structurally very distinct from the K_{590} state; during the time course of the $J_{625} \rightarrow K_{590} \rightarrow L_{550}$ transitions, as described in simulations $A1_{300}$ ... $A50_{300}$ and simulation C, retinal and the protein experience a structural relaxation, but no distinct and functionally significant transformation.

The most compelling result is the participation of water in the early stages of bR's pump cycle. A water molecule plays the pivotal role of stabilizing the early retinal intermediate in a strained geometry and furnishes a proton transfer pathway. Our investigation emphasizes the need to identify the location of water molecules in bR either through observation, e.g., through two-dimensional NMR spectra with water-amino acid cross peaks, or through improved modeling, e.g., through free energy perturbation calculations determining optimal water location in bR.

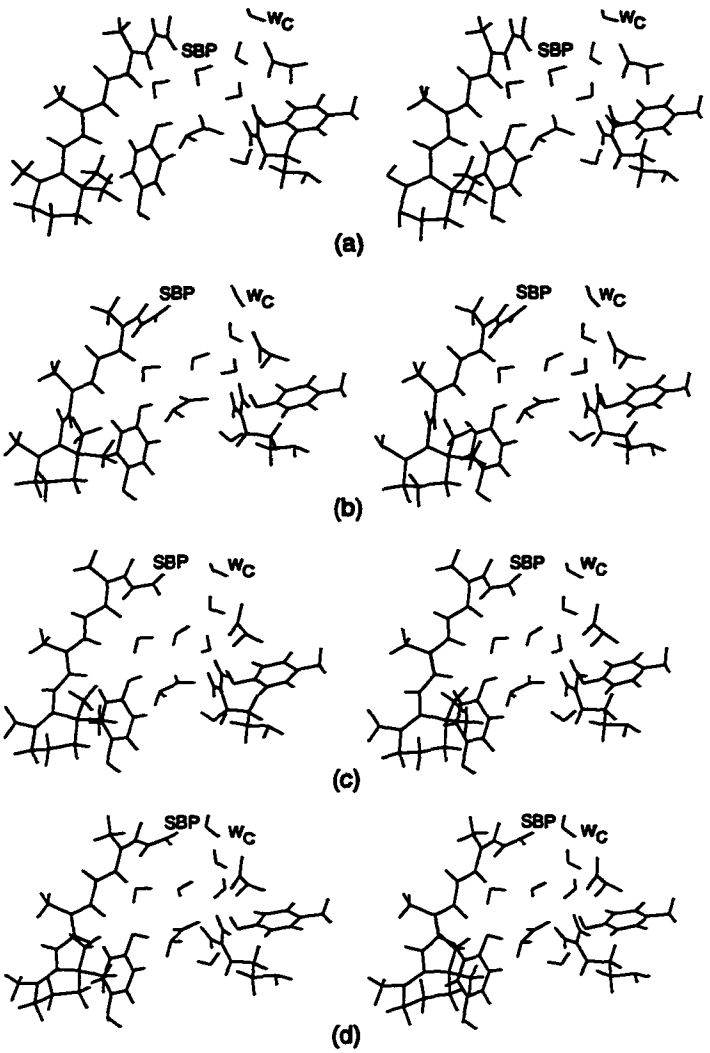


Figure 4.12: Stereo images of the retinal binding site during a case 2 photoisomerization simulation: (a) t = 0 ps. (b) t = 0.5 ps. (c) t = 1.0 ps. (d) t = 1.5 ps. During this interval, retinal converts from all-*trans* (a) to 13-*cis*, with a hydrogen-bond between the SBP and water W_C (d).

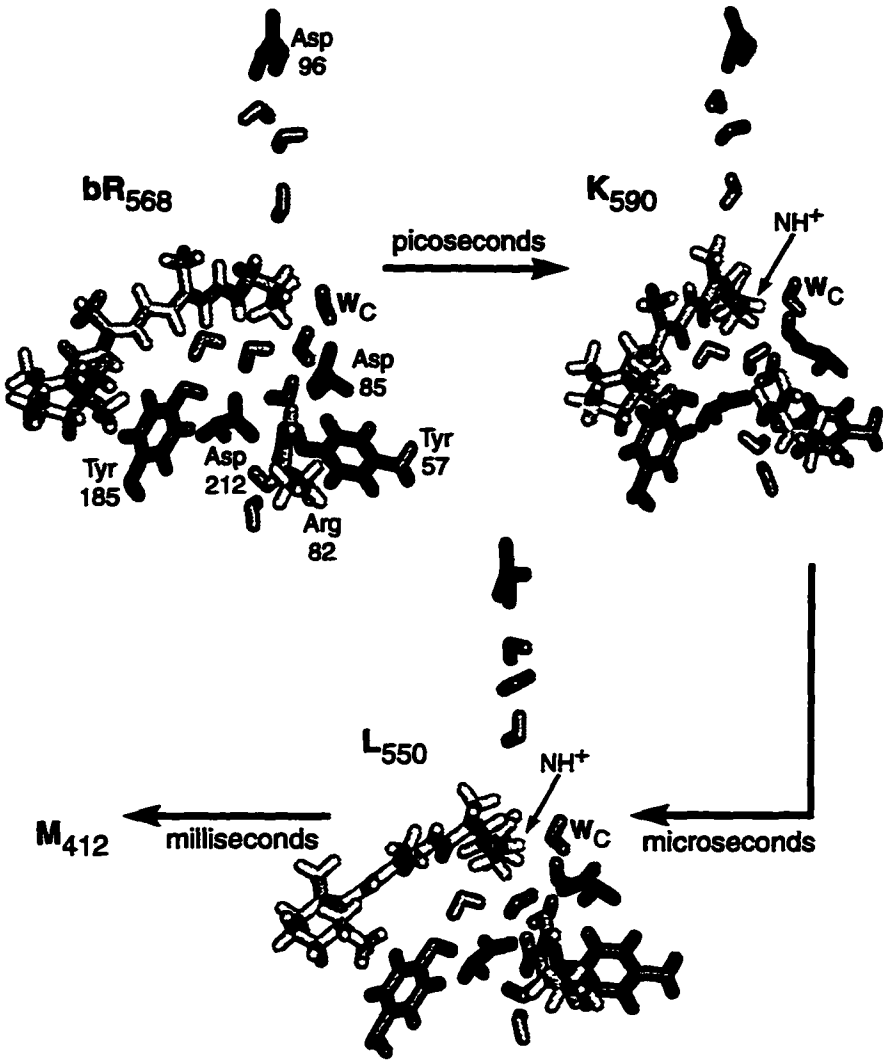


Figure 4.13: Suggested structures of early intermediates in the bacteriorhodopsin photocycle, $bR_{568} \rightarrow J_{625} \rightarrow K_{590} \rightarrow L_{550}$.

Chapter 5

Simulations of Bacteriorhodopsin Mutants[1]

5.1 Introduction

Site-directed mutagenesis experiments have contributed already much to our understanding of the function of bacteriorhodopsin. In these experiments, single amino acids are replaced by residues with distinct characteristics such as size, shape, or charge. Mutations are accomplished through modification and expression of the bR gene; several techniques for rapid and efficient production of arbitrary bR mutations have been developed [114]. In this chapter, we carry out molecular dynamics simulations of the early photoisomerization and subsequent relaxation event, eqn. (3.1), for several mutants of bR as well as for the native (wild-type) bR structure. The results are correlated with known proton pumping activity in the mutants and used to derive a model which can explain how certain mutants of bR can pump protons in the direction opposite to that of wild-type bR, as has been observed experimentally [9].

5.1.1 bR Mutagenesis Experiments

FTIR spectroscopy and observations of site-specific mutants of the bacteriorhodopsin wild-type structures have identified residues directly involved in the proton pump pathway, namely Asp-85 [107, 161, 122], Asp-212 [107, 122], Asp-96 [107, 72], and Glu-204 [23, 6, 136]. These findings

[1] Portions of this chapter have been taken from the manuscript: W. Humphrey, E. Bamberg, and K. Schulten, Photoproducts of Bacteriorhodopsin Mutants: A Molecular Dynamics Study, *submitted*.

have been corroborated by the three-dimensional structure of bR_{568} determined using electron cryo-microscopy [66] which revealed a channel involving the mentioned residues. Early observations of dehydration effects [85] as well as neutron diffraction studies [125], FTIR spectra [98] and molecular dynamics simulations [189, 74, 75] attribute also to water molecules an active role in the proton pump activity of bR. Figure 4.5 provides a view of the retinal binding site of bR_{568} which exhibits water molecules in the proton channel as well as key residues. Several water molecules are seen in Fig. 4.5 to form a network of hydrogen-bonds with charged and polar amino acids near the Schiff base linkage of retinal; three water molecules (W_A, W_B, W_C) interact directly with the Schiff base N-H^+, water W_A being directly hydrogen-bonded to the Schiff base in bR_{568}.

Even though the bR proton channel has been structurally identified, it is still an open question how protons are transferred irreversibly between the cytoplasmic and the extracellular side of bR. MD simulations of the picosecond [75, 186], nanosecond to microsecond [75], and later [187] intermediates of the pump cycle suggest that the irreversible transfer of protons requires the photoisomerization of retinal towards a torsionally deformed 13-*cis* geometry which is stabilized through a strong interaction of the Schiff base proton with a water molecule forming a bridge to Asp-85. If such a photoproduct is not formed, bR does not pump protons in the desired outward direction. Resolution of the crucial retinal geometry presently defies available experimental methods. In this chapter we resort to the study of various bR mutants by means of MD simulations to investigate in how far the occurence of the stated retinal photoproduct relates to the pump behavior observed for the mutants.

Experiments involving bR mutants D85{N,T,E,A} [107, 161, 122] have pointed to the role of the negatively charged Asp-85 as the initial acceptor of retinal's Schiff base proton, while studies of mutants D212{N,E,A} [107, 122] and R82{A,Q,K} [161, 122, 6] have shown these ionic groups near retinal to control the yield of the pump. Site-directed mutagenesis has also pointed to Glu-204 as the group responsible for accepting the proton from Asp-85, before releasing it to the extracellular domain [23, 6, 136].

Mutants of bR in which residues Asp-85 or Asp-212 are replaced by neutral or positively charged

groups, such as asparagine or threonine, exhibit dramatic differences in their proton pump behavior and, accordingly, were chosen as the subject of our investigation. At physiological pH, photoisomerization to early intermediates with close similarities to K_{590} and L_{550} of native bR arise, but proton pumping is abolished and an intermediate with a deprotonated retinal Schiff base is not formed [107, 111]. Mutation of Asp-85 in bR to a non-negative residue results in a significant change in the pK_a of the Schiff base, lowering it in many cases by 5–7 units [122]. Replacement of Asp-85 by asparagine or threonine can result in a reversal of the direction of proton transfer, when yellow and blue light is applied and two photons are absorbed in a cycle [168, 167]. The resulting pump action is similar to that of the homologous chloride-pumping protein halorhodopsin (hR), which contains a threonine residue at the position corresponding to Asp-85 in bR; employing two photons, hR can also pump protons inwardly [9].

5.1.2 Outline of Chapter

In this chapter we compare the photoisomerization products, as described by means of MD simulations, of wild-type bR and of bR's D85N, D85T, D212N, and Y57F mutants. We employ the same method as in our previous simulations of wild-type bR as described in the previous chapter [75]. In Methods we describe the construction of bR mutants studied and simulation of the photoisomerization reaction. In Results the outcomes of the photoproducts are presented. In Discussion, we compare these results to earlier simulations of the wild-type photocycle, and derive a model which can explain how the D85N and D85T mutants of bR can pump protons in the direction opposite to that of wild-type bR.

5.2 Methods

Molecular dynamics calculations described in this chapter were performed using the programs NAMD [113] and X-PLOR [26], both of which implement the CHARMm force field [21]. The simulations were carried out in vacuum at a temperature of 300 K, using a dielectric constant of $\epsilon = 1$ and a 12 Å cutoff of Coulomb forces. All calculations used the CHARMm-19 parameters and topology for the modeling of bR, with the exception of retinal. Simulations were carried out both with

and without an explicit hydrogen-bond term in the force field. The CHARMm-19 parameters were preferred over more recent releases to allow comparision with previous simulations of native bR. Simulations employing CHARMm-22 parameters did not exhibit major differences from simulations using CHARMm-19 parameters.

The previously reported structure of bR_{568} (Figs. 1.3 and 3.2), refined by means of molecular dynamics simulations [74], was used as a starting point for all simulations of bR_{568}, and served as the initial structure for the modeling of bR mutants. Following initial construction, each mutant structure underwent 1000 steps of energy minimization, followed by a 15 ps MD simulation with frequent velocity rescaling to heat the system to 300 K, and a 5 ps MD simulation at 300 K with infrequent velocity rescaling. Harmonic restraints were applied to the oxygen atoms of the water molecules during this equilibration to maintain their positions.

The retinal charges, parameters, and modeling procedures for the photoisomerization process employed presently were the same as employed in the previous study [75]. The photoisomerization procedure had been explained earlier as well in terms of ground and excited state potential energy surfaces evaluated quantum chemically [147]. Accordingly, retinal was induced to rotate about the C_{13}-C_{14} bond using the following procedure:

1. At $t = 0$, the conventional ground state energy surface governing torsion around the C_{13}-C_{14} bond was replaced by a potential surface describing the excited state with maxima at the all-*trans* and the 13-*cis* positions and a minimum at 90°.

2. At $t = 250$ fs, this potential surface was switched to one describing intersystem crossing and exhibiting a maximum at the all-*trans* position and a single minimum at 13-*cis*.

3. At $t = 500$ fs, the conventional ground state energy surface with minima at the all-*trans* and the 13-*cis* positions was re-installed for a period of 2.5 ps.

All simulations were carried out at $T = 300$ K. In all cases a barrier of 10 kcal/mol inhibited torsion around the C_{14}-C_{15} single bond.

Following the construction and equilibration of the initial (all-*trans*) structure for each mutant,

the systems were simulated for an additional 20 ps and the RMSD of each C_α atom determined. The resulting RMSD for wild-type bR and its D85N, D85T, D212N, and Y57F mutants exhibited average values ranging between 0.5 Å and 2 Å, reflecting the structural stability of the bR mutants. Similar results were obtained for the RMSDs of the positions of water molecules.

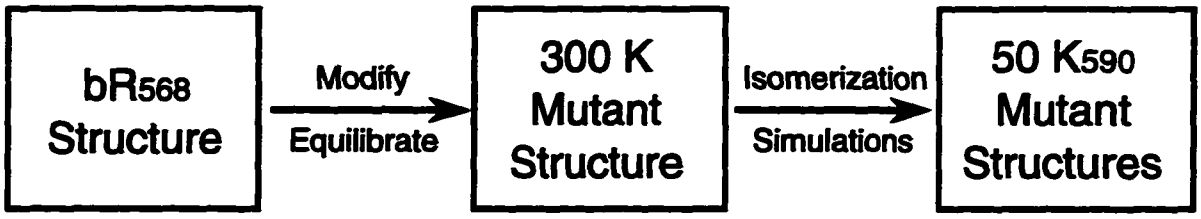


Figure 5.1: Schematic representation of the protocol used to construct and simulate the photoisomerization of bacteriorhodopsin mutants. The isomerization simulations followed the same schedule as used in Humphrey *et al.* [75]

MD simulations were then performed to model the initial photoisomerization reaction of bR_{568}. As in the previous study [75], 50 independent simulations were performed starting from the same protein configuration, but with different velocities chosen randomly from a Maxwell distribution at 300 K. Following the initial velocity assignment, 100 fs of dynamics were performed after which the simulated 3 ps photoisomerization reaction was carried out as described above. The overall procedure, summarized in Fig. 5.1, resulted in a set of 50 photoproducts for each mutant.

5.3 Results

The photoproducts resulting from each set of 50 simulated photoisomerization reactions for wild-type bR and its D85N, D85T, D212N, and Y57F mutants have been categorized based on the final configuration of the water molecules and the orientation of the N-H^+ bond. To measure the latter orientation we determined the angle θ_{SB} between the line through the N-H^+ group of retinal and the line connecting the Schiff base nitrogen with the carboxyl of Asp-96 (see Fig. 4.5). Four distinct sets of photoproducts, referred to below as case 1–4 products, were identified through the value of θ_{SB}: case 1 photoproducts contained a 13-*cis* retinal with a θ_{SB} of less than 60°, indicating a Schiff base proton oriented towards Asp-96; case 2 photoproducts also contained a 13-*cis* retinal,

but with a θ_{SB} value near 90°, indicating a twisted retinal with an N-H^+ bond oriented towards Asp-85; case 3 photoproducts contained 13,14-*dicis* retinal; case 4 photoproducts resulted when the isomerization failed to complete and retinal remained all-*trans*. The case 1 trials often contained a hydrogen-bond between the Schiff base and water W_B; all case 2 trials had a hydrogen-bond between water W_C and retinal.

	Case 1				
	bR_{568}	D85N	D85T	D212N	Y57F
w/HB Frequency[a]	.62	.78	.84	.98	.44
wo/HB Frequency[b]	.66	.86	.94	.98	.52
CW Rotation[a]	.00	.38	.67	.04	.00
E_{SB} (kcal/mol)[a]	-19.3	-6.5	-8.9	-7.7	-18.4
ΔE_{SB} (kcal/mol)[a]	15.0	14.3	15.0	14.6	15.3
	Case 2				
	bR_{568}	D85N	D85T	D212N	Y57F
w/HB Frequency[a]	.32	.00	.00	.02	.38
wo/HB Frequency[b]	.20	.02	.06	.00	.32
CW Rotation[a]	.00	-	-	1.00	.16
E_{SB} (kcal/mol)[a]	-32.2	-	-	-22.3	-31.3
ΔE_{SB} (kcal/mol)[a]	2.2	-	-	-0.4	2.2

[a] Results for simulations including an explicit hydrogen-bond interaction term.
[b] Results for simulations without an explicit hydrogen-bond interaction term.

Table 5.1: Summary of bR mutant simulation results.

Table 5.1 summarizes the results of the simulated photoisomerizations. In all mutants studied, case 1 and case 2 photoproducts (those containing a 13-*cis* retinal) accounted for at least 78% of all photoproducts, and only these cases are included in the table. The table entries state the observed fractions of case 1 and case 2 photoproducts as well as what percentage of the case 1 or case 2 photoproducts resulted from a clockwise isomerization, the rotation direction being defined looking from the Schiff base nitrogen toward retinal's β-ionone ring. A clockwise isomerization corresponds to a motion in which the Schiff base N-H^+ bond passes more closely by the Asp-85 residue rather than by the Asp-212 residue. E_{SB} given in Table 5.1 is the electrostatic interaction energy between the Schiff base proton and the rest of the protein, averaged over all photoproducts corresponding to a specific case. ΔE_{SB} measures the difference of the same electrostatic energy between the final

photoproduct and the initial state (i.e., with all-*trans* retinal).

The simulations of the bR photoisomerization reaction were carried out in two ways, the first including an explicit hydrogen-bond term in the molecular dynamics energy function, the second omitting this term. The values in Table 5.1 include the yield of case 1 and case 2 photoproducts for both sets of simulations, but provide the rotational motion and retinal energy data only for the simulations involving an explicit hydrogen-bond term. Inclusion of the hydrogen-bond term primarily affected the relative yield of case 1 and case 2 photoproducts for the bR wild-type and Y57F mutants; the remaining values reported in Table 5.1 were very similar in both series of molecular dynamics calculations.

5.3.1 bR Wild-Type

As a point of reference we consider first the results of simulated photoisomerizations for wild-type bR. We have extended previous simulations [75] in order to confirm the earlier findings and to insure proper comparison with the simulations of bR mutants discussed further below. In this study, 62% of the trials resulted in case 1 structures, while 32% of the trials resulted in case 2 structures; the remaining 6% of the cases failed to isomerize and remained all-*trans*. In the previous simulations [75], case 1 and case 2 photoproducts for the wild-type occurred with similar frequency (58% and 36%, respectively). Case 2 stuctures arise through a hydrogen-bonding with water W_C as shown in Fig. 5.2b, which helps stabilizes the twisted retinal geometry; this is highlighted by the reduction of the case 2 photoproduct yield in simulations without an explicit hydrogen-bond energy term. It the latter case, 66% of the trials form case 1 structures, while only 20% form case 2 photoproducts.

While case 1 structures occurred more frequently than case 2 structures, it was argued in the previous chapter (chapter 4, and ref [75]) that case 2 photoproducts actually represent the K_{590} intermediate which initiates a pumping cycle; case 1 structures lack a pathway for transfer of the Schiff base proton to Asp-85 and either do not arise upon excitation of bR_{568}, or they initiate a separate non-pumping (idle) cycle, characterized through the lack of a deprotonated retinal Schiff

base intermediate in the cycle. While the simulated photoisomerizations may not capture quantitatively the propensity of the various photoproducts, they have identified a viable candidate for the K_{590} structure and may, furthermore, properly describe how mutations affect the quantum yield of various photoproducts. In this respect it is important to note that combined classical/quantum mechanical simulations of the photoisomerization carried out recently (in preparation) and based on quantum chemically evaluated potential surfaces [147] yield analogous isomerization products with a significantly higher yield of case 2 products. Furthermore, the recent simulations result in more all-*trans* products, in accordance with the quantum yield of the photoreaction.

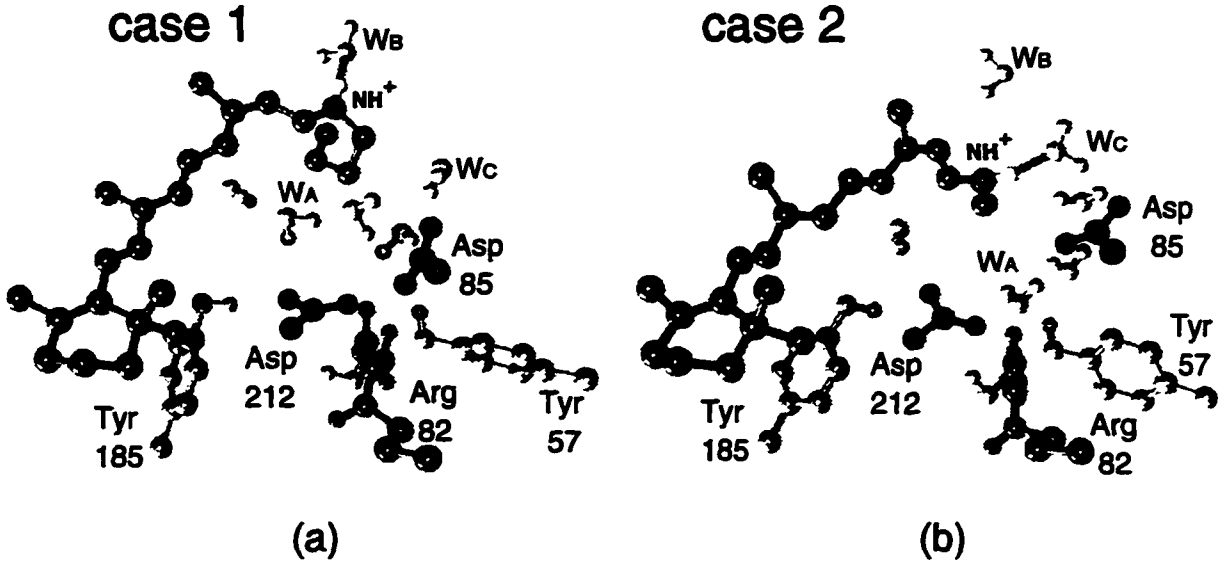


Figure 5.2: Structure of the retinal binding site of the case 1 and case 2 photoproducts of wild-type bacteriorhodopsin, including water molecules and key residues participating in the proton pump cycle. (a) Sample case 1 product, with the Schiff base proton connected to the cytoplasmic channel and hydrogen-bonded to water W_B; (b) sample case 2 product, exhibiting a Schiff base proton connected to the extracellular channel via a hydrogen-bond with water W_C.

Figure 5.2 presents two structures corresponding to case 1 and case 2 photoproducts of bR_{568}. One can recognize that in case 2 the Schiff base proton is favorably positioned to be transferred via W_C to Asp-85, whereas in case 1 the proton is pointing towards Asp-96, i.e., in the wrong direction. Energetic analysis reveals that for case 2 photoproducts retinal assumes a twisted, i.e., energetically unfavorable, geometry which is stabilized, however, through a strong hydrogen-bond

with water W_C.

The electrostatic interaction energy of the Schiff base proton (E_{SB}) for case 2 photoproducts lies 12.9 kcal/mol below that of case 1 photoproducts; ΔE_{SB}, for case 2 photoproducts, is close to zero (2.2 kcal/mol), while for case 1 photoproducts the Schiff base proton electrostatic energy increases by 15.0 kcal/mol. This difference arises from the Schiff base, in case 2, being closer to the negatively charged aspartic acids and to polar water molecules than in case 1 and also forming a stronger hydrogen-bond with water.

5.3.2 D85N and D85T Mutants

Modification of the negatively charged Asp-85, the primary proton acceptor in the wild-type photocycle, to a neutral residue such as arginine or threonine has been shown to cause a 5 to 6 unit drop in retinal's Schiff base pK_a [107, 161]. Simulations of D85 mutants described here were performed using a protonated Schiff base; thus, they relate to experiments done at sufficiently low pH to maintain the Schiff base in a protonated state.

Simulated photoisomerization of the D85N mutant resulted in a large number of case 1 photoproducts (78%), an example of which is shown in Fig. 5.3a. For this mutant, no case 2 structures arose, with the remaining 22% of photoisomerization processes resulting in either 13,14-*dicis* (case 3) or all-*trans* (case 4) retinal. When simulated without an explicit hydrogen-bond term, the D85N mutant exhibited an equally large frequency of case 1 photoproducts (86%), and an equally small yield of case 2 structures. The direction of rotation was primarily counter-clockwise, as for case 1 photoproducts of wild-type bR_{568}. Due to the loss of a negative charge on residue 85, E_{SB} is significantly more positive for the D85N case 1 photoproducts as compared to the value for wild-type case 1 photoproducts. However, the change in E_{SB} for case 1 D85N products, 14.3 kcal/mol, is very close to the corresponding value for bR_{568} case 1 products.

Simulated photoisomerization of the D85T mutant resulted in a distribution of photoproducts similar to that of the D85N mutant, with the difference being a slightly higher fraction (6%) of case 1 photoproducts. The same is true for the simulations without hydrogen-bond interactions, except

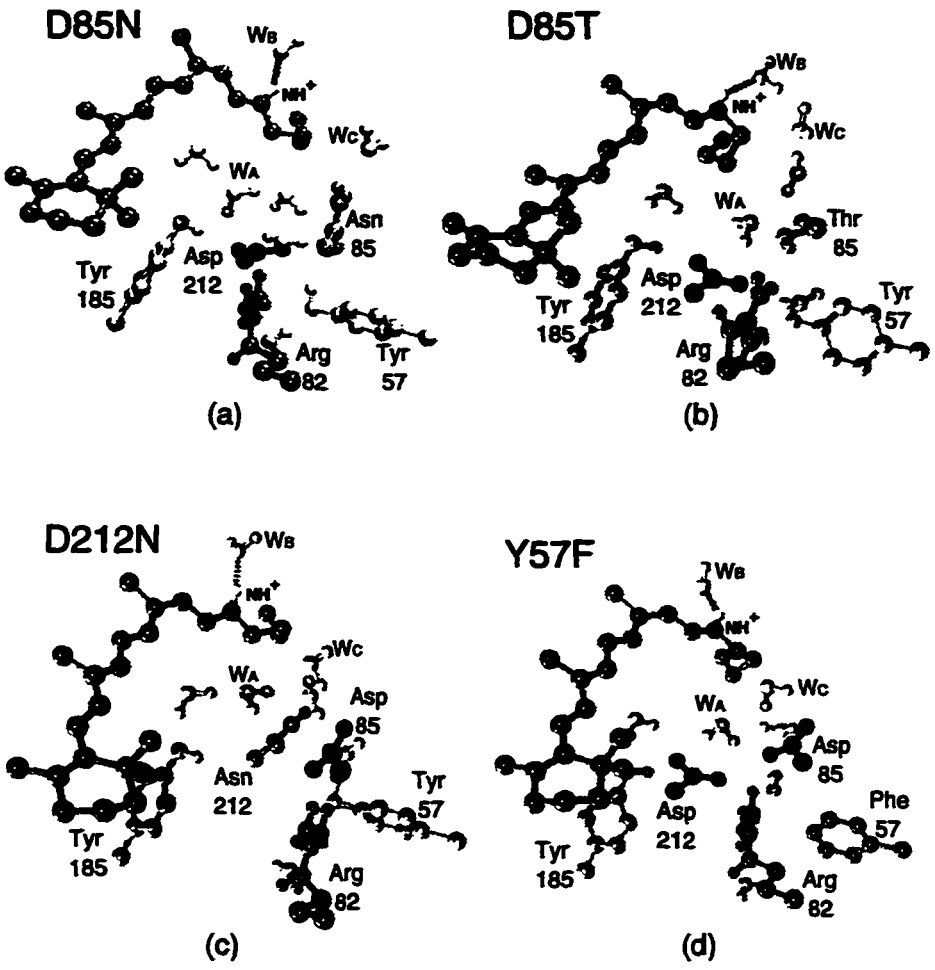


Figure 5.3: Structure of the retinal binding site of the case 1 photoproducts of several bR mutants. (a) D85N. (b) D85T. (c) D212N. (d) Y57F.

for a slight (6%) yield of case 2 D85T structures not seen in the calculations with a hydrogen-bond term. In D85T case 1 simulations, E_{SB} is more positive than the value of E_{SB} for bR_{568} case 1 photoproducts, which is due, in the same way as seen for D85N, to the loss of a negative charge on residue 85. Also for case 1 photoproducts in the D85T simulations, the value for ΔE_{SB} is close to those for bR_{568}; this result is comparable to the situation for D85N as well. The major differences between the D85T and D85N simulations are a preferred clockwise rotation direction in all D85T isomerization simulations, and a lower value for E_{SB}. Figure 5.3b shows a case 1 photoproduct of the D85T mutant.

5.3.3 D212N Mutant

Simulations of the D212N mutant of bR were carried out assuming a protonated Schiff base. This is reasonable, as mutation of Asp-212 to a neutral form does not significantly lower the Schiff base pK_a [161]. The D212N mutant is of interest since at neutral pH light absorption by the mutant leads neither to a deprotonated retinal Schiff base intermediate nor to an active pump cycle [107]. Simulated photoisomerization resulted for this mutant always in case 1 photoproducts, with the exception of a single case 2 occurence. No case 2 photoproducts arose in any of the D212N simulations which did not include a hydrogen-bond term. Similarly to wild-type and D85N photoisomerizations, but opposite to the behavior for the D85T mutant, case 1 photoproducts arose from counter-clockwise rotation. Figure 5.3c shows a case 1 photoproduct of the D212N mutant. Due to the loss of a negative charge in the counterion region as a result of the D212N mutation, the electrostatic energy of the Schiff base proton for case 1 photoproducts is comparable to that for the same photoproducts of the D85N and D85T mutants. ΔE_{SB} is also very close to the values for both the D85N and D85T mutants for the case 1 photoproducts.

5.3.4 Y57F Mutant

Of the mutations considered in this study, the Y57F mutation produces the smallest perturbation to the retinal binding site, a loss of two hydrogen-bond participants due to the removal of the tyrosine hydroxyl. Indeed, light absorption by Y57F mutants induces an unprotonated retinal Schiff base intermediate as well as proton pumping [61]. Simulated photoisomerization yielded a distribution containing 44% case 1 and 38% case 2 photoproducts; of the remaining nine trials, seven failed to complete the isomerization and remained in an all-*trans* state, while the remaining two trials produced a 13,14-*dicis* retinal. A very similar distribution of trials was seen when simulated without an explicit hydrogen-bond term. This distribution is closer to that seen for the wild-type than for any of the other mutants considered here. In regard to the rotation direction and the Schiff base proton interaction energy, Y57F acts very much like the wild-type. A case 1 photoproduct of the Y57F mutant is shown in Fig. 5.3d.

5.4 Discussion

In the simulations of bR and its mutants, the photoreactions completed with few exceptions the isomerization around the C_{13}-C_{14} bond of retinal within 500 fs. The vast majority of the trials formed case 1 and case 2 photoproducts (see Figs. 5.2, 5.3), in proportions which depend on the particular mutant studied. As pointed out above, earlier studies [75, 147] suggested that case 2 photoproducts represent the structure of the K_{590} intermediate which actually initiates the proton pump cycle. In the previous study, two model potentials governing the excited-state isomerization dynamics were considered: a surface with a 10 kcal/mol barrier against torsion of the C_{14}-C_{15} bond and a surface in which this barrier had been reduced to 2 kcal/mol [75]. The simulations reported here employed only the first potential which has a larger propensity to form case 1 and case 2 photoproducts rather than case 3 photoproducts.

5.4.1 Lack of Case 2 Products for D85N, D85T and D212N bR

For wild-type bR, 32% of the simulated photoproducts form case 2 structures. The Y57F mutant, which has been shown to undergo a functional photocycle with a deprotonated retinal Schiff base intermediate [61], forms case 2 photoproducts in 38% of the simulated isomerization reactions. The D85N, D85T, and D212N mutants, in contrast to the functional wild-type and Y57F mutants, represent non-functional systems, and do not form case 2 structures to any significant degree. In fact, replacement of Asp-85 by neutral residues has been observed to eliminate proton pumping activity [107, 161, 122], while D212N mutants engage in photocycles without a deprotonated retinal Schiff base (for pH values above 7) and form only a small fraction of deprotonated intermediates at lower pH values [111]. Thus, the occurrence of case 2 photoproducts in the mutants studied correlates with the proton pumping activity of these mutants. This supports the suggestion that case 2 structures represent the functionally important K_{590} intermediate, as proposed earlier [75, 147].

5.4.2 Energetics of Case 1 and Case 2 Photoproducts

How is the formation of case 1 and case 2 photoproducts controlled by native bR and its mutants? The lack of the primary proton acceptor group in the D85N and D85T mutants could be considered the primary reason for the lack of M_{412} formation in these mutants. However, D212N, which also fails to form a deprotonated retinal intermediate, does contain the proton acceptor Asp-85, which suggests that the reason for elimination of pumping activity in certain bR mutants is due to electrostatic interactions. ΔE_{SB}, which measures the net electrostatic interaction between the Schiff base proton and the surrounding protein, differs significantly between case 1 and case 2 photoproducts for each mutant. The values of ΔE_{SB} for case 1 products are all in the range 13-16 kcal/mol, for both functional and nonfunctional mutants, while for case 2 products the valuese of ΔE_{SB} are close to zero. The hydrogen-bond of the Schiff base proton to water W_C (see Fig. 5.2b) stabilizes case 2 structures and maintains a pathway for proton transfer to Asp-85, while the loss of electrostatic attraction when a neutral residue replaces Asp-85 or Asp-212 destabilizes case 2 structures. This can explain why mutation of Asp-212 eliminates pumping activity, even though the Asp-85 proton acceptor is available: both mutations destabilize case 2 photoproducts, and disrupt the retinal hydrogen-bond connection to the counterion vicinity.

5.4.3 Direction of Photoisomerization

We have investigated the possiblity that the direction of isomerization around retinal's C_{13}-C_{14} bond may determine bR's case 1 and case 2 photoproducts; a clockwise rotation (defined looking from the Schiff base nitrogen to β-ionone ring of retinal) may favor one type of product, a counterclockwise rotation the other. For example, water W_C may capture the Schiff base N-H^+ only during a clockwise isomerization since in this case N-H^+ approaches W_C closely, whereas it does not approach it in case of a counterclockwise all-*trans* $\rightarrow$ 13-*cis* isomerization. For this purpose we have monitored the rotational sense of all isomerizations occurring in our simulations, seeking a possible correlation of this property with the type of photoproduct formed. Very little correlation was found as can be seen from the data in Table 5.1.

5.4.4 Mechanism for Reversed Proton Pumping

At p*H* 6.7, the D85N and D85T mutants of bR exist as a mixture of both all-*trans*, protonated retinal (bR_{trans}) and all-*trans*, deprotonated retinal (M_{trans}). Illuminating this mixture with blue light results in a proton current in the same direction as for wild-type bR [168, 167]. In this case, blue light is absorbed by bR with deprotonated all-*trans* retinal, the latter absorbing near 410 nm; retinal photoisomerizes to the 13-*cis* isomer, becomes protonated from Asp-96, re-isomerizes thermally to all-*trans* and forms again a mixture of protonated and deprotonated retinal. Illumination of the D85N and D85T mutants by yellow light excites the bR with protonated all-*trans* retinal, but no steady proton current is generated. However, addition of blue light induces a proton current, albeit in a direction opposite to the current of native bR [168, 167]. In this case the protonated retinal photoisomerizes through absorption of yellow light, releases its proton towards Asp-96, absorbs in the deprotonated form a blue photon and returns to the initial retinal all-*trans* isomer forming again a mixture of protonated and deprotonated retinal. This surprising reversal of the proton pump direction of the D85N and D85T mutants of bR can be reconciled in a straightforward way with the appearance of case 1 photoproducts as demonstrated in Fig. 5.4.

The model in Fig. 5.4 extends a suggestion by Tittor and coworkers [168, 167], in that it introduces the case 1 type structure into a scheme given by these authors. We suggest that protonated all-*trans* retinal photoisomerizes to a case 1 product as shown in Fig. 5.2b. The case 1 geometry of retinal is well suited for deprotonation to Asp-96, as indicated in Fig. 5.4, leading to a deprotonated 13-*cis* retinal. In this protonation state the barrier for thermal isomerization around the C_{13}-C_{14} bond measures about 50 kcal/mol, i.e., for all practical purposes is insurmountable. However, photoisomerization can occur when light in the 400 nm range, i.e., blue light, is provided. This leads back to unprotonated all-*trans* retinal as shown in Fig. 5.4. The nitrogen pointing towards the extracellular side can receive a proton from this direction, form the initial protonated all-*trans* retinal and enter a new pump cycle, thus, establishing a continuous inward current.

It is possible, however, to have case 2 photoproducts appear in this scheme, which would then simply return to the bR_{trans} state because of the lack of a proton acceptor. However, case 2

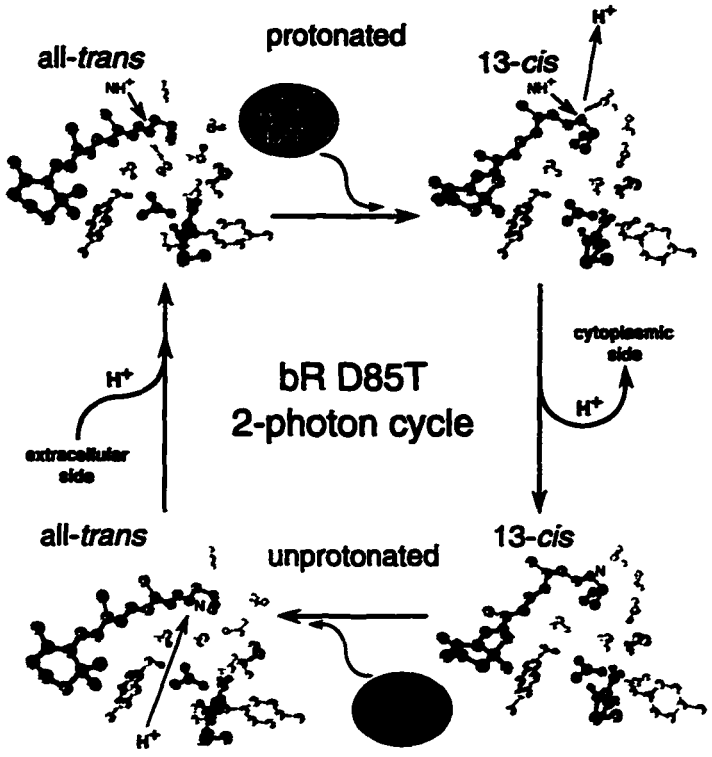


Figure 5.4: Suggested mechanism of reverse proton pumping of the bR D85T mutant. Protonated all-*trans* retinal (top, left) absorbs yellow light and photoisomerizes to a case 1 product (top,right); the N-H^+ bond points towards the cytoplasmic (top) direction and releases its proton to form the corresponding unprotonated retinal Schiff base (bottom, right); the latter absorbs blue light and photoisomerizes to unprotonated all-*trans* retinal (bottom, left) with the Schiff base nitrogen pointing towards the extracellular (bottom) direction; the chromophore is reprotonated and the initial protonated all-*trans* chromophore (top, left) is reformed.

photoproducts are much less likely to appear, due to the changed electrostatic environment of the retinal binding site, and the disruption of the water structure in this region. Conversely, it is possible for case 1 products to occur in a functional photocycle of native bR, and thermally reisomerize back to the initial ground state while case 2 photoproducts continue the cycle. The quantum yield of the bR photocycle, which is less than unity (0.64 ± 0.04 [143, 60, 166]), suggests this as a possibility. While the bR mutant simulations reported here do not model all the steps in the suggested model of reverse proton pumping, they do provide a possible mechanism for the initiation of the observed reverse pumping in D85N and D85T mutants.

5.5 Conclusions

5.5.1 Case 1 and Case 2 Photoproducts Both Can Have a Functional Role

The inverted pump cycle in Fig. 5.4 hinges on the formation of a case 1 photoproduct. The model is consistent with the suggestion that case 2 photoproducts in bR initiate an outwardly pumping photocycle and that mutations abolishing this photoproduct and favoring case 1 products lead to an inwardly pumping cycle under appropriate light conditions. The question arises why such control switching between two photoproducts and two very different cycles might be advantagous to the bacterial host.

Two answers suggest themselves. First, the protein halorhodopsin, a close analogue of bR, lacks a negatively charged side group in place of Asp-85 in the retinal binding site. The protein shows the same reverse pump behavior as bR D85T mutants [9]. Accordingly, the protein reverses the proton flow in the cell membrane of *Halobacterium halobium* under intense light conditions with both yellow and blue light present at sufficient intensities. Such reversal seems to be necessary to avoid overcharging the bacterial cell through a strictly unidirectional proton pump [9]. One can conclude that the design of bR with two possible photoproducts, namely case 1 and case 2, lends itself to the design of halorhodopsin.

A second answer is also based on the fact that case 1 photoproducts either do not initiate proton pumping in the outward direction or, for some light conditions, initiate reversed proton pumping. The bistability of bR's photoproducts can induce a slippage in the proton pump, i.e., idle photocycles, by allowing a shift to case 1 products when a sufficient membrane potential has been generated. This potential would, indeed, have the same effect as neutralizing the Asp-85 negative charge favoring formation of case 1 products. We conclude that the bistability reflected in the appearance of case 1 and case 2 photoproducts might be a desirable, if not necessary, inefficiency of biological proton pumps.

5.5.2 Need to Improve the Description of the Photoproduct Yields

A serious shortcoming of the present description is the low yield of case 2 photoproducts, which are suggested to initiate bR_{568}'s pump cycle, relative to case 1 photoproducts. The yield depends sensitively on the depth and on the range of the potential energy well corresponding to the formation of hydrogen-bonds to water W_C. Accurate calculations of potential energy surfaces within an error margin required to predict accurate yields are beyond the scope of the rough force field used in our simulation; quantitative predictions of the quantum yields of bR_{568}'s photoisomerization require essential improvements of the force field used, at least of the force field governing the photodynamics of retinal and the hydrogen-bonding properties of the retinal Schiff base and the internal water molecules. The change in case 1 and case 2 photoproduct yields between simulations with and without an explicit hydrogen-bond term indicate the importance of this type of interaction in determining and stabilizing the isomerization results.

One step in this respect should base the simulation of retinal's photoisomerization on separate ground and excited state potential surfaces as outlined in [147] rather than employing a single surface, as in the present case. The simulations reported here using a single potential energy surface enforce a complete all-*trans* $\rightarrow$ 13-*cis* isomerization and, accordingly, favor case 1 photoproducts.

Chapter 6

Bacteriorhodopsin Photodynamics Simulations[1]

6.1 Introduction

The simulations of the early steps in the bR photocycle described in chapters 4 and 5 modeled the initial photoisomerization event using classical potential energy surfaces to force rotation about the retinal C_{13}=C_{14} double bond. In this chapter, we study the initial photoisomerization reaction dynamics using a combined quantum/classical treatment of retinal. This method incorporates the effects of coherence between wavefunctions into our classical simulations, and allows us to examine the very fast (10 – 100 fs) motion of retinal following photoexcitation.

The current view of the photoisomerization process of bacteriorhodopsin, which has emerged from time-resolved spectroscopy and modeling of the bR photointermediates [99, 37, 40, 75, 186], suggests that the reaction proceeds in the sequential steps

$$bR_{568} \overset{h\nu}{\rightsquigarrow} bR^*(200\text{ fs}) \overset{500\text{ fs}}{\rightarrow} J_{625} \overset{3\text{ ps}}{\rightarrow} K_{590} \cdots \tag{6.1}$$

where photoexcitation of bR generates bR* which successively decays to the intermediates J_{625} and K_{590}, the latter being a ground state product with retinal in a 13-*cis* geometry. The observations suggest that substantial torsional motion takes place during the ~200 fs lifetime of the excited state bR*.

A schematic structure of bR together with retinal bound as a protonated Schiff base to Lys-216,

[1]Portions of this chapter have been taken from the manuscript: W. Humphrey, H. Lu, I. Logunov, H.-J. Werner, and K. Schulten, Three Electronic State Model of the Primary Phototransformation of Bacteriorhodopsin, *submitted*.

water molecules and key amino acid side groups including Asp-85, Asp-96, and Asp-212 is shown in Fig. 3.2. Asp-85 is unprotonated and negatively charged at the beginning of the proton pump cycle; this group has been shown in mutagenesis [107, 161, 122] and FTIR experiments [19] to accept a proton from the Schiff base following photoisomerization. The D85N mutant forms the K_{590} intermediate in (6.1), but exhibits no proton pumping activity [107, 161, 122]. Similarly, the D212N mutant involving another unprotonated aspartic acid side group near retinal fails to pump protons [107, 122]. Following photoexcitation, the excited state dynamics of these mutants differs markedly from the behavior for native bR. The fluorescence lifetime for D85N (D212N) is 20 (4) times longer than the 500 fs lifetime of native bR [157]. The quantum yields of these mutants, nevertheless, are close to the wild-type value of 0.64 ± 0.04 [143, 60, 92, 91].

Previous simulations of the photoisomerization of bR have employed a single potential surface for the excited state, evaluated by means of QCFF/PI methods [182], or modelled by a simple periodic function [189, 75]. However, polyenes are known to have two closely lying singlet excited states, one of which, S_{single}, in the case of pure polyenes, is optically allowed and involves mainly single excitations, while the other state, S_{double}, is optically forbidden and involves mainly two-electron excitations [148, 73, 164, 153]. In polyenes and unprotonated Schiff bases of retinal, S_{double} is the lowest state in energy, while for protonated Schiff bases S_{single} is the lowest state [146]. Photoexcitation of bR populates mainly S_{single}, i.e., the lowest excited singlet state; the higher state S_{double} is responsible for the two-photon absorption reported in [16].

The question arises in how far a torsion around retinal's C_{13}=C_{14} bond, during photoisomerization, alters the energy level ordering of S_{single} and S_{double} such that the state S_{double} becomes involved in the photodynamics of retinal. This question must be pursued since omission of the second, i.e., higher energy, excited state S_{double} in such a case could affect the photoreaction. In fact, non-adiabatic torsion around the C_{13}=C_{14} double bond promotes the two π-electrons residing in the ethylenic ground state of this bond to a double excitation which argues for an involvement of the state S_{double} in the all-*trans* $\rightarrow$ 13-*cis* reaction. The state S_{double} has been shown, indeed, to decrease in energy upon torsion around the double bond of a retinal analogue [147].

Accordingly, we employ in the present chapter model potential surfaces involving three electronic states suggested in [147]. This model implies that two level crossings are required to return, after photoexcitation, to the ground state. Level crossing is a quantum mechanical process, and for single crossing events one might apply a probabilisitic interpretation of quantum mechanics to account for such events. However, the energetic proximity of the two excited states and the scenario of two level crossings occuring within a few hundred femtoseconds requires a quantum mechanical description in which a 3×3 density matrix accounts for the occupancy of the three participating electronic states and for possible phase coherence [101, 145].

6.1.1 Outline of Chapter

In the following we introduce first in Methods the quantum chemical and combined quantum/classical mechanical simulation methods employed in our study. We present then in Results the outcome of these simulation demonstrating the participation of three electronic states in the photodynamics of retinal in wild-type bR, and for the D85N and D212N mutants of bR. From the simulations we compute the quantum yield and excited state lifetimes, and in Discussion compare the results to the experimentally observed photoisomerization reaction dynamics.

6.2 Methods

6.2.1 Excited State Potential Surfaces

The potential surfaces governing photoisomerization processes have been evaluated *in vacuo* for the retinal analogue $[CH_2–(CH)_3–(C_2H_3)–(CH)_2–NH–CH_3]^+$, at the CASSCF(8,8)/6-31G *ab initio* level by means of the program MOLPRO [184, 84]. This retinal analogue is known to be the shortest analogue that exhibits proton pump activity upon reconstitution with the apoprotein [159]. The utilized complete active space, CAS(8,8), includes the four HOMO and four LUMO orbitals of the full π-orbital space of the four-double-bond retinal analogue. The choice of this analogue was dictated by the fact that retinal itself, which contains six double bonds, proved too large for a sufficiently extensive calculation. The initial geometry of the retinal analogue was minimized at the Hartree-Fock level using the 6-31G basis set; the isomerization potentials were computed for a

gradual torsion in steps of 15° around the C=C double bond adjacent to the Schiff base, keeping all other degrees of freedom constant. Basis set improvements beyond the 6-31G level were found not to significantly affect the energetics of isomerization around the relevant double bond of the retinal analogue.

As argued above, one expects the three lowest electronic states (S_0, S_1, and S_2) to be implicated in retinal's photoisomerization. The associated potential surfaces were computed simultaneously by performing state-averaged calculations, in which molecular orbitals are optimized with respect to the average energy of the S_0, S_1, and S_2 states. In regions of close approach between any two surfaces, additional state-averaged calculations were performed to investigate the nature of the "crossing" between the respective states; in such calculations molecular orbitals were optimized with respect to the average energy of the two particular states of interest (either S_0 and S_1, or S_1 and S_2).

From the adiabatic surfaces in Fig. 6.1 we change in our simulations to a presentation in terms of a ϕ-dependent 3×3-Hamiltonian matrix H

$$H(\phi) = \begin{pmatrix} E_1(\phi) & V_{12} & V_{13} \\ V_{12} & E_2(\phi) & V_{23} \\ V_{13} & V_{23} & E_3(\phi) \end{pmatrix} \tag{6.2}$$

where ϕ represents the torsional angle about the C_{13}=C_{14} double bond (the isomerization reaction coordinate), $E_j(\phi)$ are non-adiabatic potential surfaces, and V_{ij} are the non-adiabatic coupling constants which we choose to be real and ϕ-independent. As discussed further below, the non-adiabatic surfaces $E_j(\phi)$ and coupling constants V_{ij} were chosen such that the diagonalized form of $H(\phi)$ in eqn. (6.2) approximates the *ab initio* surfaces of the states S_0, S_1 and S_2 in Fig. 6.1. In the new representation the regions of non-adiabatic interaction between the states S_0, S_1 and S_2 are represented by explicit crossing points along the $E_j(\phi)$ surfaces.

The use of a truncated retinal analogue and the neglect of the stabilizing effect of the protein environment resulted in an overestimate of the $S_0 \rightarrow S_1$ excitation energy. This deficiency was corrected through a linear scaling of E_1, E_2, and E_3 such that the energy difference between the states S_0 and S_1 at the all-*trans* position measured 50 kcal/mol, a value equal to the excitation

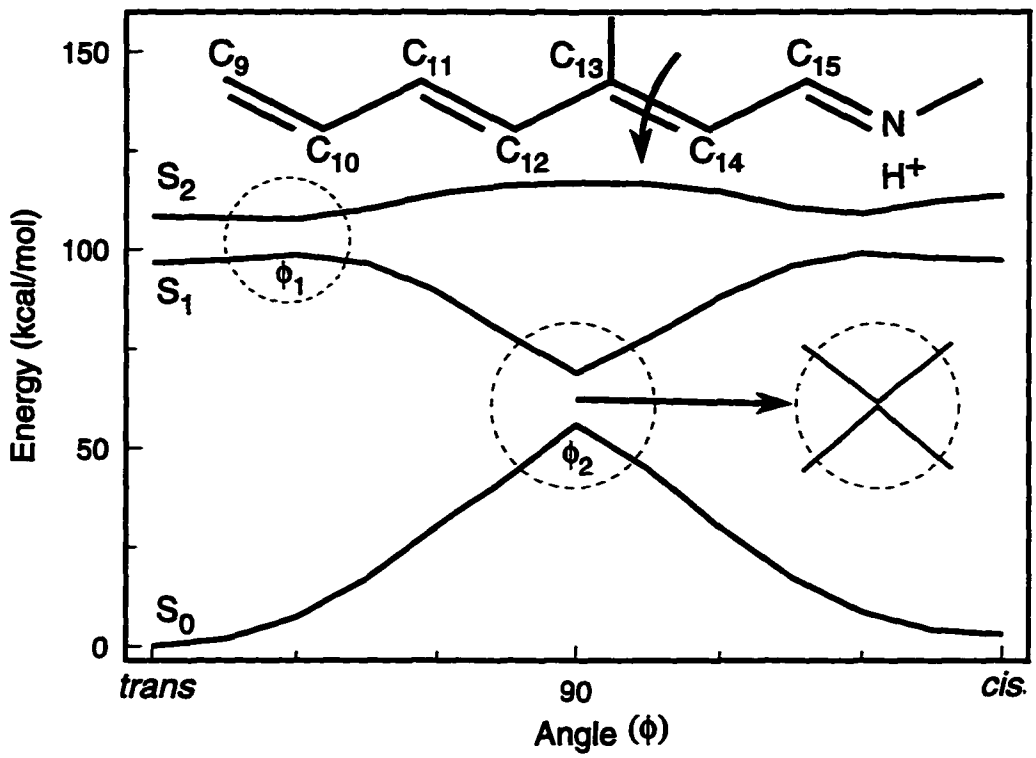


Figure 6.1: Computed ground and excited state potentials of [H_2C=(CH)–(CH)=(CH)–(CH_2-C*)=(C*H)–(CH)=NH^+–CH_3], an analogue of the protonated Schiff base of retinal. The dependence of the energies of the ground state (S_0) as well as the first (S_1) and second (S_2) excited states on the torsional angle ϕ of the bond C*=C* is shown. The bond C*=C* corresponds to the C_{13}=C_{14} double bond of retinal. State-average calculations for the S_0, S_1, and S_2 electronic states were performed at the CASSCF(8,8)/6-31G *ab initio* level (see text). ϕ_1 and ϕ_2 indicate the two regions of close interaction between the two excited states, and between the first excited state and the ground state, respectively.

energy of retinal in bR. The resulting non-adiabatic surfaces $E_j(\phi)$ are shown in Fig. 6.2.

The minimum energy separation between S_0 and S_1, as well as between S_1 and S_2, allows one to determine the coupling constants V_{ij}, i.e., the coupling constants should be half of the minimum energy splitting. Instead of adopting such values we employed coupling constants V_{12}, V_{13}, and V_{23} of 0.5, 1, and 1 kcal/mol, respectively, which are considerably smaller than the estimates due to level splittings. However, the latter splittings depend sensitively on the type of state averaging adopted, as discussed further below; certain procedures lead to very small splittings. The value of 0.5 kcal/mol for V_{12} was chosen to enforce agreement with the observed quantum yield of 0.64 [143,

60]; a choice of 1 kcal/mol would result instead in a quantum yield of 0.34. The value of 1 kcal/mol for V_{13} and V_{23} was selected to obtain agreement with the measured lifetime of wild-type bR of about 300 fs in the state S_1 [64]; a value of 5 kcal/mol would yield a crossing time (lifetime) of about 100 fs.

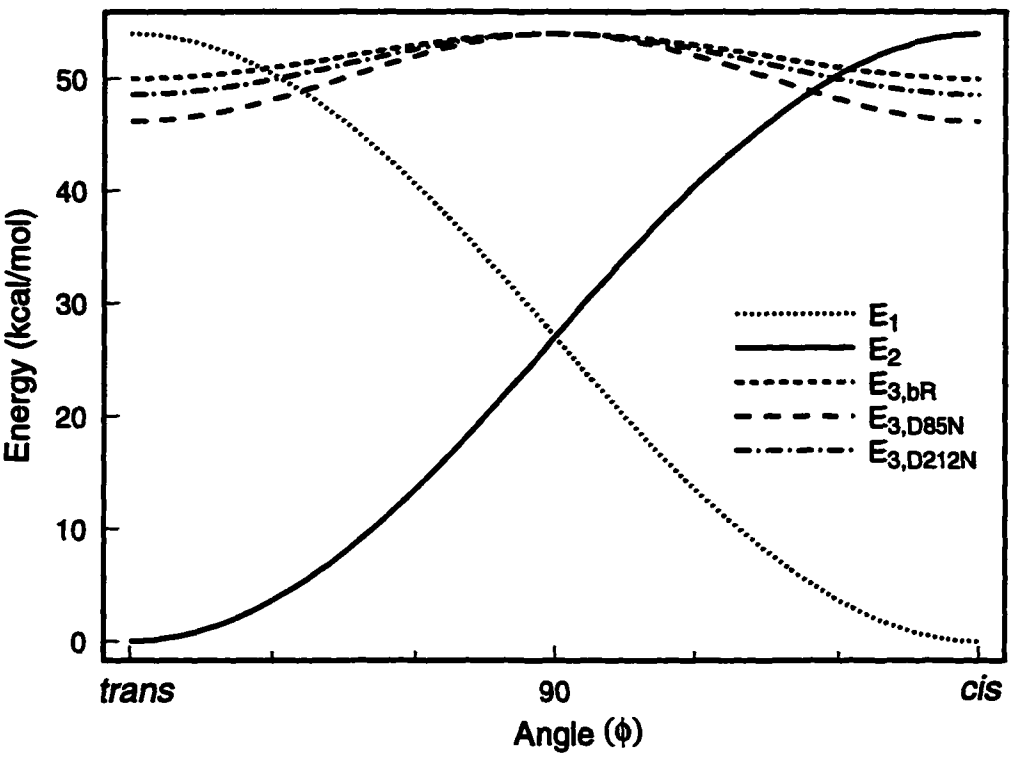


Figure 6.2: Non-adiabatic potential energy surfaces used in molecular dynamics simulations of bacteriorhodopsin and bR mutants. The minimum of the surfaces $E_{3,\mathrm{bR}}$, $E_{3,\mathrm{D85N}}$, and $E_{3,\mathrm{D212N}}$ are matched to the measured excitation energy bR and its respective mutants.

Hamiltonians similar to (6.2) were constructed for the bR mutants D85N and D212N absorbing at 615 nm [107] and at 585 nm [111], respectively. For this purpose the minimum of the E_3 surface at the all-*trans* and at the 13-*cis* position was lowered to match the respective excitation energy. The resulting surfaces $E_{3,\mathrm{bR}}$, $E_{3,\mathrm{D85N}}$ and $E_{3,\mathrm{D212N}}$ are compared in Fig. 6.2.

6.2.2 Density Matrix Evolution Method

The photodynamics simulations used as a starting point the refined bR_{568} structure reported in [74] which is based on a structure of bR obtained by electron microscopy [66]. The refined structure

includes the loop regions between the seven trans-membrane helices of bR, as well as sixteen water molecules within the protein interior. Five of these water molecules are located within the vicinity of retinal's binding site, in hydrogen-bond contact with the Schiff base proton and nearby amino acids including Asp-85 and Asp-212. D85N and D212N mutants were constructed through modification and equilibration of the refined structure. A series of simulations was carried out for each mutant, each run consisting of a brief 100 fs equilibration starting from a random velocity distribution and using a 1 fs integration time step, followed by simulation of the photoisomerization reaction employing the co-called density matrix evolution (DME) method [12] based on the potential energy surfaces in Fig. 6.2. The photoisomerization simulations lasted until retinal returned to the ground state, and varied from 1 to 12 ps. All simulations which completed a all-*trans* $\rightarrow$ 13-*cis* isomerization were continued for a further 2 ps to model the $J_{625} \rightarrow K_{590}$ transformation; for this purpose the ground state potential introduced in [75] was used to describe the C_{13}=C_{14} torsion.

The DME description of the torsion around the C_{13}=C_{14} bond couples the evolution of the quantum mechanical density matrix of the three electronic states involved in the photoisomerization with the classical equations of motion of the entire protein [12]. Using the three-state Hamiltonian H in eqn. (6.2), we introduced the density matrix $\rho(t)$, the diagonal elements of which represent the occupation of the three states as a function of time. Photoexcitation was accounted for by starting from the pure state

$$\rho_{ij}(t_0) = \delta_{33}. \tag{6.3}$$

During the subsequent simulation, the density matrix evolved according to the Liouville – von Neumann equation

$$\dot{\rho} = \frac{i}{\hbar}[\rho, H]. \tag{6.4}$$

The density matrix incorporates the effect of coherence between the quantum states during the dynamics, without the need to calculate the retinal wavefunctions. A more complete description of the DME method is provided in Appendix C.

The energy stored in the torsion around the C_{13}=C_{14} bond is

$$E_{13-14} = \mathrm{Tr}\,\rho\,H = \sum_{j=1}^{3}\rho_{jj}E_j + \sum_{\substack{j,k=1\\ j\neq k}}^{3}\rho_{jk}V_{kj}\,. \tag{6.5}$$

The corresponding forces are obtained after taking the (negative) gradient with respect to atomic cordinates. Since the couplings V_{jk} are chosen to be independent of ϕ, the forces acting on atom k with position $\vec{r}_k$ are

$$\vec{F}_{13-14,k}(\phi) = \sum_{j=1}^{3}\rho_{jj}\vec{F}_{jk} \tag{6.6}$$

where $\vec{F}_{jk} = -\partial E_j(\phi)/\partial\vec{r}_k$, i.e., the net force acting on an atom is the average over the forces due to each state, weighted by the occupancy ρ_{jj} of that state. To the resulting force on atom k are added all other forces acting on the atom from the surrounding protein environment.

Equation(6.4) was integrated by means of a fourth-order Runge-Kutta scheme to yield $\rho(t)$; rapid oscillations in the occupancy of the states required a 0.1 fs integration time step. For the classical molecular dynamics simulations a modified version of the program NAMD [112] was used and a 1 fs integration time step employed. The simulations were carried out at a temperature of 300 K. For the force field, except for the C_{13}=C_{14} bond torsion, the CHARMm force field [21] was used, modified for retinal as described previously [75].

6.3 Results

The singlet state potential surfaces for torsion around the double bond adjacent to the C=N bond of the retinal analogue $[CH_2$–$(CH)_3$–(C_2H_3)–$(CH)_2$–NH–$CH_3]^+$ are presented in Fig. 6.1. For the all-*trans* [13-*cis*] geometry the calculations predict, besides the ground state S_0(*trans*) [S_0(13-*cis*)], two closely spaced excited states, a state S_1(*trans*) [S_1(13-*cis*)] with predominantly single-excited electron configurations below a state S_2(*trans*) [S_2(13-*cis*)] with predominantly double-excited electron configurations. This level ordering is in agreement with the configuration interaction calculations on a related polyenylic ion reported in [146]. Due to the lack of symmetry, both S_1(*trans*) and S_2(*trans*) are optically allowed from the ground state S_0(*trans*); the respective transition dipole moments are $D_{0\to1}$ = 3.5 a.u., $D_{0\to2}$ = 1.4 a.u., and $D_{1\to2}$ = 1.1 a.u.

As can be seen from Fig. 6.1, the potential surfaces along the complete [0°, 180°] interval of the torsional angle ϕ exhibit three avoided crossings, two near $\phi_1 = 30°$, $\phi_1' = 150°$, and one near $\phi_2 = 90°$. Following the procedure outlined in Methods, the surfaces can be interpreted as arising from three adiabatic surfaces shown in Fig. 6.2, one surface (E_2) connecting the all-*trans* ground state S_0(*trans*) to the excited state S_2(13-*cis*), a second (E_1) connecting conversely the excited state S_2(*trans*) to the ground state S_0(13-*cis*), and a third one ($E_{3,\mathrm{bR}}$) connecting the excited state S_1(*trans*) to the excited state S_1(13-*cis*).

The identification of the two adiabatic surfaces E_1 and E_2 constitutes a key result of this chapter on which the further description is based. The surface E_2 is consistent with the simple interpretation that the state S_0(*trans*) contains two π-electrons in the bonding orbital of the double bond under consideration, and that a 180° torsion adabatically lifts these π-electrons into the anti-bonding orbital, i.e., into a doubly excited electron configuration; the occurence of such configurations is characteristic of the S_2(13-*cis*) state. The surface E_1 can be interpreted in a similar manner.

The distinctive feature of the surfaces in Figs. 6.1, 6.2 is the existence of regions of non-adiabatic interactions inducing, near ϕ_1 (ϕ_1'), the crossing $E_1 \leftrightarrow E_{3,\mathrm{bR}}$ ($E_2 \leftrightarrow E_{3,\mathrm{bR}}$) and, near ϕ_2, the crossing $E_1 \leftrightarrow E_2$. We examined the surfaces near ϕ_1 by means of the quantum chemical calculations outlined above state averaging over the S_1 and S_2 states, i.e., involving only the two states participating in the crossing $E_{3,\mathrm{bR}} \leftrightarrow E_1$. The results revealed that the interaction in this region is of the "avoided crossing" type; the potential curves obtained in state-average calculations for the two states do not differ considerably from those presented in Fig. 6.1, implying that the coupling between the S_1 and S_2 states in the ϕ_1 region is strong.

In contrast, examination of the potential surface in the region near ϕ_2, employing averaging over only the two states S_0 and S_1, leads to results which differ considerably from those obtained when all three states are involved in the average. This is demonstrated in the inset in Fig. 6.1, which shows that the separation between the surfaces becomes less than 1 kcal/mole in this case. Most likely, the crossing between S_0 and S_1 is of the "conical intersection" type [30]. Clarification of this matter requires consideration of further retinal degrees of freedom in the crossing region,

e.g., of the stretch vibration of the C_{13}=C_{14} bond [30]. Calculations performed for a smaller retinal analogue indicate, indeed, the existence of a conical intersection between S_0 and S_1 [17].

The features outlined suggest that a sufficiently strong non-adiabatic coupling between the $E_{3,bR}$ and E_1 surfaces near ϕ_1 is required for the photoisomerization process to induce the necessary transition $E_{3,bR} \rightarrow E_1$, while weak interaction between the E_1 and the E_2 surfaces are preferable, since the latter favors continuation of the torsion on the E_1 surface (see Fig. 6.2) towards the 13-*cis* geometry.

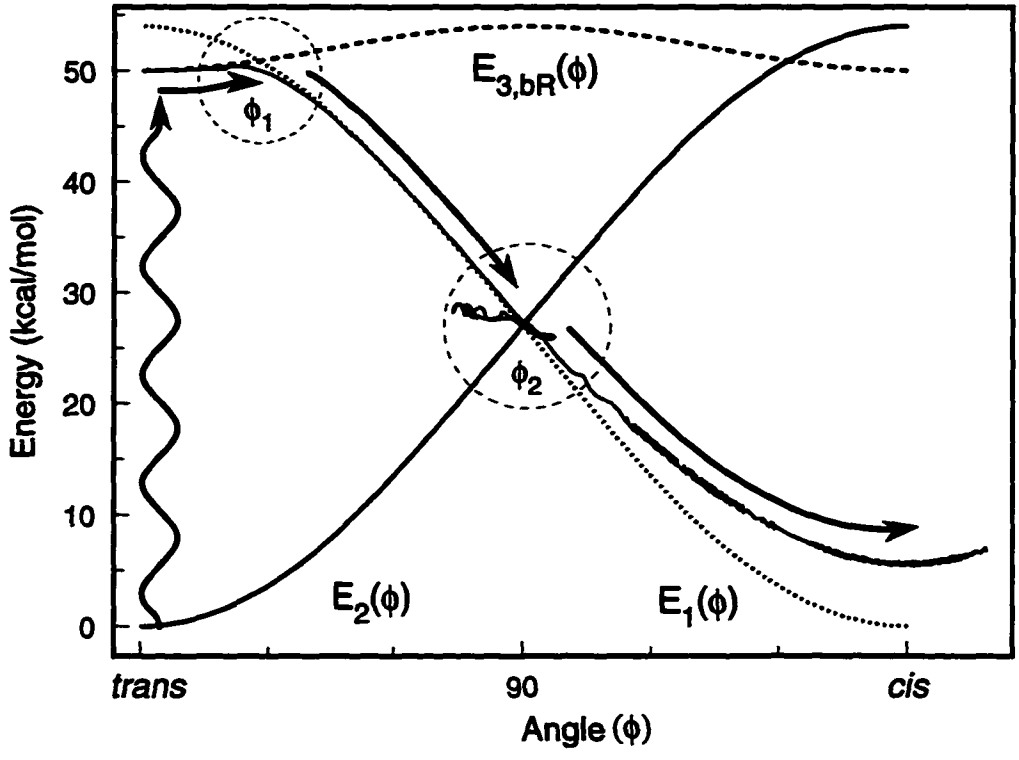


Figure 6.3: Energy of the C_{13}-C_{14} torsional degree of freedom E_{13-14} as a function of the torsional bond angle ϕ, i.e., $[E_{13-14}(t), \phi(t)]$ trajectory, for a sample molecular dynamics simulation, compared to the three non-adiabatic energy surfaces E_1, E_2, and $E_{3,bR}$ shown in Fig. 6.2. Initially retinal is promoted in the all-*trans* geometry from E_2 to $E_{3,bR}$; subsequently, retinal crosses to E_1 near $\phi = \phi_1$. At $\phi = \phi_2$ retinal may choose to continue its motion along E_1 to reach the 13-*cis* geometry as shown, or it may cross to E_2 to reform the all-*trans* state (not shown).

As outlined in Methods, the non-adiabatic potential surfaces in Fig. 6.2 served to carry out a combined quantum/classical mechanical simulation, resulting in trajectories which provide coordinates, e.g., the torsional angle of the C_{13}=C_{14} bond $\phi(t)$, as well as selected observables, e.g., the

energy E_{13-14} define in eqn. (6.5). Figure 6.3 presents the $[E_{13-14}(t), \phi(t)]$ diagram of a typical trajectory resulting from a 1 ps simulation. Retinal is excited initially in the all-*trans* geometry by a 568 nm photon, i.e., placed on the $E_{3,\mathrm{bR}}$ surface at $\phi = 0$ [c.f., eqn. (6.3)]; after a time τ_1 the system crosses to the surface E_1, where it quickly descends to the second crossing point ϕ_2 at time τ_2. Here, after a brief period of meandering near the crossing point, the system chooses to continue along E_1 to complete the photoisomerization and reaches the 13-*cis* geometry. The energy $E_{13-14}(t)$, initially, follows closely the potential surfaces, but at the end of the photoreaction deviates significantly from the surface $E_1[\phi(t)]$, reflecting the fact that the system is trapped in a mixed quantum state.

The character of the quantum state is characterized by the diagonal elements ρ_{jj}, $j = 1, 2, 3$ of the density matrix, which describe the occupancies of the three electronic states corresponding to the surfaces E_1, E_2, $E_{3,\mathrm{bR}}$. These occupancies are shown in Fig. 6.4. Following excitation, i.e., at $t_0 = 0$, the occupancies of states corresponding to E_1 and $E_{3,\mathrm{bR}}$ oscillate rapidly for about 200 fs due to the energetic proximity of the surfaces. The system is then seen to switch rapidly to the surface E_1, whereupon a fast torsion moves the system in a time $\Delta\tau$ (see Table 6.1) to the second crossing point with the surface E_2. Rapid oscillations are again discernible at this stage, mixing the states corresponding to E_1 and E_2; due to the rather weak non-adiabatic coupling $V_{12} = 0.5$ kcal/mol, the system chooses to stay on the E_1 surface. However, beyond the crossing point the states corresponding to E_2, $E_{3,\mathrm{bR}}$ remain populated to a significant degree. Due to environmental relaxation effects, not accounted for in our description, the occupancies of these states should vanish within about ten femtoseconds as demonstrated recently for a hydrated electron for which relaxation effects are expected to be stronger, however, than in the present case [151]. The neglect of this decay is a shortcoming of the approach adopted here which, as a result, can lead to unphysical "average" final states, when in reality the system would choose between one of many possible outcomes with some probability distribution.

In all our simulations of the photoisomerization event in bR and its mutants, the final product exhibited over 50% occupancy of either of the states corresponding to E_1 or E_2; in the example

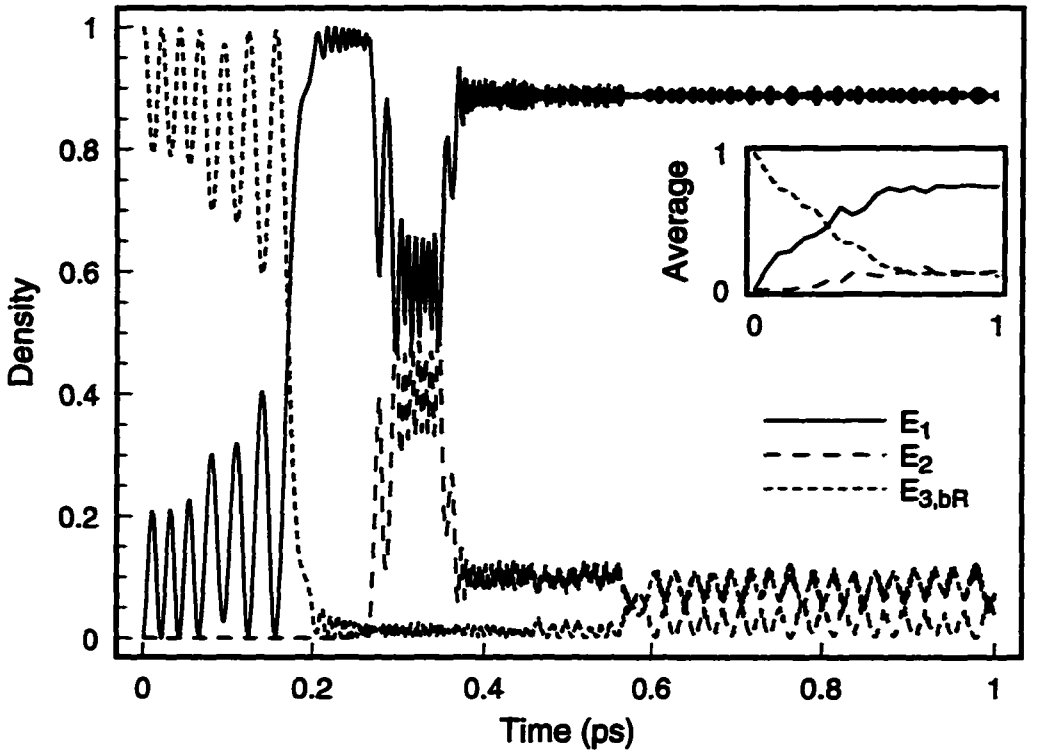


Figure 6.4: Occupancies $\rho_{jj}(t)$ of the three non-adiabatic potential energy surfaces E_1, E_2, and $E_{3,\mathrm{bR}}$ as a function of time for the $[E_{13-14}(t), \phi(t)]$ trajectory shown in Fig. 6.3. In the inset are shown the occupancies averaged over all simulations which resulted in 13-*cis* retinal.

shown in Figs. 6.3, 6.4, the occupancy is almost 90%. After completion of the photoisomerization, our simulations switched from the DME approach to the torsional potential as used in conventional simulations.

An alternate scheme which has been employed in combined quantum/classical molecular dynamics simulations is the surface-hopping algorithm [170]. In this scheme, the system is taken to occupy only one state at any given time; after each integration time step, the system may switch from the current state to another state with a probability designed to minimize the number of transitions while providing the correct statistical occupation of each state. A benefit of this method is that it will generate proper asymptotic behavior since the system cannot occupy an average state. The probabilistic nature of the method, however, requires in many cases a large number of iterations for proper convergence. Thus, application of the surface-hopping algorithm in this chapter

was not feasible due to the extensive computational requirements.

A total of 100 simulations of the photoisomerization of wild-type bR were carried out resulting in trajectories like the one shown in Figs. 6.3, 6.4. Fifty simulations were also completed for the D85N and D212N mutants of bR, by replacing the potential surface $E_{3,\mathrm{bR}}$ by the surfaces $E_{3,\mathrm{D85N}}$ and $E_{3,\mathrm{D212N}}$ shown in Fig. 6.2. The simulations for the mutants required simulation periods of 1–12 ps due to the longer time required for retinal in D85N and D212N bR to cross to the E_1 potential surface. The results of the simulations are presented in Table 6.1. For each set of simulations, the fraction of trials which completed the all-*trans* $\rightarrow$ 13-*cis* isomerization were identified and the average crossing times τ_1 and τ_2 were determined.

	Bacteriorhodopsin Mutant		
	bR_{568}	D85N	D212N
λ_{max}	568 nm	615 nm	585 nm
Simulations	100	50	50
13-*cis* Products	71	30	37
Case 1 Products	66%	83%	84%
Case 2 Products	34%	0%	5%
ϕ_1	152.2°	144.1°	148.7°
τ_1	332 fs	5.59 ps	3.24 ps
ϕ_2	90°	90°	90°
τ_2	462 fs	5.73 ps	3.42 ps
$\Delta\tau$	130 fs	140 fs	178 fs
τ_2 (mutant)/τ_2 (bR)	-	12.4	7.4

Table 6.1: Summary of photodynamics simulation results. ϕ_1 and ϕ_2 are torsional angles of the C_{13}=C_{14} bond at the first and second crossing points; τ_1 and τ_2 are the average times required to reach these points. $\Delta\tau$ is the difference between τ_2 and τ_1.

In the 100 simulations of wild-type bR, 71 13-*cis* products emerged. Trials which failed to form a 13-*cis* isomer either did not move beyond the first crossing point ϕ_1 or switched to the surface E_2 at the second crossing point ϕ_2, i.e., such trials reformed all-*trans* retinal. The simulated photoisomerization quantum yield measured 0.71 for wild-type bR, matching closely the observed quantum yield of 0.64 $\pm$ 0.04 [143, 60]. The agreement, however, is a consequence of the choice of coupling between the E_1 and E_2 surfaces as stated in Methods.

Time-resolved spectroscopy of bR has revealed two fast processes, one occuring within 200 $\pm$ 70 fs

[37], and the other occuring within 500 ± 100 fs [99, 37, 40]. The observations were interpreted to suggest that photoexcitation is followed by rapid progression along the reaction coordinate from all-*trans* to 13-*cis* during the initial 200 fs. More recently, however, femtosecond spectroscopy covering an extremely wide spectral range has indicated a lack of a spectral shift of induced absorption, indicating a flat excited state surface; the observations revealed also a relaxation processes with a decay time of 370 fs [64]. The $E_{3,\mathrm{bR}} \rightarrow E_1$ crossing at $\tau_1 = 332$ fs can explain the observed fast decay.

Figures 6.3, 6.4 show the behavior of a single, albeit typical, quantum/classical mechanical trajectory. Of interest is the average behavior of bR. The inset in Fig. 6.4 shows the time evolution of the density of the states corresponding to the surfaces E_1, E_2 and $E_{3,\mathrm{bR}}$ averaged over all 71 simulations which formed 13-*cis* photoproducts. The time at which the occupation of $E_{3,\mathrm{bR}}$ and E_1 intersect is τ_1, while τ_2 is near the time where the occupation of E_1 exceeds 50%. One can discern that retinal photoisomerizes within about 500 fs to the 13-*cis* geometry with a final populaton of the E_1 surface, on average, of about 80%.

For the D85N and D212N mutants of bR the dynamics of retinal after light excitation is significantly slower in the initial phase. This is due to the fact that the surfaces $E_{3,\mathrm{D8N}}$ and $E_{3,\mathrm{D212N}}$ cross the E_1 surfaces at higher energies than in case of the $E_{3,\mathrm{bR}} \rightarrow E_1$ crossing. As a result, for D85N bR, of which 30 of 50 simulations formed 13-*cis* isomers, the average time to the first crossing point was 5.59 ps. In case of D212N bR with 37 out of 50 completing the 13-*cis* isomerization, the first crossing occurred at 3.24 ps. Interestingly, the time between the first and the second crossing, $\Delta\tau$, measured approximately the same for native bR and for its D85N or D212N mutants; the mutants spent the major part of the reaction time before encountering the first crossing point.

In previous molecular dynamics simulations of the photoisomerization of retinal in bR [75, 147], the isomerization products assumed two distinct structures: case one structures, in which the Schiff base N-H^+ bond is oriented towards the cytoplasmic side of the protein, and case two structure, in which twists around retinal's single bonds oriented the N-H^+ bond to point parallel to the plane of the membrane and to remain connected to Asp-85 via hydrogen-bonds with an intermediate water

molecule. The latter product was identified with the K_{590} intermediate which actually initiates the proton pump cycle [75, 147, 187]. Figure 6.5 compares the retinal binding site region for a case one (Fig. 6.5b) and case two (Fig. 6.5c) with the all-*trans* retinal binding site (Fig. 6.5a). The combined quantum/classical simulation of the present chapter led to 66% case one products and 34% case two products formed. This distribution is similar to that in the earlier simulations (58% and 36%, respectively). The present work supports, therefore, the conclusions reached previously regarding a possible mechanism of bR's pump cycle [147, 187].

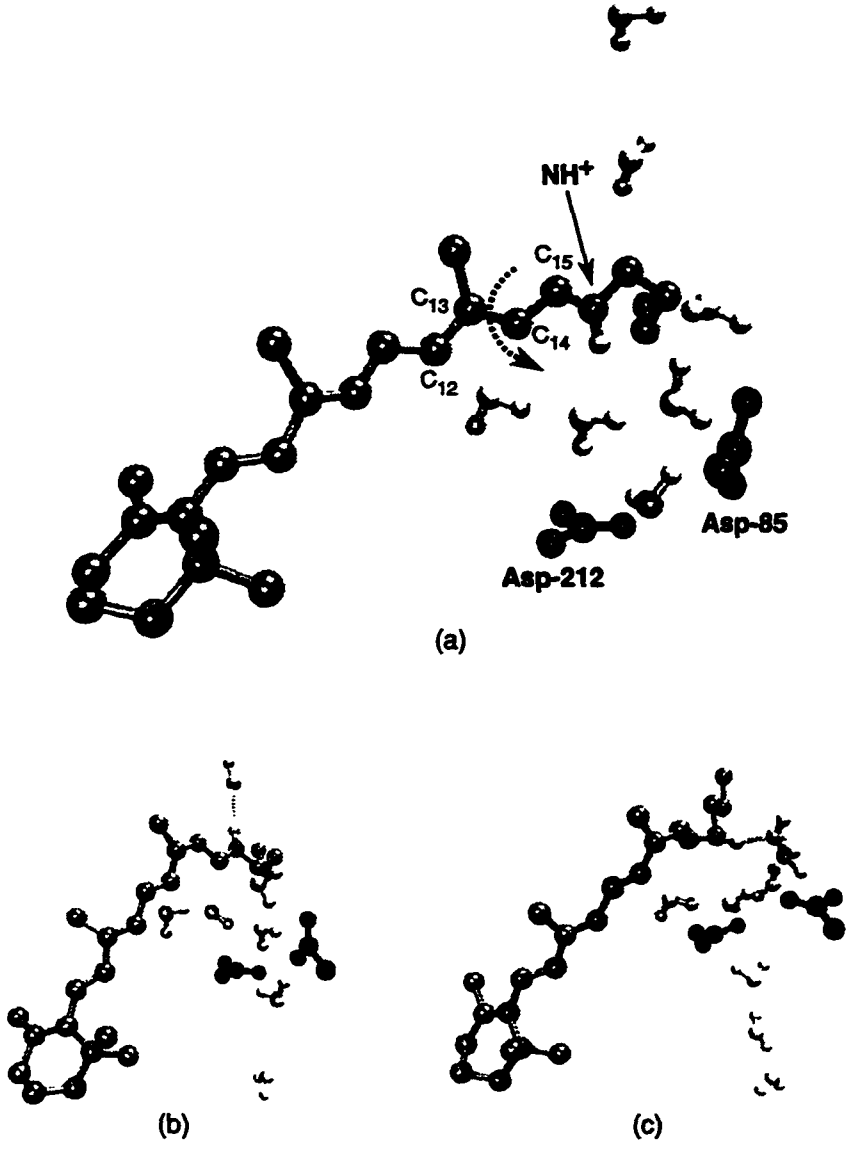


Figure 6.5: (a) Retinal and its binding site in bR, including residues Asp-85, Asp-212, and internal water molecules. (b) Case 1 13-*cis* photoproduct. (c) Case 2 13-*cis* photoproduct.

6.4 Discussion

The excited state dynamics of retinal in bR, starting from light absorption into the first excited state, proceeds rapidly as a torsion around the C_{13}=C_{14} bond and is characterized by two crossing events: one crossing involves relatively strong non-adiabatic coupling between the first and second excited states at a nearly 30° torsion 332 fs after light absorption; a second crossing occurs between the second excited state and the ground state, in which retinal moves essentially along a steep non-adiabatic surface, completing a 60° motion in another 130 fs. A choice of non-adiabatic couplings of the order of 1 kcal/mol can explain the overall quantum yield. Due to interaction with the protein, two isomerized reaction products arise, one with the Schiff base proton in hydrogen bond contact with Asp-85, the other with the Schiff base proton pointing to Asp-96 on the cytoplasmic side.

6.4.1 First Crossing

Following photoexcitation from the ground state surface E_1 to the excited state surface $E_{3,\mathrm{bR}}$, retinal rotates around its C_{13}=C_{14} bond until it reaches the first crossing point ϕ_1. In this region, interaction with the nearly degenerate excited state E_2 allows the system to overcome the slight energy barrier before crossing to E_1. For the wild-type, this crossing point was reached after 332 fs, which compares quite well with time-resolved spectroscopy measurements of this process by Hasson and coworkers if one associates the observed 370 fs fast relaxation process [64] with the crossing event. The near degeneracy of the surfaces $E_1(\phi)$ and $E_{3,\mathrm{bR}}(\phi)$ for small angles ϕ implies a continuous interaction which prepares retinal for the crossing and induces it for a relatively small non-adiabatic coupling; thus, the interaction of the system in the region of the first crossing region does not affect the quantum yield. In almost all simulations, retinal eventually surmounted the slight all-*trans* $\rightarrow \phi_1$ energy barrier of 1–3 kcal/mol. The important characteristic of the initial phase of the excited state dynamics is that retinal remains very close to its all-*trans* geometry; rapid torsion sets in only after the first crossing is completed.

6.4.2 Second Crossing

Upon reaching the second crossing point ϕ_2, near the 90° torsion, retinal makes a non-adiabatic transition to the ground state characterized by very weak coupling between the respective non-adiabatic surfaces. Essentially, retinal moves along the non-adiabatic surface $E_1(\phi)$ (see Fig. 6.2), halting briefly at the crossing point as shown in Fig. 6.3. The brief halt at the crossing points makes it unlikely that this stage of the photodynamics can be identified with the J_{625} intermediate. The crossing determines, however, the quantum yield of formation of 13-*cis*-products; the observed value of 0.71 [143, 60] is reproduced only for a weak coupling $V_{12} = 0.5$ kcal/mol. The second crossing point is reached within 462 fs for wild-type bR, while a slower decay process on the order of 1 ps has been observed by Hasson *et al.* [64]. The difference may stem from an asymmetry in the real potential energy surfaces or from an alternative interpretation of the observed relaxation.

6.4.3 Photodynamics of bR Mutants

The D85N and D212N mutants of bR exhibit strongly increased fluorescence life times without changes in the quantum yield [157], a behavior which is explained well by the three state model through an increase of the barrier for the first crossing event. The schematic potential surfaces employed in the simulations cannot be expected to reproduce exact crossing times. The close agreement for wild-type bR is due to the choice of the coupling $V_{12} = 1$ kcal/mol. Nevertheless, our results provide a possible explanation of the behavior of bR mutants. Assuming that the crossing times are governed by a Boltzmann distribution at the crossing point ϕ_1 relative to the all-*trans* position, and that the crossing points are at the same absolute height, the ratio of the mutants' excited state lifetimes to that of bR is

$$\frac{\tau_{2,mutant}}{\tau_{2,bR}} \approx \exp\left[\frac{\Delta E_{3,mutant} - \Delta E_{3,bR}}{kT}\right], \tag{6.7}$$

where $\Delta E_{3,mutant} = E_{3,mutant}(\phi_1) - E_{3,mutant}(trans)$. Accordingly, mutations which shift the retinal absorption to the red increase the barrier to photoisomerization and lead to longer excited state lifetimes. Inserting for $\Delta E_{3,mutant}$ in eqn. (6.7) the values corresponding to the spectral shifts

experienced by the D85N and D212N mutants predicts lifetime increases by factors of 20 and 1.8, respectively, for the mutants.

The quantum yield for mutants such as D85N and D212N has been found experimentally to be similar to the wild-type value [92, 91], and the yield obtained for these mutants in our simulations was likewise similar (0.60 and 0.74 for D85N and D212N, respectively, compared to 0.71 for the wild-type).

6.5 Conclusions

In summary, combined quantum/classical simulations of the bR photoisomerization supports a new model of the very early events in the bR photocycle, differing from previous models based on ultrafast spectroscopy [99, 37, 40]. In the new model, photoexcitation of retinal is followed not by rapid movement along the reaction coordinate due to excitation into a repulsive Frank-Condon region, but by motion along a relatively flat excited state surface until crossing to a region with a second excited state, which exhibits a steep potential gradient inducing rapid isomerization. A proper description of this behavior requires inclusion of at least the first two excited states. The observed spectral and fluorescent changes occuring within 1 ps after photoexcitation of bR arise due to non-adiabatic transitions between the excited states and an excited state and the ground state. A significant fraction of the excited state lifetime is spent surmounting a small energy barrier before reaching the first non-adiabatic crossing region. The fluorescence lifetime of bR is controlled by this barrier and can be altered strongly through mutations which affect the relative stability of retinal's two lowest excited states.

Appendix A

Molecular Dynamics: The Program NAMD[1]

A.1 Introduction

The primary bottleneck in molecular dynamics simulations is evaluation of the non-bonded interactions, $U_{nonbond}$ in eqn. (2.4). While processors speeds have improved dramatically in the years since the initial application of MD to biopolymer systems, still more performance is necessary to make possible simulation of large systems, e.g., proteins on the order of $10^5 \ldots 10^6$ atoms or larger. One promising method for increasing the net processor performance is to use the peak capabilities of parallel computing platforms such as dedicated large-scale parallel computers or clusters of high-performance workstations. This requires a molecular dynamics program which runs efficiently on such platforms. The program NAMD [112, 113] has been developed with this goal in mind.

The key design elements of NAMD are a spatial decomposition strategy to partition the parallel computation tasks, as well as a multi-threaded, message-driven method of execution. This design uses a number of independent computation tasks for each processor in a parallel computer, instead of one single task; control is transferred between threads based on availability of messages from other processors, which reduces inefficiency due to communication latency.

NAMD is written in C++, an object-oriented extension of the C programming language. It implements the CHARMm energy function (eqn. (2.4)), the same as that used by the programs

[1]Portions of this appendix have been taken from the manuscript: M. Nelson, W. Humphrey, R. Kufrin, A. Gursoy, A. Dalke, L. Kale, R. Skeel, and K. Schulten, MDScope — A visual computing environment for structural biology, *Comp. Phys. Comm.*, 91:111-133, 1995.

CHARMm [21] and X-PLOR [26], and understands molecular parameter, structure, and coordinate files in the same formats as used by X-PLOR. The remainder of this appendix describes the important features of NAMD, in particular, the methods used to parallelize a molecular dynamics simulation.

A.2 Parallelization Strategy

NAMD views the parallel hardware as a set of N independent processors, each with an associated amount of memory (i.e., a distributed-memory model), and each connected via an underlying interprocessor communication network. Parallelization of the MD force evaluation requires efficient partitioning of the computation tasks among these processors, with a minimum of communication between the nodes. One method used in other MD applications to parallelize the force computation involves dividing the tasks using various forms of force decomposition [22]. This method suffers, however, from poor scalability to larger number of processors.

A more efficient parallelization method for MD is spatial decomposition, implemented in NAMD by dividing the volume of space occupied by the molecule into uniform cubes referred to as *patches*. Figure A.1 shows a sample decomposition into patches of the volume surrounding a simple polypeptide. With this strategy, each patch is assigned to a processor, which is responsible for computing the force on the atoms within that patch. When a cutoff distance for the electrostatic interactions is employed, and the size of the patches to set equal to this cutoff radius, then computation of the atomic forces in a patch requires knowledge only of the coordinates of atoms in the immediate neighborhood of the patch. Generally, many more patches than processors exist, and so several patches are assigned to each node. This provides a useful mechanism for load-balancing during the simulation: patches may be reassigned among the processors to even the work load for greater efficiency.

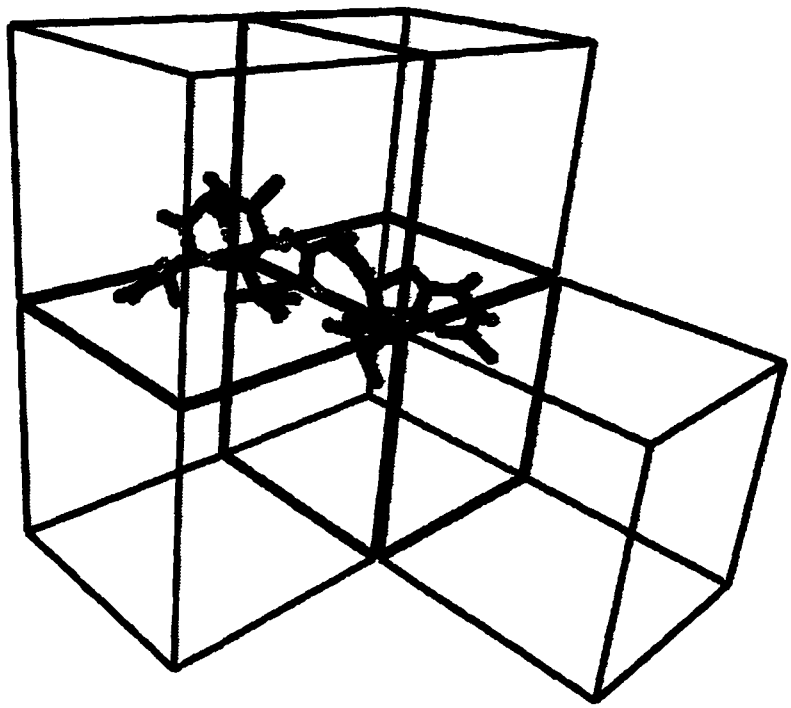


Figure A.1: Spatial decomposition of a small polypeptide in NAMD.

A.3 Message-Driven Execution

Efficient use of a parallel architecture requires limiting the amount of time each processor spends waiting for data from other nodes. In a program employing spatial decomposition, if only a single large patch were used on each CPU, significant time can be lost in gathering data from other processors. This inefficiency may be due not only to network communication latencies, but to processor load imbalances, i.e., the sending node might be busy with other tasks and thus not be able to send the data when the receiver is ready for it.

To reduce this idle time, NAMD employs a mult-threaded execution model in which each patch acts like a logical thread of control. During a simulation time step, each patch transmits messages

between itself and neighboring patches (which may reside on the same physical processor). A patch is allowed to perform some computation (scheduled), such as calculation of the pair-wise forces between atoms on that patch and on a neighbor patch, only when a message arrives for that patch. This *message-driven* design, illustrated in Fig. A.2, reduces processor idle time by increasing the amount of overlap between computation and communication.

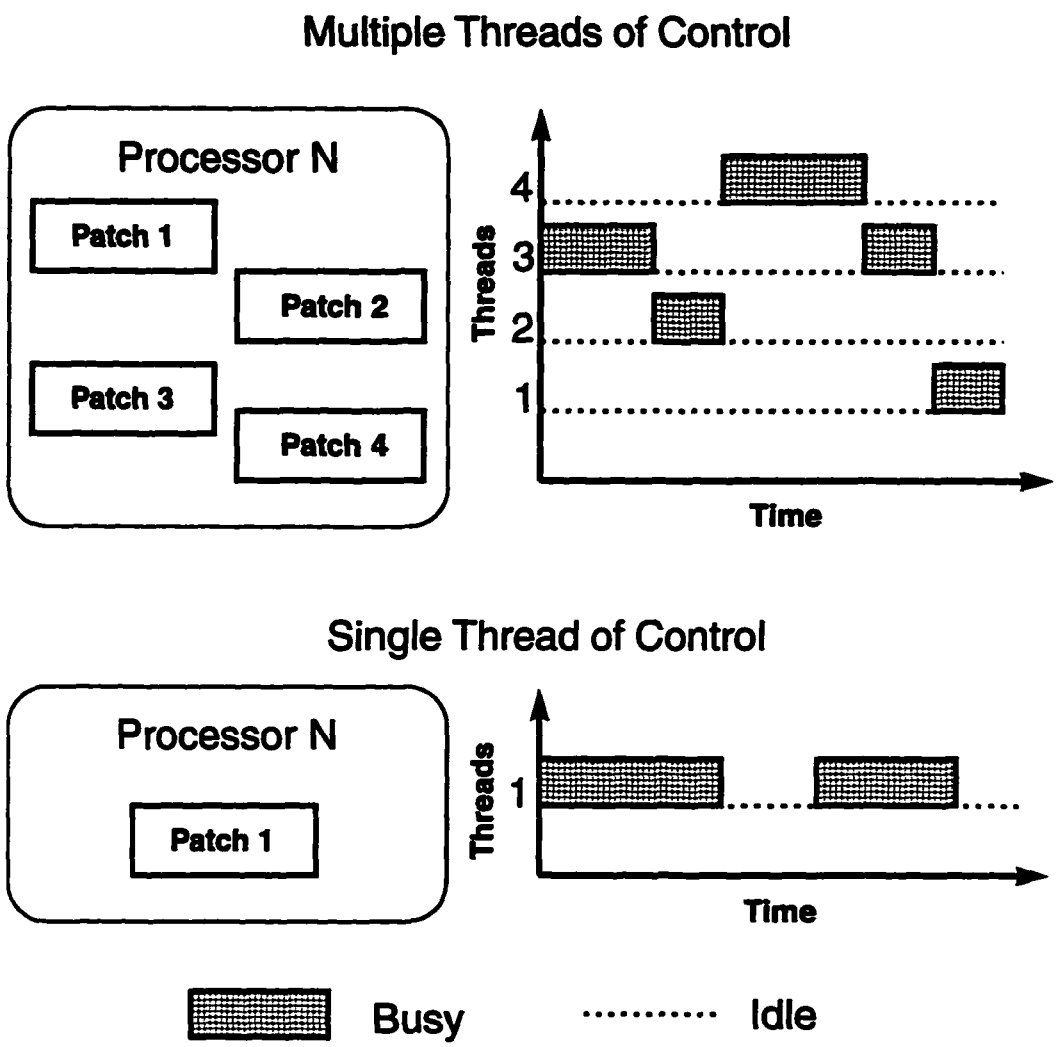


Figure A.2: Comparison of single-threaded and multi-threaded execution patterns. Multi-threaded execution allows another patch to be scheduled for execution when one is waiting for data. Single-threaded execution, without this flexibility, suffers longer periods of idle time waiting for messages from other nodes.

Appendix B

Molecular Graphics: The Program VMD[1]

B.1 Introduction

The burgeoning number of experimentally-resolved biomolecular structures has resulted in an increasing need for computational tools for molecular visualization and analysis. Several excellent packages exist for the graphical display of static molecular structures, such as RIBBONS [29], XMol [105], Midas [52, 51], SETOR [49], GRASP [115], and others. Molecular dynamics (MD) studies of the function of the structurally known biopolymers are being widely applied. The results of such studies, however, are typically large molecular trajectory files, which represent substantial amounts of dynamical data which require suitable visualization tools, e.g., to initiate molecular dynamics simulations and to display sequences of structures. The increased speed and the parallelization of computers makes it possible today to carry out molecular dynamics studies interactively, in order to probe key properties of biopolymers such as potential energy barriers along chosen reaction pathways. This approach requires a new molecular graphics user interface.

The program VMD [76] has been developed for interactive graphical display of molecular systems, in particular biopolymers such as proteins or nucleic acids. The motivation for the development of VMD has been to provide a well documented and freely available program that is easy to use and modify, and which addresses several challenges in molecular graphics including:

[1]Portions of this appendix have been taken from the manuscript: W. Humphrey, A. Dalke, and K. Schulten, VMD — Visual Molecular Dynamics, *J. Mol. Graphics*, 14:33–38, 1996.

- support for the display of *dynamic* data such as molecular trajectories generated by molecular dynamics calculations;
- direct interaction with a separate molecular dynamics application, in order to provide a graphical user interface and visualization console for the simulation program;
- the capability to work with three-dimensional, immersive display devices such as a large-screen stereo projection facility;
- text-based as well as mouse-based user interface controls, including user-customizable menus and program extensions using an interpreted scripting language;
- display of molecules in a wide range of rendering styles easily selected by the user;
- production of high-quality hardcopy images of currently displayed molecular systems.

This appendix describes the features and structure of the program VMD, first mentioning the current implementation of the program, and then discussing the major program capabilities and functionality. We then describe the use of VMD, coupled with a large-screen stereo projection system, as a collaborative tool for several researchers to employ for analysis and discussion of molecular assemblies. Finally, we discuss the use of VMD for interactive molecular dynamics modeling, and list the current availability of documentation and source code.

B.2 Implementation

VMD is written in C++ using an object-oriented design which assists maintenance of the program and the addition of new features. The distribution of VMD includes documentation describing how to compile, install, use, and modify the program for different hardware and software configurations. VMD requires either the Silicon Graphics Inc. GL library or the OpenGL library for three-dimensional graphics rendering.

A sample VMD session is shown in Fig. B.1. This snapshot illustrates the three components of the user interface: the graphics display window, the graphical user interface windows, and the VMD

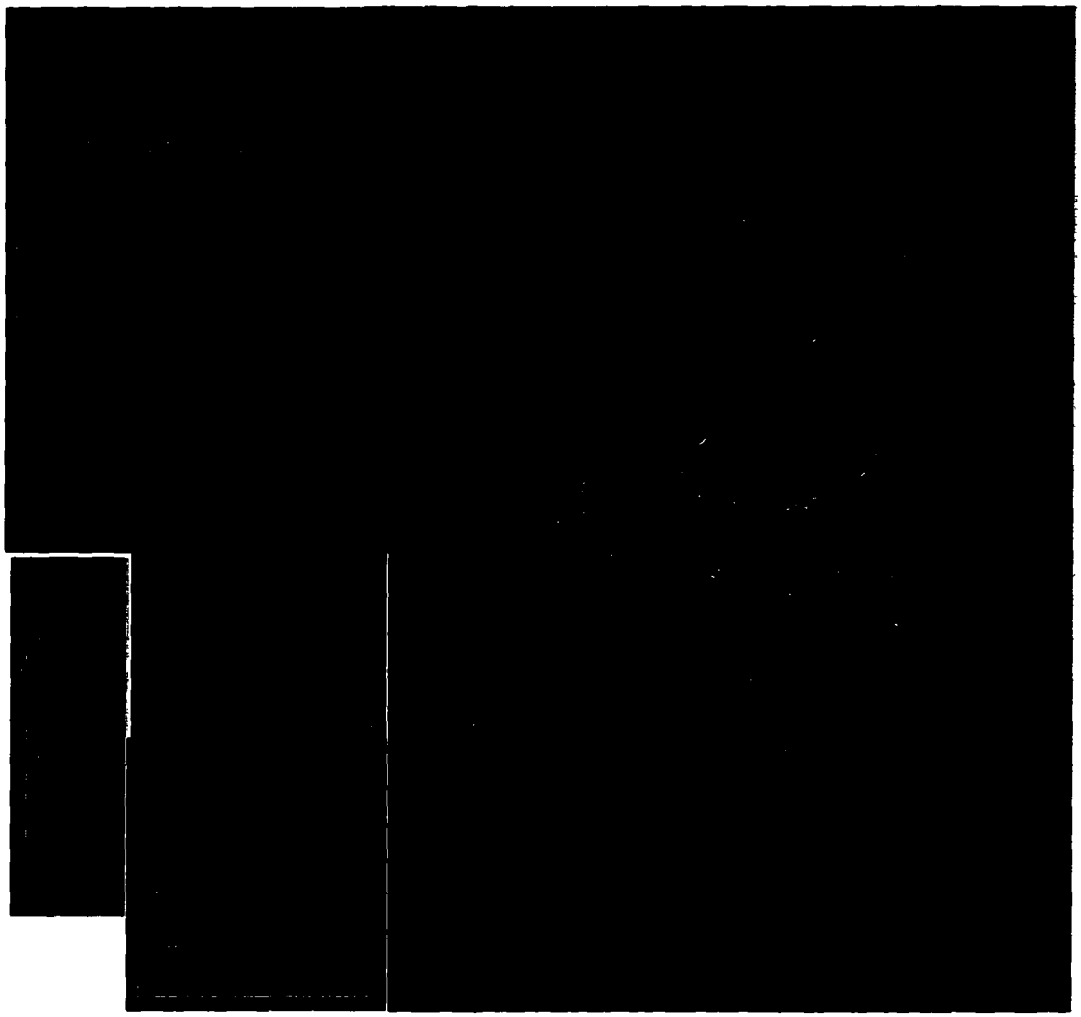


Figure B.1: Sample VMD session, illustrating the use of the graphical user interface and text console for the display of the protein BPTI bound to trypsin. Molecules are rendered in the diplay window shown in the upper-right corner of the figure, while components of the GUI are shown at the left. Below the display window is the text console interface for VMD.

console. On the right in Fig. B.1 is the graphics display window, in which molecules are rendered and interactively rotated, translated, and scaled via mouse controls. A user-customizable pop-up menu is available in this window. The image shown in the graphics display window is referred to as the current *scene*. VMD contains options for rendering the scene to a high-quality raster image file as described below.

VMD uses XForms [188] for the graphical user interface, which uses a collection of *forms* to represent the graphical controls available to the user. Examples of some of the forms available in

VMD are shown on the left in Fig. B.1. The graphical user interface (GUI) in VMD includes a toolbar which provides access to the specific forms for tasks such as changing the current molecular display characteristics or animating selected molecules using molecular dynamics trajectories.

Below the graphics display window in Fig. B.1 is the VMD text console, which displays informative messages and provides a command prompt for keyboard control of the program. All actions in VMD are available via text commands; full session logging and playback is possible. Users can write scripts which may be run at any time, and which can be executed, for example, by entering a short text command or through a user-customized pop-up menu selection. VMD employs Tcl [124], an embeddable interpreted script language parser, to process text commands. Tcl provides a full set of scripting capabilities for the program, and makes possible complex features such as variable substitution, conditional expressions, control loops, and subroutines. A user-customizable configuration file is read when VMD starts up, which may be used to define a personalized VMD working environment and to define new commands and program extensions, as well as to create customized pop-up menu controls.

B.3 Capabilities

New molecules are read into VMD from a set of molecular structure/coordinate files, or may be read from a single coordinate file. The molecular structure file contains static information about the system, such as bond connectivity and atomic mass and charge values. The molecular coordinate file contains the positions of all the atoms which make up the molecule. When a structure file is not provided, the bond connectivity for each atom is determined by VMD through a nearest-neighbor distance search. VMD understands the CHARMm [21]/X-PLOR [26] compatible PSF protein structure file format, and the Brookhaven PDB [13] coordinate file format. In addition to molecular file formats, VMD can read and display images in Raster3D [103] input file format.

Due to the large number of molecular data file formats currently in use, VMD implements an interface to the program Babel [180] to read data from formats other than PDB or PSF. If Babel is installed, it is used to convert files from alternate formats into PDB format before being read by

VMD. This also includes multiple-coordinate-set data files, such as those in XYZ format.

A key feature in VMD is the ability to work with molecular dynamics simulation programs, and to display the simulated molecule as its motion is computed (as discussed below). Thus, new structures may be loaded into VMD through connection to a running simulation, instead of being read from a file. Once loaded, however, a structure obtained from a network connection is treated in the same manner as if read from disk.

B.3.1 Animating Molecular Structures

For each molecule displayed by VMD there is an associated *animation list*, which is a collection of atomic coordinate sets for the molecule. Controls are available to play back the trajectory, with options to control the animation speed, frame increment, and animation direction. New trajectory coordinate sets may be read from PDB files, or from DCD files binary compatible with the molecular dynamics/refinement programs CHARMm [21] and X-PLOR [26]. These coordinate sets may be loaded when the molecule is initially read into VMD, or may be loaded later. A molecular trajectory editor is also available, with options to delete specific frames or sets of frames from the animation list, and to write coordinate sets to files.

B.3.2 Displaying Molecular Structures

Any number of molecules may be displayed and manipulated in the graphics display window, as shown in Fig. B.1. Each structure is drawn as several *representations*, or *views*, of the molecule. Multiple views of a molecule may be shown simultaneously to create a complex image of the system. A view is just one particular method of drawing the molecule, and consists of three characteristics:

1. An *atom subset* selection, which determines which atoms are to be included in the view.

2. A *rendering* style, which determines the primitives used to draw the atoms, bonds, and other molecule components. Table B.1 lists many of the rendering styles available in VMD.

3. A *coloring* method, which determines what color to use for the components of the view. Numerous coloring methods are available, some of which are listed in Table B.2.

Style	Description
Lines	Simple lines for bonds; points for atoms
Bonds	Lighted cylinders for bonds; no atoms
VDW	Solid, lighted van der Waals spheres for atoms; no bonds
CPK	Scaled van der Waals spheres for atoms; cylinders for bonds
Dotted	Dotted van der Waals spheres for atoms; no bonds
Licorice	Solid spheres and cylinders with equal radii
Tube	Cylindrical tube through C_α atoms[a]
Ribbon	Flat ribbon through C_α atoms[a]
Surf	Solvent-accessible surface of selected atoms[b]

[a]For nucleic acids, the P atoms are used.
[b]Using the SURF algorithm [178] developed at the University of North Carolina.

Table B.1: Selected molecule rendering styles available in VMD.

Views are created or modified via text commands or, more easily, by using the graphical interface controls. Figure B.2 shows a detailed representation of a protein dimer complexed with a segment of DNA. Molecules may be displayed using either an orthographic or a perspective projection.

The atom selection capabilities of VMD are quite extensive, and include a flexible syntax for complex selection expressions. Each atom in a molecule has several characteristics; keywords are used to select the atoms which have values matching a specified criterion. For example, each atom has a set of character string names, an (x,y,z) position and other numeric values such as charge and mass. Each atom also has boolean characteristics identifying it, for example, as a member of an amino acid or nucleic acid or as part of the protein backbone. Boolean operations may be used to select atoms based on multiple characteristics, and parentheses may be used to select the order of evaluation. For example, the selection

```
(resname ASP or resname GLY) and mass > 11
```

selects all atoms which are in aspartic acid or glycine residues, and which have a mass greater than 11. Similarly, the selection

```
backbone and within 8 of resid 100
```

selects all backbone atoms which are within 8 units (e.g. within 8 Å) of residue 100. Regular expressions may be used in strings to quickly specify many different names in a selection. When

Figure B.2: A protein-DNA complex, visualized with VMD. The double-helix structure of the DNA is clearly visible in the lower portion of the figure; above this are two identical protein segments bound to the nucleic acid sequence. The system is shown as a collection of different representations, to produce a complex image highlighting the secondary structure, location of ions, and solvent-accessible surface of subsections of the molecule.

a selection changes during a trajectory (such as the atoms within some radius of another set of atoms), the user can request an update of the selection each time a new animation timestep is displayed.

When displaying solid objects, VMD can make use of available hardware-accelerated lighting capabilities, a capability exploited in other molecular visualization packages such as SETOR [49]. Up to four independent, infinitely-distant light sources may be positioned in the scene and used to illuminate the displayed objects. Each light can be turned on or off, and may be interactively moved to new positions with the mouse.

Style	Description
Name	Color determined by atom name
Resname	Color determined by residue name
Segname	Each segment is shown in a different color
Molecule	Each molecule is shown in a different color
Chain	Color determined by one-character chain identifer
Beta	Color scale based on the beta values of the PDB file
Occupancy	Color scale based on the occupancy values of the PDB file
Mass	Color scale based on the atomic mass of each atom
Charge	Color scale based on the atomic charge of each atom
Pos	Color scale based on the distance of each atom from the center

Table B.2: Selected molecule coloring styles available in VMD.

B.3.3 Raster Image Generation

An interface to a number of image rendering packages is provided in VMD which can be used to create input scripts for use by these programs. The generated input scripts may then be read by the selected package, to create a raster output image of the graphics scene displayed by VMD suitable for publication or slides. Table B.3 lists the currently supported image rendering packages.

Name	Description
Raster3D	Fast raster file generator
POV	Persistence of Vision ray tracing package
Rayshade	Rayshade ray tracing package

Table B.3: Available rendering program output formats.

The capability to create input scripts for image rendering programs may be combined with the Tcl scripting language available in VMD to easily create high-quality movies of a molecule. As an example, the following script may be used to create a series of images showing a previously loaded molecule rotating about the 'Y' axis, by creating input scripts for the Raster3D [103] program which are processed immediately after they are created:

```
### Tcl script to create a Raster3D movie of a rotating molecule
for { set i 0 } { $i < 360 } { set i [ expr $i + 10 ] } {
  ### create and process a Raster3D script
  set scriptfile "rotate.$i.r3d"
```

```
    set imagefile "rotate.$i.rgb"
    render Raster3D $scriptfile
    catch { exec render < $scriptfile -sgi $imagefile }
    ### rotate the current image by 10 degree about the Y axis
    rot y by 10
}
```

After loading a molecule, this script would be invoked simply by typing (assuming it was in a file called *movie.tcl*): '`play movie.tcl`'. More complex transformations are performed in a similar manner.

B.3.4 Stereo Display

The scene displayed in the graphics window may be rendered in stereo, to enhance the appearance and information content of the displayed systems. The stereo viewing parameters, which include the eye separation distance and the focal length, can be interactively controlled by the user. Two stereo display formats are available in VMD:

1. A side-by-side stereo display, which splits the display window into two halves, a left-eye view and a right-eye view. This format may be used with all graphics display hardware, and is suitable for preparing images for publication.

2. A crystal-eyes stereo display, which utilizes special hardware available on many graphics workstations to display stereo images to be viewed with liquid crystal-shuttered glasses.

The stereo viewing modes may be used with standard display monitors, or with external stereo display equipment. As an example, VMD is currently being used with a three-dimensional projection system for the display and analysis of molecules viewed simultaneously by several users. This system consists of a ceiling-mounted projector, echoing onto a large (6' by 8') screen the stereo images displayed on a graphics workstation monitor. The large-screen display makes it possible for many viewers to study collaboratively at a large magnification a biopolymer system; the stereo

display environment enhances the information content of the scene through the addition of visual depth cues. VMD contains several options to configure the graphics display for use with large-screen projection systems, e.g. to set the projection screen vertical and horizontal dimensions.

One problem encountered when several people view the same computer-generated stereo image is that each person sees the scene from a different perspective. To help alleviate this problem, VMD supports the use of spatial tracking devices (such as the Polhemus Fastrak) which measure the 3D position and orientation of sensors relative to a fixed position, in order to provide a set of 3D pointers. VMD displays visual representations of the pointers on the screen, such that all viewers perceive the pointers at the same positions relative to the molecules.

New features and user interface methods are currently being added to VMD for use with this stereo projection environment to provide a natural problem solving environment for structural biology. The goal of these enhancements is to help 'untie' the users from the keyboard, to make possible manipulation and analysis of molecular structures by several collaborators at the same time in a natural working environment. The user interface controls of VMD are being updated to make use of each available 3D pointer as a 3D mouse, that would be used for rotation and control of the molecules just as the normal, 2D mouse is used. Also being explored are an audio user interface to allow users to issue spoken commands. User interface controls such as these may eventually replace or augment keyboard and menu controls.

B.3.5 Trajectory Analysis Tools

VMD includes functions for analyzing molecular structures and trajectories. These functions access the internal VMD data structures to return or modify characteristics such as charge, mass, and position for individual atoms, residues, and molecules. More complex analysis capabilities, such as computing the RMS deviation or correlation functions of a dynamics trajectory, can then be implemented as Tcl scripts without the need to modify the VMD source code.

As an example, the following commands illustrate the use of the VMD analysis language to query and change the center of mass of a specific residue. The first step is to create an atom

selection and bind it to an identifier:

```
vmd > set my_sel [atomselect top "resid 3"]
```

Information about the atoms in this selection may then be retrieved using the "get" option:

```
vmd > $my_sel get {name mass x y z}
Info) {N 14.007 1.488 2.280 -0.863} {H 1.008 0.770 1.998 -1.467}
      {CA 12.011 1.981 3.643 -0.909} {CB 12.011 1.147 4.464 -1.880}
      {C 12.011 1.865 4.326 0.444} {O 15.999 2.801 4.963 0.924}
```

Using these atom selection commands, and a set of vector manipulation procedures implemented in Tcl, the function to compute the center of mass of a selected set of atoms is then realized as follows:

```
### Tcl procedure to compute the center of mass of a set of atoms
proc com {sel} {
  set atoms [$sel get {mass x y z}]
  set totalMass 0.0
  set com [veczero]
  foreach atom $atoms {
    set mass [lindex $atom 0]
    set pos [lrange $atom 1 3]
    set totalMass [expr $totalMass + $mass]
    set com [vecadd $com [vecscale $mass $pos]]
  }
  return [vecscale $com [expr 1.0 / $totalMass]]
}
```

The application of this procedure is straightforward; executing the new command "com" with a previously defined atom selection identifier as an argument returns the center of mass of the atoms in that selection:

```
vmd > com $my_sel
Info) 4.59783 5.14667 4.275
```

This information could then be used, for example, to move the center of mass of the atoms to some other position, for example to the origin:

```
vmd > $my_sel moveby [vecinvert [com $sel]]
```

In addition to the calculation of center of mass motion, several other Tcl procedures are available in VMD for dynamic trajectory analysis including computation of RMSD values, autocorrelation functions, and Ramachandran plot ($\phi - \psi$ angle) data. New analysis functions can be easily developed and added to VMD by the user.

B.4 Interactive Molecular Dynamics

VMD is designed to act as a visualization console and graphical front end for a molecular dynamics application running on a remote supercomputer or high performance workstation. VMD uses a set of daemons and library routines, known as the MDCOMM software, to broker the communication of data and commands between VMD and a remote simulation program. The MDCOMM software interface is independent of the particular remote application; the daemons buffer data transfer between the application and VMD, and act as managers for the running jobs and interprocess communication. Once a particular simulation program has been suitably modified to work as an MDCOMM client, VMD can be used with that program without change.

VMD is the visualization component of a larger set of computational tools for structural biology known as MDScope [112]. MDScope includes not only VMD and MDCOMM, but also the parallel molecular dynamics program NAMD [113]. NAMD is a parallel, portable molecular dynamics program written in C++, which implements the CHARMm force field [21] and contains numerous simulation options. NAMD uses a spatial decomposition algorithm to distribute the computation tasks among parallel processors, which partitions the volume of space occupied by the simulated molecule into uniform cubes, known as *patches*, assigned to different processing nodes. While

NAMD and VMD may be used independently of each other, taken together, VMD, NAMD and MDCOMM constitute MDScope. Information on MDScope may be obtained from the MDScope WWW home page, `http://www.ks.uiuc.edu/Research/mdscope`.

VMD provides an interface to initialize a new simulation, and can serve also to monitor and control certain simulation parameters. After a simulation is started and a connection between VMD and the application is made, molecular structures are communicated to VMD as they are calculated. The connection between VMD and a running simulation may be dropped and later reestablished, e.g., to allow a user to check on the progress of a running simulation without having to stop and restart the calculation. Molecules displayed in this manner may also be manipulated and rendered in the same fashion as molecules loaded from data files, and transferred molecular structures may be saved in an animation list for storage and playback.

B.4.1 Example of Use

Figure B.3 illustrates an example of the use of VMD for interactive molecular dynamics. The first step in establishing a connection to a running simulation is to select a host computer, and to choose between starting a new simulation or reconnecting to a running job (Fig. B.3a). The initial connection to the host computer returns the lists of available applications and running simulations from a daemon running on the host computer. Selecting an available application in order to start a new simulation (in this case a simulation using NAMD) results in VMD requesting from the host computer the list of parameters necessary to start the job (Fig. B.3b). Once all the parameters are entered, the simulation program is launched, and the static molecular structure data are returned to VMD. As the molecular trajectory is computed, the coordinate sets are sent to VMD for display (Fig. B.3c). The available graphical user interface controls may be used to view or modify simulation parameters or display characteristics, or to terminate the simulation. Figure B.3c also shows several atoms together with the applied forces, the latter specified by the user with the mouse (the horizontal and vertical arrows). These forces have transformed the small polypeptide from an α-helix conformation to the form shown in the figure.

Once a connection to a running simulation has been established, VMD provides options to modify simulation parameters, such as the temperature, and remote connection parameters, such as the frequency with which coordinate sets are communicated from the simulation program to VMD. The user can also directly participate in the simulation through the addition of perturbative or guiding forces to selected atoms or residues, a feature developed and studied earlier in the program SCULPT [163]. Forces are added using the mouse to indicate the magnitude and direction of the additional interaction on selected atoms; these forces are communicated to the simulation program and incorporated into the dynamics.

The use of interactive guiding forces in simulations of biomolecular systems is currently being applied or considered for a number of projects. Two systems being examined with these interactive tools are the protein bacteriorhodopsin, and the binding of biotin to the protein avidin. For bacteriorhodopsin (bR), a 26 kd membrane protein used to convert light energy into a proton gradient for ATP synthesis in *Halobacterium halobium* [118], molecular dynamics simulations are being performed using interactive forces to guide the positioning of water molecules near key residues. Such water molecules could not be resolved in the experimentally determined bR structure, but are considered to play key roles in the proton transfer mechanism of bR. Interactive forces are also being considered for the study of the interaction of the 32-atom biotin vitamin with the protein avidin. Avidin is a 15.6 kd protein found in tetrameric form in animal and reptile egg-white, which has a strong binding affinity for biotin [131]. Interactively applied forces could be used to pull biotin from the binding site in avidin, in order to shed light on the molecular mechanism of how avidin accomodates the biotin structure.

B.5 Documentation and Availability

Extensive documentation on how to use the visualization features of VMD and how to modify and extend the program is available. The following documents, in PostScript format, are provided via anonymous ftp:

- an installation guide, describing how to compile and install VMD;

- a user's guide, explaining the capabilities and features of VMD;
- a programmer's guide, listing the structure and layout of VMD and indicating how to add new capabilities.

An on-line help feature is also available.

The complete set of source code and documentation files for VMD, as well as a precompiled binary for Silicon Graphics workstations, is available free of charge for noncommercial use via anonymous ftp from the ftp server `ftp.ks.uiuc.edu`, in the directory `/pub/mdscope/vmd`. Up-to-date information on VMD may be obtained by accessing the VMD WWW home page, `http://www.ks.uiuc.edu/Research/vmd`. VMD is available primarily for Silicon Graphics workstations, running version 5 or later of the IRIX operating system. The program has also been compiled and run on Hewlett-Packard PA-RISC workstations running version 9 of the HP-UX operating system, and on IBM RS-6000 workstations running AIX.

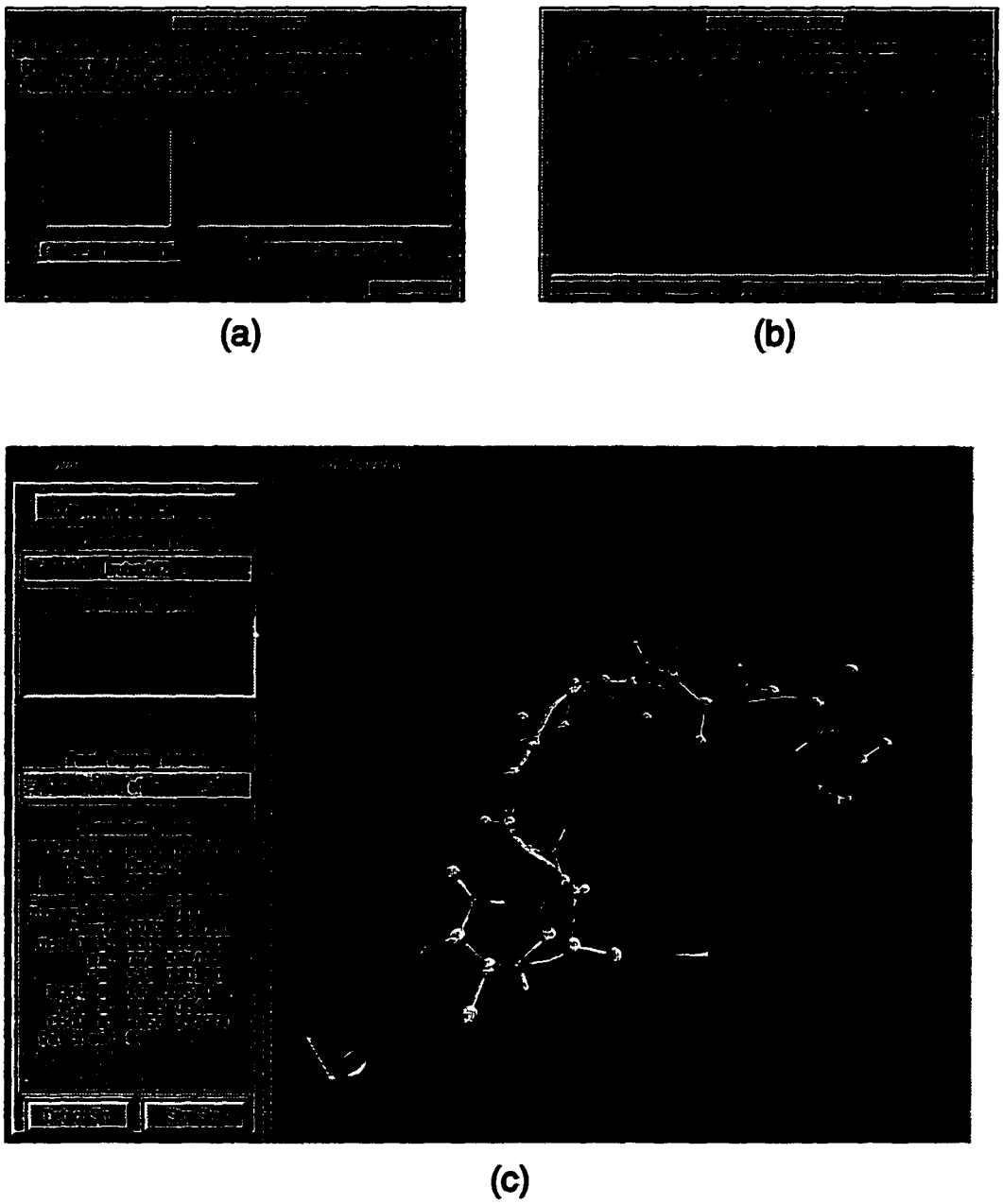


Figure B.3: VMD being used to initiate and control a molecular dynamics simulation on a remote workstation. (a) Initially, VMD connects to a daemon running on the remote computer and requests a list of available applications and a list of currently running jobs. (b) After selecting an application, a list of required parameters for the simulation is retrieved, and is filled in by the user. (c) Following the transfer of the completed parameter list to the remote host, the simulation program is launched and the molecular structure is returned to VMD. As the molecular trajectory is computed is it sent to VMD for display. The blue arrows indicate user-specified guiding forces being applied to selected atoms in the molecule, while the form at left displays the current simulation status and graphical user controls.

Appendix C

Combined Quantum/Classical Simulations

This appendix describes the derivation of the atomic forces which arise in the combined quantum/classical molecular dynamics simulations described in chapter 6. In this scheme, a single degree of freedom of the protein is treated quantum mechanically, while all other atoms in the molecule are treated classically.

C.1 The Density Matrix Evolution Method

C.1.1 Three-Level Model Hamiltonian

We consider here the case of a three-level system, dependent on a single coordinate q, which models a single molecular degree of freedom quantum mechanically. The system is assumed to be governed by the Hamiltonian

$$H(q) = \begin{pmatrix} E_1(q) & V_{12} & V_{13} \\ V_{12} & E_2(q) & V_{23} \\ V_{13} & V_{23} & E_3(q) \end{pmatrix} \tag{C.1}$$

where the diagonal elements $E_i(q)$ represent potential energy surfaces for the system, and the off-diagonal elements V_{ij} describe the coupling between surfaces i and j. In the discussion below, the V_{ij} elements are taken to be constant, although this assumption could be relaxed in a more general derivation. Given an orthonormal basis set of stationary wavefunctions $|\phi_i(q)\rangle$ for this system, the total time-dependent wavefunction $|\psi(q,t)\rangle$ is given by

$$|\psi(q,t)\rangle = \sum_{i=1}^{3} c_i(t)|\phi(q)\rangle. \tag{C.2}$$

To describe the evolution in time of this system, we employ the density matrix $\rho(t)$, where

$$\rho_{ij}(t) = c_i^*(t)c_j(t). \tag{C.3}$$

The density matrix describes the evolution of the wavefunction subject to the Hamiltonian in (C.1), where the diagonal elements of ρ give the average occupation of each state, and the off-diagonal elements contain the relative phase of each component wavefunction $|\phi_i(q)\rangle$. The density matrix has the properties of

$$\mathrm{Tr}\,\rho = \sum_i |c_i|^2 = 1 \tag{C.4}$$

and

$$\rho^\dagger = \rho, \tag{C.5}$$

and the average value of an operator A is given by

$$\langle A \rangle = \langle \psi | A | \psi \rangle = \mathrm{Tr}\,\rho A. \tag{C.6}$$

From (C.6), the expectation value of the energy of the system is then

$$\langle H(q) \rangle = V(q) = \mathrm{Tr}\,\rho H(q). \tag{C.7}$$

Inserting (C.1) in (C.7), we then have for the potential energy of the system

$$V(q) = \sum_i \rho_{ii} E_i(q) + \sum_{i \neq j} \rho_{ij} V_{ij}. \tag{C.8}$$

C.1.2 Derivation of Quantum Forces

To couple this quantum degree of freedom q to the classically-treated molecular environment, we must determine the forces acting on the atoms involved in this degree of freedom and incorporate them into the numerical integration of the classical equations of motion. The net forces acting on the atoms in the degree of freedom q are then

$$\vec{F}_{net} = \vec{F}_c + \vec{F}_{qm}, \tag{C.9}$$

where $\vec{F}_c$ represents the "classical" forces arising from interaction with the surrounding molecular environment, and $\vec{F}_{qm}$ is determined from the negative gradient of the potential energy $V(q)$ in (C.8):

$$\vec{F}_{qm}(q) = -\frac{d}{dq}V(q) \tag{C.10}$$

$$= -\sum_{i=1}^{3}\rho_{ii}\frac{d}{dq}E_i(q) \tag{C.11}$$

$$= -\sum_{i=1}^{3}\rho_{ii}\vec{F}_{qm,i}(q). \tag{C.12}$$

The second term in (C.8) drops out after taking the derivative. Here, $\vec{F}_{qm,i}(q)$ is the force which would result from just the single potential energy surface $E_i(q)$; the net force arising from the quantum degree of freedom is then just the average force from the three potential energy surfaces in (C.1), weighted by the occupation of each state (ρ_{ii}).

As an example, in the study of the retinal photoisomerization reaction in chapter 6, the quantum mechanically treated reaction coordinate q is the C_{13}=C_{14} dihedral angle rotation, which is a four-body interaction involving atoms C_{12}, C_{13}, C_{14}, and C_{15}. The potential energy surfaces $E_j(q)$ employed in the Hamiltonian matrix H, as shown in Fig. 6.2, have the form

$$E_j(q) = k_j[1+\cos(n_j q+\delta_j)], \tag{C.13}$$

which is the same functional form as the dihedral angle interaction term in the classical molecular dynamics force field. The force on each atom k involved in the C_{13}=C_{14} dihedral angle due to the potential energy surface $E_j(q)$ is then the negative gradient of this surface:

$$\vec{F}_{qm,jk}(q) = -\vec{\nabla}_{\vec{r}_k}E_j(q) \tag{C.14}$$

where q is a function of the atom positions $\vec{r}_k$.

C.1.3 Evolution of the Density Matrix

In a combined quantum/classical simulation, we must compute the time evolution of the density matrix in conjunction with the computation of the classical and quantum mechanical forces. The

equation of motion for the density matrix of the quantum system is given by the Liouville-von Neumann equation,

$$\dot{\rho} = \frac{i}{\hbar}[\rho, H]. \tag{C.15}$$

In practice, to numerically compute the solution to (C.15), we must represent the complex-valued matrix ρ as

$$H(q) = \begin{pmatrix} a_1 + a_2 & b_1 + ib_2 & c_1 + ic_2 \\ b_1 - ib_2 & 1/2 - a_1 & d_1 + id_2 \\ c_1 - ic_2 & d_1 - id_2 & 1/2 - a_2 \end{pmatrix}, \tag{C.16}$$

where $a_1 \ldots d_2$ are real quantities. Notice this representation preserves the properties of the density matrix that $\rho^\dagger = \rho$, and $\mathrm{Tr}\,\rho = 1$. Inserting (C.16) and (C.1) in (C.15), we obtain the following set of first-order, coupling linear differential equations which describe the evolution of the quantum system:

$$\dot{a}_1 = \frac{2}{\hbar}[d_2 V_{23} - b_2 V_{12}] \tag{C.17}$$

$$\dot{a}_2 = \frac{2}{\hbar}[-c_2 V_{13} - d_2 V_{23}] \tag{C.18}$$

$$\dot{b}_1 = \frac{1}{\hbar}[-d_2 V_{13} - c_2 V_{23} - b_2 E_{21}] \tag{C.19}$$

$$\dot{b}_2 = \frac{1}{\hbar}[b_1 E_{21} - d_1 V_{13} + c_1 V_{23} + (2a_1 + a_2 - 1/2) V_{12}] \tag{C.20}$$

$$\dot{c}_1 = \frac{1}{\hbar}[-c_2 E_{31} + d_2 V_{12} - b_2 V_{23}] \tag{C.21}$$

$$\dot{c}_2 = \frac{1}{\hbar}[c_1 E_{31} - d_1 V_{12} + b_1 V_{23} + (a_1 + 2a_2 - 1/2) V_{13}] \tag{C.22}$$

$$\dot{d}_1 = \frac{1}{\hbar}[-d_2 E_{32} + c_2 V_{12} + b_2 V_{13}] \tag{C.23}$$

$$\dot{d}_2 = \frac{1}{\hbar}[d_1 E_{32} - c_1 V_{12} + b_1 V_{13} + (a_2 - a_1) V_{23}]. \tag{C.24}$$

Here, $E_{ij} = E_i - E_j$, where each energy surface E is evaluated at the current value of the reaction coordinate q. The solution of this system of differential equations may be computed numerically using, e.g., a fourth-order Runge-Kutta algorithm.

C.1.4 Simulation Algorithm

The algorithm for a classical molecular dynamics simulations combined with quantum mechanical treatment of a degree of freedom q is then as follows:

1. Initialize the density matrix $\rho(t=0)$ to indicate the initial quantum state of the system;

2. For each time step in the simulation:

   (a) Calculate $H(q)$ for the current reaction coordinate value q;

   (b) Update the density matrix $\rho(t)$ through numerical solution of (C.15);

   (c) Find the forces $\vec{F}_{qm}$ for each atom in the degree of freedom q, using (app:forces) with the updated values of $H(q)$ and $\rho(t)$;

   (d) Calculate all other classical forces arising from the standard molecular dynamics force field, including the classical forces $\vec{F}_c$ acting on the atoms in the degree of freedom q;

   (e) Update the position of each atom in the molecule through numerical integration of the Newtonian equations of motion.

The quantum system couples to the classical system through the forces $\vec{F}_{qm}$. The classical equations of motion are then used to find the net force on the atoms and to update the positions of each atom, including the quantum reaction coordinate q. Thus, the external molecular environment acts essentially like a mean field which perturbs the quantum system.

Bibliography

[1] A. Albeck, N. Livnah, H. Gottlieb, and M. Sheves. C_{13} NMR-studies of model compounds for bacteriorhodopsin – facts affecting the retinal chromophore chemical-shifts and absorption maximum. *J. Am. Chem. Soc.*, 114:2400–2411, 1992.

[2] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics II. behavior of a small number of elastic spheres. *J. Chem. Phys.*, 33:1439–1451, 1960.

[3] B. Aton, A. G. Doukas, D. Narva, R. Callender, and U. Dinur. Resonance Raman studies of the primary photo-chemical event in visual pigments. *Biophys. J.*, 29:79–84, 1980.

[4] T. Baasov, N. Friedmann, and M. Sheves. Factors affecting the C=N stretching in protonated retinal Schiff-base – a model study for bacteriorhodopsin and visual pigments. *Biochemistry*, 26:3210–3217, 1987.

[5] T. Baasov and M. Sheves. Alteration of pK_a of the bacteriorhodopsin protonated Schiff base. A study with model compounds. *Biochemistry*, 25:5249–5258, 1986.

[6] S. P. Balashov, R. Govindjee, E. S. Imasheva, S. Misra, T. G. Ebrey, Y. Feng, R. K. Crouch, and D. R. Menick. The two pK_a's of aspartate-85 and control of thermal isomerization and proton release in the arginine-82 to lysine mutant of bacteriorhodopsin. *Biochemistry*, 34:8820–8834, 1995.

[7] S. P. Balashov, E. S. Imasheva, R. Govindjee, and T. G. Ebrey. Quantum yield ratio of the forward and back light reactions of bacteriorhodopsin at low temperature and photosteady-state concentration of the bathoproduct K. *Photochem. Photobiol.*, 54:955–961, 1991.

[8] S. P. Balashov, N. V. Karneyeva, F. F. Litvin, and T. G. Ebrey. Bathoproducts and conformers of all-*trans*- and 13-*cis*-bacteriorhodopsin at 90 k. *Photochem. Photobiol.*, 54:949–953, 1991.

[9] E. Bamberg, J. Tittor, and D. Oesterhelt. Light-driven proton or chloride pumping by halorhodopsin. *Proc. Natl. Acad. Sci. USA*, 90:473–486, 1992.

[10] D. Bashford and K. Gerwert. Electrostatic calculations of the pK values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.*, 224:473–486, 1992.

[11] B. Becher, F. Tokunaga, and T. G. Ebrey. Ultraviolet and visible absorption spectra of the purple membrane protein and the photocycle intermediates. *Biochemistry*, 17:2293–2300, 1978.

[12] H. Berendsen and J. Mavri. Quantum simulation of reaction dynamics by density matrix evolution. *J. Phys. Chem.*, 97:13464–13468, 1993.

[13] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rogers, O. Kennard, T. Shimanouchi, and M. Tasumi. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542, 1977.

[14] R. R. Birge. Photophysics and molecular electronic applications of the rhodopsins. *Ann. Rev. Phys. Chem.*, 41:683–733, 1990.

[15] R. R. Birge, L. P. Murray, R. Zidovetzki, and H. M. Knapp. Two-photon, ^{13}C and two-dimensional ^{1}H NMR spectroscopic studies of retinyl Schiff bases, protonated Schiff bases, and Schiff base salts: Evidences for a protonation induced $\pi\pi^*$ excited state level ordering reversal. *J. Am. Chem. Soc.*, 109:2090–2101, 1987.

[16] R. R. Birge and C. F. Zhang. Two photon spectroscopy of light adapted bacteriorhodopsin. *J. Chem. Phys.*, 92:7178–7195, 1990.

[17] V. Bonacic-Koutecky, K. Schoffel, and J. Michl. Critically heterosymmetric biradicaloid geometries of the protonated schiff bases. *Theoret. Chim. Acta*, 72:459–474, 1987.

[18] M. Braiman and R. Mathies. Resonance Raman spectra of bacteriorhodopsin's primary photoproduct: Evidence for a distored 13-*cis* retinal chromophore. *Proc. Natl. Acad. Sci. USA*, 79:403–407, 1982.

[19] M. S. Braiman, T. Mogi, T. Marti, L. J. Stern, H. G. Khorana, and K. J. Rothschild. Vibrational spectroscopy of bacteriorhodopsin mutants: Light-driven proton transport involves protonation changes of aspartic acid residues 85, 96, and 212. *Biochemistry*, 27:8516–8520, 1988.

[20] C. Branden and J. Tooze. *Introduction to Protein Structure*. Garland Publishing, Inc., New York and London, 1991.

[21] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.

[22] B. R. Brooks and M. Hodošček. Parallelization of CHARMm for MIMD machines. *Chemical Design Automation News*, 7:16–22, Dec. 1992.

[23] L. S. Brown, J. Sasaki, H. Kandori, A. Maeda, R. Needleman, and J. K. Lanyi. Glutamic acid 204 is the terminal proton release group at the extracellular surface of bacteriorhodopsin. *J. Biol. Chem.*, 270:27122–27126, 1995.

[24] A. Brünger. Simulated annealing in crystallography. *Ann. Rev. Phys. Chem.*, 42:197–223, 1991.

[25] A. Brünger, Z. Schulten, and K. Schulten. A network thermodynamic investigation of stationary and non-stationary proton transport through proteins. *Z. Phys. Chem.*, NF136:1–63, 1983.

[26] A. T. Brünger. *X-PLOR*. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 1988.

[27] Y. Cao, G. Váró, M. Chang, B. Ni, R. Needleman, and J. K. Lanyi. Water is required for proton transfer from aspartate-96 to the bacteriorhodopsin Schiff base. *Biochemistry*, 74:10972–10979, 1991.

[28] L. Carlacci, M. W. Schulz, and K. C. Chou. Geometric and energy parameters in lysine-retinal chromophores. *Protein Engng*, 4:885–889, 1991.

[29] M. Carson. RIBBONS 2.0. *J. Appl. Cryst.*, 24:958–961, 1991.

[30] L. S. Cederbaum, H. Köppel, and W. Domcke. Multimode vibronic coupling effects in molecules. *Int. J. Quantum Chem. Symp.*, 15:251–267, 1981.

[31] B. Curry, I. Palings, A. Broek, J. Pardoen, J. Lugtenburg, and R. Mathies. Vibrational analysis of the retinal isomers. *Adv. Infrared Raman Spectrosc.*, 12:115–178, 1985.

[32] V. Daggett and M. Levitt. A molecular dynamics simulation of the C-terminal fragment of the L7/L12 ribosomal protein in solution. *Chem. Phys.*, 158:501–512, 1991.

[33] Z. Dancshazy, T. G. Ebrey, and R. Govindjee. Independent photocycles of the spectrally distinct forms of bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA*, 85:6358–6361, 1988.

[34] H. J. M. de Groot, G. S. Harbison, J. Herzfeld, and R. G. Griffin. Nuclear magnetic-resonance study of the Schiff-base in bacteriorhodopsin – counterion effects on the ^{15}N shift anisotropy. *Biochemistry*, 28:3346–3353, 1989.

[35] H. J. M. de Groot, S. O. Smith, J. Courtin, E. van den Berg, C. Winkel, J. Lugtenburg, R. G. Griffin, and J. Herzfeld. Solid-state ^{13}C and ^{15}N NMR study of the low pH forms of bacteriorhodopsin. *Biochemistry*, 29:6873–6883, 1990.

[36] R. Diller and M. Stockburger. Kinetic resonance Raman studies reveal different conformational states of bacteriorhodopsin. *Biochemistry*, 27:7641–7651, 1988.

[37] J. Dobler, W. Zinth, W. Kaiser, and D. Oesterhelt. Excited-state reaction dynamics of bacteriorhodopsin studied by femtosecond spectroscopy. *Chem. Phys. Lett.*, 144:215–220, 1988.

[38] S. J. Doig, P. J. Reid, and R. A. Mathies. Picosecond time-resolved resonance Raman spectroscopy of bacteriorhodopsin's J, K, and KL intermediates. *J. Phys. Chem.*, 95:6372–6379, 1991.

[39] A. Doukas, A. Pandi, T. Suzuki, R. Callender, and M. Ottolenghi. On the mechanism of hydrogen-deuterium exchange in bacteriorhodopsin. *Biophys. J.*, 33:275–280, 1981.

[40] M. Du and G. Fleming. Femtosecond time-resolved fluorescence spectroscopy of bacteriorhodopsin: Direct observation of excited state dynamics in the primary step of the proton pump cycle. *Biol. Cybernetics*, 48:101–111, 1993.

[41] P. Dupuis, F. Harosi, C. Sandorfy, J. M. Leclercq, and D. Vocelle. 1st step in vision – proton-transfer or isomerization. *Rev. Can. Biol.*, 39:247–258, 1980.

[42] T. N. Earnest, P. Roepe, M. S. Braiman, J. Gillespie, and K. J. Rothschild. Orientation of the bacteriorhodopsin chromophore probed by polarized Fourier-transform infrared difference spectroscopy. *Biochemistry*, 25:7793–7798, 1986.

[43] T. Ebrey. Light energy transduction in bacteriorhodopsin. In M. Jacobson, editor, *Thermodynamics of Membranes, Receptors and Channels*, pages 353–387. CRC Press, New York, 1993.

[44] T. G. Ebrey, B. Becher, B. Mao, P. Kilbride, and B. Honig. Exciton interactions and chromophore orientation in the purple membrane. *J. Mol. Biol.*, 112:377–397, 1977.

[45] W. Eisfeld, C. Pusch, R. Diller, R. Lohrmann, and M. Stockburger. Resonance Raman and optical transient studies of the light-induced proton pump of bacteriorhodopsin reveal parallel photocycles. *Biochemistry*, 32:7196–7215, 1993.

[46] M. A. El-Sayed, C. T. Lin, and W. R. Mason. Is there an excitonic interaction or antenna system in bacteriorhodopsin? *Proc. Natl. Acad. Sci. USA*, 86:5376–5379, 1989.

[47] M. Engelhard, B. Hess, G. Metz, W. Kreutz, F. Siebert, J. Soppa, and D. Oesterhelt. High resolution ^{13}C-solid state NMR of bacteriorhodopsin: Assignment of specific aspartic acids and structural implications of single site mutations. *Eur. Biophys. J.*, 18:17–24, 1990.

[48] M. Engels, K. Gerwert, and D. Bashford. Computational studies of the early intermediates of the bacteriorhodopsin photocycle. *Biophys. Chem.*, 56:94–104, 1995.

[49] S. V. Evans. SETOR: Hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics*, 11:134–138, 1993.

[50] K. Fahmy, F. Siebert, M. Grossjean, and P. Tavan. Photoisomerization in bacteriorhodopsin studied by FTIR, linear dichroism and photoselection experiments combined with quantum chemical theoretical analysis. *J. Mol. Struc.*, 214:257–288, 1989.

[51] T. E. Ferrin, G. S. Couch, C. C. Huang, E. F. Pettersen, and R. Langridge. An affordable approach to interactive desktop molecular modeling. *J. Mol. Graphics*, 9:27–32, 1991.

[52] T. E. Ferrin, C. C. Huang, L. E. Jarvis, and R. Langridge. The MIDAS database system. *J. Mol. Graphics*, 6:2–12, 1988.

[53] S. P. A. Fodor, J. B. Ames, R. Gebhard, E. M. M. van den Berg, W. Stoeckenius, J. Lugtenburg, and R. A. Mathies. Chromophore structure in bacteriorhodopsin's N intermediate: Implications for the proton-pumping mechanism. *Biochemistry*, 27:7097–7101, 1988.

[54] M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople. *Gaussian 92, Revision A*. Gaussian Inc., Pittsburgh, PA, 1992.

[55] Y. Gat, M. Grossjean, I. Pinevsky, H. Takei, Z. Rothman, H. Sigrist, A. Lewis, and M. Sheves. Participation of bacteriorhodopsin active-site lysine backbone in vibrations associated with retinal photochemistry. *Proc. Natl. Acad. Sci. USA*, 89:2434–2438, 1992.

[56] Y. Gat and M. Sheves. A mechanism for controlling the pK_a of the retinal protonated Schiff base in retinal proteins. A study with model compounds. *J. Am. Chem. Soc.*, 115:3772–3773, 1993.

[57] K. Gerwert, B. Hess, J. Soppa, and D. Oesterhelt. Role of Asp-96 in proton translocation by bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA*, 86:4943–4947, 1989.

[58] K. Gerwert and F. Siebert. Evidence for light-induced 13-*cis* 14-s-*cis* isomerization in bacteriorhodopsin obtained by Fourier transform IR difference spectrometry. *EMBO J.*, 4:805–812, 1986.

[59] K. Gerwert, G. Souvignier, and B. Hess. Simultaneous monitoring of light-induced changes in protein side-group protonation, chromophore isomerization, and backbone motion of bacteriorhodopsin by time-resolved Fourier-transform infrared spectroscopy. *Proc. Natl. Acad. Sci. USA*, 87:9774–9778, 1990.

[60] R. Govindjee, S. P. Balashov, and T. G. Ebrey. Quantum efficiency of the photochemical cycle of bacteriorhodopsin. *Biophys. J.*, 58:597–608, 1990.

[61] R. Govindjee, M. Kono, S. P. Balashov, E. Imasheva, M. Sheves, and T. G. Ebrey. Effects of substitution of tyrosine 57 with asparagine and phenylalanine on the properties of bacteriorhodopsin. *Biochemistry*, 34:4828–4838, 1995.

[62] M. F. Grossjean, P. Tavan, and K. Schulten. Quantum chemical vibrational analysis of the chromophore of bacteriorhodopsin. *J. Phys. Chem.*, 94:8059–8069, 1990.

[63] G. S. Harbison, S. O. Smith, J. A. Pardoen, J. M. L. Courtin, J. Lugtenburg, J. Herzfeld, R. A. Mathies, and R. G. Griffin. Solid-state ^{13}C NMR detection of a perturbed 6-s-*trans* chromophore in bacteriorhodopsin. *Biochemistry*, 24:6955–6962, 1985.

[64] K. Hasson, F. Gai, and P. Anfinrud. The photoisomerization of retinal in bacteriorhodopsin: A new model. *Proc. Natl. Acad. Sci. USA*, 1996. Submitted.

[65] R. Henderson. The purple membrane from *Halobacterium halobium*. *Ann. Rev. Biochem. Bioeng.*, 6:87–109, 1977.

[66] R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing. Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy. *J. Mol. Biol.*, 213:899–929, 1990.

[67] R. Henderson and P. N. T. Unwin. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*, 257:28–32, 1975.

[68] M. P. Heyn, P. J. Bauer, and N. A. Dencher. A natural CD label to probe the structure of the purple membrane from *Halobacterium halobium* by means of exciton coupling effects. *Biochem. Biophys. Res. Commun.*, 67:897–903, 1975.

[69] M. P. Heyn and H. Otto. Photoselection and transient linear dichroism with oriented immobilized purple membranes: Evidence for motion of the C(20)-methyl group of the chromophore towards the cytoplasmic side of the membrane. *Photochem. Photobiol.*, 56:1105–1112, 1992.

[70] M. P. Heyn, J. Westerhausen, I. Wallat, and F. Seiff. High-sensitivity neutron-diffraction of membranes – location of the Schiff-base end of the chromophore of bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA*, 85:2146–2150, 1988.

[71] P. Hildebrandt and M. Stockburger. Role of water in bacteriorhodopsin's chromophore: Resonance Raman study. *Biochemistry*, 23:5539–5548, 1984.

[72] M. Holz, L. A. Drachev, T. Mogi, H. Otto, A. D. Kaulen, M. P. Heyn, V. P. Skulachev, and H. G. Khorana. Replacement of aspartic acid-96 by asparagine in bacteriorhodopsin slows both the decay of the M intermediate and the associated proton movement. *Proc. Natl. Acad. Sci. USA*, 86:2167–2171, 1989.

[73] B. S. Hudson, B. E. Kohler, and K. Schulten. Linear polyene electronic structure and potential surfaces. In E. C. Lim, editor, *Excited States*, volume 6, pages 1–95. Academic Press, 1982.

[74] W. Humphrey, I. Logunov, K. Schulten, and M. Sheves. Molecular dynamics study of bacteriorhodopsin and artificial pigments. *Biochemistry*, 33:3668–3678, 1994.

[75] W. Humphrey, D. Xu, M. Sheves, and K. Schulten. Molecular dynamics study of the early intermediates in the bacteriorhodopsin photocycle. *J. Phys. Chem.*, 99:14549–14560, 1995.

[76] W. F. Humphrey, A. Dalke, and K. Schulten. VMD – Visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996.

[77] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparision of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[78] T. Kakitani, H. Kakitani, H. Rodman, B. Honig, and R. Callender. Correlation of vibrational frequencies with absorption maxima in polyenes, rhodopsin, bacteriorhodopsin, and retinal analogs. *J. Phys. Chem.*, 87:3620–3628, 1983.

[79] O. Kalisky and M. Ottolenghi. Branching pathways in the photocycle of bacteriorhodopsin. *Photochem. Photobiol.*, 35:109–115, 1982.

[80] O. Kalisky, M. Ottolenghi, B. Honig, and R. Korenstein. Environmental effects on formation and photoreaction of the M412 photoproduct of bacteriorhodopsin: Implications for the mechanism of proton pumping. *Biochemistry*, 20:649–655, 1981.

[81] H. Kandori, Y. Yamazaki, J. Sasaki, R. Needleman, J. K. Lanyi, and A. Maeda. Water-mediated proton transfer in proteins: An FTIR study of bacteriorhodopsin. *J. Am. Chem. Soc.*, 117:2118–2119, 1995.

[82] L. Keszthelyi, S. Száraz, and A. D. adn W. Stoechenius. Bacteriorhodopsin and halorhodopsin: Multiple ion pumps. *Biochim. Biophys. Acta*, 1018:260–262, 1990.

[83] H. G. Khorana. Bacteriorhodopsin, a membrane protein that uses light to translocate protons. *J. Biol. Chem.*, 263:7439–7442, 1988.

[84] P. J. Knowles and H.-J. Werner. An efficient second-order MC SCF method for long configuration expansions. *Chem. Phys. Lett.*, 115:259–267, 1985.

[85] R. Korenstein and B. Hess. Hydration effects on *cis-trans* isomerization of bacteriorhodopsin. *FEBS Lett.*, 82:7–11, 1977.

[86] T. Kouyama, A. M. Kouyama, A. Ikegami, M. K. Mathew, and W. Stoeckenius. Bacteriorhodopsin photoreaction – identification of a long-lived intermediate N (P, R350) at high pH and its M-like photoproduct. *Biochemistry*, 27:5855–5863, 1988.

[87] A. Kriebel and A. Albrecht. Excitonic interaction among three chromophores: An application to the purple membrane of *Halobacterium halobium*. *J. Chem. Phys.*, 65:4575–4583, 1977.

[88] J. K. Lanyi. Proton-transfer and energy coupling in the bacteriorhodopsin photocycle. *J. Bioenerg. Biomemb.*, 24:169–179, 1992.

[89] S. W. Lin and R. A. Mathies. Orientation of the protonated retinal Schiff base group in bacteriorhodopsin from absorption linear dichroism. *Biophys. J.*, 56:653–660, 1989.

[90] N. Livnah and M. Sheves. Model compounds can mimic spectroscopic properties of bovine rhodopsin. *J. Am. Chem. Soc.*, 115:351–353, 1993.

[91] S. Logunov, M. El-Sayed, and J. Lanyi. Replacement effects of neutral amino acid residues of different molecular volumes in the retinal binding cavity of bacteriorhodopsin on the dynamics of its primary process. *Biophys. J.*, 70:2875–2881, 1996.

[92] S. Logunov, M. El-Sayed, and L. Song. Photoisomerization quantum yield and apparent energy content of the K intermediate in the photocycles of bacteriorhodopsin, its mutants D85N, R82Q, and D212N, and deionized blue bacteriorhodopsin. *J. Phys. Chem.*, 100:2391–2398, 1996.

[93] R. H. Lozier, R. A. Bogomolni, and W. Stoeckenius. Bacteriorhodopsin: A light-driven proton pump in *Halobacterium halobium*. *Biophys. J.*, 15:955–962, 1975.

[94] R. H. Lozier, W. Niederberger, M. Ottolenghi, G. Sivorinovsky, and J. W. Stoechenius. On the photocycle of light- and dark-adapted bacteriorhodopsin. In S. R. J. Caplan and M. Ginzburg, editors, *Energetics and Structure of Halophilic Microorganisms*, pages 123–141. North Holland Biomedical Press, Elsevier, 1978.

[95] A. Maeda, J. Sasaki, J. Pfefferlé, Y. Shichida, and T. Yoshizawa. Fourier transform infrared spectral studies on the Schiff base mode of all-*trans* bacteriorhodopsin and its photointermediates, K and L. *Photochem. Photobiol.*, 54:911–921, 1991.

[96] A. Maeda, J. Sasaki, Y. Shishida, and T. Yoshizawa. Water structural changes in the bacteriorhodopsin photocycle: Analysis by Fourier transform infrared spectroscopy. *Biochemistry*, 31:462–467, 1992.

[97] A. Maeda, J. Sasaki, Y. Shishida, T. Yoshizawa, M. Chang, B. Ni, R. Needleman, and J. K. Lanyi. Structures of aspartic acid-96 in the l and n intermediates of bacteriorhodopsin: Analysis by Fourier transform infrared spectroscopy. *Biochemistry*, 31:4684–4690, 1992.

[98] A. Maeda, J. Sasaki, Y. Yamazaki, R. Needleman, and J. K. Lanyi. Interaction of aspartate-85 with a water molecule and the protonated Schiff base in the L intermediate of bacteriorhodopsin – a Fourier-transform infrared spectroscopic study. *Biochemistry*, 33:1713–1717, 1994.

[99] R. A. Mathies, C. H. Brito Cruz, W. T. Pollard, and C. V. Shank. Direct observation of the femtosecond excited-state *cis-trans* isomerization in bacteriorhodopsin. *Science*, 240:777–779, 1988.

[100] R. A. Mathies, S. W. Lin, J. B. Ames, and W. T. Pollard. From femtoseconds to biology: Mechanism of bacteriorhodopsin's light-driven proton pump. *Ann. Rev. Biochem. Bioeng.*, 20:491–518, 1991.

[101] J. Mavri, H. J. C. Berendsen, and W. F. V. Gunsteren. Influence of solvent on intramolecular proton transfer in hydrogen malonate – molecular dynamics simulation study of tunneling by density matrix evolution and nonequilibrium solvation. *J. Phys. Chem.*, 97:13469–13476, 1993.

[102] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, June 1977.

[103] E. A. Merritt and M. E. P. Murphy. Raster3D version 2.0 - a program for photorealistic molecular graphics. *Acta Cryst. D*, 50:869–873, 1994.

[104] G. Metz, F. Siebert, and M. Engelhard. Asp-85 is the only internal aspartic acid that gets protonated in the M intermediate and the purple-to-blue transition of bacteriorhodopsin. *FEBS Lett.*, 303:237, 1992.

[105] Minnesota Supercomputer Center, Inc., Minneapolis, MN. *XMol, Version 1.3.1*, 1993.

[106] T. Mogi, T. Marti, and H. G. Khorana. Structure-function studies on bacteriorhodopsin .9. Substitution of tryptophan residues affect protein-retinal interactions in bacteriorhodopsin. *J. Biol. Chem.*, 264:14197–14201, 1989.

[107] T. Mogi, L. Stern, T. Marti, B. Chao, and H. Khorana. Aspartic acid substitutions affect proton translocation by bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA*, 85:4148–4152, 1988.

[108] T. Mogi, L. J. Stern, B. H. Chao, and H. G. Khorana. Structure-function studies on bacteriorhodopsin .8. Substitution of the membrane-embedded Proline-50, Proline-91, and Proline-186 – the effects are determined by the substituting amino-acids. *J. Biol. Chem.*, 264:14192–14196, 1989.

[109] Molecular Simulations Inc., Burlington, Massachusetts. *QUANTA 4.0*, 1994.

[110] J. F. Nagle and M. Mille. Molecular-models of proton pumps. *J. Chem. Phys.*, 74:1367–1372, 1981.

[111] R. Needleman, M. Chang, B. Ni, G. Varo, J. Fornes, S. H. White, and J. K. Lanyi. Properties of $Asp^{212} \rightarrow Asn$ bacteriorhodopsin suggest that Asp^{212} and Asp^{85} both participate in a counterion and proton complex near the Schiff base. *J. Biol. Chem.*, 266:11478–11484, 1991.

[112] M. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. Skeel, K. Schulten, and R. Kufrin. MDScope— A visual computing environment for structural biology. *Comput. Phys. Commun.*, 91(1, 2 and 3):111–134, 1995.

[113] M. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. Kale, R. D. Skeel, and K. Schulten. NAMD— A parallel, object-oriented molecular dynamics program. *J. Supercomputing App.*, 1996. In press.

[114] B. Ni, M. Chang, A. Duschl, J. Lanyi, and R. Needleman. An efficient system for the synthesis of bacteriorhodopsin in *Halobacterium halobium*. *Gene*, 90:169–172, 1990.

[115] A. Nicholls, K. Sharp, and B. Honig. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Structure, Function and Genetics*, 11:282–290, 1991.

[116] M. Nonella, A. Windemuth, and K. Schulten. Structure of bacteriorhodopsin and *in situ* isomerization of retinal: A molecular dynamics study. *J. Photochem. Photobiol.*, 54(6):937–948, 1991.

[117] D. Oesterhelt and W. Stoeckenius. Blue membrane formation at low pH. *Nature*, 233:149, 1971.

[118] D. Oesterhelt and W. Stoeckenius. Functions of a new photoreceptor membrane. *Proc. Natl. Acad. Sci. USA*, 70:2853–2857, 1973.

[119] D. Oesterhelt, J. Tittor, and E. Bamberg. A unifying concept for ion translocation in retinal proteins. *J. Bioenerg. Biomemb.*, 24:181–191, 1992.

[120] M. Olivucci, F. Bernardi, P. Celani, I. Ragazos, and M. A. Robb. Excited-state cis-trans isomerization of *cis*-hexatriene. a CAS-SCF computational study. *J. Am. Chem. Soc.*, 116:1077–1085, 1994.

[121] G. Orlandi and K. Schulten. Coupling of stereochemistry and proton donor–acceptor properties of a Schiff base: A model of a light-driven proton pump. *Chem. Phys. Lett.*, 64:370–374, 1979.

[122] H. Otto, T. Marti, M. Holz, T. Mogi, L. J. Stern, F. Engel, H. G. Khorana, and M. P. Heyn. Substitution of amino acids Asp-85, Asp-212, and Arg-82 in bacteriorhodopsin affects the proton release phase of the pump and the pK of the Schiff base. *Proc. Natl. Acad. Sci. USA*, 87:1018–1022, 1990.

[123] M. Ottolenghi and M. Sheves. Synthetic retinals as probes for the binding site and photoreactions in rhodopsins. *J. Membr. Biol.*, 112:193–212, 1989.

[124] J. Ousterhout. *Tcl and the Tk Toolkit.* Addison-Wesley, Reading, Massachusetts, 1994.

[125] G. Papadopoulos, N. Dencher, G. Zaccai, and G. Büldt. Water molecules and exchangeable hydrogen ions at the active centre of bacteriorhodopsin localized by neutron diffraction. *J. Mol. Biol.*, 214:15–19, 1990.

[126] L. A. Parodi, R. H. Lozier, S. M. Bhattacharjee, and J. F. Nagle. Testing kinetic models for the bacteriorhodopsin photocycle-II. inclusion of an O to M back reaction. *Photochem. Photobiol.*, 40:501–506, 1984.

[127] L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA*, 37:205–211, 1951.

[128] J. W. Petrich, J. Breton, J. L. Martin, and A. Antonetti. Femtosecond absorption spectroscopy of light-adapted and dark-adapted bacteriorhodopsin. *Chem. Phys. Lett.*, 137:369–375, 1987.

[129] H. Polland, M. A. Franz, W. Zinth, W. Kaiser, E. Kölling, and D. Oesterhelt. Early picosecond events in the photocycle of bacteriorhodopsin. *Biophys. J.*, 49:651–662, 1986.

[130] W. T. Pollard, S. L. Dexheimer, Q. Wang, L. A. Peteanu, C. V. Shank, and R. A. Mathies. Theory of dynamics absorption spectroscopy of nonstationary states. 4. Application to 12-fs resonant impulsive raman spectroscopy of bacteriorhodopsin. *J. Phys. Chem.*, 96:6147–6158, 1992.

[131] L. Pugliese, A. Coda, M. Malcovati, and M. Bolognesi. Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 Å resolution. *J. Mol. Biol.*, 231:698–710, 1993.

[132] A. Rahman. Correlations in the motion of atoms in liquid argon. *Phys. Rev. A*, 136:405–411, 1964.

[133] A. Rahman and F. H. Stillinger. Molecular dynamics study of liquid water. *J. Chem. Phys.*, 55:3336–3359, 1971.

[134] M. Renard and M. Delmelle. pH and salt effects on the slow intermediates of the bacteriorhodopsin photocycle – a flash-photolysis study. *Eur. Biophys. J.*, 12:223–228, 1985.

[135] R. Renthal and R. Regalado. Cooperativity of the dehydration blue-shift of bacteriorhodopsin. *Photochem. Photobiol.*, 54:931–935, 1992.

[136] H.-T. Richter, L. S. Brown, R. Needleman, and J. K. Lanyi. A linkage of the pKa's of Asp-85 and Glu-204 forms part of the reprotonation switch of bacteriorhodopsin. *Biochemistry*, 1996. In press.

[137] K. J. Rothschild, M. S. Braiman, and Y. W. He. Vibrational spectroscopy of bacteriorhodopsin mutants. evidence for the interaction of aspartic acid 212 with tyrosine 185 and possible role in the proton pump mechanism. *J. Biol. Chem.*, 265:16985–16991, 1990.

[138] K. J. Rothschild, Y. W. He, S. Sonar, T. Marti, and H. G. Khorana. Vibrational spectroscopy of bacteriorhodopsin mutants - evidence that Thr-46 and Thr-89 form part of a transient network of hydrogen bonds. *J. Biol. Chem.*, 267:1615–1622, 1992.

[139] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comp. Phys.*, 23:327–341, 1977.

[140] R. Sampogna and B. Honig. Environmental effects on the protonation states of active site residues in bacteriorhodopsin. *Biophys. J.*, 66:1341–1352, 1994.

[141] B. D. Santarsiero and M. N. G. James. Crystal structure of *n*-methyl-*n*-phenylretinal iminium perchlorate: A structural model for the bacteriorhodopsin chromophore. *J. Am. Chem. Soc.*, 112:9416–9418, 1990.

[142] C. Scharnagl, J. Hettenkofer, and S. Fisher. Electrostatic and conformational effects on the proton translocation steps in bacteriorhodopsin: Analysis of multiple M structures. *J. Phys. Chem.*, 99:7787–7800, 1995.

[143] G. Schneider, R. Diller, and M. Stockburger. Photochemical quantum yield of bacteriorhodopsin from resonance Raman scattering as a probe for photolysis. *Chem. Phys.*, 131:17–29, 1989.

[144] K. Schulten. An isomerization model for the photocycle of bacteriorhodopsin. In S. R. Caplan and M. Ginzburg, editors, *Energetics and Structure of Halophilic Organisms*, pages 331–334. Elsevier, 1978.

[145] K. Schulten. Curve crossing in a protein: Coupling of the elementary quantum process to motions of the protein. In D. Bicout and M. J. Field, editors, *Proceedings of the Ecole de Physique des Houches*, pages 85–118, Paris, 1995. Les Editions de Physique, Springer.

[146] K. Schulten, U. Dinur, and B. Honig. The spectra of carbonium ions, cyanine dyes, and protonated Schiff base polyenes. *J. Chem. Phys.*, 73(8):3927–3935, 1980.

[147] K. Schulten, W. Humphrey, I. Logunov, M. Sheves, and D. Xu. Molecular dynamics studies of bacteriorhodopsin's photocycles. *Israel Journal of Chemistry*, 35:447–464, 1995.

[148] K. Schulten and M. Karplus. On the origin of a low-lying forbidden transition in polyenes and related molecules. *Chem. Phys. Lett.*, 14(3):305–309, 1972.

[149] K. Schulten, Z. Schulten, and P. Tavan. An isomerization model for the pump cycle of bacteriorhodopsin. In L. Bolis, E. J. M. Helmreich, and H. Passow, editors, *Information and Energy Transduction in Biological Membranes*, pages 113–131. Allan R. Liss, Inc., New York, 1984.

[150] K. Schulten and P. Tavan. A mechanism for the light-driven proton pump of *Halobacterium halobium*. *Nature*, 272:85–86, 1978.

[151] B. Schwartz, E. Bittner, O. Prezhdo, and P. Rossky. Quantum decoherence and the isotope effect in condensed phase nonadiabatic molecular dynamics simulations. *J. Chem. Phys.*, 104:5492–5955, 1996.

[152] R. Sen, J. Carriker, V. Balogh-Nair, and K. Nakanishi. Studies of a photoaffinity label for bovine rhodopsin. *J. Am. Chem. Soc.*, 104:3214–3216, 1982.

[153] L. Serrano-Andreas, R. Lindh, B. Roos, and M. Merchan. Theoretical study of the electronic spectrum of all-tran-1,3,5,7-octatraene. *J. Phys. Chem.*, 97:9360–9368, 1993.

[154] M. Sheves, A. Albeck, N. Friedman, and M. Ottolenghi. Controlling the pK_a of the bacteriorhodopsin Schiff-base by use of artificial retinal analogs. *Proc. Natl. Acad. Sci. USA*, 83:3262–3266, 1986.

[155] M. Sheves, T. Baasov, N. Friedman, M. Ottolenghi, R. Feinmann-Weinberg, V. Rosenbach, and B. Ehrenberg. On the binding-site of bacteriorhodopsin – a study with artificial pigments. *J. Am. Chem. Soc.*, 106:2435–2437, 1984.

[156] S. O. Smith, I. Hornung, R. van der Steen, J. A. Pardoen, M. S. Braiman, J. Lugtenburg, and R. A. Mathies. Are C_{14}-C_{15} single bond isomerizations of the retinal chromophore involved in the proton-pumping mechanism of bacteriorhodopsin? *Proc. Natl. Acad. Sci. USA*, 83:967–971, 1986.

[157] L. Song, M. A. El-Sayed, and J. K. Lanyi. Protein catalysis of the retinal subpicosecond photo-isomerization in the primary process of bacteriorhodopsin photosynthesis. *Science*, 261:891–894, 1993.

[158] J. L. Spudich. Protein-protein interaction converts a proton pump into a sensory receptor. *Cell*, 79:447–450, 1994.

[159] G. Steinberg, N. Friedman, M. Sheves, and M. Ottolenghi. Isomer compostion and spectra of the dark and light adapted forms of artifical bacteriorhodopsins. *Photochem. Photobiol.*, 54:969–676, 1991.

[160] L. J. Stern and H. G. Khorana. Structure-function studies on bacteriorhodopsin. Individual substitutions of arginine residues by glutamine affect chromophore formation, photocycle, and proton translocation. *J. Biol. Chem.*, 264:14202–14208, 1989.

[161] S. Subramaniam, T. Marti, and H. G. Khorana. Protonation state of Asp (Glu)-85 regulates the purple-to-blue transition in bacteriorhodopsin mutants Arg-82 → Ala and Asp-85 → Glu: The blue form is inactive in proton translocation. *Proc. Natl. Acad. Sci. USA*, 87:1013–1017, 1990.

[162] S. Subramanian, M. Gerstein, D. Oesterhelt, and R. Henderson. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. *EMBO J.*, 12:1–8, 1993.

[163] M. C. Surles, J. S. Richardson, D. C. Richardson, and F. P. Brooks. Sculpting proteins interactively: Continual energy minimization embedded in a graphical modeling system. *Protein Science*, 3:198–210, 1994.

[164] P. Tavan and K. Schulten. The low-lying electronic excitations in long polyenes: A PPP–MRD–CI study. *J. Chem. Phys.*, 85(11):6602–6609, 1986.

[165] P. Tavan, K. Schulten, and D. Oesterhelt. The effect of protonation and electrical interactions on the stereochemistry of retinal Schiff bases. *Biophys. J.*, 47:415–430, 1985.

[166] J. Tittor and D. Oesterhelt. The quantum yield of bacteriorhodopsin. *FEBS Lett.*, 263:269–273, 1990.

[167] J. Tittor, D. Oesterhelt, and E. Bamberg. Bacteriorhodopsin mutants D85N, D85T, and D85,96N as proton pumps. *Biophys. Chem.*, 56:153–157, 1995.

[168] J. Tittor, U. Schweiger, D. Oesterhelt, and E. Bamberg. Inversion of proton translocation in bacteriorhodopsin mutants D85N, D85T, and D85,96N. *Biophys. J.*, 67(4):1682–1690, 1994.

[169] M. Tsuda, M. Glaccum, B. Nelson, and T. G. Ebrey. Light isomerizes the chromophore of bacteriorhodopsin. *Nature*, 287:351–353, 1980.

[170] J. C. Tully. Molecular dynamics with electronic transitions. *J. Chem. Phys.*, 93:1061–1071, 1990.

[171] A. S. Ulrich, I. Wallat, M. P. Heyn, and A. Watts. Evidence for a curved retinal in BR from solid-state ^{2}H-NMR. In J. L. Rigaud, editor, *Structures and Functions of Retinal Proteins*, volume 221, pages 247–249. Colloque INSERM / John Libbey Eurotext Ltd., 1992.

[172] W. F. van Gunsteren and H. J. C. Berendsen. *GROMOS Manual*. BIOMOS b. v., Lab. of Phys. Chem., Univ. of Groningen, 1987.

[173] P. J. M. van Laarhoven and E. H. L. Aarts. *Simulated Annealing: Theory and Application*. D. Reidel, Dordrecht, 1987.

[174] G. Váró and J. K. Lanyi. Pathways of the rise and decay of the M photointermediate(s) of bacteriorhodopsin. *Biochemistry*, 29:2241–2250, 1990.

[175] G. Váró and J. K. Lanyi. Protonation and deprotonation of the M-intermediate, N-intermediate, and O-intermediate during the bacteriorhodopsin photocycle. *Biochemistry*, 29:6858–6865, 1990.

[176] G. Váró and J. K. Lanyi. Kinetic and spectroscopic evidence for an irreversible step between deprotonation and reprotonation of the Schiff-base in the bacteriorhodopsin photocycle. *Biochemistry*, 30:5008–5015, 1991.

[177] G. Váró and J. K. Lanyi. Thermodynamics and energy coupling in the bacteriorhodopsin photocycle. *Biochemistry*, 30:5016–5022, 1991.

[178] A. Varshney, F. P. Brooks, and W. V. Wright. Linearly scalable computation of smooth molecular surfaces. *IEEE Comp. Graphics and Applications*, 14:19–25, 1994.

[179] L. Verlet. Computer 'experiments' on classical fluids: I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159:98–103, 1967.

[180] P. Walters and M. Stahl. *Babel Version 1.1*. University of Arizona, Tuscon, AZ, 1992.

[181] A. Warshel and N. Barboy. Energy storage and reaction pathways in the first step of the vision process. *J. Am. Chem. Soc.*, 104:1469–1476, 1982.

[182] A. Warshel, Z. T. Chu, and J.-K. Hwang. The dynamics of the primary event in rhodopsins revisited. *Chem. Phys.*, 158:303–314, 1991.

[183] P. K. Weiner and P. A. Kollman. AMBER: Assisted model building with energy refinement. a general program for modeling molecules and their interactions. *J. Comp. Chem.*, 2:287, 1981.

[184] H.-J. Werner and P. J. Knowles. A second order multiconfiguration SCF procedure with optimum convergence. *J. Chem. Phys.*, 82:5053–5063, 1985.

[185] S. Wu and M. A. El-Sayed. CD spectrum of bacteriorhodopsin. *Biophys. J.*, 60:190–197, 1991.

[186] D. Xu, C. Martin, and K. Schulten. Molecular dynamics study of early picosecond events in the bacteriorhodopsin photocycle: Dielectric response, vibrational cooling and the J, K intermediates. *Biophys. J.*, 70(1):453–460, 1996.

[187] D. Xu, M. Sheves, and K. Schulten. Molecular dynamics study of the M412 intermediate of bacteriorhodopsin. *Biophys. J.*, 69(6):2745–2760, 1995.

[188] T. C. Zhao and M. Overmars. *Forms Library for X*, 1995.

[189] F. Zhou, A. Windemuth, and K. Schulten. Molecular dynamics study of the proton pump cycle of bacteriorhodopsin. *Biochemistry*, 32(9):2291–2306, 1993.

Vita

William F. Humphrey was born on March 22, 1968, to Fowler (Chuck) and Judy Humphrey in Devil's Lake, North Dakota. He was raised in Fargo, North Dakota and attended Oak Grove High School, starting in 1982 and graduating valedictorian in 1986. His undergraduate studies were carried out from 1986 to 1990 at Drake University in Des Moines, Iowa, where he majored in physics, mathematics, and computer science, and graduated summa cum laude with a B.S. degree. He enrolled in the Ph.D. program in physics at the University of Illinois at Urbana-Champaign in 1990, where he taught classical mechanics as a teaching assistant from 1990 to 1992. In January of 1992 he completed a M.S. degree at the University of Illinois, and joined the Theoretical Biophysics Group. Until his graduation, he worked on molecular dynamics (MD) studies of the protein bacteriorhodopsin, and on development of the MD visualization program VMD. During the period 1993 to 1994, he was the recipient of a DOE Computational Science Graduate Research Fellowship, and from 1994 to the present, he has been supported by grants from the NIH and the NSF. At the time of this writing he has been offered a postdoctoral position with the Advanced Computing Lab at Los Alamos National Laboratory.

Publications:

[1] Humphrey, W., Logunov, I., Schulten, K., and Sheves, M., Molecular Dynamics Study of Bacteriorhodopsin and Artificial Pigments (1994), *Biochemistry* 33, 3668-3678.

[2] Humphrey, W., Xu, D., Sheves, M., and Schulten, K., Molecular Dynamics Study of the Early Intermediates in the Bacteriorhodopsin Photocycle (1995), *J. Phys. Chem.* 99, 14549-14560.

[3] Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kale, L., Skeel, R., Schulten, K., and Kufrin, R., MDScope - A Visual Computing Environment for Structural Biology (1995), *Comp. Phys. Comm.* 91, 111-134.

[4] Logunov, I., Humphrey, W., Schulten, K., and Sheves, M., Molecular Dynamics Study of the 13-*cis* Form (bR_{548}) of Bacteriorhodopsin and Its Photocycle (1995), *Biophys. J.* 68, 1270-1282.

[5] Reynders, J., Forslund, D., Hinker, P., Tholburn, M., Kilman, D., and Humphrey, W., OOPS - An Object-Oriented Particle Simulation Class Library for Distributed Architectures (1995), *Comp. Phys. Comm* 87, 212-224.

[6] Humphrey, W., Dalke, A., and Schulten, K., VMD - Visual Molecular Dynamics (1996), *J. Molecular Graphics* 14, 33-38.

[7] Schulten, K., Humphrey, W., Logunov, I., Sheves, M., and Xu, D., Molecular Dynamics Studies of Bacteriorhodopsin's Photocycles (1996), *Israel J. Chem.* 35, 447-464.

[8] Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kale, L., Skeel, R., and Schulten, K., NAMD – A Parallel, Object-Oriented Molecular Dynamics Program (1996), *J. Supercomputing App.*, in press.