ELSEVIER

# Molecular dynamics of bacteriorhodopsin

**James A. Lupo[1] and Ruth Pachter**

*Materials Directorate, Wright Laboratory (USAF), Wright-Patterson Air Force Base, Ohio.*

*A model of bacteriorhodopsin (bR), with a retinal chromophore attached, has been derived for a molecular dynamics simulation. A method for determining atomic coordinates of several ill-defined strands was developed using a structure prediction algorithm based on a sequential Kalman filter technique. The completed structure was minimized using the GROMOS force field. The structure was then heated to 293 K and run for 500 ps at constant temperature. A comparison with the energy-minimized structure showed a slow increase in the all-atom RMS deviation over the first 200 ps, leveling off to approximately 2.4 Å relative to the starting structure. The final structure yielded a backbone-atom RMS deviation from the crystallographic structure of 2.8 Å. The residue neighbors of the chromophore atoms were followed as a function of time. The set of persistent near-residue neighbors supports the theory that differences in $pK_a$ values control access to the Schiff base proton, rather than formation of a counterion complex. © 1997 by Elsevier Science Inc.*

## INTRODUCTION

Bacteriorhodopsin (bR) is a protein that connects the outer and inner surfaces of the purple membrane in *Halobacterium holobium*. It is composed of seven $\alpha$ helices, and has a retinal chromophore in the interior of its structure attached via a protonated Schiff base linkage to the side chain of a lysine residue (216). Details of the bR structure and its photochemistry have been reviewed [1]. Interest has focused on potential nonlinear optical applications [2].

Several previous molecular dynamics simulations of bR have been performed [3–5]. The model by Nonella et al. [5] introduced the concept of adding the missing loops to achieve dynamic stability rather than applying position constraints. The study concentrated on the conformations assumed by the retinal molecule during the course of the photocycle. The structure developed by Nonella et al. was then used by Humphrey et al., with the addition of 16 internal water molecules, [4] to examine potential structures for the Schiff base counterion. Ferrand et al. [3] also created a structure via a technique similar to that of Nonella et al., and added 21 internal water molecules. They found that only seven waters were confined within the structure, and the predictions of neutron scattering agreed well with experimental results. These studies were based on the CHARMM force field [6]. Sankara-Ramakrishnan and Vishveshwara used the AMBER force field [7] in a study of the F helix (Leu-174–Ile-198) [8].

Mutagenic studies have been performed to identify the amino acid residues involved in the absorption process [1,9,10]. Ahl et al. [9] found that Tyr-185 was directly involved with color regulation and that Trp-182, Tyr-185, Pro-186, Trp-189, and Ser-193 appeared to serve as part of the retinal-binding pocket [9]. Sankara-Ramakrishnan and Vishveshwara reported a different, although overlapping, set of residues, downplaying the participation of Tyr-185 [10]. Thus, a better understanding of the residues that play a role in the photocycle process depends on identifying those residues that neighbor the retinal atoms and the Schiff base as a function of time.

This study identifies the near-neighbor residues of the retinal atoms as a function of time during a long-duration molecular dynamics simulation. A simulation of 500 ps was performed and, at periodic intervals, all residues that approached the retinal within 4 Å were recorded time. By identifying the residues that spend time close to the retinal, we are able to restrict the set of residues that can play an active role in the photocycle, and differentiate between competing theories on how the protein controls access to the Schiff base proton. Our results tend to reject the formation of a counterion complex, and favor variation in $pK_a$ values as the controlling mechanism.

We first discuss the derivation of the full set of bR atomic coordinates, including those missing from several fragments. Those missing were mostly in the loops connecting the $\alpha$ helices. A double-iterated sequential Kalman filter estimation algorithm was used successfully to generate the missing coordinate data. Molecular dynamics simulations were then performed with this model, using the parallel UHGROMOS code. The structure reached a constant all-atom root-mean-square

Color Plates for this article are on pages 49–52.

Address reprint requests to: James A. Lupo, WL/MLPJ Bldg. 651, 3005 P Street, Suite 1, Wright-Patterson AFB, Ohio 45433-7702.
[1] Contractor: Technical Management Concepts, Inc., 2875 Rhett Drive, Beavercreek, Ohio 45434.

Received January 28, 1996; revised December 16, 1996; accepted January 10, 1997.



655 Avenue of the Americas, New York, NY 10010

1093-3263/97/$17.00
PII S1093-3263(97)00016-7

(RMS) deviation from the energy-minimized state of 2.4 Å with the average deviation stabilizing at 2.0 Å. The time-averaged structure was compared to the crystallographic structure and gave an all-atom RMS deviation of 2.7 Å. The residue neighbors of the Schiff base and retinal atoms were computed and contacts were noted as a function of time to 3-ps resolution. Thirty-one residues were found to approach within 4.0 Å of a retinal atom at one time or another. Of these, 14 made contact over 80% of the time, while 9 made contact less than 40% of the time. The deviations of the model early in the simulation run matched those found in other studies, while its continued evolution with constant deviation values and varying residue contacts demonstrated considerable flexibility in the bR macromolecule.

Overall, the conformations assumed under thermal motion were found to deviate from the crystallographic structure by less than the experimental resolution. The thermal motion was also found to stimulate *trans–cis* interconversion of the NZ–C01 and C01–C02 bonds (See Figure 1.). However, flexing in the other bonds of Lys-216 resulted in little net spatial displacement of the retinal fragment.

# COMPUTATIONAL DETAILS

## Model Construction

A procedure for constructing the missing atomic data was developed using the Protean2 molecular structure prediction tool. This method uses a double-iterated sequential Kalman filter estimation algorithm to adjust a coordinate data set until a list of applied constraints is satisfied to some specified level of accuracy [11]. The technique has proven advantageous for protein structure determination [12], and most recently for protein structure prediction at low resolution [13]. The list of constraints contains distances derived from the known bonds between pairs of atoms, as well as those implied by angular (bond and dihedral) relationships. Data sets that support Protean2 provide for automatic generation of chemistry-based constraints for amino acid residues, given only the residue sequence [14]. The system of codes also allows specification of fixed points whose coordinates are used for distance comparisons but that are not otherwise modified. These two features were integrated to enable generation of the full set of missing coordinates in the bR macromolecule.

The protocol for establishing the missing coordinates of a fragment is straightforward. First, the residue sequence of the fragment was used to generate a set of chemistry-based constraints, specifying intraresidue bonds and dihedral angles, and interresidue bonds to nearest neighbors. The atom coordinates of nearest neighbor residues at either end of the fragment were then treated as fixed points. At one end, four sets of fixed distance constraints were generated by pairing the backbone $C_\alpha$ and C atoms of the fixed neighbor residue with the backbone N and $C_\alpha$ atoms of the fragment residue. Another four fixed distance constraints were generated at the other end of the fragment using the backbone $C_\alpha$ and C atoms of the fragment terminal residue with the backbone N and $C_\alpha$ atoms of its fixed neighbor residue. Coordinates for the fixed residues were taken from the Protein Data Bank (PDB) file [15], and initial coordinates for the fragment were taken from extended linear residue sequences generated with the Quanta molecular design system [16]. The sequence data of the fragment and the coordinates generated by Protean2 were then manually patched into the PDB file. In the case of the terminal serine residue, only one fixed residue neighbor was present, hence only four fixed distance constraints were generated. Table 1 summarizes the fragments generated by this technique, and their associated fixed neighbors.

## Molecular Dynamics

Molecular dynamics simulations are well served by parallel processing techniques, particularly for such a large molecular system. The UHGROMOS molecular dynamics code promd, based on GROMOS promd [17,18], has been the focus of efforts to port it to a variety of different massively parallel machines [19]. Its use of Pfortran [20] allows for rapid porting to new architectures simply by changing the underlying interprocessor communications library without affecting the main body of scientific code. Pfortran, and hence UHGROMOS promd, were ported to an IBM SP2. Comparison benchmarking runs with promd showed the IBM SP2 to be the fastest of several available massively parallel supercomputers for this particular model [19].

The GROMOS support tools provide an automatic method for generating a molecular topology file based on a specified residue sequence [21]. To use these tools, a synthetic amino acid residue consisting of lysine with the retinal chromophore attached as an extended side chain (LysR) was created and added to the GROMOS database (see Figure 1). The GROMOS tools were then used to generate a complete structure for bR. The structure was minimized with the GROMOS energy minimizer, proem (see Color Plate 1). A comparison of the PDB structure with the resulting energy-minimized structure found an all-atom RMS deviation of 3.2 Å and an average deviation of 2.6 Å. A comparison of only the protein backbone atoms yielded an RMS deviation of 2.6 Å, with an average deviation of 2.2 Å.

The simulations of the bR model were run with the IBM SP2 parallelized version of UHGROMOS promd. A constant temperature of 293 K was specified, and no initial velocity scaling was performed. The time step was set to 1 fs. The system reached the specified temperature at about 50 ps. The time step was decreased to 0.1 fs for the remainder of the run in order to

**Table 1. List of Bacteriorhodopsin Fragments and Fixed Neighbor Residues Used to Generate Atom Coordinates Using the PROTEAN2 Procedure[a]**

| Fixed neighbor | Fragment | | Fixed neighbor |
|---|---|---|---|
| | Start | End | |
| Met-32 | Gly-33 | Pro-37 | Asp-38 |
| Leu-62 | Gly-63 | Gly-73 | Glu-74 |
| Leu-100 | Val-101 | Gln-105 | Gly-106 |
| Ala-110 | Leu-111 | Leu-111 | Val-112 |
| Leu-127 | Thr-128 | Val-136 | Trp-137 |
| Thr-157 | Ser-158 | Pro-165 | Glu-166 |
| Ile-191 | Gly-192 | Leu-201 | Asn-202 |
| Arg-225 | Ser-226 | | |

[a] Sequence numbers are according to PDB.

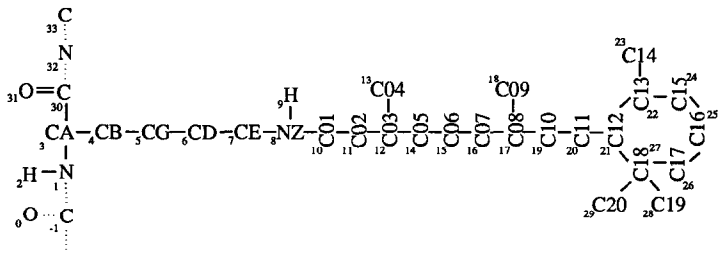

*Figure 1. Symbolic diagram of the synthetic residue Lysr (lysine with retinal attached), showing the GROMOS atom names and sequence numbers. Sequence numbers 1–30 comprise the atoms of Lysr. The remainder are used in the GROMOS automatic linking process.*

maintain structural stability due to the presence of explicit hydrogen atoms.

At 3-ps intervals, the distances of the retinal and Schiff base atoms to all other atoms in the structure were computed. If the distance was less than or equal to 4 Å, the residue to which the atom belonged was scored as being contacted. Contact analysis as a function of time was then made for the retinal fragment as a whole and for three localized components, identified by the following subgrouping of atoms: (1) the Schiff base, atoms CE, NZ, H, and C01; (2) the midsection, atoms C02–C10; and (3) the β-ionone ring, atoms C11–C20.

## RESULTS AND DISCUSSION

The total energy of the model rose smoothly with heating until about the 50-ps point. It then exhibited a slow decrease until it stabilized at about 200 ps. The dynamic structure was also compared to the energy-minimized structure as the run progressed. As seen in Figure 2, the average, maximum, and RMS deviations rise steadily until they level off after approximately 200 ps. The behavior of the maximum deviation is consistent with rapid, early changes in locations of the free ends of outer side chains, while the gradual changes in average and RMS deviation are as expected for a large system undergoing thermal adjustment. The system was assumed to be equilibrated after the 200-ps point. The final dynamic structure was found to have an all-atom RMS deviation of 2.4 Å relative to the energy-minimized starting conformation. We note that the average and RMS deviations are within the experimental resolution limits [5].

Humphrey et al. reported an RMS deviation of approximately 1.0 Å after 25 ps of simulation [4]. We saw a similar

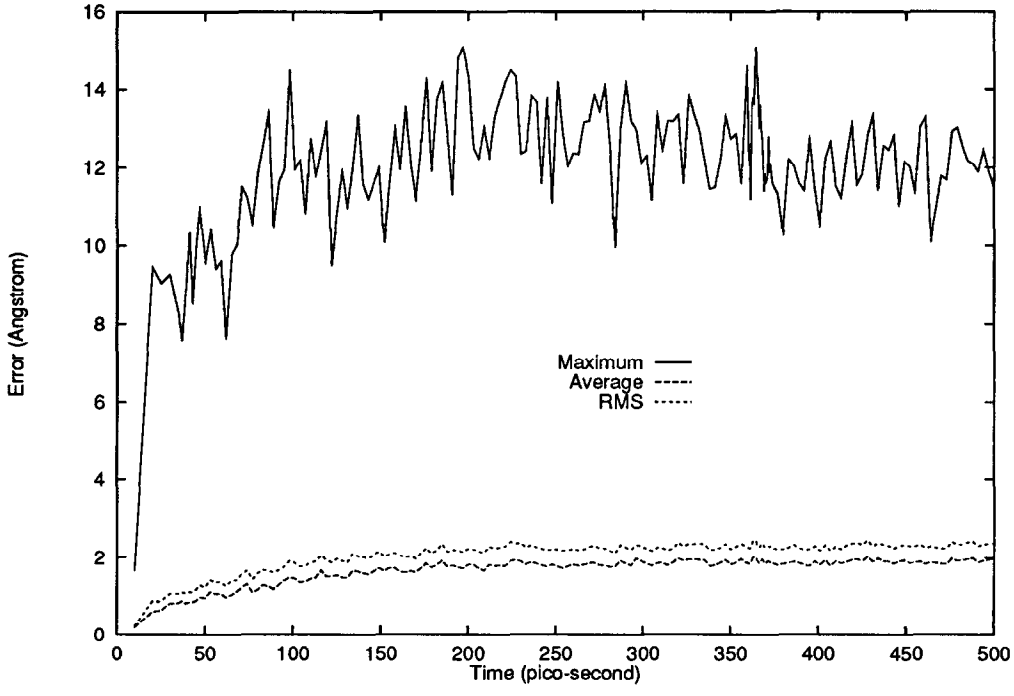


*Figure 2. A plot of the maximum, average, and RMS all-atom deviations from the minimized state (in angstroms) as a function of time for the bR model.*

change over the first 25 ps of our simulation, only to find that the RMS deviation continued to rise slowly until well after 200 ps. This suggests that there is a fair degree of flexibility in the molecule, which can be captured only on long time scales.

The resulting structures of the bR model were compared with those specified in the PDB file [15]. Comparisons of the complete molecular structure were performed using the initial energy-minimized structure and a structure formed by averaging the atomic coordinates found in the 200- to 500-ps interval. Comparisons of the individual helices were also made. All atoms in the dynamic structure that had corresponding atoms defined in the PDB file were used, while those in the added loops were ignored. Similar comparisons were made using only the backbone atoms. Table 2 summarizes the results and shows little change between the initial and final values, again indicating that the structure is stable.

It is interesting that the deviations are essentially identical to the 2.9 Å found by Nonella et al. [5]. Humphrey et al. produced a constrained structure with an overall backbone deviation of 1.8 Å, and $C_\alpha$-only deviation of 2.3 Å in the G helix (Asn-202–Arg-225) [4]. Ferrand et al. [3] averaged the backbone-atom RMS deviations over the course of their simulation, yielding 2.15 Å. We note that our G helix shows a backbone-only RMS deviation of only 1.0 Å, and positional constraints were not used.

The difference between the time-averaged dynamic structure and the crystal structure is shown in Color Plate 2 by overlaying the structures (rendered as ribbon traces) of the backbone atoms. The retinal fragment is displayed as a ball-and-stick figure to show its position relative to the helices. Closer examination of the helices indicates the source of the deviations seen in Table 2. The time-averaged helices B, C, E, and G were found to be nearly identical to their crystallographic counterparts. In particular, the G helices show a slight twist in the retinal position (Color Plate 3), but the helix and retinal essentially occupy the same spatial region. The A helices are tilted slightly with respect to one another, but in other respects are nearly identical (see Color Plate 4). These five helices therefore contribute bulk displacement errors rather than fundamental changes in the helical structures.

Helices D and F, on the other hand, show much larger differences. The time-averaged helix D appears to have lost much of its helical structure. The unwinding is clearly seen in Color Plate 5. The differences in the F helices, shown in Color Plate 6, are much less severe, consisting of a slight unwinding of the helix near the Val-167 end of the helix, and a tilting of the Ile-191 end. These differences make the largest contributions to the deviation results. The unwinding of the helices should not be considered a defect. Such behavior is a common event in molecular dynamics simulation [22].

The distances of the retinal atoms (including the Lys-216 CE, NZ, and H atoms) to the atoms of all other residues were computed at 3-ps intervals. Those neighbors that were separated by less than 4.0 Å were recorded by retinal atom number and neighbor residue number. The neighbor residue was scored as being in contact. Over the course of the simulation run, 31 residues had at least 1 or more retinal atoms approach within the 4.0-Å range. The number of contacts observed, divided by the number of observations, was taken as a rough estimate of the contact time percentage. Percentages were computed for the entire 500-ps duration, and for the final 250 ps. The results for all contacted residues are shown in Figures 3 and 4. The contact percentages were also computed for three subdivisions of the retinal as described above: the Schiff base region (Figure 5), the midsection (Figure 6), and the β-ionone ring (Figure 7). Not surprisingly, the heaviest contacts occur over the β-ionone ring segment, as this end has the most freedom of motion, being at the end of an extended side chain.

Of particular interest is the contact data for Asp-85, assumed to be the proton acceptor in the photocycle, and Asp-96, assumed to be the proton donor. Asp-85 shows very little time in contact with the Schiff base region; in fact, the contact decreases at later times in the simulation. Asp-96, on the other hand, is never seen in contact. Asp-96 approached Lys-216 by no closer than 6.1 Å, averaging 7.5 Å. In particular, it approached the Schiff base region by no closer than 8.6 Å, averaging 10.0 Å. This tends to support the idea that the protein controls accessibilities to the Schiff base proton by providing different $pK_a$ values, rather than formation of a Schiff base counterion complex (see, for instance, Kataoka et al. [23]). The heavy contacts made by the β-ionone ring suggest that it is strongly constrained; in fact, the retinal shows very little spatial displacement. This is seen in Figure 8, which is a composite of retinal positions at 30-ps steps.

**Table 2. Comparisons of Various Model Structures with Bacteriorhodopsin Crystal Structure**

| | All atoms | | | Backbone only | | |
|---|---|---|---|---|---|---|
| Structure | Maximum | Average | RMS | Maximum | Average | RMS |
| Initial | 12.3 | 2.7 | 3.3 | 8.3 | 2.2 | 2.7 |
| All atoms | 12.9 | 2.9 | 3.4 | 8.3 | 2.4 | 2.7 |
| Helix A | 9.8 | 2.7 | 3.1 | 4.2 | 2.0 | 2.2 |
| Helix B | 9.8 | 2.2 | 2.9 | 5.6 | 1.5 | 1.9 |
| Helix C | 5.4 | 1.3 | 1.6 | 2.0 | 0.7 | 0.8 |
| Helix D | 11.0 | 3.9 | 4.4 | 7.1 | 3.4 | 3.8 |
| Helix E | 7.5 | 1.8 | 2.3 | 2.6 | 0.9 | 1.0 |
| Helix F | 10.0 | 3.0 | 3.6 | 6.4 | 2.4 | 2.7 |
| Helix G | 5.0 | 1.5 | 1.8 | 2.2 | 0.9 | 1.0 |

[a] The all-atom and helix structures are based on atomic positions averaged over the final 300 ps of the dynamics run. Helix G includes the retinal chromophore.

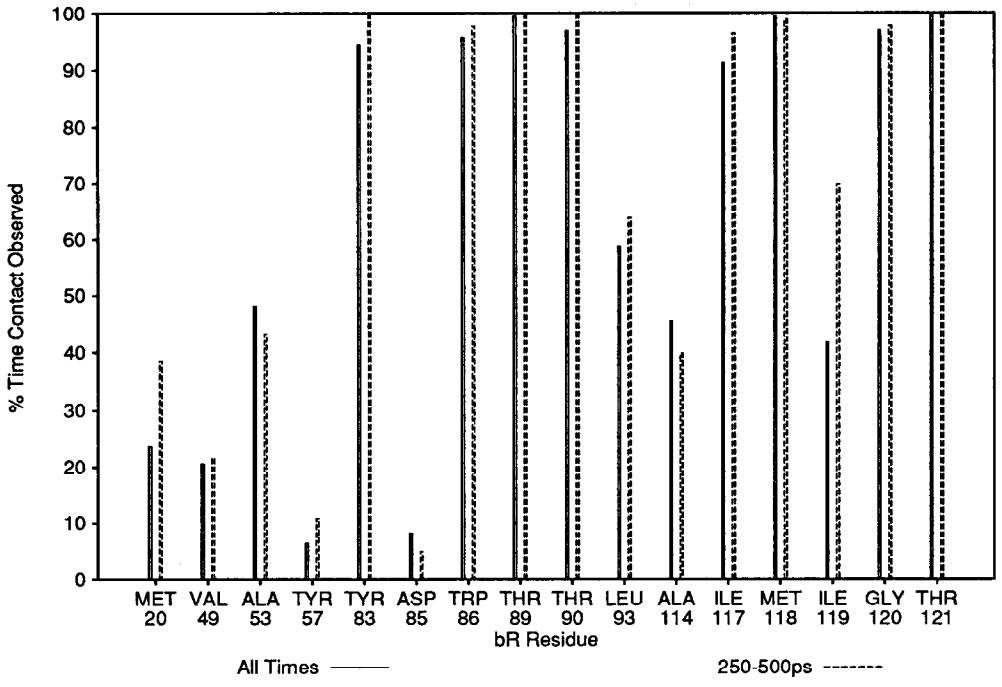


*Figure 3. Percentage of time at least one retinal atom was in contact with a neighbor residue (one of two).*

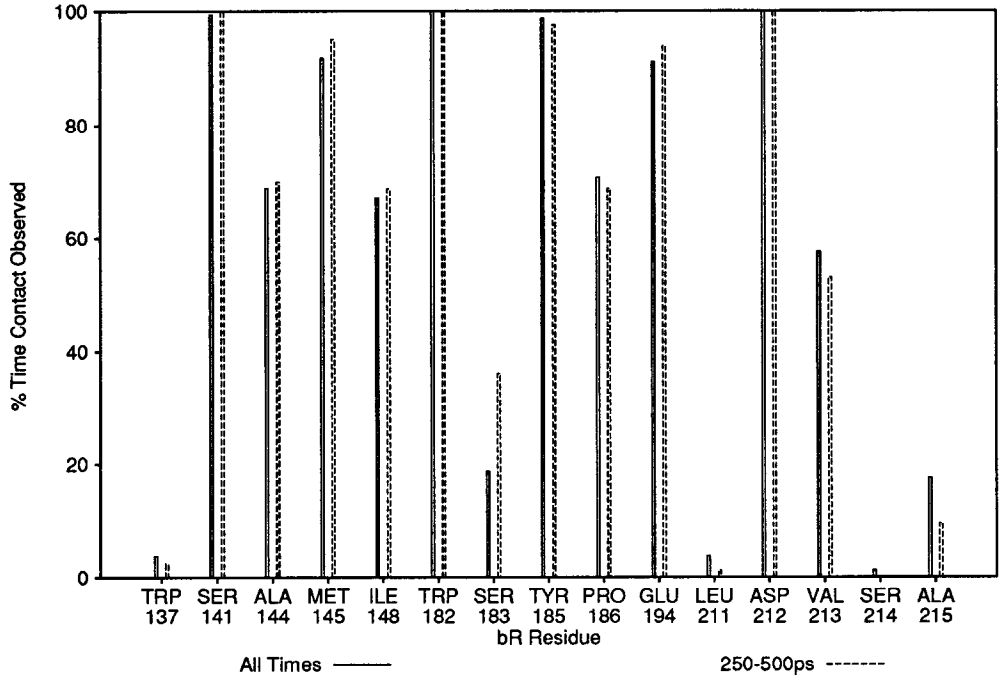


*Figure 4. Percentage of time at least one retinal atom was in contact with a neighbor residue (one of two).*

Thermal *trans–cis* interconversion of the NZ–C01 and C01–C02 bonds have been proposed as a mechanism for moving the Schiff base proton close to Asp-96, along with possible conformational changes within Lys-216 [10]. The dihedral angles within Lys-216 were computed and analyzed for variations. All but two of the angles monitored showed motion in the range of ±30°, with RMS deviations from the average values of 12° or less. The NZ–C01 and C01–C02 bonds, however, were found to vary much more. The NZ–C01 bond varied from −145° to +173° about its normal *trans* orientation, with an RMS deviation about its average position of 54°. The C01–C02 bond also varied about its *trans* orientation with a range of motion from −171° to +163°, with an RMS deviation of 60°. The entire Lys-216 residue thus appears quite flexible, allowing

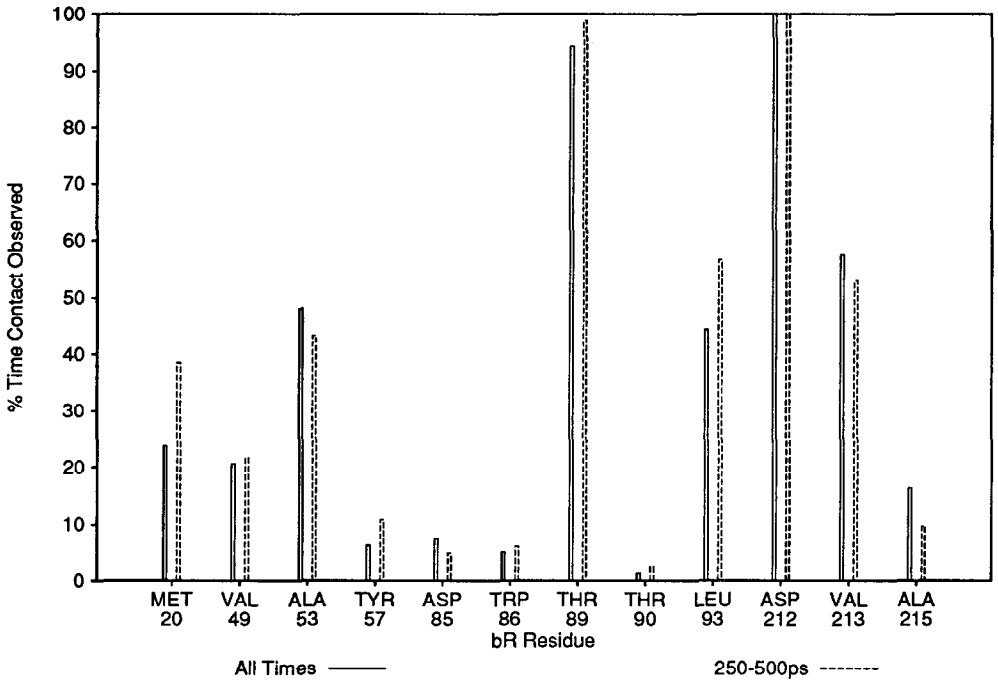


*Figure 5. Percentage of time at least one Schiff base atom was in contact with a neighbor residue.*

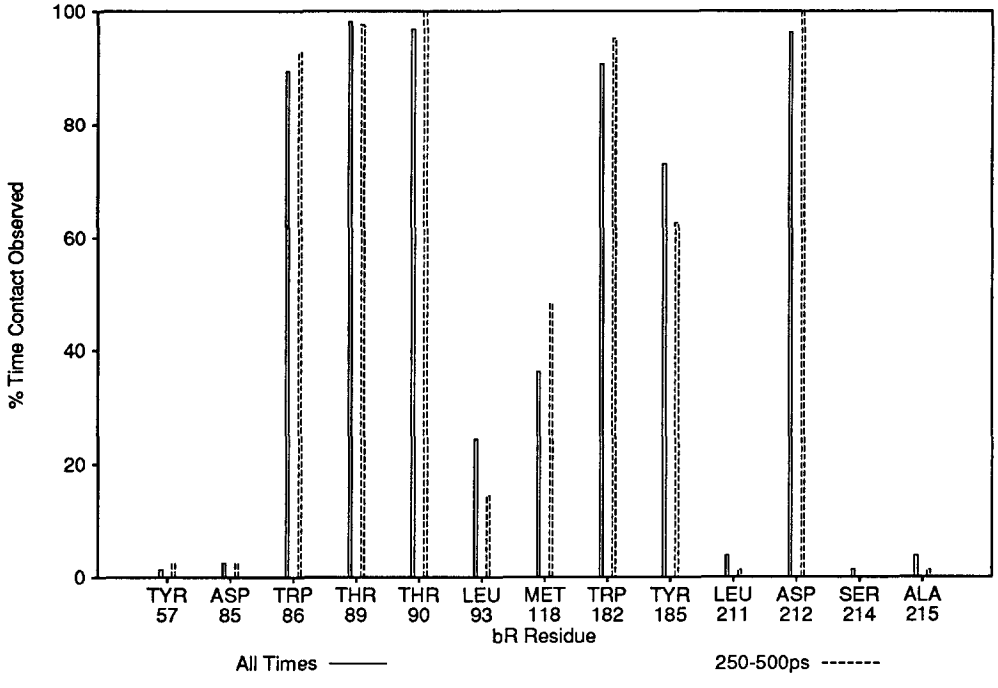


*Figure 6. Percentage of time at least one midsection retinal atom was in contact with a neighbor residue.*

cumulative changes in dihedral angles along the length of the side chain to absorb one or two *trans–cis* thermal isomerizations without significant spatial displacement in the retinal.

## CONCLUSIONS

The process of reconstructing atomic coordinates with Protean2 has proved to be a rapid method of building complete molecular structures, yielding results comparable to other methods. The initial energy-minimized structure shows deviations from the crystal structure similar to those derived in previous studies. The long-term evolution of the structure strongly suggests that a stable structure was produced. The fact that deviations seen early on in the model run match those found in other studies, coupled with the stable deviations and contact variations over the later portions of

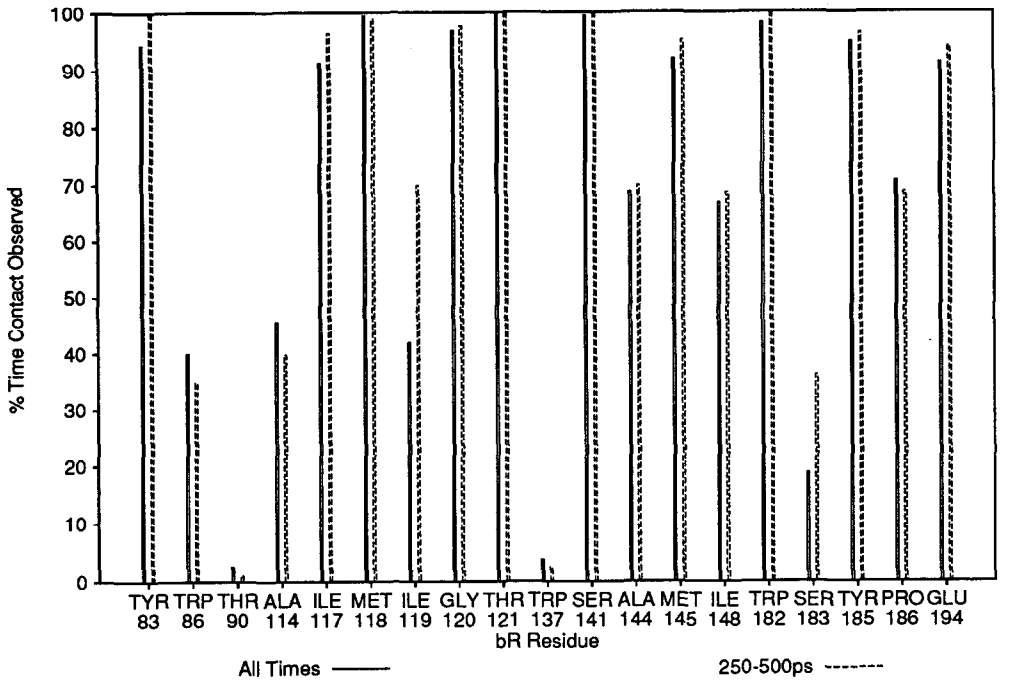


*Figure 7. Percentage of time at least one β-ionone ring atom was in contact with a neighbor residue.*

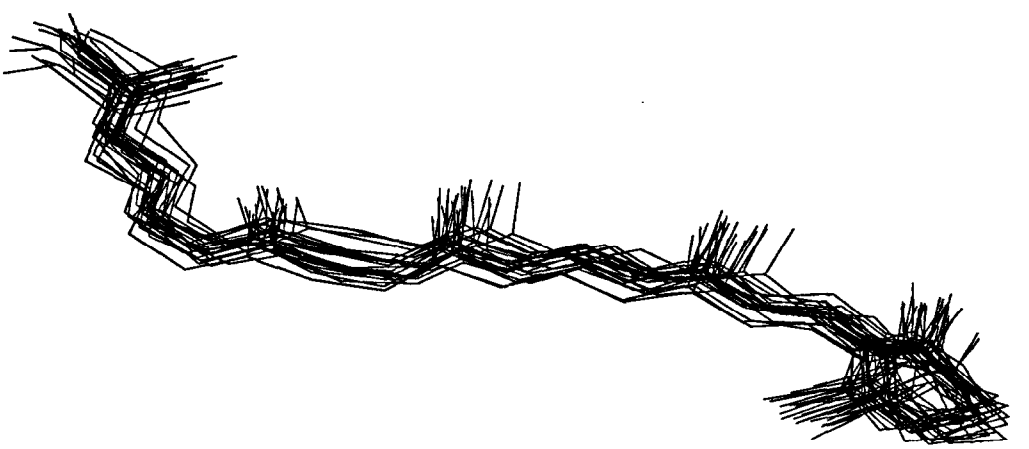

*Figure 8. Time-elapsed positions of Lys-216 at 30-ps intervals.*

the run, suggest considerable flexibility in the bR macromolecule.

This study emphasizes the importance of performing long-duration dynamic models for systems as large or larger than bR. Previous studies, which generally relate only energy-minimized structures to a standard structure, fail to capture time-dependent details and attempt to draw conclusions from a small sample of static structures. A system undergoing realistic thermal motion can sample many more possible states. The resulting flexibility demonstrated by the system easily accounts for thermal isomerization changes, while maintaining an overall spatial orientation. In particular, the position of the retinal need not undergo significant reorientation as is required if all but one bond are allowed to transform and energy minimization is performed.

## ACKNOWLEDGMENTS

This work was supported in part by grant of computer time from the DoD High Performance Computing Shared Resource Centers. IBM SP2 time was provided by the Maui MSRC (Maui, HI). Cray C-90 time was provided by the Waterways Experimental Station MSRC (Vicksburg, MS). This work was performed under Contract F33615-94-C-5803. The prime contractor was Ogden Professional Services, Lawrence Associates (Dayton, OH) in support of the Materials Directorate, Wright Laboratory, Air Force Material Command (ASC), USAF.

## REFERENCES


1 Birge, R.R. *Biochim. Biophys. Acta* 1990, **1016,** 293
2 Birge, R.R., et al. *Mol. Cryst. Liq. Cryst. Sci. Technol. Sec. B Nonlinear Optics* 1992, **3,** 133
3 Ferrand, M., et al. *FEBS Lett.* 1993, **327,** 256
4 Humphrey, W., Logunov, I., Schulten, K., and Sheves, M. *Biochemistry* 1994, **3,** 3668
5 Nonella, M., Windemuth, A., and Schulten, K. *J. Photochem.* 1991, **54,** 937
6 Brooks, B.R., et al. *J. Comput. Chem.* 1983, **4,** 187
7 Weiner, S.J., et al. *J. Am. Chem. Soc.* 1984, **106,** 765
8 Sankara-Ramakrishnan, R., Vishveshwara, S. *J. Biomol. Struct. Dyn.* 1989, **7,** 187
9 Ahl, P.L., et al. *J. Biol. Chem.* 1988, **263,** 13594
10 Sankara-Ramakrishnan, R., Vishveshwara, S. In: *Biomembrane Structure and Function—the State of the Art* (B. P. Gaber and K. Easwaran, eds.). Adenine Press, Schenectady, New York, 1992, pp. 375–386
11 Altman, R., Pachter, R., Carrara, E., and Jardetzky, O. *QCPE Bull.* 1990, **10,** 596
12 Pachter, R., Altman, R., and Jardetzky, O. *J. Magn. Reson.* 1990, **89,** 578
13 Pachter, R., Fairchild, S.B., Lupo, J.A., and Adams, W.W. *Biopolymers* 1996, **39,** 377
14 Pachter, R. Support programs for PROTEAN, Part II (unpublished)
15 pdb1brd.ent, Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Long Island, New York, 1990
16 Molecular Simulations. *Quanta 4.0.* Molecular Simulations, Inc., Burlington, Massachusetts, 1994
17 van Gunsteren, W., and Berendsen, H. Technical report. Laboratory of Physical Chemistry, University of Groningen, Nijenborgh, The Netherlands (unpublished)
18 van Gunsteren, W., et al. J. Comput. Chem. 1984, **5,** 272
19 Lupo, J.A. In: *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences* (T. N. Mudge and B. D. Shriver, eds.). IEEE, Los Alamitos, California, 1995, Vol. 5, p. 132
20 Clark, T., Scott, L.R., and Bagheri, B. *Pfortran: A Parallel Extension of Fortran: Reference Manual,* release 1.0. University of Houston, Houston, Texas, 1991
21 van Gunsteren, W., and Berendsen, H. *GROMOS Manual.* Laboratory of Physical Chemistry, University of Groningen, Nijenborgh, The Netherlands, 1987
22 Perahia, D., Levy, R., and Karplus, M. *Biopolymers* 1990, **29,** 645
23 Kataoka, M., et al. *J. Mol. Biol.* 1994, **243,** 621