Large Scale Motions in a Biosensor Protein Glucose Oxidase: A Combined Approach by QENS, Normal Mode Analysis, and Molecular Dynamics Studies

Sujit S. Tatke,[1] Chun K. Loong,[2] Nicolas D'Souza,[2] Richard T. Schoephoerster,[1] M. Prabhakaran[1]
[1] *Biomedical Engineering Department, FL International University, Miami, FL*
[2] *Argonne National Laboratory, Argonne, IL*

Received 17 September 2007; revised 21 December 2007; accepted 18 January 2008
Published online 13 February 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.20956

ABSTRACT:

The characteristics of the glucose oxidase were studied using a combination of experimental and theoretical techniques. Quasi elastic neutron scattering experiments were used to obtain the vibrational frequencies of the protein. These were compared to theoretical results obtained by normal mode analysis. Results indicate a good match between the experimental and theoretical values. Molecular dynamic simulation with covariant analysis was used to study the structure and dynamics of glucose oxidase. Various parameters like the radius of gyration, root mean square fluctuations, solvent accessibility were studied for evaluating the structural stability of the protein. The frequency of vibration calculated from the three methods is used to derive the large scale motions. Theses studies were used to predict the suitable lysine residues for linkage with carbon nanotubes. © *2008 Wiley Periodicals, Inc. Biopolymers 89: 582–594, 2008.*

Keywords: quasi neutron elastic scattering; molecular dynamics; normal mode analysis; glucose oxidase

Correspondence to: M. Prabhakaran; e-mail: prabha@prodigymeter.com

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

Glucose oxidase from *aspergillusniger* or *Penicillium* is a dimeric protein with a molecular weight of 160 kDa. Each monomer (Figure 1a) has a flavin adenine dinucleotide (FAD) cofactor. The FAD is not covalently bonded and is released when the protein is denatured. It is a strong oxidizing agent that participates in the electron transfer reaction. This reduced form is re-oxidized by molecular oxygen. Glucose oxidase is a glycoprotein. About 20% of the molecule consists of carbohydrate by weight. The enzyme shows a high degree of specificity for β-D-glucose and can hence be utilized for the detection of glucose in body fluids, beverages, and foodstuffs. Glucose oxidase catalyses the oxidation of β-D-glucose to D-glucono-1,5-lactone (gluconic acid) and hydrogen peroxide using molecular oxygen as the electron acceptor. The reaction is shown below;

$$\beta\text{-D-glucose} + O_2 \rightarrow \text{D-glucono-1,5 lactone} + H_2O_2$$

FAD cycles between the oxidized and reduced state while remaining attached to the glucose oxidase molecule. β-D-glucose, is an excellent indicator of the blood sugar level, found in the human saliva, sweat, urine, and serum. Detection of glucose is one of the most frequently performed tasks in a clinical chemistry laboratory. Diabetes patients also need to monitor their blood sugar levels for effective control.[1] Glucose biosensors monitor the glucose content in the blood. Glucose oxidase is an excellent enzyme for this purpose since it can detect β-D-glucose, the principle component of blood

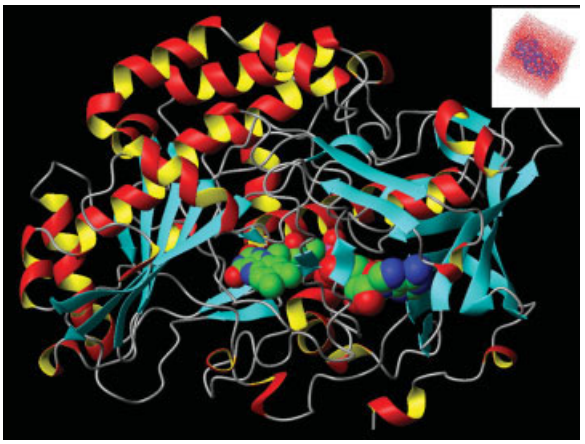

FIGURE 1 Glucose oxidase model & glucose in a box of water and ions for MD simulation (1b).

sugar, with a high degree of specificity. The focus of this study is characterization of the biosensor protein molecule, glucose oxidase to help in better understanding its structure thereby improving the signal to noise ratio during sugar detection. The unique approach involves a combination of experimental (quasi elastic neutron scattering [QENS]) and theoretical methods (molecular dynamic [MD] simulations and normal mode [NM] analysis). The study also provides a path to choose the covalent linking sites on the surface of the protein with carbon nanotube.[2]

The results obtained from experimental study are interpreted from NM analysis and compared with MD simulations. QENS has been used to study the dynamics of proteins.[3–5] It uses the large incoherent scattering of hydrogen nuclei, to probe the diffusive motions in the proteins in the picosecond to nanosecond time scale.

MD has been used to study a number of different problems. These include the effect of solvent and temperature on protein structure and dynamics, structural refinement of X-ray data using the simulated annealing method, NMR structure determination.[6] In the present study, a combination of MD principal component analysis and NM calculations will be used to describe the protein in terms of its various parameters, structural properties, mass distribution properties over time, distances between structures. Knowledge of these properties will shed light on the various aspects such as backbone flexibility, mean square fluctuations, and packing density. These can be helpful in tuning the protein for the biosensor application. MD studies will be handy in the interpretation of neutron scattering data. QENS has been used before to study the dynamic of different functional parts of Bacteriorhodopsin from *halobacterium salinarium*[7] and study its internal molecular motions as a function of the hydration level.[8] QENS studies combined with MD simulations has been used to study the amplitude of atomic motions in staphylococal nuclease. Our group successfully characterizes Cytochrome c, a heme protein using a similar combination of MD and QENS.[9]

Neutron scattering[10–14] and MD Simulations are independently capable of probing the atomic structure. Both these methods are especially suited for the study of biological samples like proteins. They complement with each other due to their similar time and space ranges.[4,15–17] The experimental results from neutron scattering can be interpreted by NM analysis and verified theoretically by MD simulation studies.[6] The frequencies of vibration obtained from QENS experiment can be compared with those from the NM Analysis. The knowledge of these frequencies and their corresponding eigen vectors will assist in delineating regions within the protein that show large scale collective motions. QENS has been used to study the glucose oxidase sample. In a neutron scattering experiment we measure the intensity of the scattered neutrons as a function of the momentum and energy transferred to the sample during scattering.

In the present study, two simulation techniques covariance analysis of MD and NM have been used to characterize the glucose oxidase molecule. NM is a rather approximate method that can be used to interpret the vibration spectra of molecules and show collective motions.[18,19] There is no record in literature that documents the combination of MD, NM studies and neutron scattering on glucose oxidase.

METHODS AND MATERIALS

QENS

The sample glucose oxidase from *Aspergillusniger* was obtained from Sigma Aldrich chemicals (Sample A). This sample contains potassium gluconate as a stabilizing agent and also contains traces of impurities like amylase and galactase etc. For experimental purposes, the contribution of only the gluconate with glucose oxidase is considered since it is 61% weight of the whole sample. For practical purposes the contribution of other impurities is neglected. Potassium gluconate was the sample B. The experiment was carried out in two parts. First the sample A is run followed by the sample B. Subtraction of the signal B from signal A will give the scattering from the protein alone. The contribution of the remaining impurities is neglected. The weight of the glucose oxidase sample used for the experiment was ~6 g. The weight of the potassium gluconate buffer was ~11 g. The relative contribution made by either sample to the measured spectra was obtained from the mass fraction of both, Sample A and Sample B. The glucose oxidase sample (A), the run was carried out for 3,642,724 pulses (~33.7289 h). The run was repeated using the gluconate sample (B). The data from all the arms of the spectrometer were grouped together and a signal was obtained. The data was analyzed using the QENSH workbench Soft-

ware. The subtracted spectra were used to obtain the frequencies associated with the low energy transfers.

Using QENSh, the ASCII data for the runs were obtained. Then, data plots were generated. The data obtained from the QENS experiment is in meV. It was converted to wave numbers (cm^{-1}).

The MD Simulation

The high resolution crystal structure of glucose oxidase solved to 1.9 Å,[20] (PDB entry code 1CF3) was used as the starting model for the MD simulation. The globular protein contains 583 residues and is glycosylated at various aspargine residues. The structure contains both α helices and β secondary structures. It contains a cofactor FAD. The MD simulation was performed using the GROMACS suite of programs.[21] The topology of the FAD structure was obtained using the web based program PRODRG.[22] This was followed by the parameterization of the molecule. The process of parameterization was done so as to generate a comprehensive topology file for the model. The compatibility of these parameters with the existing gromacs parameters was checked manually. New parameters, if any, were added to the gromacs parameter library before simulation. This program also adds polar hydrogens to the nonamino residues. Glucose oxidase is a glycosylated protein due to the presence of sugars. The presence of glycosylated part makes the process of parameterization difficult. In the case of FAD, the parameterization was done on an independent system within the complex, whereas, the protein itself had to be modified to accommodate protein glycosylation. In order to ensure conformational geometry during the MD simulations, the covalent connectivity with the amino acid residue ASN was parameterized. The glycosylated residues were added as two modified aspargine residues in the existing gromacs residue topology file. Initial MD simulations were run to test the newly modified asparagines residues.

Periodic boundary conditions were applied using a rectangular simulation cell, filled with single point charge (SPC) water molecules taken from an equilibrated liquid configuration. The net charge of (+12) on the system was compensated for by adding chlorine ions at the twelve most electropositive positions by replacing twelve water molecules. Each structures contains 5782 protein atoms, 56 FAD atoms, 8782 water molecules, and 12 negatively charged chlorine atoms. The total system size was 23,2176 atoms. The starting model (Figure 1b) was subjected to steepest descent energy minimization. The energy minimization runs were repeated until local minima were achieved. This is done to relax possible strain within the molecule. A pair list cut off of 10 Å was used together with a 20 Å long range electrostatic cut-off and it was updated every 10 steps. Velocities were then generated using a Maxwellian distribution at 300 K. The system was pre equilibrated by keeping the coordinates of the heavy atoms fixed for 10 ps. The energy minimization was followed by the MD simulation run, using the leap-frog algorithm, at 300 K, for a period of 1.5 ns (1500 ps). The simulation is a constant temperature constant volume simulation performed by coupling the system to a Berendsen thermostat.[23] The time step used in the simulation was 2 fs using SHAKE algorithm to constrain the bond lengths (bond lengths fixed).[24] Coordinate trajectory frames (1500) were stored at the rate of one structure per ps.

Following the MD study a principle component analysis was undertaken to study the large oscillations in the glucose oxidase molecule. It is known that the 10–20 principle components are responsible for around 90% of the large scale oscillations in the protein and hence a knowledge of these components can reveal information on a large percentage of the conformational fluctuations occurring within the protein. Hence, the method of principle components analysis on the protein configurations[25] from MD simulations was used to delineate large motions. An essential dynamics analysis of the protein motions was performed.[26] In the present analysis, a covariance matrix that contains information on positional fluctuations was constructed from the C-α atoms.

The C-α atom trajectory will provide sufficient information on large scale oscillations which is of interest to us. By using only the C-α atoms, the covariance matrix will be smaller in dimension thus making the diagonalization process computationally less intensive. The resulting diagonal matrix carries information on the eigen values. The eigen vectors are themselves written to a trajectory file. The eigen values are a measure of the mean square positional fluctuation along the corresponding eigen vector. The eigen values are then sorted in decreasing order. The eigen vectors with largest eigen values represent the collective motions that best represent the sum of actual fluctuations whereas the smallest values represent the most constrained or restricted degrees of freedom. In order to study the characteristics of these collective fluctuations, the ensemble of structures was projected onto a single eigen vector. The atomic displacements connected with that eigen vector were then visualized by translating these projections to 3D space. From the principle component analysis, the extreme structures (the protein conformations with largest root mean square deviation) in the trajectory were obtained along with the eigen values, eigen RMSF (root mean square fluctuation) values and eigen components.

NM analysis is a computationally intensive process and requires plenty of resources with respect to the processing capacity and the storage capacities of the system. Hence in this study, the glucose oxidase molecule which consists of 583 residues in its original form was truncated to match the memory capacities of the system. Hence, it was truncated to a value of 450 residues. The size of the structure was arrived at using a trial and error method. Initially, the NMA was started with the full glucose oxidase molecule. It was observed that the system is incapable of supporting this size. Hence, the molecule size was reduced and another analysis was tried. This process was continued until we hit upon a size that was within the memory capacities of the system. Finally, the simulation was done for 450 residues protein whose original size is 583 residues. While truncating the residues care was taken that only the inconsequential residues are cut making sure that the active site is not disturbed. Similar efforts were done by using different NM analysis protocols. A block normal mode (BNM) algorithm, originally proposed by Tama et al.[27] was implemented for Ca^{2+} -ATPase (994 residues).[28] A truncated approach was also developed by Bahar et al.[29] Once the system size was finalized the molecule was energy minimized using the double precision method. Energy minimization was done successively using both conjugate gradient and steepest descent methods. The combination of these steps results in a structure with the lowest possible energy. This was followed by the MD simulation using GROMACS with NM option with double precision. The system now contained only protein and not the solvent. This simulation resulted in the generation of a NM matrix also called as the Hessian Matrix. The covariance matrix, from original MD simulations, gives the square of distances between the atoms and the simulated trajec-

tory of structures, whereas the NM matrix generated above, contains the energy of interaction of the atoms of the well minimized protein in vacuum. The obtained eigen vectors indicate directions in the $3\times N$ configurational space. The corresponding eigen values indicate the amplitude of these correlated motions. The eigen values were converted to frequency units for comparison with experiments. The eigen values and eigen vectors were then analyzed, to reveal parts of the molecule that show more flexibility and conformational changes.

RESULTS AND DISCUSSION

The QENS experiment was done on glucose oxidase and gluconate samples. The QENS experiment generated data in terms of energy transfers in meV. The scattering intensity values are plotted in terms of either in energy (meV) or wave number (cm^{-1}).

Figure 2 shows the spectrum from the Protein sample A which consists of gluconate and the protein glucose oxidase. Figure 2a shows the two spectra from sample A & B along with the difference spectrum. The scattering from the sample A and B appears to be of equal strength.

The QENS experiment generated data on energy transfers occurring from 0 to 200+ meV. However, for the actual analysis, only the energy transfers in the low energy region were analyzed (0–25 meV) since large scale oscillations occur in this energy range and the declining intensity values are noticeable. This corresponds to a range of ~0–200 cm^{-1} in terms of wave numbers. From the difference signal shown in Figure 2a significant peaks were isolated and the corresponding energy transfer values were obtained. The presence of gluconate provides a stable and natural environment for the enzyme glucose oxidase. Since the primary goal is to understand the dynamics of protein in its natural environment, the initial experiments have been carried out on the complex. The co-factor FAD is part of the complex, whereas gluconate can be considered as substrate.

Protein vibrational frequencies were also derived from the NM analysis. From the energy minimized structure the Hessian matrix was generated by the NM run in double precision. Diagonalization of the Hessian resulted in generation of the energy values (eigen values and eigen vectors). These values were converted into frequency values (cm^{-1}) by the following formula: Frequency (cm^{-1}) = 109 × sqrt (energy) (meV).

Then these values were sorted and compared to the QENS frequencies. The Table I shows the comparison. From the above values it is clear that there is a close match between the frequencies derived by both the methods. Hence, the eigen vectors corresponding to these eigen values derived from the NM analysis were further analyzed. RMSF values for the

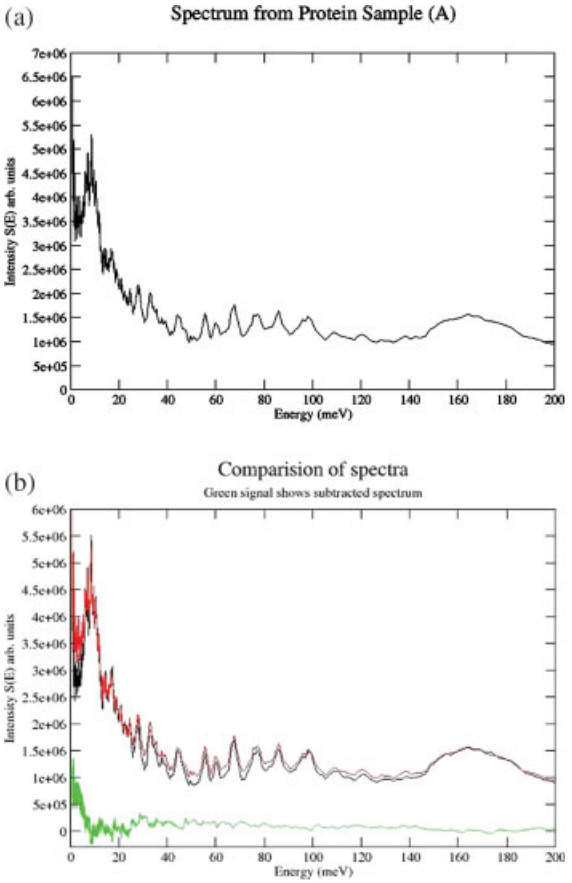

FIGURE 2 The Spectrum from protein + gluconate sample. (a) The difference and the original spectra.

atoms, and the eigen components (vector components per atom) were plotted.

From the RMSF graphs flexible regions in the molecule were identified. All the vectors from 1 to 200 were analyzed for RMS fluctuations and corresponding components. The RMSF for vectors 4 and 103 (Figures 3 and 4) that correspond to the Eigen frequencies (6.38 and 91.88 cm^{-1}) are presented as examples. For the eigen vectors 4 (Figure 3a) the displacement vectors along the three principle axes are also shown. The displacement occurs in all the axes around atom number 1000 (104 THR O) whereas the displacement pertains to z direction at the region surrounding atom number 3500 (360 ASP H). From the RMSF plots highly flexible regions in the molecule were identified. All the Vectors from 1 to 200 were analyzed for RMS fluctuations and corresponding components.

Using the knowledge of the RMSF values and the eigen values derived from NM analysis, the regions of flexibility in the protein were predicted in terms of residue number. A detailed analysis of 10 eigen vectors was done. Each vector reveals a set of residues that show high flexibility. As an

Table I Comparison of QENS & NMA Frequencies

QENS Frequencies (cm^{-1})	NMA Frequencies (cm^{-1})
6.78	6.38
7.66	8.71
13.79	13.91
16.13	15.84
30.25	30.46
31.13	31.67
37.91	38.06
40.33	39.64
45.90	45.53
47.02	47.02
51.14	50.55
54.85	55.33
56.62	56.96
60.50	60.13
66.54	65.64
74.37	72.45
76.30	75.71
77.27	78.53
80.66	79.90
87.52	85.49
91.39	91.88
96.79	96.69
99.78	99.49
105.02	104.84
106.96	106.82
116.96	115.54
125.35	125.40
132.12	132.89
145.19	145.89
147.20	148.06
151.48	151.20
158.90	158.27
171.81	170.74
175.68	175.67
182.86	182.49
190.03	190.69
199.47	201.12

example, analysis of RMSF of the fourth vector revealed the following residues to be of high flexibility with collective motions: 258ASN, 22TYR, and 280TYR. Analysis of each such vector thus generates data on all flexible residues. These are shown in Figure 5 below. Each different color represents the residues of collective motion obtained by analyzing a particular vector. Only two vectors have been shown in the Figure 5 to ensure clarity.

MD Simulation and Principle Component Analysis

The stability of the simulation was checked by monitoring both structural properties (number of hydrogen bonds, root mean square displacement, and solvent accessibility surface area) and mass distribution properties over time (mean square displacement and radius of gyration).

The starting model of glucose oxidase is shown in Figure 1 and with water molecules during simulation is shown in Figure 1a. The skeleton model of the two structures with largest displacement between them during the simulation is shown in Figure 6. The two extreme conformations are similar which suggests that the overall tertiary structure of the system is stable and preserved over the simulation period of 1.5 ns. This is a desirable characteristic for a biosensor protein since it means that the protein can be stable in the solvent environment. The RMS deviation between the two extreme conformations was observed to be 2.238 Å.

The changes in the packing arrangement can be best understood by studying the radius of gyration (R_g) (see Figure 7). The R_g gives information about the changes in the molecular specific volume which is the reciprocal of density. The fluctuation in the R_g values shows the breathing motions during the simulation (both contraction and expansion) with the average value of 23.2 Å.

The second parameter of analysis was the root mean square fluctuation (see Figure 8). The RMSF is an indicator of the flexibility and rigidity of the protein under thermal equilibrium. Because of SHAKE algorithm, the bond lengths were constrained. Hence the fluctuation is mainly due to side chain and tertiary motions in solvent environment. A visual inspection of the trajectory also confirms well equilibrated fluctuation of the system. The average value of the RMSF was found to be 0.277 nm.

The number of hydrogen bonds is an indicator of the intactness of the protein. During the course of the simulation the number of hydrogen bonds is well maintained as shown from the trajectory analysis of the hydrogen bonds (see Figure 9). There was an initial decrease in the number at the initial stages at the starting period, which was compensated with the increase in the number of hydrogen bonds of polar atoms of protein with solvent molecule (not shown in the fig). The density of hydrogen bonding does correlate well with the changes in structural fluctuations. The RMSF and R_g plots also verify this trend. The average value of the hydrogen bond number is 423.

The other tertiary parameter total solvent accessibility of the protein was plotted with respect to time (see Figure 10). The change in the surface shape during the simulation was brought out from this study. An initial decrease due to protein contraction, as revealed from radius of gyration is also nicely brought out from the observation of this parameter. Both hydrogen bond formation and surface exposure shows steady oscillations validating the simulation proce-

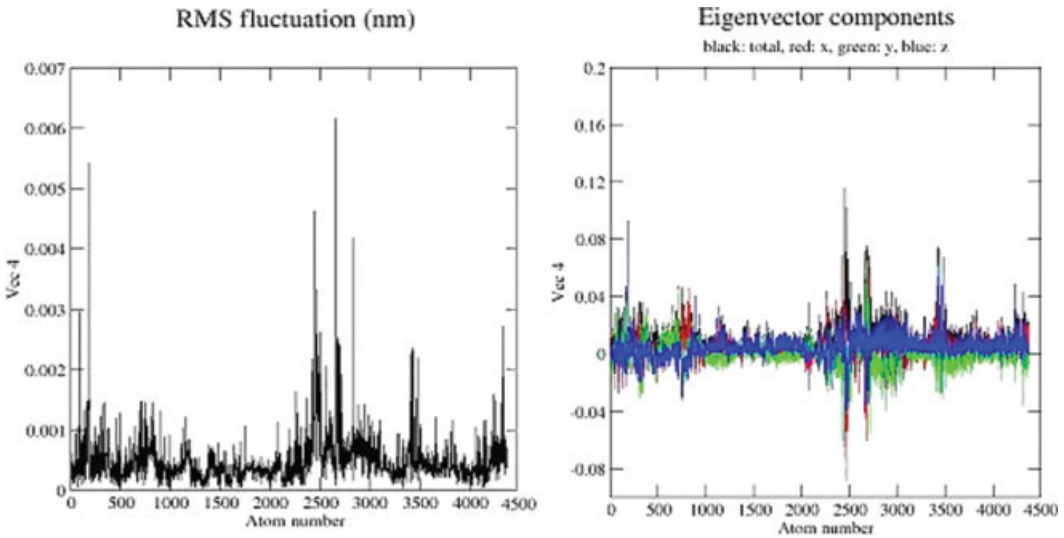

FIGURE 3 Eigen components for vector 4 (color code x red y blue z green).

dure. This also indirectly validates the NM analysis because the starting models for both simulations were energy minimized by identical procedures and they were subjected to same GROMACS force field. The average values of the solvent accessible surface area was found to be 125.60 nm^2.

We also studied the electrostatic interactions between the residues during the simulation. As an example, the distances between the two ARG residues during the simulation were plotted (see Figure 11). The effect of repulsive forces were clearly seen around 1 ns, these two residues drift apart from

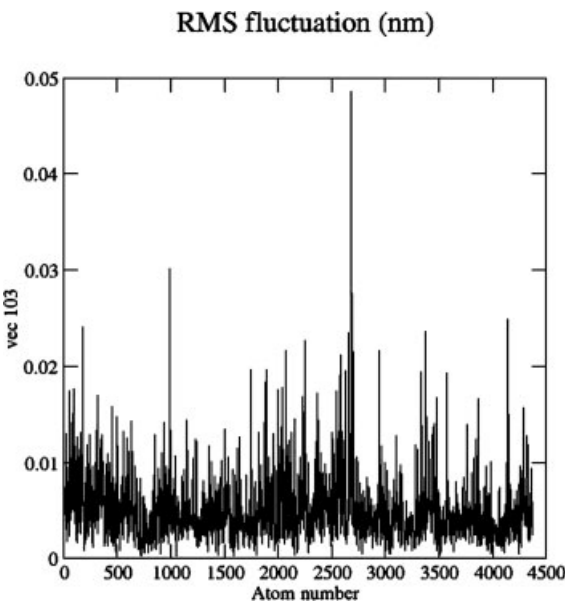

FIGURE 4 Eigen RMSF for vector 103.

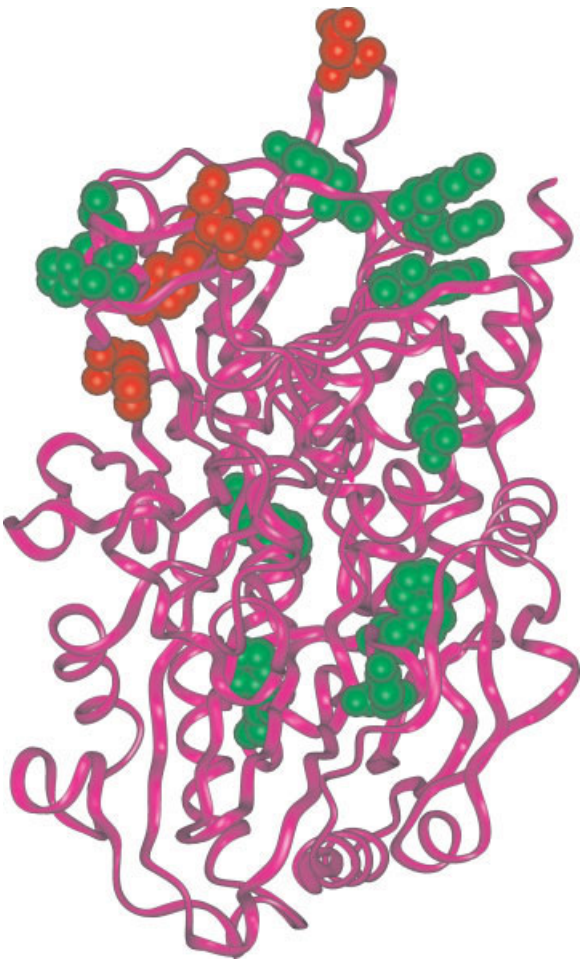

FIGURE 5 High Flexibility regions in glucose oxidase predicted from Normal Mode Analysis.

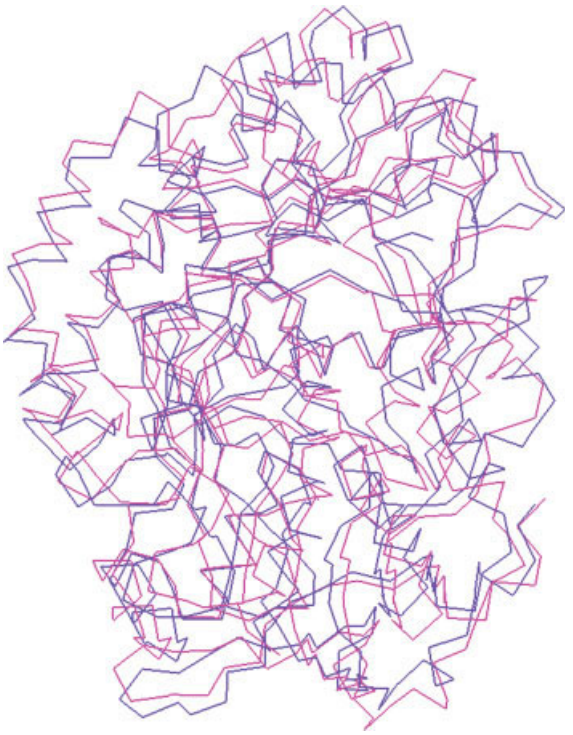

FIGURE 6 Simulated structures with largest root mean square deviation.

4.0 Å and stays stationary around 6.0 Å. This kind of shift in the equilibrium is noticeable in other oscillations also. On the other hand, the attractive forces between opposite charged residues (see Figure 12) are shown nicely in the following plot between the acidic glutamic residue and positively charged and lengthy ARG residue. At 750 ps, they move from 10 to 4 Å. The average value of the distance between the two basic residues was found to be 4.24 nm and between the Lys-Glu residues was found to be 7.78 nm.

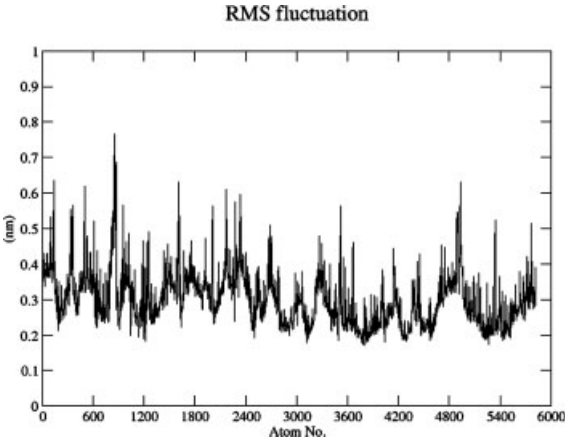

FIGURE 8 Root Mean Square fluctuation.

Flexibility of Glycosylated Part and FAD

The glucose oxidase molecule consists of a cofactor FAD and four glycosylated residues. The flexibility of the glycosylated part and the cofactor FAD was also studied separately. Figures 13 and 14 show the RMSF of the glycosylated residue and FAD, respectively. When compared to the RMSF values for the protein it is observed that the glycosylated residues show higher flexibility values whereas the FAD shows flexibility similar to the protein residues. The average value of the RMSF for the glycosylated residue, FAD and the entire protein is 0.48, 0.305, and 0.277 nm, respectively. The surface glycosylated residues show high flexibility due to large conformational freedom. It should be borne in mind that the individual rings are rigid units, mainly due to the application of SHAKE algorithm. The flexibility of the cofactor FAD is crucial from the point view of its association with active site. The simulation clearly exhibits that the motion of FAD is on par with the protein, though there is no covalent linkage

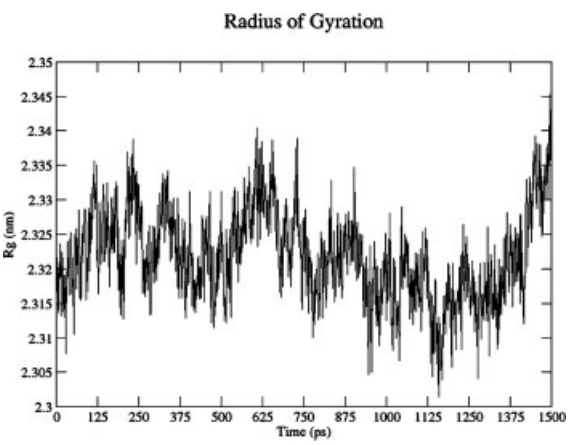

FIGURE 7 Radius of gyration.

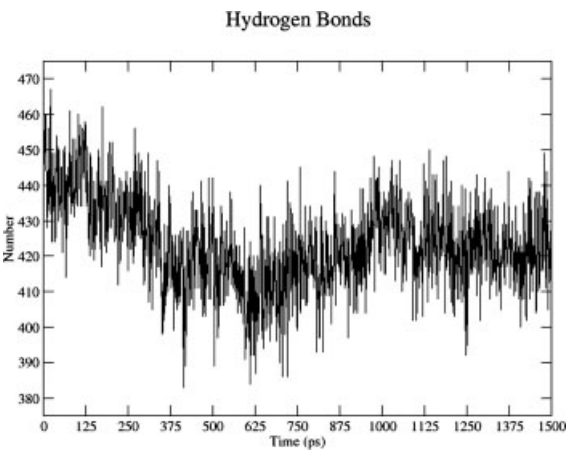

FIGURE 9 Number of hydrogen bonds during simulation.

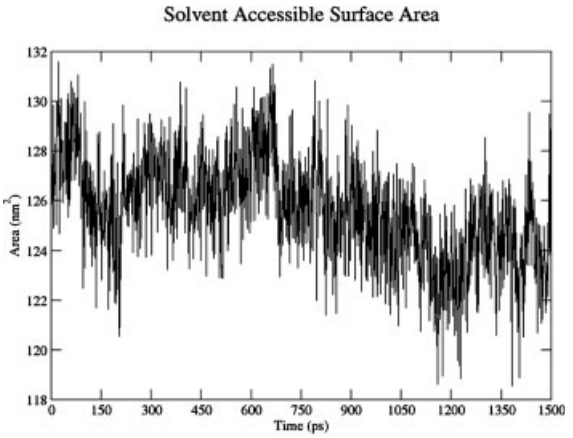

FIGURE 10 Solvent accessible surface area.

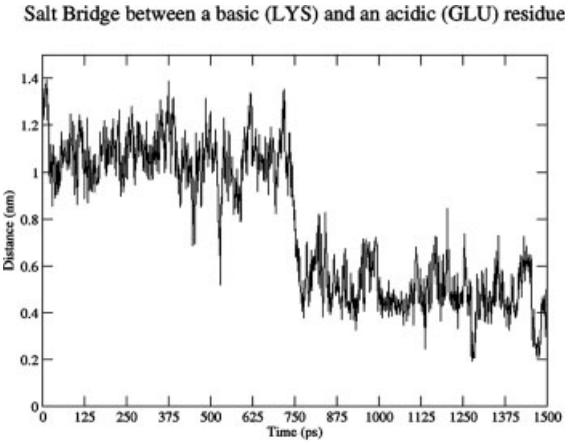

FIGURE 12 Distances between acidic and basic residue.

with the protein. It also validates the choice of parameters and topology file for FAD. Our results on FAD are in agreement with femto second dynamics of glucose oxidase, exhibiting a gradual change.[30]

The comparison of theoretical (black) and experimental (red) RMSF values is shown in the Figure 15. The values shown here are the average values of the RMSF for every residue. The values for the *B* factor are the experimentally derived values from the crystallographic study. The units have been altered accordingly. The crystallographic *B* factors include a variety of contributions (e.g., crystal contacts and static disorder in the crystal) in addition to that from the dynamical fluctuations, and these additional contributions are expected to be substantial for a structure of modest resolution.[31] Generally, the crystallographic *B* factors were larger than those observed in MD because the simulation sampled far fewer conformations of the protein than crystallography; the termini, however, exhibited larger *B* factors in the MD simulation where the protein is solvated in water rather than packed in a crystal. The red curve shows the values that the simulation produced. The two plots match each other closely which suggest confidence in the simulation and the stability and equilibration of the simulated system. The secondary structure was also preserved during the simulation as shown from the movie of the dynamic trajectory.

Principal Component Analysis

Principal component analysis was also used to derive large scale motions in the protein. An analysis of the essential dynamics of the protein enables a classification of the large scale movements of the enzyme. The diagonalization of the covariance matrix, built from the equilibrated portion of the trajectory, provided the principal direction of the large-amplitude concerted motions, i.e., the eigen vectors with largest eigen values that characterize the essential subspace of the internal dynamics of a protein. The first eigen values provide a large part of the overall fluctuations. The large-scale movements

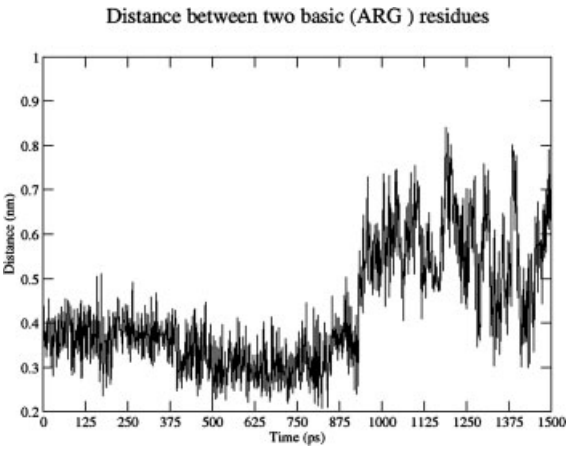

FIGURE 11 Distances between two basic residues.

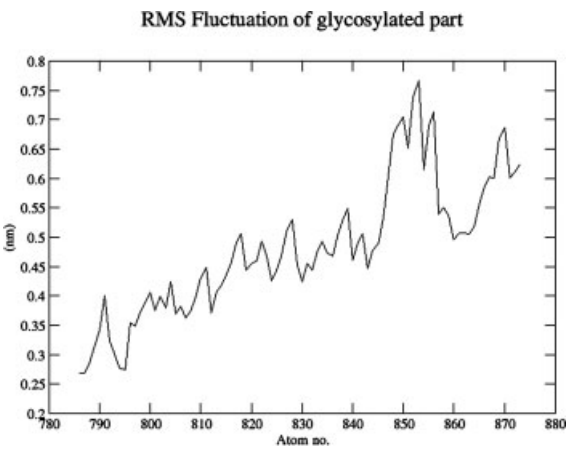

FIGURE 13 RMSF of the glycosylated residues.

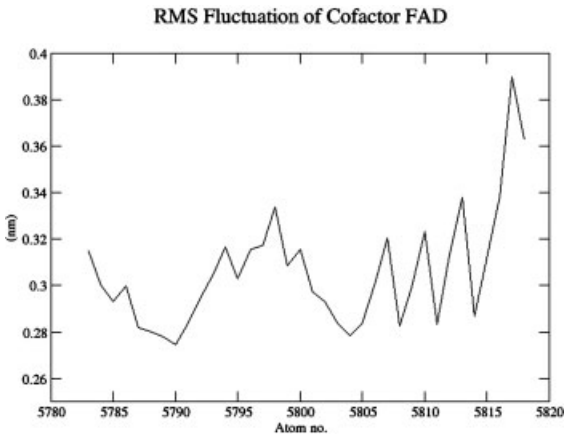

FIGURE 14 RMSF of cofactor FAD.

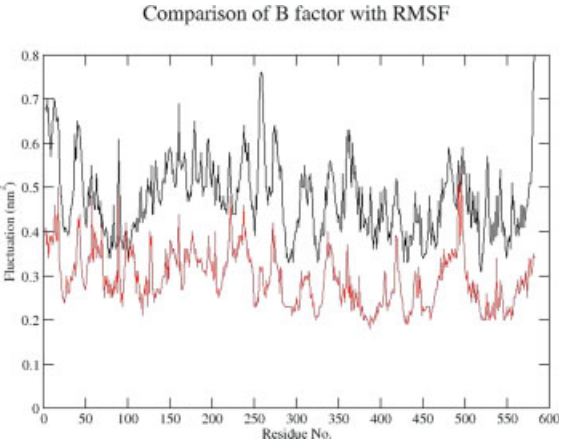

FIGURE 15 Comparison of simulation and crystallography.

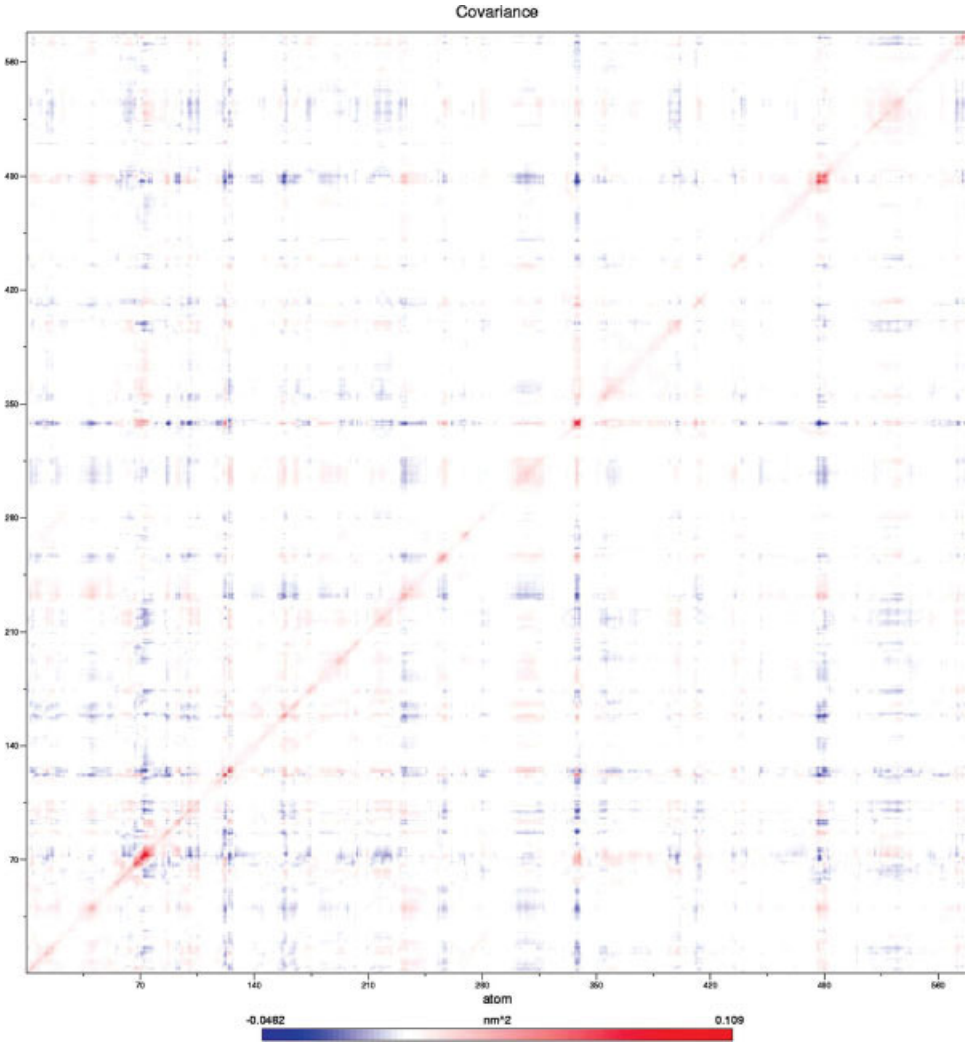

FIGURE 16 The covariance matrix of correlations.

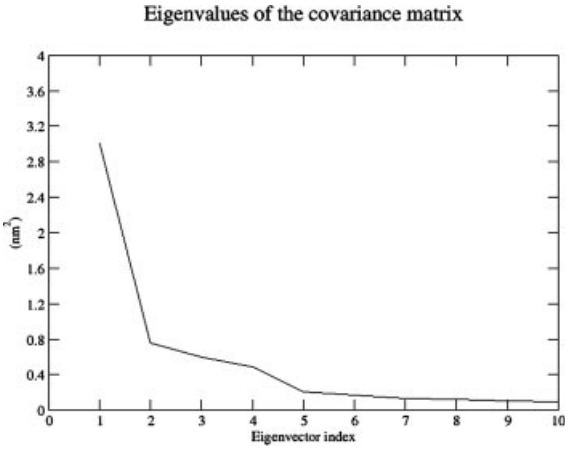

FIGURE 17 Eigen values spectrum of the covariance matrix.

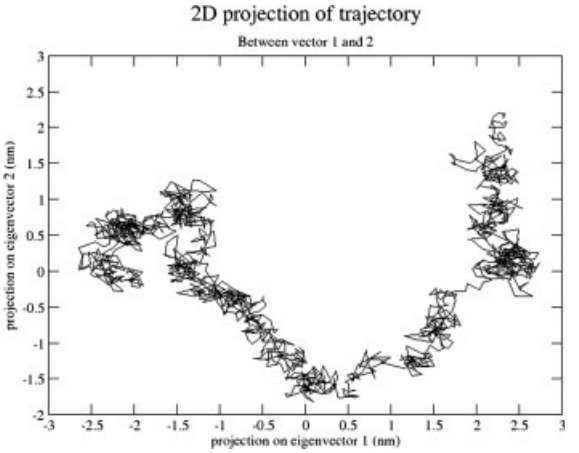

FIGURE 19 The 2D projection of trajectory between vector 1 and 2.

(i.e. those in the essential subspace) are represented by the eigenvectors of the correlation matrix representing the position vectors of the C-α atoms. The analysis is limited to the first few eigen vectors because they take into account the major part of the protein motion (the first 10 eigen values correspond to 50% of the global motion) as already reported for numerous other cases.[26] Each structure along the trajectory is fitted to the starting configuration, eliminating global translational and rotational degrees of freedom. The fitting is performed using only the C-α atoms. The covariance matrix generated from the C-α atoms is shown in Figure 16. It shows the correlation between the pairs of atoms. The color code is as follows: Red indicates a positive correlation which means that the pairs of atoms move together in the same direction. Blue indicates a negative correlation which means that the pairs of atoms move in opposite directions.

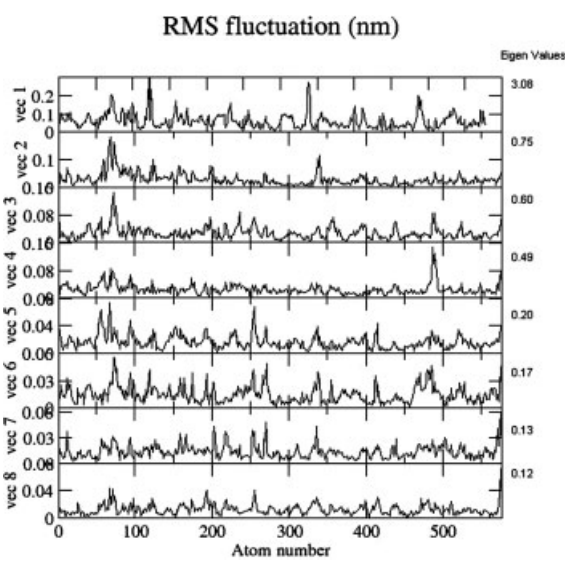

FIGURE 18 RMS fluctuation from the covariance matrix.

The eigen values spectrum, obtained from the covariance matrix, which represents the principle motion of the protein C-α atoms, is shown in Figure 17 as a function of the eigen vector index. Only the first 10 values are shown since the remaining values are approximately equal to zero. The first eigen values provide a large part of the overall fluctuations. The eigen vector components are shown in Figure 18. It shows the RMS fluctuations of the first 8 vectors. The collective motion is shown in the plots with decreasing amplitudes (eigen values) from 3.06 of vector 1 to 0.12 for vector 8. As an example from the top lot, the collective motion exhibited at residue regions 72, 128, 325, and 431 with high flexibility. His 516 is also in the collective motion of vector 1 and 5 which plays an important role in function.[32] Using the values of the residue numbers the corresponding region in the protein was identified and color coded. Although the single essential eigenvectors are likely to be not the equilibrium ones, i.e., the eigenvectors obtained by a completely converged statistics, the corresponding atomic collective motions are probably significant as they belong to a good approximation of the equilibrium essential subspace. As part of the essential dynamics analysis, the projection of the C-α atoms on the plane between the first and second and between the first and eighth vector derived from the principal component analysis was also studied. This is shown in Figures 19 and 20, respectively. These two figures give an estimate of the variation in projection. The variation is 6 nm in Figure 19 and 1.5 nm in the Figure 20. The collective motion vectors move in different paths as exhibited in those figures. Both NM analysis and principle component analysis were used to derive the flexibility regions and identify large scale motions in the protein. Both techniques revealed the similar regions of flexibility in the protein thus providing a validation of the simulation procedure. The high flexibility regions (16,70,74,75,128,258,274,275,277,336,353,354) that

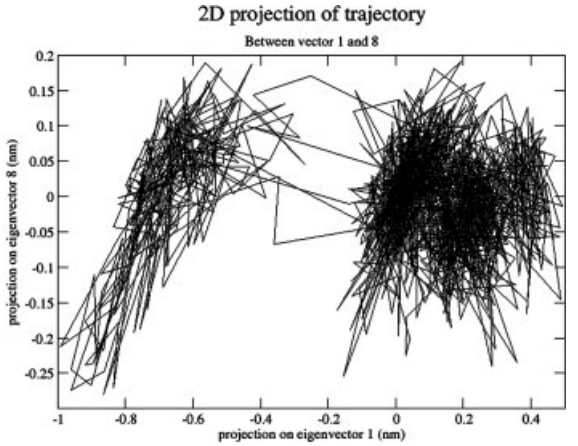

FIGURE 20 The 2D projection of trajectory between vector 1 and 8.

were predicted by both the methods are shown below in Figure 21. It is noted here that each method did indicate some high fluctuation regions that could not be predicted by the other. Hence, it is important to use a combination of both types of studies: NMA and PCA to identify the flexibility regions in the protein with cross validation. The frequency correspondence between the NMA and QENS further validates the results. From the agreed regions, the eigen vectors of the corresponding eigen values was studied thoroughly to bring out the large scale collective oscillations. The correspondence between theory and experiment was shown by earlier investigators.[4,17,33–35]

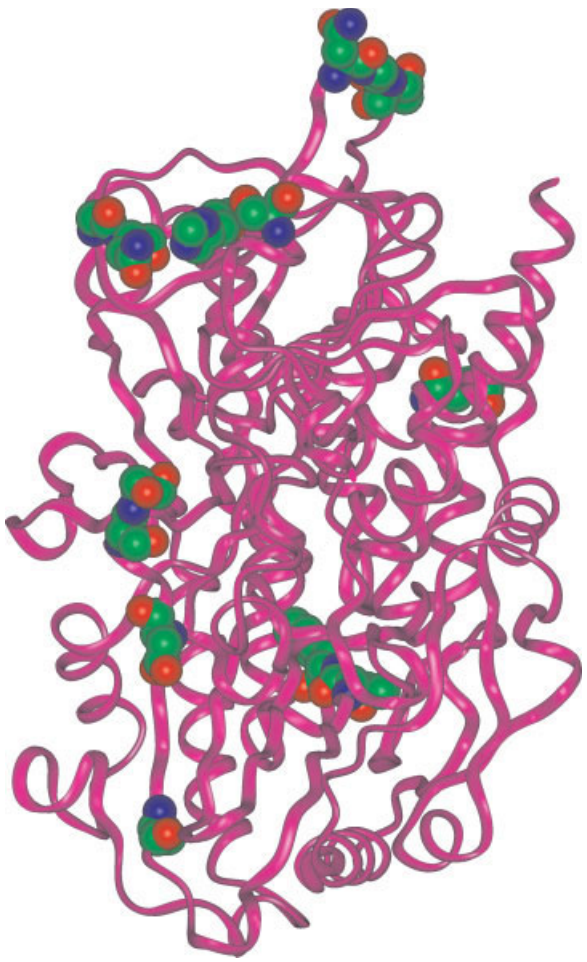

FIGURE 21 Regions of high flexibility common to both analyses.

The Interfacing of Biological Macromolecules to Carbon Nanotubes

The above characterization suggests that the protein is stable in the solvent environment. It is robust and not perturbed during the course of the simulation. The protein shows no conceivable distortion as evidence from the extreme structures. Thus, we can conclude that the protein is well suited and stable enough for the proposed application in biosensors. Using this knowledge we proceed with the next step of interfacing the protein to the transducer.

Prediction of Residues for Interfacing to the Nanotube

The transducer will be hooked to the protein at one of the exposed Lysine (LYS) residues. The glucose oxidase molecule has 14 LYS residues. The choice of LYS residue for interface to the nanotube was made by studying each of these residues separately in terms of their RMSF and solvent accessibility. The RMSF data reveals information on flexibility whereas the accessibility data reveals information on the surface exposure to the solvent. The Table II shows the flexibility and accessibility of the LYS residues. Ideally a combination that shows

Table II RMSF and Accessibility of LYS Residues

LYS Residue No.	RMS Flexibility	SAS
11	0.37	1.34
116	0.25	0.83
152	0.38	0.75
187	0.33	0.97
201	0.33	1.25
202	0.31	0.88
252	0.23	0.74
273	0.35	1.14
282	0.31	1.05
306	0.23	0.92
364	0.27	1.24
372	0.27	0.68
441	0.27	1
526	0.2	1.36
570	0.28	0.59

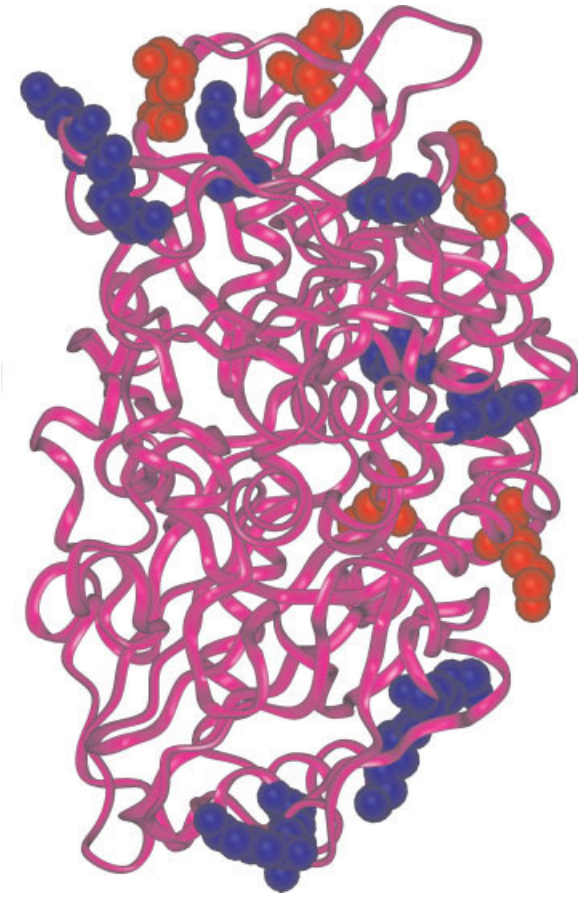

FIGURE 22 LYS residues with the predicted interface residues shown in red.

minimum flexibility and maximum accessibility is desired. Using this criterion the following residues appear to be good candidates for interface with the nanotube: 116, 306, 364, 441, and 526. These residues are shown in bold font in Table II. Figure 22 shows the LYS residues of glucose oxidase with the predicted linkage residues colored in red. Five such residue sites have been proposed for interfacing to the carbon nanotube shown.

CONCLUSIONS

A general breathing motion was observed without any bending or hinge movements during this study. This suggests that the protein is stable in the solvent environment and does not denature. The experimental (QENS) and theoretical (NMA) frequencies show excellent match. The knowledge of these frequencies was used to delineate large scale motions by observing the RMSF that correspond to those frequencies. The flexibility and rigid regions were bought out from this study and compared with the *B* factor crystallographic data. The results show good agreement between the values.

Glucose oxidase is a unique protein because it contains glycosylated residues and a cofactor FAD. The glycosylated residues show a high flexibility compared to the protein. However it is not high enough to perturb the protein structure. The cofactor follows the general dynamics of the protein even though it is not covalently attached. Using a combination of QENS experiment, NMA and principal component analysis the large scale motions in the protein were delineated. It was observed that these motions are confined mainly to the surface residues. Because of the presence of FAD the active site with its exposed part remains fairly stationary. The FAD plays a role in stabilizing the exposed active site. From RMSF and accessibility calculations, residues for interface to the nanotube were proposed. This study has a few limitations. First, from the QENS experiment only wave numbers in the range of 0–200 cm^{-1} were analyzed. High energy transfers were neglected. Second, due to computational limitations truncated system of 450 residues was run for NM calculation.

REFERENCES

1. Delvaux, M.; Champagne-Demoustier, S. Biosens Bioelectron 2003, 18, 943–951.
2. Tatke, S. S.; Renugopalakrishnan, V.; Prabhakaran, M. Nanotechnology 2004, 15, S685–S690.
3. Caronna, C.; Natali, F.; Cupane, C. Biophys Chem 2005, 116, 219–225.
4. Hayward, J. A.; Smith, J. C. Biophys J 2002, 82, 1216–1225.
5. Lechner, R. E. Physica, B. 301:83–93, 2001.
6. Karplus, M.; McCammon, J. A. Nat Struct Biol 2002, 9, 646–652.
7. Reat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterhelt, D.; Zaccai, G. Proc Natl Acad Sci USA 1998, 95, 4970–4975.
8. Fitter, J.; Lechner, R. E.; Buldt, G.; Dechner, N. A. Proc Natl Acad Sci USA 1996, 93, 7600–7605.
9. Prabhakaran, M.; Gursahani, S H.; Verma, C. S.; Guarduno-Juarez, R.; Renugopalakrishnan, V. J Phys Chem Solids 2004, 65, 1615–1622.
10. Bee, M. Quasielastic Neutron Scattering: Principles and Applications in Solid-State Chemistry, Biology and Materials Science; Hilger: Bristol, 1988.
11. Trouw, F. R.; Price, D. L. Annu Rev Phys Chem 1999, 50, 571–601.
12. Byron, O.; Gilbert, R. J. C. Curr Opin Biotechnol 2000, 11, 72–80.
13. Filabozz, I. A.; Di Bari, M.; Deriu, A.; Di Venere, A.; Andreani, C.; Rosato, N. J Phys Condens Matter 2005, 17, S3101–S3109.
14. Tehei, M. Æ.; Daniel, R Æ.; Zaccai, G. Eur Biophys J 2006, 35, 551–558.
15. Smith, J.; Cusack, S.; Tidor, B.; Karplus, M. J Chem Phys 1990, 93, 2974–2991.
16. Tarek, M.; Neumann, D. A.; Tobias, D. J. Chem Phys 2003, 292, 435–443.
17. Smith, J. Q Rev Biophys 1991, 24, 227–291.

18. Tnama, F.; Sanejouand, Y. H. Protein Eng 2001, 14, 1–6.
19. Hinsen, K. Proteins 1998, 33, 417–429.
20. Wohlfahrt, G.; Witt, S.; Hendle, J.; Schomburg, D.; Kalisz, H. M.; Hecht, H. J. Acta Crystallogr 1999, D55, 969–977.
21. Lindahl, E.; Hess, B.; van der Spoel, D. J. Mol Mod 2001, 7, 306–317.
22. van Aalten, D. M. F.; Bywater, R.; Findlay, J. B. C.; Hendlich, M.; Hooft, R. W. W.; Vriend, G. J Comput Aid Mol Design 1996, 10, 255–262.
23. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684–3690.
24. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327–341.
25. Ichiye, T.; Karplus, M. Proteins 1991, 11, 205–217.
26. Amadei, A.; Linssen, A. B.; Berendsen, H. J. Proteins 1993, 17, 412–425.
27. Tama, F.; Gadea, X.; Marques, O.; Sanejouand, Y. Proteins 2000, 41, 1–7.
28. Li, G.; Qiang Cui, Q. Biophys J 2002, 83, 2457–2474.
29. Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Phys Rev Lett 1998, 80, 2733–2736.
30. Zhong, D.; Zewail A. H. Proc Natl Acad Sci USA 2001, 98, 11867–11872.
31. Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Adv Chem Phys 1988, 71, 1–259.
32. Meyerd, M.; Wohlfahrtb, G.; Knäbleinc, J.; Schomburg, J. D. J Comput Aid Mol Design 1988, 12, 425–440.
33. Cusack, S.; Smith, J.; Finney, J.; Tidor, B.; Karplus, M. J Mol Biol 1988, 202, 903–908.
34. Jennifer, A. H.; Smith, J. C. Biophys J 2002, 82, 1216–1225.
35. Balog, E.; Smith, J. C.; Perahia, D. Phys Chem Chem Phys 2006, 8, 5543–5548.

Reviewing Editor: Laurence Nafie