Published in final edited form as:

Curr Top Med Chem. 2011 ; 11(2): 211–220.

Thermal Motions of the E. Coli Glucose-Galactose Binding Protein Studied Using Well-Sampled Semi-Atomistic Simulations

D. J. Cashman, A. B. Mamonov, D. Bhatt, and D. M. Zuckerman

Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh

Abstract

The E. coli glucose-galactose chemosensory receptor is a 309 residue, 32 kDa protein consisting of two distinct structural domains. We used two computational methods to examine the protein's thermal fluctuations, including both the large-scale interdomain movements that contribute to the receptor's mechanism of action, as well as smaller-scale motions. We primarily employ extremely fast, "semi-atomistic" Library-Based Monte Carlo (LBMC) simulations, which include all backbone atoms but "implicit" side chains. Our results were compared with previous experiments and all-atom molecular dynamics (MD) simulation. Both LBMC and MD simulations were performed using both the apo and glucose-bound form of the protein, with LBMC exhibiting significantly larger fluctuations. The LBMC simulations are in general agreement with the disulfide trapping experiments of Careaga & Falke (*J. Mol. Biol.*, 1992, Vol. 226, 1219-35), which indicate that distant residues in the crystal structure (i.e. beta carbons separated by 10 to 20 angstroms) form spontaneous transient contacts in solution. Our simulations illustrate several possible "mechanisms" (configurational pathways) for these fluctuations. We also observe several discrepancies between our calculations and experimental rate constants. Nevertheless, we believe that our semi-atomistic approach could be used to study fluctuations in other proteins, perhaps for ensemble docking or other analyses of protein flexibility in virtual screening studies.

Keywords

Computational Simulations; Coarse-Grain; GGBP; LBMC; Molecular Dynamics; Monte Carlo; Protein dynamics; Protein fluctuations

Introduction

Proteins are tiny molecular machines that carry out biochemically relevant functions within all cells. The latter half of the twentieth century has given us over 50,000 individual protein structures, primarily from X-ray crystallography and NMR sources (www.rcsb.org). When analyzing this vast assortment of proteins, it becomes easy to think of the solved structure as the "correct" structure. However, it is important keep in mind that proteins are parts of living organisms, and protein motions are critical to life.[1,2] These inherent movements play a critical role in locomotion and enzyme catalysis, as well as in protein-ligand interactions, and serve as the basis for many biological processes, including, but not limited to, muscle contraction, cellular metabolism, antigen-antibody interactions, gene regulation, and virus assembly.1-3 Additionally, these movements can be both large-scale, domain movements (e.g. hinge movements), occurring on the order of milliseconds to microseconds, or rapid, small-scale atomic movements, occurring on the nanosecond to picosecond timescale.[3-5] Large conformational changes may also be coupled to catalytic function,[6] especially in motor proteins.[7,8] Protein conformational fluctuations are also of vital importance in rational drug design, and the development of new methods of incorporating protein flexibility into docking and scoring studies is a topic of much recent research.[9-11]

In this study, we investigate whether our recently developed semi-atomistic simulation approach[12] is suitable for modeling large, as well as small scale, thermal motions involved in the *Escherichia coli* D-Glucose/D-Galactose binding protein (GGBP). This protein is a 32 kDa globular protein that is organized into two distinct structural domains, each consisting of a α/β folding motif. A well-refined X-ray crystal structure (PDB accession code: 2GBP) provides a detailed view of the sugar binding site, located in the cleft between the two primary domains.[13] Some recent studies of this protein have looked at its possible use as a glucose biosensor, which could be used in diabetic patients.[14-16]

A 1992 study by Careaga & Falke experimentally investigated the thermal motions of the alpha-helices on the surface of GGBP by a series of disulfide trapping experiments on pairwise cysteine mutants.[17] Each experiment examined disulfide bond formation between residue 26 and a cysteine mutant placed at one of several locations in the protein, as shown in Figure 1. The cysteine sites included several residues on two helices of the N-terminal domain, and one residue on the surface of the C-terminal domain. These cysteine residues – some with a beta carbon (C_β) distance of more than 10 Å from residue 26 – were then observed to form disulfide bonds with it, in the presence of catalyst, which was quantified by variable mobility in SDS/polyacrylamide gel electrophoresis. The disulfide trapping data was used to analyze the protein's intra- and inter-domain thermal motions.[17,18]

In this study, we used computational methods to generate an approximate equilibrium ensemble of GGBP in order to obtain insight into the mechanism by which the experimentally observed fluctuations occur. Are they a result of complete or partial unfolding and refolding of the protein? Or are there rigid-body movements that result in the observed fluctuations? Our simulations connect the experimental disulfide trapping measurements, which provide data as to which residue pairs are interacting, with the crystal structure, which provides us with a detailed, yet static, picture of what the protein looks like. We performed simulations of both the apo and holo forms of the GGBP protein, using both all-atom molecular dynamics (OPLS-AA force field[19] with the NAMD software package[20]), as well as a coarse-grained Library-Based Monte Carlo (LBMC) simulation method developed in our laboratory.[12] The LBMC method performs Boltzmann sampling of molecular systems based on precalculated statistical libraries of molecular-fragment configurations, energies, and interactions. It is a “coarse-grained” model with fully atomistic backbones, using simplified Gō-like interactions among residues, which allows for the stabilization of the native state of the protein while allowing large fluctuations.[12] While all-atom molecular dynamics provides the most detailed computational simulation analysis available, it is also computationally demanding, and requires a large amount of single processor CPU time to reach full convergence.[21] In particular, our data suggest that the all-atom simulations are well short of exhibiting motions observed in the Careaga/Falke experiments.[17] Because our coarse-grained LBMC approach uses much less CPU time, it can exhibit larger-scale motions where residues approach each other close enough for disulfide bond formation to occur.

This paper provides details of our LBMC simulations and compares them to MD in an effort to study the overall protein fluctuations involved. By tracking the C_β distances between the residues studied by Careaga and Falke, we are able to make some observations on the mechanism by which “disulfide-capable” interactions are taking place. Using these C_β distances, we were also able to calculate the rate at which the residues drop below a particular threshold, and to compare this to experimental rate constants for further validation of our models.

Results

This study of the E. Coli D-Glucose/D-Galactose Chemoreceptor protein centers primarily on four Library-Based Monte Carlo (LBMC) simulations, each of 3 billion steps in length. The simulation method is described below and in Ref. [12]. One simulation modeled the unbound protein ensemble, and three simulations used different models of the glucose-bound protein ensembles. Additionally, two molecular dynamics (MD) simulations totaling 31 nanoseconds in length (one of the unbound protein and one of the glucose-bound protein) were also performed. Each of the LBMC simulations took approximately 30 days of single CPU time on a 3.6 GHz Intel Xeon system with 2 GB of RAM, while the MD simulations took approximately 141 days of single CPU time on the same system.

Three models of glucose were used in our holo-GGBP simulations to test for artifacts resulting from our simplified modeling. First, the glucose-bound protein was modeled using a “virtual glucose” representation, in which an extra G δ -type interaction was added to those residues that formed either hydrophobic or hydrogen-bonding interactions with glucose in the crystal structure (termed holo-GGBP “virtual glucose”). We next used explicit but coarse-grained representations of glucose for the other two LBMC studies – one representing glucose as a single “atom” of 1.8 Å in diameter (holo-GGBP 1-C glucose), and the other representing glucose as three “united carbon” atoms of 1.5 Å in diameter (holo-GGBP 3-C glucose).

Overall Fluctuations

A key strength of the LBMC method is its ability to sample large-scale fluctuations frequently while maintaining the overall stability of the protein. Figure 2 shows the RMSD vs. Frame for the LBMC simulations (A – D), and RMSD vs. Time for the molecular dynamics simulations (E, F). The RMSD is referenced to the starting X-ray crystal structure in all cases.[13] In all four LBMC simulations, we observe that the overall trajectory is reasonably stable, with no global unfolding events.

We observe significant fluctuations throughout the trajectory, from ~2 Å, up to ~6 Å, most abundantly in the apo simulation (cross reference Figure 3 with Figure 2). If we count the number of instances in which the RMSD is higher than 4.5 Å, we find that there are 897 counts during the apo-GGBP trajectory, versus only 74 counts in the holo-GGBP (“virtual glucose”) trajectory, 423 counts in the holo-GGBP (1-C glucose) trajectory, and 656 counts in the holo-GGBP (3-C glucose) trajectory. The upper limit of RMSD in the apo-GGBP trajectory exceeds 5.0 Å 117 times, versus 0 times in the holo-GGBP (“virtual glucose”) trajectory, 11 counts in the holo-GGBP (1-C glucose) trajectory, and 87 counts in the holo-GGBP (3-C glucose) trajectory. In sum, there are significantly more large-scale fluctuations in the apo-GGBP trajectory versus the holo-GGBP trajectories. The largest events in the apo-GGBP trajectory, in which the RMSD is above 5.0 Å, are, in fact, domain opening events, in which the protein’s “mouth” opens up, and then closes again. We also observed hinge opening movements in both the apo and holo MD simulations (Figure 2: E, F), although in both cases, the protein opens but does not close again. The limited timescale of the all-atom simulations does not permit us to know whether these fluctuations are merely transient.

Our apo-GGBP LBMC trajectory also “finds” a semi-open configuration from an apo-GGBP X-ray study.[22] In particular, the apo-GGBP trajectory approaches 21 times within 3 Å RMSD of this alternative X-ray structure, which was not used in our modeling. Note from Figure 2 that 3 Å RMSD is well within the range of thermal fluctuations around the native state. The RMSD between the LBMC starting structure and the semi open structure is 4.6 Å.

Beta Carbon Distances Among Key Residues

To analyze the fluctuations in closer detail, as well as to compare with experimental results, [17] we tracked the C_β distance between five residues and Gln26, at intervals of 200 MC steps throughout the simulations. These five residue pairs, Gln26/Met182, Gln26/Asn260, Gln26/Lys263, Gln26/Asp267, and Gln26/Asp274, were each mutated to Cysteine residues in separate experiments and observed to form disulfide bonds by Careaga/Falke.[17] Four of the residues are on an alpha helix located immediately adjacent to the alpha helix containing the Gln26 residue (see Figure 1). Met182, on the other hand, is located in the opposite domain of the protein, so tracking the Gln26/Met182 C_β distance should give us a reasonably good picture of the hinge opening.

Figure 4 shows histograms of the distribution of these C_β distances for each residue pair (Figure 4: A-E). First, as expected, we observed much broader peaks for the histograms generated from the four LBMC simulations compared to the histograms generated from the MD simulations. This is to be expected, as evident from the RMSD plots, because the LBMC simulations exhibit more large fluctuations than MD. The hinge opening in the MD simulations is clearly evident in Figure 4A for residues Gln26/Met182 – instead of two tall, narrow peaks, we see two broad peaks extending from 26 Å to beyond 40 Å. Note that neither the LBMC nor the MD simulations considered explicit mutants to cysteine, but rather monitored wild-type fluctuations, which is our primary focus.

Our LBMC simulations clearly suggest the plausibility of “disulfide-capable” conformations in GGBP. Figure 4 shows an example based on three configurations spaced 6×10^6 MC steps apart in the simulation, in which we can see that there is a considerable range of movement. Based on the minimum distances recorded in Table 1, there are several cases where the C_β distance approaches a very short range, which could make disulfide bond formation possible. Based on a survey of 50 PDB structures in the Protein Data Bank that contain one or more disulfide bonds, we found that the shortest disulfide bond was 3.0 Å and the longest disulfide bond was 4.2 Å, with an average of 3.8 Å. So it is certainly possible that, for Gln26/Asn260, Gln26/Lys263, and Gln26/Asp267, a disulfide bond could be formed. The C_β distances were too great in the LBMC simulations for Gln26/Met182 and Gln26/Asp274. We also did not observe any of the C_β distances in the MD simulations to be close to “disulfide-capable”; see Table 1. See also our discussion in the later in this paper of possible experimental artifacts based on the catalyst used.

Unexpectedly, we did not observe any notable difference in the C_β distributions between apo-GGBP and holo-GGBP. Table 1 shows the minimum, 1 percentile (1% of C_β distances are below this value), and standard deviation of these C_β distance distributions, allowing a closer examination of its extremities. We expected that the minimum C_β distance observed would be lower for the apo-GGBP simulations compared to the holo-GGBP simulations, since the glucose molecule is supposed to stabilize the protein, thereby resulting in decreased intramolecular fluctuations. What was observed in our data is, in fact, the opposite – overall, the lower minimum C_β distance observed is in the holo-GGBP simulations, versus the apo-GGBP. Although the C_β distance between Gln26 and Met182 is consistent with experiment, with lower C_β distances for the apo-GGBP versus holo-GGBP. However, the standard deviation of these C_β distances for Gln26/Met182 is relatively high, so it is difficult to say for certain. We also observe better consistency with the experimental data overall when comparing the apo-GGBP and holo-GGBP (3-C glucose) LBMC simulations.

Mechanism of Achieving “Disulfide-Capable” Conformations

In the previous section, we observed that the C_β distance between Gln26 and three of the residues of interest (Asn260, Lys263, and Asp267) did approach “disulfide-capable”

distances at points during our simulation. How is this occurring? If we look at the C_{β} distance distributions in Figure 4, the fact that most of these distributions are roughly centered around the C_{β} distance in the original X-ray crystal structure suggests that there is no major unfolding events occurring in the protein. This is also evident in the relatively stable RMSD plots in Figure 2 (A-D) – if there was unfolding occurring, the RMSD for the unfolding event would be much higher. However, also evident from the RMSD, we can see significant localized distortions, as the RMSD fluctuates in a range from approximately 2 Å to more than 4.5 Å. We can see a small snapshot of some of these localized distortions in Figure 4 – note, in particular, the distortions of the alpha helices that the Gln26 and four of the other residues reside on. So the mechanism by which these “disulfide-capable” interactions are occurring is clearly not due to a global unfolding of the protein, but in fact, to localized protein fluctuations. In particular, we observe from animated trajectories (movies) that the helices remain intact and relatively rigid as they shift; see Figure 4. The large local fluctuations could also be termed “cracking”, similar to previous observations of adenylate kinase by Whitford and colleagues.[23]

Rates of Formation of “Disulfide-Capable” Distances

The rate of formation of “disulfide-capable” C_{β} distances was analyzed for the LBMC simulations and compared to the experimental rate constants determined by Careaga/Falke. [17] While time in Monte Carlo simulations is somewhat arbitrary, we can calculate the average number of Monte Carlo steps between instances where the C_{β} distance falls below a certain predetermined threshold, which we would like to be as close to 4.2 Å as possible, since that is the longest disulfide bond length identified by our survey (*vide supra*). While the minimum C_{β} distances for three of the residue pairs studied (Gln26/Asn260, Gln26/Lys263, Gln26/Asp267) were in the vicinity of this 4.2 Å value (Table 1), the frequency that this occurred throughout our simulations was not enough to be able to calculate a meaningful rate. So we gradually raised our threshold, and found that at 7 Å, there were enough instances of close contacts between these residue pairs to calculate a rate which could be compared qualitatively to the experimental rate constants from the experiment. Note from Table 1 that no MD configurations approached close enough.

Overall, we observed that the rate of formation of “disulfide-capable” distances was greater for Gln26/Lys263, followed by Gln26/Asp267 and Gln26/Asn260, which is consistent with the overall trend of the experimental data. We observe that the rates of formation were slightly greater in the holo-GGBP LBMC simulations versus the apo-GGBP simulation, which was consistent with our C_{β} distance distribution data, but not with the experimental rate constants. We believe this discrepancy is probably an artifact of the simplicity of our model, which will be improved upon in subsequent studies.

Discussion and Conclusion

Overall, the rapid semi-atomistic LBMC simulations provide us with a good picture of the conformational changes and protein fluctuations of the E. Coli D-Glucose/D-Galactose Chemosensory Receptor, and are consistent with previous observations of the open and closed conformations of the protein.[18,22] Because of the protein’s size (309 residues) and the low cost of our simulations (~30 days of single CPU time), we believe that the promise of our semi-atomistic approach has been clearly demonstrated. Our previous statistical analysis showed that the LBMC simulations generated approximately 20-30 statistically-independent configurations in the trajectories, or approximately one statistically-independent configuration every 24 to 36 hours.[12] However, there is certainly room for improvement: discrepancies between our calculations and the experimental data raise some important cautionary notes.

We see reasonably good agreement between our simulations and the experimental data with respect to the plausibility of “disulfide-capable” interactions. We can also see that those interactions are not caused by a global unfolding in the protein, but by more localized fluctuations. We did not see any major differences in the C_β distributions between the apo and holo LBMC or MD simulations. Perhaps this is evidence that our simplified potential, while good enough to sample the overall conformational space reasonably well, is simply too rough to properly sample the smaller scale fluctuations involved with the presence of a small molecule such as glucose. This is one of the reasons why we performed multiple holo-GGBP simulations – first with the “virtual glucose”, then with one carbon representing the glucose molecule, followed by three carbons representing the glucose molecule. Additionally, it was also recently observed that the presence of Ca^{+2} in this protein enhances its thermal stability – neglecting this in our model could have an effect as well.[24] It is also possible that by adding additional chemistry to our coarse-grained model, such as Ramachandran potentials, hydrogen bonding, or residue-specific interactions, we might be able to observe a more realistic sampling of the fluctuations. We are currently working on such improvements.

Another possibility is that some of the buffers and/or catalysts used in the original experiment[17] could be inadvertently having an effect on promoting disulfide bond formation. For example, the chemical that was used to catalyze the formation of disulfide bonds was $Cu(II)(1,10\text{-phenanthroline})_3$, a fairly large molecule ($\sim 8 \text{ \AA}$ across), not incorporated into any of our models, and it is possible that there could be some “molecular crowding” or other non-trivial interactions with this reagent, thereby promoting disulfide bond formation between the residues that are far apart, like Gln26/Met182 and Gln26/Asp274.[25]

In conclusion, we can clearly see that our coarse-grained LBMC method is a fast and robust method for generating approximate protein ensembles. A coarse-grained LBMC simulation of 3 billion MC steps is completed in about one month of single-processor CPU time, compared to approximately five months of CPU time for the poorly-sampled 31 ns all-atom MD simulations. This represents a significant increase in the rate of observing large fluctuations that lends itself quite well to an investigation of a structurally diverse protein ensemble for use in high-throughput or ensemble docking calculations involved in rational drug design. To help facilitate this, as well as for studies of other protein fluctuations such as allostery, our laboratory maintains an Ensemble Protein Database (www.epdb.pitt.edu) as a repository for protein ensembles.

Methods

We studied the thermal motions of the E. Coli Glucose-Galactose Chemoreceptor using two different computational methods. The first method employs our previously developed Library-Based Monte Carlo (LBMC) method with a semi-atomistic protein model.[12] The second method is molecular dynamics (MD) with the OPLS-AA force field[19] and explicit solvent. We only studied the wild-type protein, and did not consider mutants to cysteine. This is partly because of our interest in native/wild-type fluctuations and partly to avoid complications of the catalyst used in [17].

Library-Based Monte Carlo

In LBMC, a molecule is divided into non-overlapping fragments and an ensemble of configurations – called a library – is generated in advance for each fragment. During LBMC simulation, fragment configurations in the molecule are swapped with configurations in the libraries. The new state is accepted according to the corresponding acceptance criterion, as described before.

In further detail, for a molecule divided into M fragments with coordinates denoted by $\vec{r} = \{\vec{r}_1, \dots, \vec{r}_M\}$ the total potential energy can be decomposed as:

$$U^{tot}(\vec{r}_1, \dots, \vec{r}_M) = \sum_{i=1}^M U_i^{frag}(\vec{r}_i) + U^{rest}(\vec{r}_1, \dots, \vec{r}_M) \quad \text{Eq. 1}$$

where U_i^{frag} is the potential energy of individual fragments and U^{rest} represent all other interactions between fragments.

The acceptance criterion of LBMC can be derived from the detailed balance condition.[12] A trial move in LBMC consists of swapping one or several fragment configurations in the molecule with configurations in the corresponding libraries. The old state will be denoted by o and the new one by n . The acceptance criterion for LBMC swap move can be written as:

$$p_{acc}(o \rightarrow n) = \min \left[1, \exp \left(-\beta U^{rest} \right) \right] \quad \text{Eq. 2}$$

where $\Delta U^{rest} = U^{rest}(n) - U^{rest}(o)$.

Our previous study showed that, for large proteins, LBMC can have a very small acceptance rate.[12] To cope with this problem, we developed a “neighbor list” trial move, in which trial configurations are selected from the neighbor lists of similar configurations, instead of the whole library, thereby increasing the acceptance rate. Fragment configurations in the libraries can be classified into the neighbor lists based on some similarity criterion like RMSD, or the sum of absolute differences for all bond angles and dihedrals within a fragment. When using a neighbor list trial move, the acceptance criterion should be modified to account for the introduced bias. The acceptance criterion for the LBMC swap move from a neighbor list can be written as:

$$p_{acc}(o \rightarrow n) = \min \left[1, \exp \left(-\beta \Delta U^{rest} \right) \frac{k_i^o}{k_i^n} \right] \quad \text{Eq. 3}$$

where k_i^o and k_i^n is the number of neighbors for the old fragment configuration and the new one, respectively. Note that when the number of neighbors is the same for all configurations in the library, the acceptance criterion of Eq. 2 simplifies to a simpler form of Eq. 3.

Fragment libraries

LBMC is flexible with regard to how a molecule can be divided into fragments. In this study, we use the same peptide-plane fragments as used in our previous work.[12] Specifically, three types of peptide-planes are employed corresponding to Alanine, Glycine and Proline residues. The peptide-planes span from the alpha carbon of one residue to the alpha carbon of the next residue and include all of the backbone degrees of freedom except ψ . To allow for the incorporation of Ramachandran potential, the peptide-plane fragments were modified to be conditional on the ϕ dihedral (i.e. to be uniformly distributed in ϕ with a suitable energy correction).

In all of the LBMC simulations, the library size was 2.9×10^5 for Alanine, 2.0×10^5 for Proline, and 3.6×10^5 for Glycine. For all libraries, neighbor lists were generated to contain 10 configurations. All of the details of neighbor list construction are described in Ref. 14.

Libraries of peptide-plane configurations were generated using molecular dynamics as implemented in the Tinker v. 4.2 software package[26] with the OPLS-UA force field[27] and implicit GB/SA solvent at 298 K.

Protein Model

LBMC is flexible in the choice of U^{rest} in Eq. 1, which can correspond to standard force field energy terms (e.g., Coulomb and van der Waals interactions) or more approximate interactions, such as Gō potential.[28,29] Following our previous work here, we chose U^{rest} to correspond to Gō interactions, which stabilize the native state while at the same time allowing large fluctuations.[12,30,31] All the details of Gō interactions are provided in Ref. [28,29].

LBMC Simulation Details

The starting structure used for all calculations is the X-ray crystal structure of GGBP with glucose bound (PDB accession coordinates: 2GBP), containing 309 residues.[13] This starting structure was chosen to match that used in Careaga and Falke's analysis.[17] While the wild type protein, including all residue types was input as the starting structure, the LBMC method treats all peptide plane configurations as that of Alanine, Glycine, or Proline (*vide supra*). The side chains of each residue are not included in our coarse-grained model, but the C_α and C_β are included throughout the simulation. The glucose molecule was removed for the apo-GGBP simulations, and all simulations were run for 3×10^9 MC steps, which followed an equilibration phase of 3×10^8 MC steps. The simulation temperature was chosen to be slightly below the unfolding temperature based on 13 short simulations of 3×10^8 MC steps. We thus selected $k_B T/\epsilon = 0.8$, where ϵ is the depth of the Gō potential.[28,29] Frames were saved after every 10^4 MC steps. Trial moves consisted of swapping three consecutive peptide-planes per step and/or changing the corresponding ψ angles. Due to the large size of this protein, we found that it was optimal to tune the acceptance rate to approximately 20-25% by adjusting two parameters. The first parameter, controlling the fraction of local moves from the neighbor lists to the “global” ones in which configurations are randomly selected from the neighbor lists of neighbor configurations, was set to 10%. The second parameter, controlling the fraction of ψ -only moves to the full peptide plane moves, was set to 30%.

The glucose-bound form of the chemoreceptor protein was modeled in three different ways. All three models are simplified, consistent with our protein model. The first method was a “virtual glucose” method, in which glucose itself wasn't physically added to the model, but an additional Gō type interaction was added to each residue that was observed to make either a hydrogen bond or hydrophobic contact with glucose in the crystal structure (the complete list of residues that additional Gō interactions were added is available in Supplementary Information). These additional interactions are supposed to mimic interactions of the glucose molecule and stabilize the binding site.

The glucose-bound form of the chemoreceptor was also modeled by adding one or three atoms – a coarse-grained representation – of the glucose molecule. The radius of the carbon atom used to represent glucose was 1.8 Å when one atom was used and 1.5 Å when three atoms were used. Gō type interactions were added between each atom in the coarse-grained representation of glucose, and every alpha carbon of the protein within 8 Å in the crystal structure.

For the disulfide analysis, the C_β distances between Gln26 and each of Met182, Asn260, Lys263, Asp267, and Asp274, was recorded every 200 MC steps during all of the LBMC simulations. These C_β distances are used to determine distributions (Figure 4) and rates

(Table 2) relative to the crystal structure. To determine a distance threshold for rate estimation, we first surveyed 50 disulfide bonds in published X-ray crystal structures from the Protein Data Bank (www.rcsb.org) and determined that the average disulfide bond length was 4.2 Å. While most of our LBMC simulations did produce minimum C_β distances between several residue pairs in the vicinity of this 4.2 Å length, the frequency of these occurrences was not enough to produce meaningful rate calculations. Therefore, a threshold of 7 Å was used.

The LBMC source code is available as a free download from our website: <http://www.ccbb.pitt.edu/Faculty/zuckerman/software.html> The modified source code used for the “virtual glucose” and docked glucose calculations is available upon request.

Molecular Dynamics Simulations

For the MD simulations, the crystal structure of GGBP (2GBP) was also used as the starting structure. The glucose molecule was removed to simulate apo-GGBP, and it was kept in place to simulate holo-GGBP. The protein was solvated using 11,636 TIP3P water molecules,[32] and simulated using the NAMD software package (version 2.6)[20] with the CHARMM27 all atom force field at 298 K.[33] The integration step was set to 2 fs and all of the distances in the system involving hydrogen atoms were constrained to equilibrium values.

Both systems were initially minimized using the conjugate gradient algorithm for 500 steps with the backbone atoms fixed and then for another 500 steps with all the atoms allowed to move. After minimization, the systems were heated to 298 K with harmonic restraints applied to C_α atoms first for 10 ps with constant volume, followed by another 10 ps at constant pressure. The restraints were removed and the systems were equilibrated for the final 1 ns.

The production runs for both systems were simulated for 30 ns with frames saved every 1 ps resulting into 30,000 frames. The C_β distance between Gln26 and Met182, Asn260, Lys263, Asp267, and Asp274 were calculated for each saved frame.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Bin Zhang, Xin Zhang, Ying Ding, and Andrew Petersen for helpful discussions. Funding was provided by the NIH through grants GM070987 and GM076569, as well as by the NSF through grant MCB-0643456.

List of Abbreviations

C _β	Beta Carbon
GGBP	Glucose-Galactose Binding Protein
LBMC	Library-Based Monte Carlo
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
OPLS-AA	Optimized Potentials for Liquid Simulations – All-Atom [force field]

OPLS-UA	Optimized Potentials for Liquid Simulations – United-Atom [force field]
RMSD	Root-Mean-Square Deviation

References

- [1]. Alberts, B.; Bray, D.; Hopkin, K.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Essential Cell Biology. 2nd Edition. Garland Science; New York & London: 2004. p. 119-159.
- [2]. Berg, JM.; Tymoczko, JL.; Stryer, L. Biochemistry. 5th Edition. W.H. Freeman & Company; New York: 2002. p. 41-73.
- [3]. Columbus L, Hubbell WL. A new spin on protein dynamics. Trends in Biochem. Sci 2002;27:288–295. [PubMed: 12069788]
- [4]. Doniach S, Eastman P. Protein dynamics simulations from nanoseconds to microseconds. Curr. Opin. in Struct. Biol 1999;9:157–163. [PubMed: 10322213]
- [5]. Ho BK, Agard DA. Probing the Flexibility of Large Conformational Changes in Protein Structures through Local Perturbations. PLoS Comput. Biol 2009;5:1–13.
- [6]. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature 2007;450:913–6. [PubMed: 18026087]
- [7]. Howard J. Molecular motors: structural adaptations in cellular functions. Nature 1997;389:561–567. [PubMed: 9335494]
- [8]. Howard, J. Mechanics of Motor Proteins and the Cytoskeleton. Sinauer Associates; Sunderland, Massachusetts: 2001.
- [9]. Cozzini P, Kellogg GE, Spyraakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sottriffer CA. Target flexibility: an emerging consideration in drug discovery and design. J. Med. Chem 2008;51:6237–55. [PubMed: 18785728]
- [10]. McCammon JA. Target flexibility in molecular recognition. Biochim. et Biophys. Acta 2005;1754:221–224.
- [11]. Lin J-H, Perryman AL, Schames JR, McCammon JA. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. J. Am. Chem. Soc 2002;124:5632–5633. [PubMed: 12010024]
- [12]. Mamonov AB, Bhatt D, Cashman DJ, Ding Y, Zuckerman DM. General Library-Based Monte Carlo Technique Enables Equilibrium Sampling of Semi-atomistic Protein Models. J. Phys. Chem. B 2009;113:10891–904. [PubMed: 19594147]
- [13]. Vyas NK, Vyas MN, Quirocho FA. Sugar and signal-transducer binding sites of the Escherichia coli galactose chemoreceptor protein. Science 1988;242:1290–5. [PubMed: 3057628]
- [14]. Amiss TJ, Sherman DB, Nycz CM, Andaluz SA, Pitner JB. Engineering and rapid selection of a low-affinity glucose/galactose-binding protein for a glucose biosensor. Protein Sci 2007;16:1–10.
- [15]. Tolosa L, Gryczynski I, Eichhorn LR, Dattelbaum JD, Castellano FN, Rao G, Lakowicz JR. Glucose sensor for low-cost lifetime-based sensing using a genetically engineered protein. Anal. Biochem 1999;267:114–20. [PubMed: 9918662]
- [16]. Taneoka T, Sakaguchi-Mikami A, Yamazaki T, Tsugawa W, Sode K. The construction of a glucose-sensing luciferase. Biosens. Bioelectron 2009;25:76–81. [PubMed: 19559587]
- [17]. Careaga CL, Falke JJ. Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor. Detection by disulfide trapping. J. Mol. Biol 1992;226:1219–35. [PubMed: 1518053]
- [18]. Careaga CL, Sutherland J, Sabeti J, Falke JJ. Large amplitude twisting motions of an interdomain hinge: a disulfide trapping study of the galactose-glucose binding protein. Biochem 1995;34:3048–55. [PubMed: 7893717]
- [19]. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc 1996;118:11225–11236.

- [20]. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J. Comput. Chem* 2005;26:1781–1802. [PubMed: 16222654]
- [21]. Lyman E, Zuckerman DM. On the structural convergence of biomolecular simulations by determination of the effective sample size. *J. Phys. Chem. B* 2007;111:12876–82. [PubMed: 17935314]
- [22]. Borrok MJ, Kiessling LL, Forest KT. Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded and ultra-high-resolution ligand-bound structures. *Protein Sci* 2007;16:1032–1041. [PubMed: 17473016]
- [23]. Whitford PC, Miyashita O, Levy Y, Onuchic JN. Conformational Transitions of Adenylate Kinase: Switching by Cracking. *J. Mol. Biol* 2007;366:1661–1671. [PubMed: 17217965]
- [24]. Herman P, Vecer J, Barvik I, Scognamiglio V, Staiano M, de Champdore M, Varriale A, Rossi M, D'Auria S. The Role of Calcium in the Conformational Dynamics and Thermal Stability of the D-Galactose/D-Glucose-Binding Protein From *Escherichia coli*. *Proteins: Struct., Funct., Bioinf* 2005;61:184–195.
- [25]. Ellis RJ. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci* 2001;26:597–604. [PubMed: 11590012]
- [26]. Ponder JW, Richards FM. An efficient newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem* 1987;8:1016–1024.
- [27]. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc* 1988;110:1657–1666.
- [28]. Taketomi H, Ueda Y, Nobuhiro G. Studies on protein folding, unfolding and fluctuations by computer simulation. *Int. J. Pept. Protein Res* 1975;7:445–459. [PubMed: 1201909]
- [29]. Ueda Y, Taketomi H, Nobuhiro G. Studies on protein folding, unfolding and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers* 1978;17:1531–1548.
- [30]. Zuckerman DM. Simulation of an Ensemble of Conformational Transitions in a United-Residue Model of Calmodulin. *J. Phys. Chem. B* 2004;108:5127–5137.
- [31]. Zhang BW, Jasnow D, Zuckerman DM. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proc. Natl. Acad. Sci. U. S. A* 2007;104:18043–18048. [PubMed: 17984047]
- [32]. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys* 1983;79:926–935.
- [33]. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Proffhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 1998;102:3586–3616.

Figure 1.

X-ray crystal structure of the *E. coli* chemoreceptor protein, with several of the key residues of interest labeled and highlighted as CPK representations: A. Red – Gln26, B. Yellow – Met182, C. Green – Asn260, D. Orange – Lys263, E. Cyan – Asp267, F. Blue – Asp274. The C_{β} distance between Gln-26 (A) and each of the following residues is as follows: B. 27.8 Å, C. 12.9 Å, D. 9.1 Å, E. 13.2 Å, F. 19.8 Å.

Figure 2. RMSD vs. Frame for Monte Carlo Simulations of GGBP: A. Unbound protein, B. Bound “virtual glucose” protein, C. Bound protein using single carbon atom representation of glucose, D. Bound protein using three carbon atom representation of glucose. RMSD vs. Time for Molecular Dynamics Simulations of GGBP: E. Unbound protein, F. Bound protein.

Figure 3.

Illustration of the range of positions of the beta carbons of interest at snapshots in the apo-GGBP simulation trajectory. The inset on the left shows the positions of C_β at frames spaced every 6×10^6 MC steps apart in the trajectory. At the right, are snapshots from various points along the trajectory; starting from the X-ray crystal structure conformation, going to a point with the Gln26/Lys263 C_β distance at 4.4 Å and the Gln26/Asp267 C_β distance at 4.1 Å, and ending at a point with the Gln26/Asn260 C_β distance at 7.6 Å and the Gln26/Lys263 C_β distance at 5.1 Å. C_β are colored as follows: Red – Gln26, Yellow – Met182, Green – Asn260, Orange – Lys263, Cyan – Asp267, Blue – Asp274.

Figure 4.

Histograms of the beta-carbon distances from the LBMC and MD simulations for each of the five residue pairs studied: A. Gln26/Met182, B. Gln26/Asn260, C. Gln26/Lys263, D. Gln26/Asp267, E. Gln26/Asp274. The vertical black line in the center of each graph corresponds to the beta-carbon distance of the 2GBP X-ray crystal structure. The probability density is plotted on the y-axis.

Table 1

The minimum and 1 percentile beta-carbon distances (in Å), as well as the standard deviation, from the LBMC and MD trajectories

X-Ray Crystal Value	MET182 27.8			ASN260 12.9			LYS263 9.1			ASP267 13.2			ASP274 19.8		
	MIN	1 %ile	S.D.	MIN	1 %ile	S.D.	MIN	1 %ile	S.D.	MIN	1 %ile	S.D.	MIN	1 %ile	S.D.
Simulation															
apo-GGBP MC	17.5	23.0	2.1	5.7	9.7	1.6	2.2	5.9	1.5	3.0	8.3	1.6	10.2	14.5	1.9
holo-GGBP MC "virtual glucose "	19.7	23.7	1.9	5.1	9.1	2.2	0.0	4.9	2.1	3.3	7.5	2.0	6.1	11.8	2.3
holo-GGBP MC 1-C glucose	18.0	23.2	2.2	4.6	9.3	1.6	1.0	4.7	1.7	2.0	7.3	1.7	8.7	13.6	2.2
holo-GGBP MC 3-C glucose	17.6	23.1	1.9	6.4	9.5	1.6	2.4	5.7	1.7	3.7	7.3	1.8	9.9	14.1	1.9
apo-GGBP MD	25.8	26.9	4.3	10.7	11.9	0.8	7.4	8.2	0.7	11.0	11.9	0.6	17.0	17.9	0.8
holo-GGBP MD	26.6	27.6	4.4	11.0	12.0	0.7	7.5	8.5	0.7	11.6	12.4	0.6	17.6	18.6	0.9

Table 2
Calculated “Rates” for Disulfide-capable interactions in which the C_β distance falls below 7 Å

Residue Pair	apo-GGBP	holo-GGBP “virtual glucose”	holo-GGBP 1-C G6 center	holo-GGBP 3-C G6 center	apo-GGBP exp. k _{ss} [‡] (s ⁻¹)	holo-GGBP exp. k _{ss} [‡] (s ⁻¹)
Gln26 / Met182	0	0	0	0	1.40 × 10 ⁻²	0
Gln26 / Asn260	7.59 × 10 ⁻⁹	1.70 × 10 ⁻⁸	1.95 × 10 ⁻⁸	4.37 × 10 ⁻⁹	1.60 × 10 ⁻²	1.30 × 10 ⁻²
Gln26 / Lys263	9.13 × 10 ⁻⁷	6.22 × 10 ⁻⁷	1.06 × 10 ⁻⁶	8.35 × 10 ⁻⁷	7.00 × 10 ⁻¹	2.10 × 10 ⁻¹
Gln26 / Asp267	6.98 × 10 ⁻⁸	1.23 × 10 ⁻⁷	1.58 × 10 ⁻⁷	2.00 × 10 ⁻⁷	7.30 × 10 ⁻²	5.60 × 10 ⁻⁴
Gln26 / Asp274	0	2.28 × 10 ⁻⁹	0	0	1.50 × 10 ⁻²	1.90 × 10 ⁻⁴

[‡]The calculated “rates” are shown in (MC Step)⁻¹.

[‡]The Experimental Rate Constant (k_{ss}) used is from Careaga, C.L.; Falke, J.J. 1992. *J. Mol. Biol.*, Vol. 226, 1219-35.