COMPUTATIONAL PREDICTION OF CHEMICAL SHIFTS OF APO-STATE OF THE PROTEIN CALBINDIN D$_{9K}$

Mahendra Thapa* and Dr. Mark Rance**

* Department of Physics, University of Cincinnati, OH

**Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati, OH

Abstract: Calbindin D$_{9k}$ is a small protein. It is found predominantly in tissues involved in the uptake and transport of calcium such as cells of the intestinal brush border membrane. Molecular dynamics simulation (MD) of the protein also helps to study motion and its development with time which may not be studied experimentally. Chemical shifts are also calculated from the data generated by MD using various chemical shift prediction tools such as Sparta+. Chemical shifts are most often used to predict regions of secondary structure of proteins, to aid in the refinement of complex structures, and for the characterization of conformational changes associated with partial unfolding or binding. We used AMBER 12.0 software on GPU to simulate the apo-state of the protein using water model TIP3P and the force field ff12SB and then Sparta+ was used to predict chemical shifts values for backbone atoms. These values are then studied with the values obtained from its initial structure.

Keyword: Protein; Calbindin D$_{9k}$; Amino acids; Molecular Dynamics Simulation (MD); Chemical shift; Sparta+; Protein Data Bank; AMBER12; TIP3P water model; Langevin dynamics.

INTRODUCTION

Calbindin D$_{9k}$ is a single domain protein of the EF-hand family with molecular weight 8700 Dalton and 75 amino acids. It consists of a pair of EF-hands connected by a flexible linker; the EF-hand is a helix-loop-helix motif. It is found predominantly in tissues involved in the uptake and transport of calcium such as cells of the intestinal brush boarder membrane, also in the kidney and uterus in some mammalian species. It in encoded in humans by the S100G gene. It is one of the vitamin D-dependent calcium-binding proteins. It serves as an attractive model system for computational investigation due to its small size. Its structure in various calcium-loaded states has been extensively characterized experimentally [NMR/X–rays].The apo structure of the protein was determined by high resolution ^{1}H NMR technique and refined with restrained molecular dynamics in vacuo[1].

The effective magnetic field experienced by a nucleus is the result of the local magnetic fields induced by currents of electrons in the molecular orbitals and the externally applied magnetic field in nuclear magnetic resonance (NMR) spectroscopy. Since some atomic nuclei possess nuclear spin, the resonance frequency of the nucleus shifts with the change in the effective magnetic field acting on it. The magnitude of the shift depends on the type of the nucleus and the details of the electron motion in the nearby atoms and molecules. It is called a chemical shift associated with the nucleus in the given sample. In structural biology, chemical shifts are most often used to predict regions of secondary structure of proteins, to aid in the refinement of complex structures, and for the characterization of conformational changes associated with partial unfolding or binding.[11, 12,13]

Molecular Dynamics (MD) Simulation is a computer simulation of physical movements of atoms and molecules. This method is used to calculate the time dependent behavior of a molecular system in terms of classical Newtonian dynamics of nuclei. All electronic structure and other quantum mechanical properties are implicitly described using empirical potential functions. MD simulations of a protein produce very large amounts of data which are analyzed using the concept of statistical mechanics. So, these data cannot easily be studied and utilized without automated data mining. The algorithms and force fields used in MD simulations have become more powerful and accurate over the last years and the steadily rising amount of computational resources available to researchers have made them popular and widely applied tools. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. MD studies, when combined with NMR data, allow for detailed analysis of protein

Author for Correspondence: Mahendra Thapa, Department of Physics, University of Cincinnati, OH. E-mail: mahendrab2007@gmail.com.

dynamics and examining the contributions of such dynamics to protein function and allosteric processes. Comparing MD results to NMR data will also provide important information concerning the validation of force fields currently in use in studies of proteins by MD simulations. It should be kept in mind that MD makes some severe approximations on (i) the quality of force field , (ii) Size and time (atomistic simulations can be performed only for systems of a few tenths of Å on the length scale and typically sub-microseconds on the time scale) and (iii) conformational freedom of the molecule (the number of possible conformations a molecule can adopt is enormous, growing exponentially with the number or rotatable bonds) .This method can only be applicable to systems that have been parameterized . It is also assumed that the connectivity of atoms cannot change during dynamics (i.e., no chemical reaction).[14, 15]

Sparta+ [16] is one of the semi-empirical chemical shift prediction tools. It employs a well-trained neural network algorithm to make rapid chemical shift prediction on the basis of known structure. It was trained on a large carefully pruned database containing 580 proteins for which high resolution x-ray structures and nearly complete backbone and $^{13}C_\eta$ chemical shifts are available.

Many computational studies have already been completed to determine the chemical shifts of atoms of various proteins using the techniques of MD and chemical shift prediction tools. The structural dynamics of the proteins are successfully interpreted on the basis of the results so obtained. In the present study, 50 ns explicit solvent all atoms MD simulation of the apo-state of the protein Calbindin D_{9k} was performed. Sparta+ was then used on each of the large number of snapshots (approx. 50000 pdb files) so generated by MD. The average chemical shift values for backbone atoms (C_α, C, N) of the protein was determined by linear averaging of the predicted chemical shift for each snapshots saved during the trajectory[9] .These values were then compared with the starting structure, which was a median static structure of the protein from the NMR –derived ensemble of 33 structures. The main aim of this study was to introduce the relevant concepts to individuals who are just entering in this field or planning such research endeavors.

MATERIALS AND METHODS

The P43G apo structure of the Protein Calbindin was taken directly from the Protein Data Bank; the PDB code is 1CLB[17]. When the Theseus program [2] was run on the ensemble of structures consisting of 33 models, it determined model 10 as the median structure. With the help of the Chimera software [3], the amino acid methionine (Met) was added at the beginning of the sequence of model 10. The hydrogen atoms of the model were at first removed and then were added back to the model at the appropriate positions. This job was

done using the software reduce[4] which is a component of the molecular dynamics suit AMBER12[5]. This fixes hydrogens in the model in standardized geometry with optimization of the orientations of OH, SH, NH_3^+, Met methyls, Asn and Gln side-chain amides, and His rings. At this point, the model consisted of 1212 atoms.
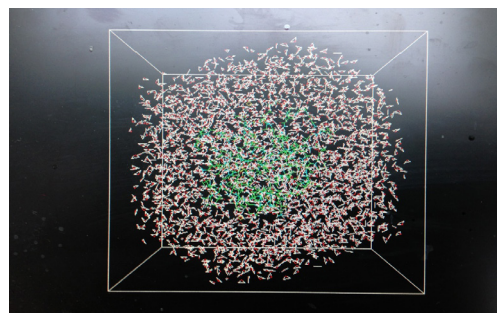

Figure: 1 A protein molecule in octahedron box of water.

When the leap command was performed using AMBER12 with the AMBER force field ff12SB and explicit TIP3P water model[6], 2914 water molecules were added to the model to form a truncated octahedron box of volume 118223.227 $Å^3$ and 7 Na-ions were added to the model to make the system neutral such that no solute atom was within 10 Å of a box edge. The total mass and density were 61083.320 amu and 0.858 g/cc respectively. The resulting system was as shown in the Fig-1. The leap command produces the necessary files prmtop (this file consists of the description of the molecular topology and the necessary force field parameters) and inpcrd (this file consists of the atomic coordinates of the model) for the simulation in the AMBER suit[7]. Before starting simulations, we had to minimize and equilibrate the system (so generated by the leap command) under particle mesh Ewald periodic boundary conditions and 10 Å cutoff. Langevin dynamics was also introduced.

The minimization step minimizes the energy of the system to keep it in the most favorable configuration and it also removes the overlaps between atoms. Minimization with sander or pmemd of the AMBER12 suit will take the system to the nearest minima. Our minimization procedure for the solvated protein consisted of two stages: (i) In the first stage, the protein was held fixed using a force constant of 500 K Cal mol^{-1} $Å^{-2}$ and just minimized the positions of water and Na-ions. There were 1000 steps of minimization in which 500 steps of steepest descent algorithms followed by 500 steps of conjugate gradient algorithms. (ii) In the second stage, the entire system was minimized using 2500 steps of steepest descent followed by 2500 steps of conjugate gradient algorithms.[7]

The equilibration step makes sure that the structure, pressure, temperature and energy of the system is stable with respect to time. For this, the system was heated to 50K through 10 ps of canonical (NVT) MD with a 2fs time step and SHAKE constraints on bonds involving

hydrogen atoms followed by heating to 150K and 301K in successive 20 ps steps under the same condition. Then the system was allowed to equilibrate in a 1000 ps isothermal-isobaric (NPT) simulation to equilibrate the solvent density to 1.0 g/cc. A 50 ns of production MD was performed under NVE conditions. Snapshots of each trajectory were saved every 1.0 ps during both the equilibration and production steps[8]. Finally, we generated 50000 pdb files using cpptraj module of AMBER 12. These were analyzed using various tools like mathematica/matlab, Perl, FORTRAN, etc.

Sparta+ was used to predict the chemical shifts of backbone atoms of the protein. In the prediction process, waters were removed from each trajectory and each snapshot was saved as an individual PDB coordinate file [9,10]. Instead of taking the default initial structure (which was used to start the leap command in above step), its minimized structure (which we call here as initial structure or starting structure) was chosen for the comparison purpose and the chemical shifts of corresponding atoms of the protein were predicted using Sparta+.

RESULTS AND DISCUSSION

Before using the simulated data to extract various physical quantities like chemical shifts and crystallographic B factors, the consistency check of the data was done by some plots. The plots (not shown here) for the temperature looked fine, the temperature had equilibrated around 301 K as expected. The heating phase also showed a nice gradual increase in temperature with no sudden jump. The simulation seemed to be fairly stable, as the potential energy was fluctuating around a constant mean value. Similarly, pressure, volume, density, RMSd were also monitored[5].
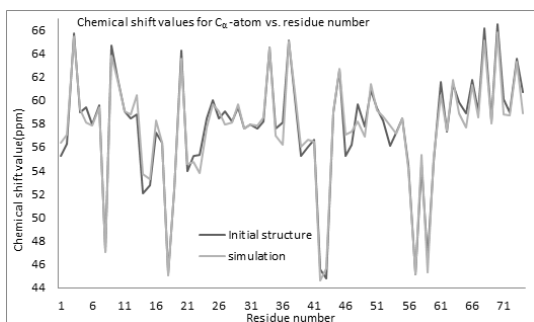

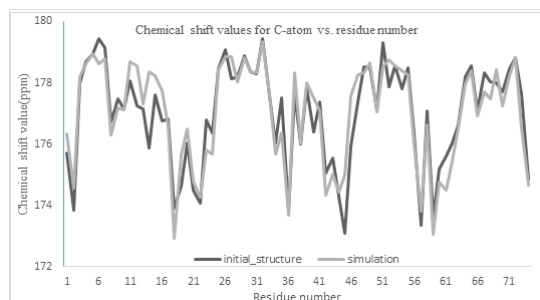**Figure : 2 A plot of chemical shift of C_α - atom vs. residue number.**

Figure : 3 A plot of chemical shift of C - atom vs. residue number.
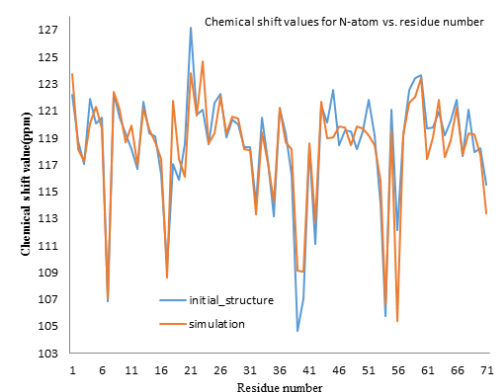

Figure : 4 A plot of chemical shift of N - atom vs. residue number.

Plots of chemical shift value versus residue number of the protein are shown in Fig -2, Fig-3 and Fig-4. These values obtained from MD data match well with corresponding values obtained from the starting structure data for each of backbone atoms C_α, C and N of the protein. If the difference of 1ppm between the chemical shift values obtained from MD data and that obtained from starting structure for each corresponding residue is ignored, following facts are observed in the case of the backbone C_α atom (see Fig-2): For the residue 'Ala 15' (alanine residue in position 15) the simulation data gives a higher value of the shift than that from initial structure. The same is true for the residues 'Leu47' and 'Leu54'. But, for the residues 'Leu24', 'Glu36' and 'Glu49', the chemical shift values obtained from the initial structure data are higher than that from the simulation data. For rest of the residues, these values are seen to either overlap or be very close. In the case of the backbone C-atom (see Fig-3), residues 'Ala15' and 'Thr46' give higher chemical shift values for the simulation data than that for the initial structure data. But, the reverse is true for residues 'Leu24', 'Glu36' and 'Val62'. For the remaining residues, these values are seen to either overlap or be nearly equal. For the backbone N-atom (see Fig-4), only the residue 'Gly43' gives a higher chemical shift value for the simulation data than that for the initial structure and for the residues 'Leu24' , 'Glu49', 'Asp55', 'Gly60' and 'Lys72', the chemical shift values for the initial structure data are higher than that for the simulation data. For the remaining residues, these values are seen either to overlap or be nearly equal.

On the average, the chemical shift values obtained from the initial structure have higher than that obtained from the simulation data in the case of backbone atoms C_α, C and N. It is also noticed that the chemical shift value for 'Ala15' is higher for the simulation than that for initial structure in both backbone atoms C_α and C but the case is the opposite (i) for 'Leu24' in the backbone atoms C_α, C and N (ii) for 'Glu36' in C_α and C atoms and (iii) for 'Glu49' in C_α and N.

Table for Rmsd of chemical shifts

Rmsd between chemical shift prediction from MD simulation and energy minimized median NMR structure with Sparta+ using ff12SB force field

Water model used	C_α-atom	C-atom	N-atom
TIP3P	0.81	0.68	1.73

While plotting the graphs and calculating the Rmsd, the first and last residues of the protein were not considered as demanded by the tool Sparta+. Likewise, proline (Pro) residues were not taken into account for the graph plotting and finding Rmsd for N-atom of the residues.

Conclusions and Future directions

The chemical shift values predicted from MD simulations with the AMBER force field ff12SB and water model TIP3P using Sparta+ for backbone atoms C, C_α and N of the apo state of the Calbindin D_{9k} matches well with the corresponding values obtained from the energy minimized median NMR structure so obtained using Theseus for the protein. The variation of chemical shift values in the range of 1 ppm is observed for most residues in these backbone atoms except for some residues. The sources of errors for these variations may be due to empirical nature of the force field ff12SB and chemical shift determining tool Sparta+. So, these results also demand for further possible modification in the force field parameters. If it is possible to compare results with the experimental chemical shift values of the protein, the picture will be more robust. Furthermore, if such simulations were repeated for different combination of different force fields, water models and chemical shift determining tools, it will be more helpful. Instead of simulating the protein at NVE condition, it will be good to observe the results at NPT and NVT conditions.

REFERENCES

1] PROTEINS: Structure, Function and Genetics. 1998. **33:**265-284.

2] Bioinformatics. 2006. Vl1 **22** (17)**:** 2171-2172.

3] http://www.cgl.ucsf.edu/chimera/docs/UsersGuide/midas/addaa. html).

4] Word, et. al. 1999. J. Mol. Biol. **285:** 1735-1747.

5] http://ambermd.org/

6] J.Chem. Phys. 1983. **79:** 926-935.

7] http://ambermd.org/tutorials/basic/tutorial1/

8] Theory Comput. J. Chem. 2007. **3:** 961-975.

9] J. Am. Chem. Soc. 2012. **134:** 6365 – 6374.

10] J Biomol NMR. 2010. **48:** 13 – 22.

11] http://en.wikipedia.org/wiki/Chemical_shift

12] http://hyperphysics.phy-astr.gsu.edu/hbase/nuclear/nmrcsh.html

13] http://orgchem.colorado.edu/Spectoscopy/nmrtheory/chemshift. html

14] J Comput Chem. 2013. **34** (3)**:** 198-205.

15] http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html

16] J. Biomol. NMR. 2010. **48:** 13-22.

17] www.rcsb.org

■