Residual Dipolar Couplings (RDCs) are one of the most recent data types to have emerged from Nuclear Magnetic Resonance (NMR) spectroscopy. They have been utilized in simultaneous protein structure determination and characterization of protein dynamics. The challenge of this particular data type is in separating structural and motional information – an undertaking that is further complicated in the case of proteins experiencing large conformational fluctuations. To assist with this task we have combined a number of published software tools in the form of a user-friendly pipeline, designed to test whether a given protein is fixed in a unique conformational state or undergoes conformational transition (Figure 1). In our investigations we used the well-studied *E. coli* dihydrofolate reductase (ecDHFR) enzyme (Figure 2), which displays conformational flexibility by rotations between two subdomains. Four distinct conformations have been observed by X-ray crystallography (open, closed, occluded and disordered); three of the states are classified by whether the active site is open, closed or occluded by the Met20 loop, while the disordered state describes the case in which the loop is rendered crystallographically invisible.

Drs. Homayoun Valafar (Professor, Department of Computer Science & Engineering, College of Engineering and Computing) and Sondra Berger (Professor Emeritus, Department of Drug Discovery & Biomedical Sciences, College of Pharmacy) from the University of South Carolina have served as sponsor and co-sponsor, respectively. The objectives of our research were: 1. Dihydrofolate reductase (DHFR) sample preparation and Nuclear Magnetic Resonance (NMR) data acquisition (to successfully obtain experimental RDC data) and 2. Development and implementation of a novel pipeline for RDC analysis (to establish the feasibility of RDC data in addressing both the structure and dynamics of DHFR in solution state). Utilizing a multidisciplinary approach which integrates computational and physical (training in protein purification methods) sciences, we have been able to implement and evaluate the pipeline using simulated RDC data and some currently available experimental RDC data.

Our work has the potential to address challenges in the field of structural biology. Our method combines programs that test hypotheses regarding a protein’s structure and dynamics and may encourage others to adopt a similar approach in efforts to provide more complete information for existing structures or full characterization of challenging proteins. Additionally, the software programs in our analysis pipeline are free and available to the public at [http://ifestos.cse.sc.edu](http://ifestos.cse.sc.edu), which we expect will encourage the use of experimental approaches acquiring non-traditional data for the study of proteins. Our project highlights the importance of a multidisciplinary approach, incorporating sponsors from the computational and biomedical sciences and partnerships with instrumentation and training facilities to establish an infrastructure for continued collaborative efforts. The presentation of our research findings at annual SC INBRE symposiums has fostered networking and research activities between faculty and students, including under represented faculty and students, from academic institutions throughout the state. Research findings have also been incorporated into computational structural biology courses at the University of South Carolina and in the form of guest lectures at a local high school to inspire pursuit of STEM disciplines by students, improve the public's perception of cyberinfrastructure, and to demonstrate that complex problems can be solved by using technologies from multiple disciplines in new ways.
The following sections contain a summary of our results for each objective.

Figure 1. Pipeline for analysis of existing structural information and solution-state NMR data (RDC data)

Figure 2. X-ray structure of ecDHFR in the closed conformation. The adenosine-binding domain is green (residues 38-88). Loops of the major domain are colored red (Met20, residues 9-23), yellow (F-G, residues 117-131), and orange (G-H, residues 142-149). The figure was rendered using PDB ID: 1RX2 and VMD

Results Specific to Aim 1:

The first major activity was DHFR expression. This involved obtaining the pET-22b(+) derived expression plasmid (provided by Dr. Stephen J. Benkovic's lab) which contained wild type \textit{E. coli} DHFR (ecDHFR). We confirmed that the plasmid contained DHFR by performing a plasmid digest experiment (Figure 3). This was followed by transformation into BL21(DE3) cells. This allowed for the overexpression of \textit{E. coli} DHFR. Transformants expressing high levels of DHFR activity were identified by a spectrophotometric assay that measures the reduction of dihydrofolic acid to tetrahydrofolic acid using NADPH as the reductant (Figure 4A).

The next major step involved purification. Our initial purification protocol utilized cells grown in Luria Bertani broth (LB). As a pre-purification step, we utilized polyethyleneimine (PEI) precipitation and ammonium sulfate (AS) precipitation. PEI precipitation experiments using different concentrations of PEI allowed us to determine that 2% PEI is sufficient to precipitate nucleic acids and acidic proteins, without precipitating ecDHFR (Figure 4B). Removal of PEI and partial purification were achieved by differential ammonium sulfate precipitation (50% to 80%) (Figure 4C).

Following PEI and AS precipitations, we decided to utilize a Methotrexate (MTX) Affinity column in place of the FPLC Cibacron blue affinity column and obtained more specificity with regard to ecDHFR purification (Figures 4D and 4E). This was motivated by the fact that spectral analysis by another laboratory reported ecDHFR samples contaminated with Tetrahydrofolate (THF), a folic acid derivative (produced from dihydrofolic acid by dihydrofolate reductase) endogenous to the \textit{E. coli} cells that binds the enzyme during overexpression and remains bound throughout purification. Utilizing the MTX affinity column eliminated \textit{E. coli} DHFR complexes with bound THF, since ecDHFR bound to THF cannot bind the MTX affinity column.

After establishing an initial purification protocol with LB, cells expressing ecDHFR were grown in minimal M9 medium to facilitate isotopic labeling of the protein. Due to the metabolic restraints that result from the change in medium, cells had to be induced by isopropyl β-D-thiogalactoside (IPTG). In pilot studies, an elution protocol was developed in which ecDHFR remained bound to the matrix in the presence of high salt and was eluted by addition of folic acid and elevation of pH. Fractions eluting from the MTX affinity column were then concentrated and defolated prior to FPLC anion-exchange chromatography.

This protocol consistently resulted in the successful purification of ecDHFR which was analyzed in 1D NMR experiments to acquire 1D-$^1$H spectra as the starting point for the analysis and behavior of ecDHFR in solution state (Figure 5A). Subsequent protein samples were uniformly enriched with $^{15}$N as the sole nitrogen source in the M9 growth medium for 2D-$^{15}$N studies. At this point, the purification process became unreliable. After several months of troubleshooting, it is apparent that a mold contamination affected the water source, leading to proteolysis during the purification process (Figures 5B and 5C).
Figure 3. Results of plasmid digest
Figure 4. Diagram summarizing ecDHFR expression and purification protocol: (A) spectrophotometric assay for analysis of ecDHFR activity in crude extracts of BL21(DE3)-pET22b transformants (B) spectrophotometric assay for analysis of ecDHFR activity in cell extracts treated with PEI (C) spectrophotometric assay for analysis of ecDHFR activity in cell extracts treated with PEI following ammonium sulfate precipitation (D) Fast Protein Liquid Chromatography (FPLC) chromatogram following methotrexate affinity (E) spectrophotometric assay for analysis of ecDHFR activity following methotrexate affinity and anion exchange chromatography in concentrated fractions.
Figure 5. NMR experiment results: (A) Proton 1D- $^1$H spectra of purified ecDHFR; (B) Proton 1D- $^1$H spectra of RNA contaminant (C) 15 N- $^1$H HSQC spectra of purified ecDHFR
**Results Specific to Aim 2:**

Preliminary pipeline analyses were previously conducted utilizing experimentally available RDC data from a DHFR structure obtained from the *L. casei* species in the closed conformation and compared to synthetically generated RDCs from a DHFR structure obtained from the *E. coli* species in the occluded conformation. Due to our inability to acquire experimental RDC data, development and implementation of the software pipeline for RDC analysis was limited to the use of synthetic RDC data and currently available experimental RDC data. This allowed for construction of an initial RDC analysis pipeline, identification of file input/output compatibility issues, and pipeline performance evaluation.

To summarize relationships among existing structures, msTALI was utilized for alignment of the following *E. coli* DHFR structures obtained from X-ray Crystallography (PDB IDs: 1RA2, 1RX2, 1RX3, 1RX5, 1RX7), representing three different conformational states (please refer to Figure 6). Although the structures appear similar, as evidenced by the msTALI RMSD calculation (0.871322; Figure 6A), the resulting extended alignment scores identified a region within the Met20 loop in which the conformations differed (residues 15-23; Figure 6B). Moreover, the resulting tree (Figure 6C) confirmed what has been reported in the literature - that the closed and occluded conformations lie on either end of the spectrum, while the open conformation resembles an intermediate state between closed and occluded. Therefore our initial structure consolidation step of the pipeline enabled us to identify conserved regions, as well as points that accommodate conformational changes which are similar to those previously reported in the literature; it also provided us with a suitable representative structure for the closed and occluded conformations (PDB ID: 1RX2 and 1RX7, respectively) which would be used to accomplish intermediate steps for further evaluation of the pipeline.

Redcat has allowed us to both generate and analyze simulated RDC data and perform RDC analyses with currently available experimental RDC data. We have previously demonstrated the success of Redcat's error analysis in identifying regions of structural inconsistency (or the presence of motion), consistent with those in the literature, by generating in Redcat synthetic RDC data from the closed conformation (PDB ID: 1RX2) and then comparing it to structures in the closed (PDB ID: 1RX2) and occluded (PDB ID: 1RX7) conformations (Refer to Figure 7). Redcat analysis utilizing currently available experimental RDC data of wild-type ecDHFR from one alignment media (NH) revealed that RMSD and Q-Factor scores (Refer to Figure 7) are lower when the wild-type data are compared to the closed (PDB ID: 1RX2) conformation (RMSD= 3.30742, Q-Factor= 0.352914), as opposed to the occluded (PDB ID: 1RX7) conformation (RMSD= 4.5026, Q-Factor= 0.480444), suggesting that the wild-type is closest to the closed conformation. However, the value of the Q-Factor scores suggest that more data (additional vectors and/or alignment media) is required to confirm whether the structure is in one of the states observed by X-ray crystallography, or that the wild-type structure is representing an averaged conformational state. To address this question we are currently examining synthetic RDCs generated from Redcat for various weighted occupancies, of each of the four conformations of ecDHFR, to see if the weights can be successfully recovered. We have also performed a simulation in which the closed form of ecDHFR was guided toward an occluded conformation; synthetic RDCs were generated for each resulting trajectory and Redcraft was then utilized to assemble and characterize the dynamical portions of the protein (Figure 8). The latest software package, as well as a user's guide, for each program in our computational pipeline is available for download at [http://IFESTOS.cse.sc.edu](http://IFESTOS.cse.sc.edu).
Figure 6. msTALI results using Xray structures in the closed (PDB IDs: 1RX2 and 1RX3), open (PDB ID: 1RA2), and occluded (PDB IDs: 1RX5 and 1RX7) conformations: (A) Amino acid structure alignment with resulting core identification, calculated msTALI score and calculated RMSD; (B) Graphical representation of the extended alignment scores at each residue; (C) Tree produced from structure alignment
Figure 7. Results of Redcat Error Analysis using different comparisons
Figure 8. Redcraft dynamic profiles utilized to assemble and characterize the dynamical portions, resulting from synthetic RDCs generated for each trajectory of a simulation in which the closed conformation of ecDHFR (1RX2) was guided toward the occluded conformation (1RX7); in order to identify discrepancies at the end of the protein, Redcraft reconstructed the structure starting from the end of the protein (refer to chart with Residues 159 to 148)