Investigation of intrinsic dynamics of enzymes involved in metabolic pathways using coarse-grained normal mode analysis

Sarah M. Meeuwsen, An N. Hodac, Lauren M. Adams, Ryan D. McMunn, Maxwell S. Anschutz, Kari J. Carothers, Rachel E. Egdorf, Peter M. Hanneman, Jonathan P. Kitzrow, Cynthia K. Keonigsberg, Oscar Lopez-Martinez, Paul A. Matthew, Ethan H. Richter, Jonathan E. Schenk, Heidi L. Schmit, Matthew A. Scott, Eva M. Volenec and Sanchita Hati

Cogent Biology (2017), 3: 1291877
Investigation of intrinsic dynamics of enzymes involved in metabolic pathways using coarse-grained normal mode analysis

Sarah M. Meeuwsen1, An N. Hodac1, Lauren M. Adams1, Ryan D. McMunn1, Maxwell S. Anschutz1, Kari J. Carothers1, Rachel E. Egdorf2, Peter M. Hanneman1, Jonathan P. Kitzrow1, Cynthia K. Keonigsberg1, Oscar Lopez-Martinez3, Paul A. Matthew4, Ethan H. Richter1, Jonathan E. Schenk1, Heidi L. Schmit1, Matthew A. Scott1, Eva M. Volenec1 and Sanchita Hati1*

Abstract: Intrinsic dynamics of proteins are known to play important roles in their function. In particular, collective dynamics of a protein, which are defined by the protein's overall architecture, are important in promoting the active site conformation that favors substrate binding and effective catalysis. The primary sequence of a protein, which determines its three-dimensional structure, encodes unique dynamics. The intrinsic dynamics of a protein actually link protein structure to its function. In the present study, coarse-grained normal mode analysis was performed to examine the intrinsic dynamic patterns of 24 different enzymes involved in primary metabolic pathways. We observed that each metabolic enzyme exhibits unique patterns of motions, which are conserved across multiple species and functionally relevant. Dynamic cross-correlation matrices (DCCMs) are visibly identical for a given enzyme family but significantly different from DCCMs of other protein families, reinforcing that proteins with similar function exhibit a similar pattern of motions. The present work also reasserted that correct identification of unknown proteins is possible based on their intrinsic mobility patterns.

ABOUT THE AUTHORS
Our research is focused on the simple question - how protein motions influence substrate binding and catalysis. The relationship between protein dynamics and its role in function (molecular recognition, catalysis, and allostery) is an evolving perspective in enzymology. For the past several years, we have been investigating the interplay of protein dynamics and enzymatic processes using computations and experiments. In particular, we are exploring how dynamics of different domains modulate substrate recognition, catalysis, and allosteric communication in multi-domain enzymes. Our overall goal is to advance the drug design and screening process, by taking into consideration proteins' intrinsic flexibility.

PUBLIC INTEREST STATEMENT
In the present work, the intrinsic dynamics of enzymes from 24 different families, which are involved in key metabolic pathways, were analyzed. The goal of this study is to better understand the connection between intrinsic dynamics and enzymatic function. As each enzyme exhibits a unique pattern of motions, these intrinsic dynamic patterns can be used to characterize the function of unknown proteins. This work demonstrates that (i) dynamic-based functional characterization is a highly reliable method and (ii) computationally less expensive coarse-grained normal mode analysis (NMA) is extremely useful to explore intrinsic dynamics of a vast number of proteins in relatively short duration of time. Overall, this work reasserted the importance of dynamics in protein function and suggested that coarse-grained NMA could be used for protein identification and classification. These findings can be applied by researchers working in the fields of protein chemistry and enzymology.
1. Introduction

Proteins play crucial roles in nearly all biological processes. Each protein has a primary amino acid sequence that dictates the tertiary structure of the protein, and ultimately determines the protein’s overall function. In other words, the amino acid sequence → structure → function relationship exists for proteins and enzymes. Although the static structure of proteins is commonly correlated with their function, proteins are intrinsically dynamic in nature, and their internal motions, over a wide-range of timescales, assist their function. In fact, it has been suggested that the protein’s dynamics is the fundamental link between its structure and function (Hensen et al., 2012). However, the local dynamical character of an enzyme active site does not alone produce the functionally/catalytically competent state necessary to perform the protein’s biological function. In fact, the whole protein’s dynamism is essential for its overall function. Several studies revealed the influence of protein dynamics on enzyme catalysis, substrate binding, allosteric signal propagation, and protein–protein interactions (Agarwal, 2005; Agarwal, Billeter, Rajagopalan, Benkovic, & Hammes-Schiffer, 2002; Bhabha et al., 2011; Chennubhotla & Bahar, 2007; Chennubhotla, Yang, & Bahar, 2008; Eisenmesser et al., 2005; Hammes-Schiffer & Benkovic, 2006; Henzler-Wildman & Kern, 2007; Henzler-Wildman et al., 2007; Klinman & Kohen, 2013; Lisi & Patrick Loria, 2016; Liu et al., 2016; Singh, Francis, & Kohen, 2015; Vöhringer-Martinez & Dörner, 2016; Zen, Micheletti, Keskin, & Nussinov, 2010). In addition, it has been observed that the dynamics of residues that are not directly involved in catalysis are also important for enzyme function (Roston, Kohen, Doron, & Major, 2014; Tousignant & Pelletier, 2004).

For a majority of known protein sequences and structures, functional information is lacking (Erdin, Lisewski, & Lichtarge, 2011). To date, the functional classification of proteins has been largely dominated by sequence- or structure-based methods. However, both methods have strict limitations. Sequence-based methods are limited due to evolutionary divergence, as a protein structure is more likely to remain conserved than its sequence (Chothia & Lesk, 1986), and effective sequence homology requires a minimum of 40% identity over significant portions of the sequence (Devos & Valencia, 2000; Hensen et al., 2012; Park, Teichmann, Hubbard, & Chothia, 1997; Whisstock & Lesk, 2003). Structure-based methods have revealed more direct functional relationships when coupled with sequence data. Unfortunately, structure-based prediction methods are limited due to diverse structures possessing similar functions or similar structures carrying out a diverse range of functions. Structure-based methods alone do not produce reliable functional predictions (reliability ≤ 30%) (Hensen et al., 2012). Hensen et al. (2012) found that the intrinsic dynamics of a protein provide a more accurate depiction of protein function due to the large correlation between protein dynamics and function. They observed that functionally similar proteins tend to exhibit similar dynamic patterns and thereby, protein functional annotations could be accomplished based on dynamics (Hensen et al., 2012). Their work demonstrated that the functional classification of proteins based solely upon protein dynamics has a success rate of 46% within the broad functional classes, which is much higher than the structure-based prediction rate of 32%. Other researchers have also explored protein dynamics-function relationships by performing dynamics-based comparisons of proteins (Micheletti, 2013; Munz, Lynsae, Hein, & Biggin, 2010; Pandini, Mauri, Bordogni, & Bonati, 2007; Tobi, 2013). Here, we have explored the dynamics of enzymes involved in some important metabolic pathways in an attempt to further our understanding of the relationships between protein structure, dynamics, and function.
Earlier, Hensen et al. have performed long timescale (100 ns) atomistic molecular dynamic (MD) simulations and derived 34 dynamic descriptors to characterize dynamics of proteins (Hensen et al., 2012). Picosecond to nanosecond timescale atomic fluctuations can be simulated using all-atom MD simulations, and important information about functional dynamics could be obtained. Unfortunately, it is computationally expensive to investigate functionally significant displacements occurring in microsecond to millisecond timescale using MD simulations. Alternatively, normal mode analysis (NMA) can be utilized to examine protein dynamics at a slower timescale (Bahar & Rader, 2005; Hinsen, 1998; Hinsen, Thomas, & Field, 1999). Both low-frequency vibrations that represent the protein’s collective motions as well as the higher frequency vibrations, which represent local deformations, can be studied by NMA. Earlier studies have shown that low-frequency vibrational modes or normal modes (large fluctuations) of proteins are functionally relevant (Brooks & Karplus, 1983; Go, Noguti, & Nishikawa, 1983; Marques & Sanejouand, 1995). Despite a few limitations such as local fluctuations upon substrate binding being poorly captured by the coarse-grained NMA simulations (Strom, Fehling, Bhattacharyya, & Hati, 2014), NMA is not limited by the system size and can be used to analyze functionally relevant conformational dynamics from protein structure, making it an incredibly powerful method of analysis. Additionally, the NMA method that employs course-grained elastic network model (ENM) offers a dependable yet computationally inexpensive way to evaluate large displacements (conformational changes) of proteins comprised of thousands of residues (Frank & Agrawal, 2000; Strom et al., 2014; Tama & Brooks, 2005; Weimer, Shane, Brunetto, Bhattacharyya, & Hati, 2009).

In the present study, ENM-based coarse-grained NMA was employed to characterize intrinsic dynamics of metabolically relevant enzymes and formulate a protein functional characterization method emphasizing dynamical importance. The dynamics of 24 enzymes involved in primary metabolic pathways were studied and sequence, structure, and dynamic similarities were calculated across six species to determine the degree of conservation of sequence, structure, and dynamics within a protein group. We examined the relationship among these three properties, as well as developed a catalog of dynamic cross-correlation matrices (DCCMs). We then attempted to match a set of proteins to the catalog of DCCMs for their functional identification. The present study demonstrated that each metabolic enzyme exhibits unique patterns of motions, which are conserved across multiple species; proteins with similar function exhibit a similar pattern of motions. Moreover, this work also demonstrated that the coarse-grained NMA is a faster method to develop protein-specific DCCMs, which could be used for functional characterization. Also, it is established from our study that dynamic-based functional characterization is highly predictable. This study could be extended to include all known protein families and generate a database of DCCMs of proteins for functional annotation.

2. Methods

2.1. Protein sequences and structures
Twenty-four proteins from common metabolic pathways, as listed in Table 1, were selected from the Protein Data Bank (PDB) (Berman et al., 2000) based upon structural data availability. We attempted to select proteins from six different species for which crystal structures are known. Also, proteins were screened for similar amino acid chain length. A total of 141 proteins were considered in this study because only 5 crystal structures were available for 3 of the 24 protein families. The symmetry and stoichiometry analyses revealed that all but one of the selected protein families are homopolymers. Therefore, a single polypeptide chain (chain A) was chosen to compare the sequence and structure, as well as intrinsic dynamics between proteins (from different species) within a given family. For the heterodimer protein, succinyl-CoA synthetase, both chains were used. In the present
Table 1. Comprehensive list of the 24 protein groups and the six selected proteins for each group with the organism name, PDB ID, and number of amino acid residues (sequence length) in chain A of the protein

| Protein group                          | Organism                        | PDB ID | Length |
|----------------------------------------|---------------------------------|--------|--------|
| Citrate synthase (citric acid cycle)   | *Vibrio vulnificus*             | 4E6Y   | 426    |
|                                        | *Escherichia coli*              | 4G6B   | 432    |
|                                        | *Acetobacter aceti*             | 2H12   | 436    |
|                                        | *Francisella tularensis*        | 3MSU   | 427    |
|                                        | *Pyrrobaculum aerophilum*       | 2IBP   | 409    |
|                                        | *Salmonella typhimurium*        | 3O8J   | 404    |
| Fumarase (citric acid cycle)           | *Escherichia coli*              | 1YFE   | 467    |
|                                        | *Mycobacterium tuberculosis*    | 4APA   | 474    |
|                                        | *Sinorhizobium meliloti*        | 4HGV   | 495    |
|                                        | *Rickettsia prowazekii*         | 3GTD   | 482    |
|                                        | *Thermus thermophilus*          | 1VDK   | 466    |
|                                        | *Homo sapiens*                  | 3E04   | 490    |
| Isocitrate dehydrogenase (citric acid cycle) | *Ruminiclostridium thermocellum* | 4ADB   | 402    |
|                                        | *Escherichia coli*              | 9ICD   | 416    |
|                                        | *Sulfolobus tokodai*            | 2DHT   | 409    |
|                                        | *Desulfotalea psychrophila*     | 2UXR   | 402    |
|                                        | *Burkholderia pseudomallei*     | 3DMS   | 427    |
|                                        | *Aeropyrum pernix*              | 1V94   | 435    |
| Malate dehydrogenase (citric acid cycle) | *Escherichia coli*              | 1EMD   | 312    |
|                                        | *Thermus thermophilus*          | 2CVQ   | 327    |
|                                        | *Haloarcula marismortui*        | 1O6Z   | 303    |
|                                        | *Chloroflexus aurantiacus*      | 4CL3   | 309    |
|                                        | *Picrophilus torridus*          | 4BGV   | 323    |
|                                        | *Archaeoglobus fulgidus*        | 2X0I   | 294    |
| Succinyl CoA synthetase (citric acid cycle) | *Methanocaldococcus janaschii* | 2YV1   | 294    |
|                                        | *Thermus thermophilus*          | 1O17   | 288    |
|                                        | *Sus scrofa*                    | 2FP4   | 305    |
|                                        | *Escherichia coli*              | 1CQI   | 286    |
|                                        | *Aeropyrum pernix*              | 2YV2   | 297    |
|                                        | *Thermus aquaticus*             | 3UFX   | 296    |
| Alcohol dehydrogenase (alcohol metabolism) | *Mus musculus*                  | 1E3L   | 376    |
|                                        | *Homo sapiens*                  | 1HTB   | 374    |
|                                        | *Equus caballus*                | 6ADH   | 374    |
|                                        | *Thermus sp*                    | 4CPD   | 347    |
|                                        | *Arabidopsis thaliana*          | 4RQT   | 375    |
|                                        | *Thermotoga maritima*           | 3IP1   | 404    |
| Lactate dehydrogenase (lactate metabolism) | *Plasmodium falciparum*         | 1CET   | 316    |
|                                        | *Squalus acanthias*             | 3LDH   | 330    |
|                                        | *Plasmodium falciparum*         | 1T25   | 322    |
|                                        | *Staphylococcus aureus*         | 3D0O   | 317    |
|                                        | *Cryptosporidium parvum*        | 4ND5   | 321    |
|                                        | *Enterococcus mundtii*          | 3WSV   | 322    |
(Continued)
Table 1. (Continued)

| Protein group                                      | Organism                        | PDB ID | Length |
|---------------------------------------------------|---------------------------------|--------|--------|
| Enolase (glycolysis)                              | Homarus gammarus                | 1PDZ   | 434    |
|                                                  | Homo sapiens                    | 3B97   | 433    |
|                                                  | Saccharomyces cerevisiae        | 1ELS   | 436    |
|                                                  | Bacillus subtilis               | 4A3R   | 430    |
|                                                  | Enterococcus hirae              | 1IYX   | 432    |
|                                                  | Trypanosoma brucei              | 2PTX   | 436    |
| Fructose bisphosphate aldolase (glycolysis)       | Encephalitozoon cuniculi       | 3MBF   | 342    |
|                                                  | Thermus caldophilus             | 2FJK   | 305    |
|                                                  | Mycobacterium tuberculosis      | 4LV4   | 349    |
|                                                  | Babesia bovis                   | 3KX6   | 379    |
|                                                  | Toxoplasma gondii               | 4TU1   | 355    |
|                                                  | Oncitologus cuniculus           | 3DFN   | 353    |
| Glyceroldehyde phosphate dehydrogenase (glycolysis)| Geobacillus stearothermophilus  | 1NQO   | 334    |
|                                                  | Streptococcus agalactiae        | 4OQ6   | 356    |
|                                                  | Toxoplasma gondii               | 3STH   | 361    |
|                                                  | Borrelia burgdorferi            | 3HJA   | 356    |
|                                                  | Bacillus anthracis              | 4DIB   | 345    |
|                                                  | Bartonella henselae             | 3LD    | 356    |
| Hexokinase (glycolysis)                           | Schistosoma mansoni             | 1BDG   | 451    |
|                                                  | Arabidopsis thaliana            | 4Q58   | 474    |
|                                                  | Xenopus laevis                  | 3WOL   | 458    |
|                                                  | Homo sapiens                    | 4IWV   | 456    |
|                                                  | Kluyveromyces lactis            | 4JAX   | 485    |
|                                                  | Saccharomyces cerevisiae        | 3BBA   | 485    |
| Phosphoglycerate kinase (glycolysis)              | Escherichia coli                | 12MR   | 387    |
|                                                  | Francisella tularensis          | 4FEY   | 395    |
|                                                  | Thermus thermophilus            | 1V6S   | 390    |
|                                                  | Geobacillus stearothermophilus  | 1PHP   | 394    |
|                                                  | Thermotoga maritima             | 1VPE   | 398    |
| Phosphoglycerate mutase (glycolysis)              | Saccharomyces cerevisiae        | 3PGM   | 244    |
|                                                  | Bacillus anthracis              | 3R7A   | 237    |
|                                                  | Toxoplasma gondii               | 4OQI   | 281    |
|                                                  | Plasmodium falciparum           | 1XQ9   | 258    |
|                                                  | Homo sapiens                    | 1YFK   | 262    |
|                                                  | Saccharomyces cerevisiae        | 1BQ4   | 246    |
| Pyruvate decarboxylase (glycolysis)               | Saccharomyces cerevisiae        | 1PYD   | 556    |
|                                                  | Vibrio parahaemolyticus         | 2QMA   | 497    |
|                                                  | Saccharomyces pastorianus       | 1QPB   | 563    |
|                                                  | Azospirillum brasilense         | 2NXW   | 565    |
|                                                  | Enterobacter cloacae            | 10VM   | 552    |

(Continued)
### Table 1. (Continued)

| Protein group                                      | Organism                        | PDB ID | Length |
|---------------------------------------------------|---------------------------------|--------|--------|
| **Pyruvate kinase (glycolysis)**                   | Homo sapiens                    | 3SRF   | 551    |
|                                                   | Felis catus                     | 1PKM   | 530    |
|                                                   | Oryctolagus cuniculus           | 1AOQF  | 530    |
|                                                   | Toxoplasma gondii               | 3EOE   | 511    |
|                                                   | Cryptosporidium parvum          | 4DRS   | 526    |
|                                                   | Leishmania mexicana             | 3EOV   | 539    |
| **Triose phosphate isomerase (glycolysis)**        | Escherichia coli                | 4IOT   | 255    |
|                                                   | Tenebrio molitor                | 219E   | 259    |
|                                                   | Gallus gallus                   | 8TIM   | 247    |
|                                                   | Saccharomyces cerevisiae        | 1YPI   | 248    |
|                                                   | Homo sapiens                    | 1WYI   | 250    |
|                                                   | Rhizophagus micropus            | 3TH6   | 249    |
| **Phosphoenolpyruvate carboxykinase (gluconeogenesis)** | Thermus thermophilus            | 2PC9   | 529    |
|                                                   | Escherichia coli                | 1OS1   | 540    |
|                                                   | Trypanosoma cruzi               | 1I12   | 524    |
|                                                   | Escherichia coli                | 1AYL   | 541    |
|                                                   | Anaerobiospirillum succiniciproducens | 1YTM   | 532    |
|                                                   | Actinobacillus succinogenes     | 1YGG   | 560    |
| **Arginase (urea cycle)**                          | Thermus thermophilus            | 2EF4   | 290    |
|                                                   | Homo sapiens                    | 4IXU   | 306    |
|                                                   | Leishmania Mexicana             | 4ITY   | 330    |
|                                                   | Rattus norvegicus               | 1ZPE   | 314    |
|                                                   | Bacillus caldovelox             | 1CEV   | 299    |
|                                                   | Deinococcus peraridilator       | XXXX   | 294    |
| **Argininosuccinate synthetase (urea cycle)**      | Campylobacter jejuni            | 4NZP   | 409    |
|                                                   | Thermotoga maritima             | 1VL2   | 421    |
|                                                   | Mycobacterium thermoresistibile| 4XFJ   | 408    |
|                                                   | Escherichia coli                | 1KP2   | 455    |
|                                                   | Homo sapiens                    | 2NZ2   | 413    |
|                                                   | Thermus thermophilus            | 1J1Z   | 400    |
| **Ornithine transcarbamoylase (urea cycle)**       | Pseudomonas Aeruginosa          | 1ORT   | 335    |
|                                                   | Escherichia coli                | 1AKM   | 333    |
|                                                   | Pyrococcus furiosus             | 1A1S   | 314    |
|                                                   | Thermus thermophilus            | 2EFO   | 301    |
|                                                   | Brucella Melitensis             | 4OH7   | 320    |
|                                                   | Gloeobacter violaceus           | 3GD5   | 323    |
| **Amine oxidase (urea cycle, amino acid metabolism)** | Escherichia coli                | 2WO0   | 727    |
|                                                   | Pisum sativum                   | 1KSI   | 642    |
|                                                   | Arthrobacter globiformis        | 1JUV   | 638    |
|                                                   | Ogataea angusta                 | 3SX1   | 692    |
|                                                   | Aspergillus nidulans            | 3PGB   | 797    |
|                                                   | Homo sapiens                    | 3H17   | 731    |

(Continued)
study, structures of proteins were visualized using the Visual Molecular Dynamics (VMD) program (Humphrey, Dalke, & Schulten, 1996).

### 2.2. Sequence comparison

The sequence identity of proteins in each group was examined. These pairwise sequence alignments were carried out with the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1997), which aligns two sequences and provides the percentage of identical residues. The sequence identity is the percentage of identical residues between protein sequences being compared. A mean and standard deviation was calculated for each protein group.

### 2.3. Structure comparison

The pairwise DALI Lite server (Holm & Rosenstrom, 2010) was used to determine the degree of structural similarity between two proteins. Pairwise structural alignments were performed between all proteins of a given group. Root-mean-square deviation (RMSD) was used to address structural (protein backbone) comparison, which represents the variation of a particular locus between two superimposed structures. The mean and standard deviation of the sample were calculated for each protein group.

### 2.4. Single and comparative NMA

In the coarse-grained ENM, developed by Hinsen, Petrescu, Dellerue, Bellissent-Funel, and Kneller (2000), the protein is simplified to a string of beads, where each bead represents a C\textalpha\ atom. The interactions between C\textalpha\ atoms are expressed as a harmonic pair potential as follows:

\[
U_{ij}(r) = \frac{k_{ij}}{2} \left( \left| r_i - r_j \right| - \left| r_i^0 - r_j^0 \right| \right)^2
\]  

where \( r_i \) and \( r_j \) are the positions of residues \( i \) and \( j \) in the current conformation of the protein, and the superscript \( 0 \) denotes the equilibrium conformation; \( k_{ij} \) is the force constant for the spring connecting residues \( i \) and \( j \). The force constant \( k_{ij} \) is distance dependent and expressed as
where $a$, $b$, $c$, and $d$ are determined by fitting the equation to an all-atom model as discussed by Hinsen et al. (2000) and Tiwari et al. (2014). The overall potential of the system, which is the sum of the pair potential over all pairs, is expressed as follows:

$$U(r) = \sum_{i,j} U_{ij}(r)$$

Eigenvectors and eigenvalues of the mass weighted matrix of the second derivative of $U(r)$ correspond to normal modes and the squares of the frequencies for each normal mode, respectively (Tiwari et al., 2014). In the present study, normal mode calculations were carried out using the WEBnm@ server (Fuglebakk, Tiwari, & Reuter, 2014; Hollup, Salensminde, & Reuter, 2005; Tiwari et al., 2014). WEBnm@ uses a molecular modeling toolkit and employs the coarse-grained ENM to calculate the low frequency normal modes. Individual PDB files were submitted to the online WEBnm@ server (Hollup, Fuglebakk, Taylor, & Reuter, 2011; Hollup et al., 2005; Skjaerven, Hollup, & Reuter, 2009; Tiwari et al., 2014) for single NMA analysis. The correlated/anticorrelated motions between residues ($C_\alpha$ atoms) were identified from the DCCMs. The DCCM was calculated from the normal modes using the method described by Ichiye and Karplus (Ichiye & Karplus, 1991). Each element in the DCCM quantifies the coupling between two $C_\alpha$ atoms $i$ and $j$ as:

$$C_{ij} = \frac{\sum_{m=1}^{M} \frac{1}{\tau_m} |X_m|^2 |Y_m|^2 }{ \left( \sum_{m=1}^{M} \frac{1}{\tau_m} |X_m|^2 \right)^{1/2} \left( \sum_{m=1}^{M} \frac{1}{\tau_m} |Y_m|^2 \right)^{1/2} }$$

In Equation (3), $X_m$ and $Y_m$ represent eigenvectors and eigenvalues of the $m$th normal mode, respectively. In the present study, the DCCMs were obtained using the default setting of WEBnm@ i.e. the combined 200 modes were used to generate each DCCM. The DCCMs of the 24 protein families were compared and a catalog of all DCCMs was assembled for identification of unknown (test) proteins. The single WEBnm@ analysis also provides atomic ($C_\alpha$ atoms) fluctuation profiles (vide infra) and vector coordinates files, which allow for the visualization of the protein’s dynamics in different modes. In WEBnm@, modes 1–6 correspond to the rotational and translational motions and are ignored. Therefore, mode 7 actually represents the lowest frequency mode of a protein system.

The comparative NMA was performed for each enzyme to examine the conservation of dynamics across species (Fuglebakk, Echave, & Reuter, 2012). In the present study, dynamics of a given enzyme from five to six different species were compared. Aligned FASTA sequence (See Supplementary Table S1) and corresponding PDB coordinate files of each protein group were used for the comparative NMA (Sievers et al., 2011). From the comparative analysis, the dynamic similarity, in terms of Bhattacharyya coefficient (BC) (Bhattacharyya, 1943), as well as the heat map showing the theoretical evolutionary relatedness between proteins solely based upon their intrinsic dynamics were obtained.

### 2.5. Bhattacharyya coefficient

The BC measures the dynamical similarity between proteins by comparing their covariance matrices, which are obtained from the normal modes of the conserved parts of proteins under consideration (Tiwari et al., 2014). The BC calculations were performed following the method described by Tiwari et al. (2014). The BC ranges from 0 to 1, and represents the amount of overlap between the collective dynamics of the aligned proteins; BC of 1 represents maximum overlap or dynamical similarity between proteins. For a given protein group, the average BC with the sample standard deviation is calculated for comparative purposes.
2.6. The Cα atom fluctuations

The normalized squared Cα atom fluctuations for each protein were calculated as the sum of the displacement of each Cα atom along the lowest modes (Tiwari et al., 2014). The fluctuations are the sum of the atomic (Cα atoms only) displacements in each mode weighted by the inverse of their corresponding eigenvalues. In the present study, the default setting of WEBnm®, which includes the first 200 modes, was used. The x-axis of a fluctuation profile represents residue index in the protein sequence, while the y-axis represents the normalized displacement corresponding to each amino acid. Peaks in the fluctuation profile correspond to flexible regions of proteins.

2.7. Root-mean-squared inner product

For quantitative comparison of atomic (Cα atoms) fluctuations between proteins, the RMSIP was computed for the lowest normal modes using the following equation as described in Amadei, Ceruso, and Di Nola (1999), Carnevale, Pontiggia, and Micheletti (2007), Fuglebakk et al. (2014), Tiwari et al. (2014) and Yang, Song, Carriquiry, and Jernigan (2008):

Table 2. The number of species studied (N), the sequence identity, structural similarity, and dynamic similarity (in terms of BC) of the 24 protein groups studied

| Proteins                                  | N  | Sequence identity | Structural similarity (RMSD, Å) | Dynamic similarity (BC) |
|-------------------------------------------|----|-------------------|----------------------------------|--------------------------|
| Citrate synthase                          | 6  | 45 ± 14           | 2.4 ± 0.4                        | 0.81 ± 0.04              |
| Fumarase                                  | 6  | 56 ± 5            | 1.6 ± 0.4                        | 0.87 ± 0.02              |
| Isocitrate dehydrogenase                  | 6  | 38 ± 17           | 2.9 ± 1.0                        | 0.78 ± 0.06              |
| Malate dehydrogenase                     | 6  | 31 ± 6            | 2.5 ± 0.7                        | 0.82 ± 0.03              |
| Succinyl-CoA synthetase                   | 6  | 59 ± 11           | 1.1 ± 0.3                        | 0.89 ± 0.02              |
| Alcohol dehydrogenase                    | 6  | 39 ± 17           | 2.1 ± 0.7                        | 0.82 ± 0.07              |
| Lactate dehydrogenase                    | 6  | 32 ± 6            | 1.8 ± 0.3                        | 0.81 ± 0.03              |
| Enolase                                   | 6  | 58 ± 7            | 1.3 ± 0.3                        | 0.89 ± 0.03              |
| Fructose bisphosphate aldolase           | 6  | 42 ± 14           | 2.7 ± 1.0                        | 0.77 ± 0.09              |
| Glyceraldehyde phosphate dehydrogenase   | 6  | 51 ± 5            | 1.1 ± 0.1                        | 0.88 ± 0.01              |
| Hexokinase                               | 6  | 42 ± 14           | 3.0 ± 1.0                        | 0.82 ± 0.04              |
| Phosphoglycerate kinase                   | 5  | 51 ± 8            | 2.5 ± 1.2                        | 0.89 ± 0.02              |
| Phosphoglycerate mutase                   | 6  | 51 ± 17           | 1.8 ± 0.7                        | 0.83 ± 0.08              |
| Pyruvate decarboxylase                    | 6  | 36 ± 8            | 3.1 ± 1.0                        | 0.75 ± 0.09              |
| Pyruvate kinase                           | 6  | 56 ± 21           | 1.9 ± 0.7                        | 0.87 ± 0.03              |
| Triose phosphate isomerase                | 6  | 59 ± 13           | 1.0 ± 0.2                        | 0.88 ± 0.03              |
| PEP carboxylase                           | 6  | 54 ± 13           | 1.8 ± 0.5                        | 0.87 ± 0.03              |
| Arginase                                  | 5  | 42 ± 8            | 1.5 ± 0.3                        | 0.78 ± 0.08              |
| Arginosuccinate synthetase                | 6  | 42 ± 11           | 1.7 ± 0.4                        | 0.78 ± 0.03              |
| Ornithine transcarbamoylase               | 6  | 44 ± 7            | 1.6 ± 0.3                        | 0.84 ± 0.03              |
| Amine oxidase                             | 5  | 24 ± 4            | 2.6 ± 0.6                        | 0.73 ± 0.05              |
| Methionine adenosyltransferase            | 6  | 61 ± 11           | 1.3 ± 0.3                        | 0.86 ± 0.02              |
| Pantathernate synthetase                  | 6  | 46 ± 7            | 2.8 ± 1.1                        | 0.84 ± 0.02              |
| Alkaline phosphatase                      | 6  | 36 ± 12           | 2.2 ± 0.8                        | 0.80 ± 0.05              |
where $X_i$ and $Y_j$ represent the eigenvectors of a pair of proteins being compared and $i$ and $j$ represent the mode numbers. In the present study, the default setting of WEBnm®, which includes the first 10 lowest energy nontrivial modes (Amadei et al., 1999; Tiwari et al., 2014), was used. The RMSIP values range from 0 to 1; RMSIP of 1 represents maximum similarity in $C_\alpha$ atom fluctuations between proteins being compared.

2.8. Identification of test proteins

PDB coordinate files of 12 separate proteins, which either belong to one of the 24 protein groups included in the present study or were randomly chosen proteins, were selected from the PDB. The coordinate files of these proteins were modified to remove any identifying evidence, leaving only the Cartesian coordinates of each atom in the PDB file to run a single NMA analysis. The DCCMs of these test proteins were then compared to the catalog of DCCMs (obtained from the NMA of 24 protein groups) in an attempt to correctly match a test protein to a known protein group for functional identification. Wherever possible, quantitative analyses were performed using the comparative NMA analysis option of WEBnm®. The comparative analysis provides the extent of dynamic similarities between proteins in terms of BC and RMSIP values.

\begin{equation}
\text{RMSIP} = \left( \frac{1}{n} \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} (X_iY_j)^2 \right] \right)^{\frac{1}{2}}
\end{equation}
3. Results and discussion

3.1. Protein sequences and structures
A total of 141 protein sequences and structures were collected from the PDB files. The proteins are from metabolic pathways including glycolysis, gluconeogenesis, fermentation, the citric acid cycle, and the urea cycle (Table 1). These 24 protein groups along with the name of organisms, PDB codes, and the number of amino acids in polypeptide chain A are listed in Table 1. Each protein group has six different organisms associated with it, with three exceptions; arginase, phosphoglycerate kinase, and ornithine transcarbamoylase, which only have five structures. Each protein group possessed a similar number of residues per chain.
Table 3. A list of proteins with PDB codes, DCCM of the protein represented in bold, and the group BC heat map (Continued)

| Protein                      | DCCM | Heat Map |
|------------------------------|------|----------|
| Citrate synthase             | ![DCCM Image](image1) | ![Heat Map Image](image2) |
| 2H12, 2IBP, EMSU, 3O8I, 4E6Y, 4G6B |      |          |
| Fumarase                     | ![DCCM Image](image3) | ![Heat Map Image](image4) |
| 1YFE, 3E04, 4APA, 3GTD, 4HGV, 1VDK |      |          |
| Isocitrate dehydrogenase     | ![DCCM Image](image5) | ![Heat Map Image](image6) |
| 4AOU, 9ICD, 2DIHT, 2UXR, 3DMS, 1V94 |      |          |
| Malate dehydrogenase         | ![DCCM Image](image7) | ![Heat Map Image](image8) |
| 2X01, 1O6Z, 2CVQ, 4BGV, 4CL3, 1EMD |      |          |
### Table 3. (Continued)

| Protein                      | DCCM                      | Heat Map                      |
|------------------------------|---------------------------|-------------------------------|
| Succinyl-CoA synthetase      |                           |                               |
| 1CQI, 2YV1, 1O17, 2FP4, 2YV1, 2YV2, 3UFX | |                               |
| Alcohol dehydrogenase        |                           |                               |
| 1E3L, 1HTB, 6ADH, 4CPD, 4RQT, 3IP1 | |                               |
| Lactate dehydrogenase        |                           |                               |
| 1CET, 3LDH, 1T25, 3D00O, 2ND5, 3WSV | |                               |
| Enolase                      |                           |                               |
| 1ELS, 1PDZ, 1IYX, 3B97, 4A3R | |                               |

(Continued)
| Protein                                      | DCCM | Heat Map |
|---------------------------------------------|------|----------|
| Fructose-bisphosphate aldolase              |      |          |
| 3MBF, 2FJK, 4LV4, 3KX6, 4TU1, 3DPN          |      |          |
| Glyceraldehyde phosphate dehydrogenase      |      |          |
| 1NQO, 4QX6, 3STH, 3HJA, 4DIB, 3L0D          |      |          |
| Hexokinase                                  |      |          |
| 1BDG, 3B8A, 3W0L, 4JAX, 4QS8, 4IWV         |      |          |
| Phosphoglycerate kinase                     |      |          |
| 1PHP, 1V6S, 1VPE, 1ZMR, 3UWD, 4FEY          |      |          |

(Continued)
Table 3. (Continued)

| Protein                                      | DCCM | Heat Map |
|----------------------------------------------|------|----------|
| Phosphoglycerate mutase                      |      |          |
| 3PGM, 3R7A, 4ODI, 1XQ9, 1YFK, 1BQ4           |      |          |
| Pyruvate decarboxylase                       |      |          |
| 1PYD, 2QMA, 1QPB, 2NXW, 10VM                 |      |          |
| Pyruvate kinase                              |      |          |
| 1AQF, 1PKM, 3E0V, 3EOB, 3SRF, 4DRS          |      |          |
| Triose phosphate isomerase                   |      |          |
| 1WYI, 1YPI, 2I9E, 3TH6, 4IOT, 8TIM           |      |          |

(Continued)
Table 3. (Continued)

| Protein                                | DCCM       | Heat Map   |
|----------------------------------------|------------|------------|
| PEP carboxykinase                      | ![Image]   | ![Image]   |
| 2PC9, 1OS1, 1II2, 1AYL, 1YTM, 1YGG     | ![Image]   | ![Image]   |
| Arginase                               | ![Image]   | ![Image]   |
| 1CEV, 1ZPE, 2EF4, 4ITY, 4IXU, XXXX     | ![Image]   | ![Image]   |
| Argininosuccinate synthetase            | ![Image]   | ![Image]   |
| 4NZP, 1VL2, 4XPI, 1KP2, 2NZ2, 1J1Z     | ![Image]   | ![Image]   |
| Ornithine transcarbamoylase             | ![Image]   | ![Image]   |
| 1ORT, 1AKM, 1A1S, 2EF0, 3GD5, 4OH7*    | ![Image]   | ![Image]   |

(Continued)
3.2. Sequence comparison

The sequence identities between proteins of a given family are computed and listed in Table 2. The sequence identities range from about 24 to 61% with an average of 46% among all protein groups. Nine of the protein families had sequence identities greater than 50%, which means that the individual protein structures have at least 90% of their residues within the common cores (Chothia & Lesk, 1986). For the other 15 protein families with sequence identities between 20 and 50% are found to have 42–98% of individual protein residues within the common cores (Chothia & Lesk, 1986).

### Table 3. (Continued)

| Protein                          | DCCM  | Heat Map |
|----------------------------------|-------|----------|
| Amine oxidase                    | ![DCCM](image) | ![Heat Map](image) |
| 1IVU, 1KS1, 2WO0, 3HI7, 3PGB, 3SX1 |       |          |
| Methionine adenosyl-transferase  | ![DCCM](image) | ![Heat Map](image) |
| 2OBV, 1FUG, 1QM4, 3IML, 3S82, 3SO4 |       |          |
| Pantothenate synthetase          | ![DCCM](image) | ![Heat Map](image) |
| 2EJC, 2A86, 1UFV, 1IHO, 3AG5, 3Q10 |       |          |
| Alkaline phosphatase             | ![DCCM](image) | ![Heat Map](image) |
| 1EW2, 1SHN, 2X98, 3BDF, 3E2D, 4KJG |       |          |

Note: Chain A of all enzymes were used for generating DCCMs and heat maps, except for succinyl-CoA synthetase, which has both chain A and chain B.
3.3. Structure comparison

The RMSD values that represent structural similarities between proteins within an enzyme family are also shown in Table 2. The RMSD represents the deviation between two superimposed structures. The observed RMSD values ranged from 1.0 to 3.1 Å, with the lower numbers representing more similar structures. The RMSD is usually small (<3 Å) for homologous proteins (Chothia & Lesk, 1986; Kosloff & Kolodny, 2008). The RMSD between homologous proteins within an enzyme family studied in this work were observed to be ≤3 Å, which suggested that these homologous proteins possess considerable extent of structural similarities. For higher sequence identities, we expected to see a low RMSD for the protein group, which was satisfied (Table 2).

3.4. The Cα atom fluctuations

The fluctuation profiles for Cα atoms of all 24 groups of proteins were obtained from the NMA study. The calculated normalized squared fluctuations of Cα atoms matched satisfactorily with the experimental results for these enzymes. For example, the structure of the *Escherichia coli* methionine adenosyltransferase (MAT) (Figure 1) revealed a dynamic loop region shown in red (Fu, Hu, Markham, & Takusagawa, 1996). Both the loop region (residues 100 to 112 for *E. coli* MAT) and the adjacent residues displayed high flexibility as observed in the normalized fluctuation (Cα atoms) profile (Figure 2). This observation confirmed that the experimentally observed flexible regions of MAT correlate well.

Figure 4. A dynamic cross-correlation matrix for *E. coli* Methionine Adenosyltransferase (1FUG) showing the correlated (red) and anticorrelated (blue) motions between Cα atoms.

Note: The black oval on DCCM represents the anti-correlated motions between the active site (residues 247–258) and the flexible loop (residues 101–109).

Figure 5. Methionine Adenosyltransferase heat map of BC showing the dynamic relatedness between proteins, with the color key at the bottom.

Notes: Red describes identical dynamics, and blue describes the most unrelated dynamics within the group. Cluster data is shown on the x- and y-axis, grouping 1FUG, 3IML, 3S82 together with 3IML being more similar to 1FUG than to 3S82. Similarly, 2P02, 3SO4, and 1QM4 form another cluster, with 1QM4 and 3SO4 being more related to each other than to 2P02.
with peaks in the fluctuation profile. Also, high flexibility was noticed from the active site motions, which can be seen around residues 232 to 254.

3.5. Correlated and anticorrelated motions

The single NMA for each protein within a particular group revealed strikingly similar correlation matrices. Slight differences were seen with respect to the extent of correlated and anticorrelated motions; however the overall patterns remained consistent. As an example, Figure 3 shows the six DCCM files for the MAT group. All six protein matrices appear identical, as was the case with all other protein groups. Therefore, only one representative DCCM was chosen for each enzyme group and listed in Table 3. Because DCCMs of homomultimeric proteins revealed that each polypeptide chain displays identical patterns of motions (Figure S1), only one polypeptide chain of multimeric proteins was used to represent the intrinsic dynamic patterns (DCCMs) of all protein families, except for succinyl-CoA synthetase. For succinyl-CoA synthetase, a heteromultimeric protein, both polypeptide

Figure 6. A list of proteins with corresponding PDB code and their structural visualization, the functionally relevant Cα atom displacement graph, and DCCM.

Notes: Chain A and chain B are shown in blue and red color, respectively. The relevant structural elements are shown in different colors. The substrate or functionally relevant active site residues are shown in green van der Waals space filling model. The dashed lines differentiate between two polypeptide chains. Red and blue boxes correlate to specific textual references.

| Protein            | Normalized squared Cα displacement | DCCM                      |
|--------------------|------------------------------------|---------------------------|
| Citrate synthase   |                                     |                           |
| 2H1Z (Chains A-B)  |                                    |                           |
| Chain A            | Residues 274 to 364                 |                           |
| Chain B            | Residues 274 to 364                 |                           |
| Fumarase           |                                     |                           |
| 4ADL (Chains A-B)  |                                    |                           |
| Chain A            | Residues 403 to 460                 |                           |
| Chain B            | Residues 403 to 460                 |                           |
| Isocitrate dehydrogenase |                               |                           |
| 2DHT (Chains A-B)  |                                    |                           |
| Chain A            | Residues 185 to 193                 |                           |
| Chain B            | Residues 185 to 193                 |                           |
| Fumarase           |                                     |                           |
| 4ADL (Chains C-D)  |                                    |                           |
| Chain A            | Residues 315 to 325                 |                           |
| Chain B            | Residues 315 to 325                 |                           |
| Arginine residues  | 115, 125, 149 in binding pocket     |                           |
| Citrate synthase   |                                     |                           |
| 2H1Z (Chains A-B)  |                                    |                           |
| Chain A            | Residues 200 to 300                 |                           |
| Chain B            | Residues 200 to 300                 |                           |
| Fumarase           |                                     |                           |
| 4ADL (Chains A-B)  |                                    |                           |
| Chain A            | Residues 1 to 100                   |                           |
| Chain B            | Residues 200 to 300                 |                           |
| Arginine residues  | 115, 125, 149 in binding pocket     |                           |

Notes: Chain A and chain B are shown in blue and red color, respectively. The relevant structural elements are shown in different colors. The substrate or functionally relevant active site residues are shown in green van der Waals space filling model. The dashed lines differentiate between two polypeptide chains. Red and blue boxes correlate to specific textual references.
Figure 6. (Continued).
Figure 6. (Continued).

Protein structure

Glyceraldehyde-3-phosphate dehydrogenase
4QX6 (Chains A, B)

Hexokinase
1HDG (chain A)

Fructose 1,6-bisphosphate aldolase
2FK (Chains A-B)

Normalized squared Cα displacement

DCCM

Mode 7 – Chains A-D

Residue index

Residue index

Residue index in sequence

Residue index in sequence

Residue index in sequence

Residue index in sequence

Residue index in sequence
Figure 6. (Continued).
Figure 6. (Continued).

Trisphosphate isomerase
8TIM (Chains A-B)

PEP carboxykinase
1OS1 (Chain A)

Arginosuccinate synthetase
1HZ (Chains A-D)

Ornithine transcarbamylase
4OH7 (Chains A-B)

Pantothenate synthetase
2A86 (Chains A-B)

Protein structure

Normalized squared $C_\alpha$ displacement

DCCM

Mode 11 – Chains A-B

Mode 10 – Chain A

Residue index in sequence

Residue index

Residue index

Residue index

Residue index
chains were used to represent the intrinsic dynamic patterns (Table 3). However, to examine the entire dynamical nature of a protein system and the functional relevance of those intrinsic dynamics, the other chains were also taken into account. For example, MAT (Figure 1) functions as a homodimer and as shown in Figure 1, the flexible loop appears to gate the active site. The majority of the active site consists of residues 247 to 258, and the flexible loop consists of residues 101–109. Interestingly, the flexible loop shows anticorrelated motion with the active site (Figure 4). This anticorrelated motion suggests that the loop would open, which will allow the substrate to have proper access to the active site. The anticorrelated motion between active site pocket and the loop region also suggests that the two subunits move apart to make room for substrate molecules to enter active sites. These motions are seen across all species, suggesting that these functional dynamics have been conserved throughout MAT evolution.

3.6. Bhattacharyya coefficients
The average BC of the 24 protein groups remained remarkably consistent (Table 2), with the lowest score of 0.73 and the highest score of 0.89, where BC of 1.0 represents the maximum similarity in dynamics of proteins used for comparison. It appears that even for the lower sequence identities and similarities, the dynamic similarity of each protein group remains relatively high. Similar observations were made in the case of the cytochrome P450 family of enzymes, where these enzymes share a strong dynamic similarity (BC > 80%) despite low sequence identity (<25%) (Dorner et al., 2015).

In the present study, the comparison of the dynamic similarities among proteins within a group was also made. Table 3 provides a representative heat map of the BCs for each enzyme. Heat map data describe the pairwise comparison of protein structures using the Bhattacharyya coefficients, with red shading representing identical dynamics between proteins being compared, and blue shading showing the least similar intrinsic dynamics. The heat map also groups the individual proteins into clusters, based upon the relatedness of enzyme dynamics. This provides additional information to enable the investigation of protein evolution with regard to protein dynamics. For example, the heat map for the BC of MAT is shown in Figure 5 (the sequence alignment file used for dynamic comparison is shown in Supplementary Table S2). The cluster data is shown on the x- and y-axis, grouping the dynamics of E. coli (1FUG), B. pseudomallei (3IML), and M. avium (3S82) MATs together. The dynamics of 3IML are more similar to the dynamics of 1FUG, than to the dynamics of 3S82. The dynamics of 2P02 (H. sapiens), 3SO4 (E. hystolytica), and 1QM4 (R. norvegicus) form another cluster, with 1QM4 and 3SO4 being more related to each other than to 2P02.

3.7. Functional relevance of the existing intrinsic dynamics
Close analysis of the intrinsic dynamics of the metabolic enzymes studied here revealed some functionally relevant motions, which are described below. For this study, depending upon the protein system, single or multiple polypeptide chains were considered to analyze the functional dynamics. In particular, an effort has been made to consider as many polypeptide chains as possible to maintain the dynamic integrity of the given protein system and show their functional relevance. Also, for simplicity, the Cα atom displacement graph of a specific mode(s) that was functionally more relevant has been considered here. The online NMA server failed to calculate and analyze the normal modes of five enzyme families when the multimeric proteins, instead of single polypeptide chains, were used. Therefore, intrinsic dynamics and their functional relevance of 19 out of 24 enzyme systems were analyzed here. Information regarding the general enzymatic properties and catalytic roles of various enzymes studied here were taken from Biochemistry 7th Edition (Berg, Tymoczko, & Stryer, 2012).

Citrate synthase (PDB code: 2H12; polypeptide chains A and B of the tetrameric protein were visualized for structural and dynamical analyses) is responsible for condensing acetyl-CoA and oxaloacetate to form citrate. To accomplish this, the enzyme undergoes an “open” to “closed” conformational change (Wiegal & Remington, 1986). First, the substrate, oxaloacetate, enters the enzyme in its “open” conformation; the coenzyme, acetyl-CoA, further interacts with nearby residues in the “closed” conformation. These residues are located too far to facilitate the reaction while in the
Figure 7. The six test proteins’ DCCMs are shown along with their matched protein representative DCCM.

| Test Protein | Actual Identity (PDB) |
|--------------|-----------------------|
| Test 1       | Phosphoglycerate kinase (3OZA) |
| Test 2       | Pyruvate Kinase (1E0U) |
| Test 3       | Enolase (5ENL)         |
| Test 4       | Arginase (4Q3Q)        |

Notes: Each test protein was correctly identified using the created catalog of DCCMs obtained from the single NMA calculations. The x- and y-axis represent the residue index of the protein.
The enzyme accomplishes this by having residues 274 to 364, in conjunction with surrounding residues, exhibit an anticorrelated motion with greater flexibility (boxes in the fluctuations plot and DCCM, Figure 6) to the active site and residues 423 to 436, as it opens and closes around the active site of citrate synthase. Residues 423 to 436 on both subunits are shown to interact with the opposing subunit, suggesting that these residues assist in conformational changes of each other, as well as facilitate substrate binding and catalysis.

**Fumarase** (PDB code: 4ADL, a tetrameric protein; in the present study chains A and B and chains C and D are visualized separately for structural and dynamical analyses) catalyzes the reversible reaction of fumarate to malate. This reaction occurs with a conformational change after the substrate enters the active site, which promotes the enzyme to transition into its “closed” conformation (Stoddard, Dean, & Koshland, 1993). For fumarase to carry out its function, it exists as a tetramer with two active sites that interact with its substrates. However, only three of the four chains interact with each substrate (chains A, B, C and chains A, B, D interact with each substrate). For chains A and B, interactions with the substrate occur between residues 185 to 193 and 403 to 460, along with surrounding residues (red boxes in the displacement graph; Figure 6). These two groups of residues, 185 to 193 and 403 to 460, are also highly flexible and anticorrelated to each other, allowing for the greatest range of motion, so that the substrate can both easily diffuse in or out, in addition to binding tightly to the enzyme (boxes in DCCM, Figure 6). Chains C and D only interact with one of the two substrates through residues 315 to 325, which are shown in the Cα_{atom} displacement graph as having the greatest movement (blue boxes in displacement graph; Figure 6).

**Isocitrate dehydrogenase** (PDB code: 2DHT, chains A and B was visualized for structural and dynamical analyses) is a homodimeric protein that is responsible for catalyzing the oxidative...
decarboxylation of isocitrate. The substrate interacts with the active site of isocitrate dehydroge-
nase via the three arginine residues, R115, R125 and R149 as well as five other residues within the
active site (arginine residues shown in green space filled spheres, Figure 6) (Fedøy, Yang, Martinez,
Leiros, & Steen, 2007). The two active sites of isocitrate dehydrogenase are on opposite sides of each
other, meaning that when one active site is in the “open” state, the other is in its “closed” state
conformation (Figure 6). The Cα atom displacement graph of isocitrate dehydrogenase shows gen-
eral fluctuations throughout the whole enzyme, which is consistent to the overall motion. The active
site for isocitrate dehydrogenase is unique in the fact that it is located between the two subunits.
The mechanism for catalysis then relies on the concerted movement of both subunits. This is re-
flected by residues 150 to 196 on both subunits, that interact with each other, acting as a hinge site
for both subunits to pivot around. Residues 1 to 100 (red box in displacement graph; Figure 6) on
either chain interacts with the residues 200 to 300 (blue box in displacement graph; Figure 6) on the
opposing chains, portraying a pincer-type mechanism around the active site by moving in an anti-
correlated motion to each other (DCCM, Figure 6).

Malate dehydrogenase (PDB code: 4CL3; chain A of the tetrameric protein was visualized) has the
role of converting malate to oxaloacetate. To do so, the enzyme is comprised of two or four chains,
forming a homodimer or homotetramer complex that moves anticorrelated to each other. Malate
dehydrogenase has a highly flexible loop between residues 80 and 90 (Figure 6) that opens and
closes around the active site, allowing entry and exit of reactants and products as well as excludes
surrounding solvents during the reaction. This flexible loop region is highly conserved within malate
dehydrogenases, emphasizing its importance in catalysis (Goward & Nicholls, 1994). The overall ob-
served anticorrelated movement within a chain appeared to be quite weak because the chain re-
volves in a circular pattern, making the cross-correlation map appear whiter (DCCM, Figure 6). There
is a weak anticorrelated motion between residues 1 to 75 and residues 80 to 90, since they are on
opposing sides due to the enzyme’s helical movement (boxes in DCCM, Figure 6).

Succinyl-CoA synthetase (PDB code: 1CQI; visualized chains A and B only for structural and dy-
namical analyses) is a heterotetramer made up of two identical heterodimers. The role of the en-
zyme is to convert succinyl-CoA to succinate. This reaction is carried out through two different
binding pockets, one on each distinct subunit within a heterodimer (Wolodko, Fraser, James, &
Bridger 1994). Chain A contains the succinyl-CoA binding pockets, while chain B contains the ATP/
GTP binding pockets. For succinate to form, ADP needs to be present in the binding pocket of chain
A. The enzyme accommodates this by having chain A and chain B rotate in an anticorrelated motion
to each other (DCCM, Figure 6), allowing ATP to become closer to the chain A, which is responsible for
dephosphorylating ATP to ADP and the reaction to occur.

Alcohol dehydrogenase (PDB code: 6ADH; both subunits of the dimeric protein were visualized for
structural and dynamical analyses) is a homodimeric protein with one active site on each identical
subunit. Alcohol dehydrogenase undergoes a conformational change, not in the presence of alco-
hols, but rather the coenzyme NAD (Eklund et al., 1976). Within each subunit, there is a highly flexi-
ble loop between residues 240 and 250 that interacts with the active site of the enzyme, indicative
of a lid-type mechanism (boxes in Cα atom displacement graph, Figure 6). The loop and surrounding
residues move in an anticorrelated motion (DCCM, Figure 6). This anticorrelated motion results in a
wide opening that allows substrate binding and catalysis.

Enolase (PDB code: 1IYX; both subunits of the dimeric protein were visualized for structural and
dynamical analyses) is an enzyme that is part of the glycolysis pathway, converting 2-phosphoglyc-
erate to phosphoenolpyruvate. The mechanism for this reaction is rather simple. Enolase has a very
generic opening and closing motion, using flexible loops to interact with the active site (Lebioda &
Stec, 1991). The first set of residues, residues 1 to 140, move in an anticorrelated motion with the
other residues, residues 141 to 300 (boxes in DCCM, Figure 6). Residues 250 to 255 have the greatest
flexibility (boxes in Cα atom displacement graph, Figure 6), suggesting a lid-type mechanism
protecting the substrate in the active site. This simple motion is needed for fast turnover rates in the glycolysis pathway to supply the organism with energy.

Fructose 1–6 bisphosphate aldolase (PDB code: 2FJK; visualized chains A and B of the tetrameric protein) catalyzes the reversible step in glycolysis that converts fructose 1,6-bisphosphate, which is a six carbon molecule, into dihydroxyacetone phosphate and glyceraldehyde 3-phosphate, which are both three carbon molecules. Fructose 1–6 bisphosphate aldolase is a homotetramer with four identical active sites, one within each monomer. The monomers of the enzyme shift between an “open” and “closed” conformation (Hester et al., 1991). Within the tetramer, chains A and B work in concert and chains C and D mirror the motions of chains A and B. Chains A, B and chains C, D have similar helix-turn-helix residues (residues 258 to 289) where they overlap with each other and with the active sites on opposing sides. This allows the overlapping helix-turn helix residues to assist each other in facilitating the conformational change from “open” to “closed”. For example, when chain A opens, its helix-turn-helix moves out of the way and allows the helix-turn-helix of chain B to close, and vice versa. The overall circular motion of the enzyme takes place because residues 135 and 150 (red boxes in \( C_\alpha \) atom displacement graph, Figure 6) rotate with a greater range of motion than the rest of the enzyme, allowing them to act as a lid over the active site. The motion of this lid, which is constituted of residues 135 and 150, is important for substrate binding and catalysis.

Glyceraldehyde 3-phosphate dehydrogenase (PDB code: 4QX6) is an important enzyme necessary for glycolysis. This enzyme converts glyceraldehyde 3-phosphate into 1,3-bisphosphoglycerate. Glyceraldehyde phosphate dehydrogenase is a homotetramer made up of 4 identical subunits, each with their own active sites. Although they are identical subunits, each chain interacts differently with each other. Interactions between chains A and B are mirrored by chains C and D, while interactions between chains A and C are mirrored by chains B and D. The openings of the active site pockets on chain A and chain B overlap, so the opening and closing of both chains occur simultaneously. Additionally each chain has a loop within residues 299 to 307 that interacts with the same residues in another chain (chains A, C and chains B, D), that further facilitates this simultaneous open and closing motion required for substrate binding (Lin et al., 1990).

Hexokinase (PDB code: 1BDG; the monomeric protein is visualized) is the first enzyme in the glycolysis pathway and also catalyzes the rate limiting reaction for glycolysis. The enzyme is responsible for converting glucose to glucose-6-phosphate, using one molecule of ATP in the process. The enzyme is a monomer with one active site where the enzyme opens and closes with a hinge like motion around the substrate. The flexibility of the enzyme allows the residues 50 to 260 (red boxes in \( C_\alpha \) atom displacement graph, Figure 6) to move anticorrelated with the rest of the enzyme to ensure a tight “closed” conformation and the loose “open” conformation (boxes in DCCM, Figure 6). This flexibility assists in the mechanism of the enzyme to tightly bind the substrate and prevent wasteful hydrolysis of ATP (Bennett & Steitz, 1978).

Phosphoglycerate kinase (PDB code: 1PHP) is a homodimer protein that is responsible for catalyzing the transfer of phosphate group from 1,3-biphosphoglycerate to ADP in order to form 3-phosphoglycerate and ATP. The two substrates bind in two separate pockets of phosphoglycerate kinase. The ADP binds at the site formed by residues 181 to 384, whereas the second binding pocket is formed by residues 1 to 180, which binds 1,3-biphosphoglycerate. After both substrates are bound, it allows the enzyme to undergo a conformational change in an anticorrelated rotating motion that brings the two substrate binding sites close enough for the catalysis to take place, as well as to exclude water molecules to prevent unwanted ATP hydrolysis (Banks et al., 1979).

Phosphoglycerate mutase (PDB code: 1YFK; visualized chains A and B of dimer for structural and dynamical analyses) is involved in glycolysis and is responsible for catalyzing the transfer of a phosphate group from \( C_3 \) to \( C_2 \) to form 2-phosphoglycerate from 3-phosphoglycerate. Phosphoglycerate mutase is a homodimer with two identical active sites within each subunit that allows phosphate and 3-phosphoglyceric acid to bind (Winn Watson, Harkins, & Fothergill 1981). The enzyme catalyzes
the transfer of a phosphate group from a specific histidine residue, to the C\textsubscript{1} carbon, taking the phosphate from the C\textsubscript{3} carbon of the 3-phosphoglycerate, through a formation of a 2, 3-bisphosphoglycerate intermediate. At the active site of phosphoglycerate mutase, is a very flexible loop (residues 100 to 125) that moves in an anticorrelated motion to the rest of the active site, acting like a lid (boxes in C\textalpha atom displacement graph, Figure 6). The anticorrelated movement of the lid allows the entering of substrate and exiting of product from the active site.

Pyruvate decarboxylase (PDB code: 1QPB; chains A and B of tetrarimers were visualized for structural and dynamical analyses) catalyzes the last step in glycolysis where phosphoenolpyruvate is converted into pyruvate. Pyruvate decarboxylase is a homotetramer with identical active sites within each subunit (Lu, Dobritzsch, Baumann, Schneider, & König, 2000). For catalysis, a pair of subunits, A and B functions together as residues 1 to 194 and 351 to 560 of opposing chains anchor to each other. This allows residues 195 to 350 to have greater movement and flexibility since the rest of the enzyme is stationary. Residues 195 to 350 then move anticorrelated to the other residues allowing it to open and close around the active site of the enzyme. This suggests that residues 195 to 350 can act as a lid mechanism to the active site (boxes in the C\textalpha atom displacements, Figure 6).

Pyruvate kinase (PDB code: 3SRF; chains A and B of the tetrarimers were visualized for structural and dynamical analyses) catalyzes the final step in glycolysis. The enzyme has identical active sites within each subunit of its homotetramer, where it catalyzes the conversion of phosphoenolpyruvate into pyruvate. A lid-type mechanism is employed by each subunit to allow for the catalysis to occur (Mattevi, Bolognesi, & Valentini, 1996). Residues 117 to 217 (boxes in C\textalpha atom displacement graph, Figure 6) are highly flexible as they move from an “open” to “closed” conformation relative to the active site, as shown in the box in the C\textalpha atom displacement graph. The DCCM (see regions in boxes) further shows that residues 117 to 217 are involved in an anticorrelated motion with respect to the rest of the enzyme to have the greatest movement during its conformational changes.

Triose phosphate isomerase (PDB code: 8TIM; both chains of the homodimer were visualized) catalyzes the reversible interconversion of different triose phosphate isomers. Triose phosphate isomerase exists as a homodimer with identical active sites within each subunit. The mechanism of catalysis is indicative of a pincer-type mechanism, with loops opening and closing over the active site (Davenport et al., 1991). Residues 170 to 220 of the enzyme open and close around the active site, and move in an anticorrelated motion to the rest of the enzyme (boxes in the DCCM, Figure 6). This anticorrelated motion alone does not capture the full catalytic mechanism, as the displacement graph shows a general flexibility exists throughout the whole enzyme to assist in catalysis.

Phosphoenolpyruvate carboxykinase (PDB code: 1OS1; only chain A was visualized) catalyzes the reaction from oxaloacetate to phosphoenolpyruvate (PEP) and CO\textsubscript{2}. PEP carboxykinase does this by employing an induced-fit mechanism. The enzyme opens and closes around the substrates, oxaloacetate and ATP, taking on a different conformation in the presence of ATP (Dunten et al., 2002). In doing so, the enzyme is able to exclude water molecules that could potentially hydrolyze ATP. This mechanism also allows residues on the enzyme that could not originally reach the active site to interact with the substrates and favor catalysis. Residues 442 to 510 (box in displacement graph, Figure 6) and surrounding residues move anticorrelated with the active site and the rest of the enzyme, as shown in the DCCM (Figure 6).

Argininosuccinate synthetase (PDB code: 1J1Z; all four chains of the tetrameric protein were visualized for structural and dynamical analyses) catalyzes the synthesis of argininosuccinate, which is a necessary precursor for arginine and fumarate. Argininosuccinate synthetase is a homotetramer with identical active sites within each subunit. The motion of the enzyme is indicative of a lid-type mechanism with “open” and “closed” conformations, depending on substrate presence around the active site (Lemke & Howell, 2001). Residues 1 to 165 show a greater range of flexibility (boxes in C\textalpha atom displacement graph, Figure 6) as it moves in an anticorrelated motion to the rest of the enzyme and...
active site (black square boxes in DCCM, Figure 6) to assist in the “open” and “closed” conformations of the enzyme.

Ornithine transcarbamylase (PDB code: 4OH7; chains A and B of the homotrimer were visualized for structural and dynamical analyses) is an enzyme necessary to catalyze the reaction between carbamoyl phosphate and ornithine to form citrulline. Visualization of the structure suggests that the two subunits have little interaction with one another. Overall, ornithine transcarbamylase is very flexible and the mechanism of catalysis is quite simplistic, where half the residues, residues 150 to 280, move in an anticorrelated motion to the rest of the enzyme (black square boxes in the DCCM, Figure 6). This indicates that the enzyme has a pincer-type mechanism around the active site, which might favor substrate binding and product release (Allewell, Shi, Morizono, & Tuchman, 1999).

Pantothenate synthetase (PDB code: 2A86; both chain A and chain B of the dimer were visualized for structural and dynamical analyses) functions as a homodimer with identical domains and active sites where it carries out the catalysis of pantothenate from condensation of pantoate and β-alanine. Pantothenate synthetase also has a flexible active site cavity to allow substrate and product entry and exit to assist catalysis (Wang & Eisenberg, 2003). The dynamics of pantothenate synthetase exhibit a circular motion where all the domains rotate back and forth around the active site, which allows residues 200 to 290 to rotate with a greater extent of flexibility (boxes in displacement graph, Figure 6), to dictate the opening and closing of the active site. This lid-type mechanism of residues 200 to 290 is further confirmed based on its anticorrelated motion with the rest of the enzyme and active site, which lies within residues 30 to 50 (black square boxes in DCCM, Figure 6).

### 3.8. Identification of test proteins

A set of 12 proteins (test No. 1–12) were used to examine if dynamic features of a protein could be used for its functional identification. Six of them (test No. 1–6) belong to the protein families reported in Figure 7 but are from different species, and six of them (test No. 7–12) were chosen randomly. Using the Cartesian coordinates of these 12 proteins, DCCMs from single NMA were obtained and compared with the catalog of DCCMs reported in Table 3. We were able to correctly identify six proteins (test No. 1–6) by simple comparison of DCCMs. The test proteins No. 1–6 were matched to phosphoglycerate kinase, pyruvate kinase, enolase, arginase, fumarase, and amine oxidase, respectively (Figure 7). A difference in the extent of correlation/anticorrelation can be seen in some cases such as in test protein No. 4 and arginase, while the DCCMs for test protein No. 6 and amine oxidase appear nearly identical. As expected, the DCCM of the six proteins, which is chosen randomly, did not match with any of the DCCMs reported in Table 3, suggesting that they belong to different families of proteins. As these test proteins (No. 7–12) exhibit different patterns of motions (Table S3) compared to the 24 proteins studied here, they must also possess different function. For example, we observed that aminoacyl-tRNA synthetases (Warren et al., 2014) and cytochrome P450 enzymes (Dorner et al., 2015) display very unique patterns of motion, and their functions are distinctly different from the set of proteins considered in the present study.

### 4. Conclusion

The NMA study of 24 important enzymes has demonstrated that each enzyme has signature dynamical patterns and the intrinsic dynamics of proteins are conserved across species, from simple organisms like *E. coli* to the complex *H. sapiens*. As expected, it was observed that for a given family of proteins, as protein sequences and structures become more similar, the protein dynamics become more related as well. We successfully created a catalog of DCCMs for the 24 protein groups. These patterns are unique to each protein and have the potential to act as fingerprints of identification for unknown proteins. This catalog of intrinsic dynamics assisted us in the correct identification of six test proteins. These test proteins were successfully identified solely by comparing their respective dynamical patterns. Similar attempt to compare DCCMs of proteins was made earlier for dynamic comparison of proteins (Munz et al., 2010). However, Munz et al. have performed MD simulations and essential dynamic analysis to develop and compare DCCM of proteins. The present approach of using coarse-grained normal mode simulation is relatively faster than MD
simulation-based protein dynamic comparison and could be useful for functional identification of uncharacterized proteins. Currently, one of the main challenges in the study of the proteome is the lack of an effective and timely method of protein function identification. A majority of known protein sequences and structures have remained largely unidentified. The Protein Structure Initiative (PSI), established in 2000 by the National Institutes of General Medical Sciences, resulted in thousands of structures of proteins whose functions remain unknown. Experimental investigation of protein function is costly and time-consuming. Here we have presented a computational method for predicting protein function. Coarse-grained NMA is a fast approach and effective way to identify and characterize the function of unknown proteins. Although coarse-grained NMA has some limitations, as it provides less detailed information about the local fluctuations compared to that obtained from all-atom NMA or MD simulations (Strom et al., 2014), the method is very useful for establishing this unique dynamical fingerprint, and it only requires the PDB coordinates of a protein. Therefore, this work could be extended to the greater scope of identifying the function of uncharacterized proteins.

In the present study, DCCMs of various proteins were only visually compared; however the advancement of computational analytical methods opens the possibility of an online catalog of dynamics. This study could be extended to include more metabolic enzymes and proteins from various functional classes. For the future, we anticipate the use of NMA and the role of protein dynamics to play a larger part in protein identification and classification.

Abbreviations and symbols

| Abbreviation | Description |
|--------------|-------------|
| BLAST        | Basic local alignment search tool |
| BC           | Bhattacharyya coefficient |
| DCCM         | Dynamic cross-correlation matrix |
| ENM          | Elastic network model |
| MAT          | Methionine adenosyltransferase |
| NMA          | Normal mode analysis |
| PDB          | Protein Data Bank |
| RMSD         | Root-mean-square deviation |
| RMSIP        | Root-mean-square inner product |
| VMD          | Visual molecular dynamics |

Supplementary material

Supplementary material for this article can be accessed here http://dx.doi.org/10.1080/23312025.2017.1291877.

Funding

The authors received no direct funding for this research.

Competing Interests

The authors declare no competing interest.

Author details

Sarah M. Meeuwsen1
E-mail: meeuwsen.sarah@gmail.com
ORCID ID: http://orcid.org/0000-0002-8178-0517
An N. Hodac1
E-mail: hodacan@uwec.edu
Lauren M. Adams1
E-mail: ADAMSLM@uwec.edu
Ryan D. McMunn1
E-mail: mcmunnrd@uwec.edu
Maxwell S. Anschutz1
E-mail: anschutz.maxwell@gmail.com
Kari J. Carothers1
E-mail: karjocarothers@gmail.com
Rachel E. Egdrorf1
E-mail: r.egdorfo@mson.com
Peter M. Hanneman1
E-mail: phanneman7@gmail.com
Jonathan P. Kitrow1
E-mail: kitzrojp@gmail.com
Cynthia K. Keonigsberg1
E-mail: cynthiakoenigsberg@gmail.com
Oscar Lopez-Martinez2
E-mail: osclopez12@gmail.com
Paul A. Matthew1
E-mail: Paul.Mattchews@gmail.com
Ethan H. Richter1
Jonathan E. Schenk1
E-mail: Jschenk@umd.edu
Fuglebakk, E., Tiwari, S. P., & Reuter, N. (2014). Comparing the intrinsic dynamics of multiple protein structures using elastic network models. Biochimica et Biophysica Acta, 1850, 911-922.

Go, N., Noguti, T., & Nishikawa, T. (1983). Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proceedings of the National Academy of Sciences, 80, 3696-3700. http://dx.doi.org/10.1073/pnas.80.12.3696

Goward, C. R., & Nicholls, D. J. (1994). Malate dehydrogenase: A model for structure, evolution, and catalysis. Protein Science, 3, 1883-1888. http://dx.doi.org/10.1002/pro.1003188301

Hammes-Schiffer, S., & Benkovic, S. J. (2006). Relating protein motion to catalysis. Annual Review of Biochemistry, 75, 519-541. http://dx.doi.org/10.1146/annurev.biochem.75.103004.142800

Hansen, U., Meyer, T., Haas, J., Rex, R., Vriend, G., & Grubmüller, H. (2007). Exploring protein dynamics space: The dynamosome over the missing link between protein structure and function. PLoS ONE, 7, e33931. doi:10.1371/journal.pone.0033931

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities. Nature, 450, 913-916. http://dx.doi.org/10.1038/nature06407

Hollup, S. M., Fuglebakk, E., Taylor, W. R., & Reuter, N. (2011). Exploring the factors determining the dynamics of different protein folds. Proteins: Structure, Function, and Genetics, 75, 103004.142800

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. Journal of Molecular Graphics, 14, 33-38. http://dx.doi.org/10.1016/0267-8594(96)00018-5

Ichiye, T., & Karplus, M. (1991). Collective motions in proteins: A model for structure, evolution, and catalysis. Science, 251, 964-972. http://dx.doi.org/10.1126/science.251.4998.964

Ichiye, T., & Karplus, M. (2004). Mechanism of enolase: The crystal structure of enolase-magnesium-2-phosphoglycerate/phosphoenolpyruvate complex at 2.2 Å resolution. Biochemistry, 33, 417-428. http://dx.doi.org/10.1021/bi034018x

Ichiye, T., & Karplus, M. (2013). Hydrogen Tunneling links protein dynamics to enzyme catalysis. Annual Review of Biochemistry, 82, 471-496. http://dx.doi.org/10.1146/annurev-biochem-053110-150652

Lebodi, L., & Stec, B. (1991). Mechanism of enolase: The crystal structure of enolase-magnesium-2-phosphoglycerate/phosphoenolpyruvate complex at 2.2 Å resolution. Biochemistry, 30, 2817-2822. http://dx.doi.org/10.1021/bi00225sd12

Lemke, C. T., & Howell, P. L. (2001). The 1.6 Å crystal structure of E. coli argininosuccinate synthetase suggests a conformational change during catalysis. Structure, 9, 1153-1164. http://dx.doi.org/10.1016/S0969-2126(01)00683-9

Lin, Y. Z., Liang, S. J., Zhou, J. M., Tsou, C. L., Wu, P., & Zhou, Z. (1990). Comparison of inactivation and conformational changes of α-glycerophosphate-1-phosphate dehydrogenase during thermal denaturation. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1038, 247-252. http://dx.doi.org/10.1016/0167-4838(90)90212-X

Lisi, G. P., & Patrick Loriø, J. P. (2016). Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function. Projections of remote mutants of dihydrofolate reductase. Magnetic Resonance Spectroscopy, 92-93, 1-17. http://dx.doi.org/10.1002/prms.2015.11.001

Liu, X., Speckhard, D. C., Shepherd, T. R., Sun, Y. J., Hengel, S. R., Yu, L., ... Fuentes, E. J. (2016). Distinct roles for conformational dynamics in protein-ligand interactions. Structure, 24, 2053-2066. http://dx.doi.org/10.1016/j.str.2016.08.019

Lu, G., Dobritszsch, D., Baumann, S., Schneider, G., & König, S. (2000). The structural basis of substrate activation in yeast pyruvate decarboxylase. European Journal of Biochemistry, 267, 861-868. http://dx.doi.org/10.1046/j.1462-5822.2000.01030.x

Marques, O., & Seneouaud, Y. H. (1995). Hinge-bending motion in citrate synthase arising from normal mode calculations. Proteins: Structure, Function, and Genetics, 23, 557-560. http://dx.doi.org/10.1002/1056-8795(19960101)23:1<557::AID-PRE2>3.0.CO;2-L

Mattivi, A., Bolognesi, M., & Valentini, G. (1996). The allosteric regulation of pyruvate kinase. FEBS Letters, 389, 15-19. http://dx.doi.org/10.1016/0014-5793(96)00462-0

Micheletti, C. (2013). Comparing proteins by their internal dynamics: Exploring structure–function relationships beyond static structural alignments. Physics of Life Reviews, 10(1), 1-26. http://dx.doi.org/10.1016/j.plrev.2012.10.009

Munz, M., Lyngsø, M., Hein, J., & Bignon, P. C. (2013). Dynamics based alignment of proteins: an alternative approach to quantify dynamic similarity. BMC Bioinformatics, 11, 188. http://dx.doi.org/10.1186/1471-2105-11-188

Micheletti, C. (2013). Comparing proteins by their internal dynamics: Exploring structure–function relationships beyond static structural alignments. Physics of Life Reviews, 10(1), 1-26. http://dx.doi.org/10.1016/j.plrev.2012.10.009

Pandini, A., Mauri, G., Bordogna, A., & Bonati, L. (2007). Detecting similarities among distant homologous proteins by comparison of domain flexibilities. Protein Engineering Design and Selection, 20, 285-299. http://dx.doi.org/10.1093/protein/gbm014

Park, J., Teichmann, S. A., Hubbard, T., & Chothia, C. (1997). Intermediate sequences increase the detection of homology between sequences. Journal of Molecular Biology, 273, 349-354. http://dx.doi.org/10.1006/jmbi.1997.1288

Pandini, A., Mauri, G., Bordogna, A., & Bonati, L. (2007). Detecting similarities among distant homologous proteins by comparison of domain flexibilities. Protein Engineering Design and Selection, 20, 285-299. http://dx.doi.org/10.1093/protein/gbm014

Sievers, F., Wilms, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7, 1411–1417. http://dx.doi.org/10.1002/mcb.201103519

Skjøver, L., Skjøver, S., & Reuter, N. (2009). Normal mode analysis for proteins. Journal of Molecular Structure: THEOCHEM, 898, 42-48. http://dx.doi.org/10.1016/j.theochem.2008.09.024
Stoddard, B. L., Dean, A., & Koshsland, Jr., D. E. (1993). Structure of isocitrate dehydrogenase with isocitrate, nicotinamide adenine dinucleotide phosphate, and calcium at 2.5-Å resolution: A pseudo-Michaelis ternary complex. Biochemistry, 32, 9310-9326. http://dx.doi.org/10.1021/bi00008a008

Strom, A. M., Fehling, S. C., Bhattacharyya, S., & Hati, S. (2014). Probing the global and local dynamics of aminoacyl-tRNA synthetases using all-atom and coarse-grained simulations. Journal of Molecular Modeling, 20, 2245. http://dx.doi.org/10.1007/s00894-014-2245-1

Tarn, F., & Brooks, C. L. (2006). 3rd diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. Journal of Molecular Biology, 345, 299-314. http://dx.doi.org/10.1016/j.jmb.2004.10.054

Tiwar, S. P., Fuglebak, E., Hollup, S. M., Skjærven, L., Cragnolini, T., Grindhaug, S. H., ... Reuter, N. (2016). WEBnm@ v2.0: Web server and services for comparing protein flexibility. BMC Bioinformatics, 15, 6597.

Tobi, D. (2013). Large-scale analysis of the dynamics of enzymes. Proteins: Structure, Function, and Bioinformatics, 81, 1910-1918. http://dx.doi.org/10.1002/prot.24335

Tousignant, A., & Pelletier, J. N. (2006). Protein motions promote catalysis. Chemistry & Biology, 11, 1037-1042. http://dx.doi.org/10.1016/j.chembiol.2004.06.007

Vöhringer-Martinez, E., & Dörner, C. (2016). Conformational substrate selection contributes to the enzymatic catalytic reaction mechanism of pin1. The Journal of Physical Chemistry B, 120, 12444-12453. http://dx.doi.org/10.1021/acs.jpcb.6b09187

Wang, S., & Eisenberg, D. (2003). Crystal structures of a pantothenate synthetase from M. tuberculosis and its complexes with substrates and a reaction intermediate. Protein Science, 12, 1097-1108. http://dx.doi.org/10.1110/ISSN1469-896X

Warren, N., Strom, A., Nicolet, B., Albin, K., Albrecht, J., Bausch, B., ... Gunderson, A. (2014). Comparison of the Intrinsinc Dynamics of Aminoacyl-tRNA Synthetases. The Protein Journal, 33, 184-198. http://dx.doi.org/10.1007/s10930-014-9548-z

Weimer, K. M., Shan, B. L., Brunetto, M., Bhattacharyya, S., & Hati, S. (2009). Evolutionary basis for the coupled domain motions in thermus thermophilus leucyl-tRNA synthetase. Journal of Biological Chemistry, 284, 10088-10099. http://dx.doi.org/10.1074/jbc.M807361200

Whisstock, J., & Lesk, A. (1999). Prediction of protein function from protein sequence and structure. Quarterly Reviews of Biophysics, 36, 307-346. http://dx.doi.org/10.1017/S0033583503003901

Wiegand, G., & Remington, S. J. (1986). Citrate synthase: Structure, control, and mechanism. Annual Review of Biophysics and Biophysical Chemistry, 15, 97-117. http://dx.doi.org/10.1146/annurev.bb.15.060186.000525

Winn, S. L., Watson, H. C., Hanks, R. N., & Fothergill, L. A. (1981). Structure and activity of phosphoglycerate mutase. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 293, 121-130.

Wolodko, W. T., Fraser, M. E., James, M. N., & Bridger, W. A. (1996). The crystal structure of succinyl-CoA synthetase from Escherichia coli at 2.5-Å resolution. The Journal of Biological Chemistry, 269, 10883-10890.

Yang, L., Song, G., Carriquiry, A., & Jernigan, R. L. (2008). Close correspondence between the motions from principal component analysis of multiple HIV-1 Protease structures and elastic network modes. Structure, 16, 321-330. http://dx.doi.org/10.1016/j.str.2007.12.011

Zen, A., Micheleit, C., Keskin, O., & Nussinov, R. (2010). Comparing interfacial dynamics in protein-protein complexes: an elastic network approach. BMC Structural Biology, 10, 26. http://dx.doi.org/10.1186/1472-6807-10-26