**WEBnm@ v2.0:**

**Web server and services for comparing protein flexibility**

Sandhya P. Tiwari 2;4,Edvin Fuglebakk 1;4, Siv M.Hollup 1;4, Lars Skjaerven 3;4, Tristan Cragnolini 2;4, Svenn H.Grindhaug 4, Kidane M. Tekle4 and Nathalie Reuter\* 2;4

1Department of Informatics, University of Bergen, Norway,

2Department of Molecular Biology, University of Bergen, Norway,

3Department of Biomedicine, University of Bergen, Norway,

4Computational Biology Unit, Department of Informatics, University of Bergen, Norway

\* To whom correspondence should be addressed. Tel: +47 55 58 40 40; Fax: +47 555 89683; Email: Nathalie.Reuter@mbi.uib.no

*The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors*

Present Address: [Tristan Cragnolini], Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, Paris, 75005, France

**ABSTRACT:**

**BACKGROUND:** Normal mode analysis (NMA) using elastic network models is a reliable and cost-effective computational method to characterize protein flexibility and by extension dynamics. Further insight into the dynamics–function relationship can be gained by comparing protein motions between protein homologs and functional classifications. This can be achieved by comparing NMA results from sets of evolutionary related proteins.

**RESULTS:** We have developed an automated method for NMA comparative analysis of a set of pre-aligned protein structures. The user can submit a sequence alignment in the FASTA format and the corresponding coordinate files in the Protein Data Bank (PDB) format. The computed normalized squared atomic fluctuations and atomic deformation energies of the submitted structures can be easily compared on graphs provided by the web user interface. The web server provides pairwise comparison of the dynamics of all proteins included in the submitted set using two measures: the Root Mean Squared Inner Product and the Bhattacharyya Coefficient. The Comparative Analysis has been implemented on our web server for NMA, WEBnm@, which also provides recently upgraded NMA of single protein structures. In addition, programmatic access to WEBnm@ is now available through a SOAP-based web service. Webnm@ is available at http://apps.cbu.uib.no/webnma

**CONCLUSION:** WEBnm@ provides an easy to use tool to analyse how protein structures compare to each other in terms of flexibility. Several analyses are offered and, in combination with the single analysis option, they can give a good view of how the structures move and how the flexibility is conserved over the different structures.

**KEYWORDS:**

Elastic Network Models; Normal Mode Analysis; Comparative Analysis; Web-tool; TIM Barrels; Adenylate Kinase; Bhattacharyya Coefficient

**AVAILABILITY**

Web server: [http://apps.cbu.uib.no/webnma](http://apps.cbu.uib.no/webnma#http://apps.cbu.uib.no/webnma)

Web services: <http://cbu.bioinfo.no/wsdl/webnma-single.wsdl>, <http://cbu.bioinfo.no/wsdl/webnma-comparative.wsdl>

**INTRODUCTION**

Protein dynamics has been described to be the time-dependent changes in the structure of a protein, which includes equilibrium fluctuations governing biological function [1]. The internal deformations of protein structures have been used successfully to describe components of these time-dependent fluctuations. The mechanisms of protein function exploit their structural flexibility at all levels, from the localized fluctuations of amino acid side chains, to the reorientation of large rigid bodies such as protein domains. In the past, much of these studies have been limited to qualitative descriptions of a limited number of static structures, but there is growing evidence that structure is linked to function via dynamics and that these may be structurally and evolutionarily conserved. Thus, we see a need for comparing protein dynamics in a systematic, quantitative manner.

Although the quantity of dynamical experimental data is ever increasing e.g. from the introduction of the time dimension in X-ray and NMR studies, the information gained from them is still sparse. Several computational methods have been developed to complement experimental structural biology data and provide dynamical models of biomolecules. Molecular Dynamics (MD) simulations are widely used to simulate the dynamics of a protein on time scales of up to the microsecond. It remains computationally expensive to perform MD simulations and the time scales necessary to satisfactorily sample large motions such as domain reorientation are typically not feasible to obtain. Normal Mode Analysis (NMA) using Elastic Network Models (ENM) is a computationally efficient and reliable method for calculating properties of slow protein dynamics, in terms of their internal deformability. Their quick computation also makes them particularly well suited for automated comparative analysis between multiple structures.

NMA has been increasingly used to capture the inherent flexibility of proteins [2-5]. NMA models the motions of atoms in a protein as a coupled harmonic oscillator and associates each mode of motion to a frequency of oscillation. Low frequency modes correspond to global or domain motions and have been found to correlate well to functionally relevant protein motions [6-9]. Moreover it has been shown to yield results in good agreement with molecular dynamics simulations [5, 10-12]. ENMs are a class of models for that can be used to perform NMA efficiently. They model the protein as a network of springs capturing the density of particles and strength of interaction between them. They generally employ very simplified descriptions of the interactions and commonly represent the protein with a reduced number of atoms, typically one for each residue. ENMs are efficiently treated in computations, and despite their simplicity predict collective motion in good agreement with more detailed computational models.

Comparative analysis is a promising approach for exploring the connection between internal dynamics and structural and functional similarities between protein structures. It has been used to validate using ENM-based NMA as a method and developing measures to compare homologues [13-20]. Moreover, some aspects of the similarity between internal dynamics are determined by the protein fold [21] and are very robust to mutations or other local perturbations [22-24]. Comparing homologous proteins, or even structural variants of identical proteins, can therefore provide a useful check for whether properties of protein intrinsic motions ascribed to the shape or fold of the protein generalises to other structures with similar fold. Some examples of successful applications include comparison of homologues to understand allostery [25, 26], oligomerisation [27] and enzymatic mechanisms [28-31]. For more insights into such analyses refer to [32, 33].

WEBnm@ [34] has served structural biologists in exploring and analysing the intrinsic flexibility of single protein structures for almost a decade [35-43]. We now extend its functionalities to support the exploratory comparative analysis of sets of protein structures for the sufficiently advanced user. In addition to the new Comparative Analysis, the new web server still provides access to the tools for analysing single protein structures (the Single Analysis).

The Comparative Analysis section of the web server includes three key analyses; the deformation energies, normalized fluctuation profiles, and comparisons of the lower frequency modes (Root Mean Squared Inner Product, RMSIP) and covariance matrices (Bhattacharyya coefficient, BC) calculated from the normal modes. In addition, web services have been set up for programmatic access to WEBnm@ v2.0.

In addition we have upgraded the Single Analysis server by adding Jmol-based animations for the six lowest energy modes, a tool for calculating and visualizing correlation matrices and an overlap analysis comparing conformational changes to the proteins normal modes. The Jmol [44] application allows the user to manipulate the animation and visualise the vector-field of the movements. A correlation matrix is calculated from all the modes to show correlated movements within the structure and this can also be visualised both as a heat map and in PyMOL[45] on the input structure via a generated script. The overlap analysis captures the modes that correspond to a conformational change of the same protein.

AD-ENM [46] using the model described in Zheng et al.[20], KOSMOS [47], NMsim[48], ElNémo[49], NOMAD-Ref [50] and the ANM webserver [51] are examples of web tools for analysing single proteins (and/or RNA/DNA in the case of KOSMOS) that are currently available, most of which use a form of ENM. As the first of its kind, WEBnm@ v2.0 introduces automated comparative NMA analyses to the WEBnm@ web server, making it possible to examine relationships between proteins in terms of protein flexibility.

We present WEBnm@ v2.0 and its potential uses here with two case studies, the first on the TIM barrel superfamilies and the second on the Adenylate Kinases. In the first example, we find that we are able to discriminate structures that are related at the family and superfamily levels using the BC measure. In the second example, we find that the comparative analysis can be used to study the ligand-binding effects on the Adenylate Kinase. In a set of 8 homologous structures with conserved domains yet different conformations, we were able to easily cluster these structures based on their state, and scrutinise their differences further at the residue-level. We were able to identify changes in flexibility in one of the domains that could be a key difference between the ligand-free and ligand-bound structures. The analysis was able to capture these differences without the explicit modelling of the ligands. These examples demonstrate the ease and reliability with which large-scale NMA analysis can be performed via WEBnm@ v2.0 and potential applications to studying a set of protein structures with varying levels of homology.

**MATERIAL AND METHODS**

**Coarse-grained normal modes calculations**

WEBnm@ uses the Elastic Network Model (ENM) called the C-alpha force field developed by Hinsen et al. [52], available in the Molecular Modelling ToolKit (MMTK) [53]. Each amino acid is represented by a mass at the position of its Cα atom. The interaction between two Cα atoms is described by the pair potential,

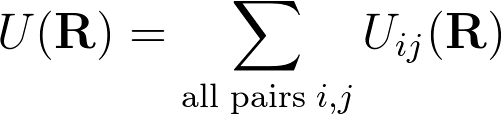
potential_300dpi.gif

where:



The following parameters for the force constant have been determined by Hinsen et al., fitting to MD simulation results [52]: a = 8.6x105 kJ.mol-1.nm-3; b = 2.39x105 kJ.mol-1.nm2; c = 128 kJ.mol-1.nm4 and d = 0.4 nm. Here, ***Rij*** is the pair distance vector between two Cα atoms and ***Rij0*** is the corresponding pair distance vector in the input configuration. Since the distance between Cα atoms adjacent in sequence cluster just short of 0.4 nm in typical protein structures, the potential can be regarded as having almost uniform force-constants for these interactions, with other interactions proportional to an inverse power of six of the equilibrium distance between interacting atoms. We do not provide an interface for adjusting these parameters as the parameterisation has been found to be transferable between proteins [52]. We prefer this approach to less detailed models that require parameterisation for each protein. Protein specific parameterization is a concern, because data for validation is scarcely available and one usually has to resort to parameterise against crystallographic B-factors, which is not a reliable source of thermal fluctuations [12, 54-56].

The potential energy of a configuration **R** of the ENM is then:



The normal modes are eigenvectors of the mass weighted matrix of second partial derivatives of the potential U. They describe deformations intrinsic to the protein structure. The eigenvalues correspond to the squares of the frequencies for each mode.

For the comparative analysis, all the modes are calculated for each structure submitted. In contrast, in the single analysis, we calculate up to 1200 modes for a speedier performance, using the method proposed by Hinsen and Kneller [57], meaning that structures with less than 300 Cα atoms have their full set of modes calculated. The full set of modes is calculated for structures where convergence is not reached for the subset. All the analyses performed under single analysis, regardless of the total number of modes calculated, use only the first 200 non-trivial modes calculated. The vectors of these 200 modes are available for download in a text file. To ensure that the potential as defined above is minimal at the input structure, any anomalous bond-lengths shorter than 2.78 Å between two Cα raise an error, this distance being defined by the parameters used in the potential described above.

**Comparative analysis**

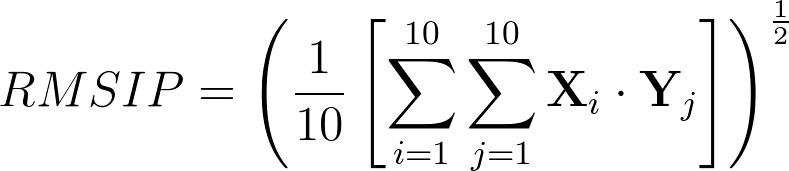
*Profile alignment analysis.* Fluctuations are calculated for each Cα atom from the obtained set of modes calculated. These can be described as the sum of each atom’s displacement along each mode (excluding the trivial ones), weighted by the reciprocal of the eigenvalues. The deformation energy is a normalized measure of the energy contributed from individual atoms of the model to the intrinsic deformation patterns[58]. Both the fluctuations and the deformation energy values reported are averaged over all modes and they are tabulated following the sequence alignment so that the results are ready for comparison purposes. The corresponding plot is generated using R[59] and made available on the web server as a Portable Document Format (PDF) file, along with the raw data values as a text file.

*Conformation overlap comparison.* For comparing large sets of aligned protein structures, it is useful to obtain a single score characterizing the level of similarity in the intrinsic motion. We implemented two measures for that purpose: the root mean squared inner product (RMSIP) [60] and the Bhattacharyya coefficient (BC) [18]. Both measures are calculated on amino acids that are conserved in the alignment of the whole dataset. Normal modes for only the conserved part of the alignment, which is derived from the following:



where the Hessian (http://latex.codecogs.com/gif.latex?%5Cdpi%7B300%7D%20%5Ctilde%7B%5Cbold%7BH%7D%7D**H**) of the full potential is partitioned so that **H***aa* reflect interactions in conserved parts of the alignment, *a*, **H***ab* reflect interactions between *a* and non-conserved parts, *b*, and **H***bb* reflect interactions in *b*.

The results are presented as a heat map with a dendrogram obtained from a complete linkage clustering. This can be used to check whether the similarity of normal modes agree with functional or evolutionary classifications. The root mean squared inner product (RMSIP) is defined as:



where **Xi** and **Yj** refer to the eigenvectors of a pair of proteins being compared, and i,j are the mode numbers. Following Amadei et al. [61] it is customary to let the sums run over the ten lowest energy non-trivial modes. The BC measure is based on the Bhattacharyya distance [62] that we adapted earlier for the purpose of comparing protein flexibility [18]. It compares the covariance matrices (**** and ) obtained from the normal modes of the conserved parts of the proteins to be compared.

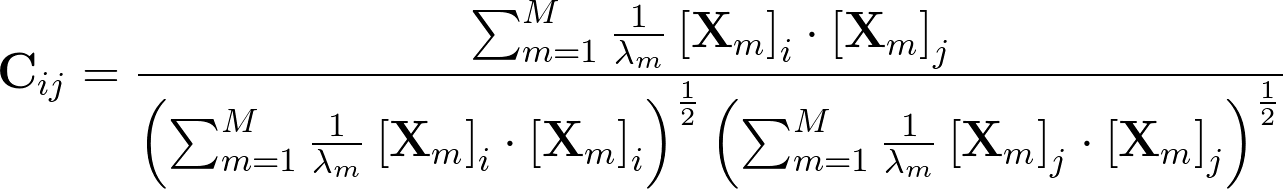


Here |**X**| denotes the determinant of **X** and the rank of the matrices are reduced in two steps. First, **An** and **Bm** are obtained from the *n* and *m* lowest frequency modes of their respective proteins and normalized by dividing by their trace. Then **** and  are obtained by projecting **An** and **Bm** on to the *s* eigenvectors of (**An**+**Bm**)/2 that explain most of its variance. For each comparison *n* and *m* are chosen so that 95% of the variance of each protein is retained and *s* so that 75% of the variance of (**An**+**Bm**)/2 is retained. The initial rank reduction, obtained by **An** and **Bm** is introduced in the web server for computational efficiency.

**Single analysis**

*Visualization of the lowest frequency modes.* Animations of the six lowest frequency modes are made using Jmol. They display vector field arrows, which show the magnitude and direction of the movements characterized by each mode. The Jmol applet takes standard Jmol commands for modifying the visual representation, and thereby allows e.g. change of representation, distance measurements, etc. All of the modes eigenvectors and eigenvalues are also available as text files for further manipulation by a more advanced user.

*Heat map of dynamically correlated regions*. The correlation matrix as defined by Ichiye and Karplus[63] is calculated from the normal modes. Each element in the matrix quantifies the coupling between two atoms i and j as:



where **X**m and λm are eigenvectors and eigenvalues of the *m*th normal mode respectively and the i and j indices denote the component of the mode corresponding to individual atoms. Cij is the expected inner product of displacements of atoms i and j, and ranges from -1 to 1, where -1 and 1 are maximal anti-correlations and correlations.

A visual representation of these correlated regions is available as a downloadable PyMOL script, where significant correlations are represented as sticks on the cartoon representation of the structure. These correlations are chosen such that:

1. They are above the chosen threshold for positive correlations (represented as red sticks), and below the negative of this threshold for the negative correlations (represented as blue sticks). The correlation threshold is picked at a percentile that changes according to the number of Cα atoms in the input structure; the percentile lies within the range of 95, for less than 200 atoms, to 99.9, for more than 2000 atoms. The percentile chosen for a given structure and the resulting threshold values are provided as a header comments in the downloadable PyMOL script.
2. Only the correlations between atoms that have a minimum distance of 0.8 nm are considered, to focus on the pairs of Cα atoms that have a limited influence from the peptide backbone and strong force constants in the ENM.
3. The network of residues that satisfies the score threshold has a minimum size of 1 pair of Cα atoms.

Based on the above criteria, the visual representation is especially informative with structures that possess more than one domain.

*Analysis of conformational changes*. The overlap analysis compares two conformations of the same protein (e.g. A and B), submitted by the user, and identifies the modes that contribute the most to the structural difference; these modes are likely to be involved in the movements leading from conformation A to B [64]. The web server calculates the overlap between the modes calculated and the structural difference between the two submitted structures. Values of the squared overlap and the cumulative overlap are plotted against mode numbers.

**Implementation**

The Webnm@ back-end is implemented using the Molecular Modelling Toolkit (MMTK) [53] and runs on a dedicated server currently utilizing 24 2.4 GHz cores with 256GB of available memory. The Application Programming Interface (API) is written in Python and the web interface is built on the Zope ToolKit [65] - based Grok framework 1.0 [66]. Management of multiple jobs is taken care of by an in-house job scheduler, Hodman [67].

**Web services**

We have developed two web services, one for the single analysis and one for the comparative analysis. Both are implemented as document/literal wrapped SOAP [68] web-services. The web-services provide only the raw data points as output and exclude the animation files. WSDL files (Web Service Definition Language) for accessing document/literal wrapped SOAP web services are available at http://cbu.bioinfo.no/wsdl/webnma-single.wsdl (Single Analysis) and http://cbu.bioinfo.no/wsdl/webnma-comparative.wsdl (Comparative Analysis). Example scripts detailing the use of the services can be found on the website.

**RESULTS AND DISCUSSION**

**Description of the web server**

For protein structures, WEBnm@ requires Cα coordinates as input, which it gathers from a PDB ID or coordinate file. Users are welcome to bookmark a URL or provide their email address to receive the link that would allow the user to retrieve their results when convenient, within two weeks of the submission date. On the WEBnm@ v2.0 page, a user is able to choose between Single Analysis and Comparative Analysis through two different tabs.

A flowchart illustrating the steps and functionality of the Comparative Analysis server is provided on Figure 1 and the main steps are described here:

(A) The input for the comparative analysis can be submitted under the ‘Comparative’ tab, where the structural alignment in the FASTA format can be uploaded under the ‘FASTA file’ field, and the corresponding coordinate files in the ‘PDB files’ field. We provide format description and examples on the web user interface (webUI), as well as advice on how to best align the structures. An e-mail address can be provided (but is not required) for larger jobs so that the user can be receive the URL to his/her results.

B) After the initial submission, all jobs go through a pre-analysis computation phase where the normal modes are calculated and initial analyses are performed. From this stage on, users can bookmark the URL that will allow them to retrieve their results at their convenience, within two weeks of the submission date.

C) The user can choose the appropriate analysis to be performed: profile alignment analysis for deformation energies and atomic fluctuations, and covariance similarity analysis for calculations of the RMSIP and BC. Warnings appear (red fonts) if the structure(s) submitted contain unrecognized heteroatoms or non-standard amino acids to inform the user how these have been taken into account in the modes calculation. From this point onwards, brief descriptions and references for the results provided upon expanding the “See more…” button in the results pages.

D) The ‘Profile Alignment Analysis’ results in a plot of the deformation energy for each protein in the dataset with respect to the sequence number in the alignment. The same format is used for the atomic fluctuation profiles. The results are also provided as raw data (text files, with no values set for positions corresponding to gaps in the alignment).

E) The ‘Covariance Similarity Analysis’ provides as output BC and RMSIP heat maps. Links to their full PDF versions at higher resolution are provided as well as the raw data should the users wish to produce their own plots.

**Case study 1: TIM-barrel proteins fold**

We have previously studied the conservation of intrinsic protein motions by comparing the normal modes between diverged protein structures [18]. We present here an example of the same type of study as exploratory analysis. We aligned 20 structures from the TIM beta/alpha-barrel fold as classified by SCOP [69] using MUSTANG[70]. The structures were selected such that two superfamilies are equally represented, and from each of these two families are equally represented. We then submitted this alignment and the corresponding protein structures to the web server, which calculated the similarity of the normal modes for all pairs of structures in the data set. Having ordered the alignment file by family and superfamily the resulting plots are presented in Figure 2. We report both RMSIP (Figure 2A) and BC (Figure 2B). It can be seen at a glance that the BC similarity measure discriminates between both families and superfamilies. Here, the dendrogram on one side matches the SCOP classification on the other. With the RMSIP, this classification is not captured as well. Figure 2B also indicates the selection of structures that are least similar to other structures sharing its classification. This illustrates how this analysis can be used to choose representative structures of a class for a more detailed study.

We believe that this type of analysis is useful for exploring the conservation of protein intrinsic motions and to check assumptions about the variability of such motions in data sets [13, 27, 29 , 31, 71, 72]. Other potentially useful applications could be contrasting isolated subunits with subunits in a protein complex [26], or proteins with and without a bound ligand [25, 28, 30]. For more detailed characterisation of the structural flexibility of any class of proteins, the visualisation in Figure 2 can be used to identify good representatives of the class for further investigation.

**Case study 2: Adenylate Kinases**

The comparative analysis is also useful for analysing conformational changes in identical and distantly related orthologues. We describe here the example of Adenylate Kinases (AdKs), which are a well-studied family of enzymes that transfers a phosphoryl group from an ATP to an AMP to create 2 ADP molecules with the aid of a magnesium ion. These enzymes are considered to be critical in the regulation of diverse cellular processes such as metabolic monitory and cell signalling (for a full review: Zhang et al.[73]). These monomeric enzymes consist of a CORE domain and two flanking mobile domains called LID (which is the site of ATP-binding) and NMP-binding (which is the site of the AMP-binding). The conformational changes that occur upon substrate binding involve large-amplitude, hinge-like movements of the LID and NMP-binding domains, which falls within the nanosecond timescale[74-78]. These domains are also known to move in a correlated manner[79] where the closing and opening of the NMP-binding domain is enhanced by the closing and opening of the LID domain. Some studies have been suggesting that the domains also undergo partial unfolding as part of this process [80, 81], whereas Daily et al. [82], amongst others, have suggested that many local motions are involved in the large conformational change seen.

We analysed Adenylate Kinase in varying states, bound to different ligands (Table 1). Their sequences vary in their similarity, ranging from 52% to 100% (Supplementary Table 2). The states include fully open and ligand-free, partially open and ligand-free, NMP-binding domain closed (AMP bound), LID domain closed (ATP bound), both closed (AMP, ADP bound or AP5 bound with or without cofactor). For all of these structures, the calculations were done on only the C-alpha atoms of the structures, disregarding the presence of the ligands. Nevertheless, since the backbones of these structures are influenced by the conformation they are in, the spatial arrangement of these atoms is captured by their respective ENMs. We report here the BC scores for the 8 enzymes (Figure 3A), which show clear separation between the states of the enzymes, with the clustering showing the relative closeness of these states to each other. The separation between these structures follows that of the clustering done by Snow et al. [74] based on essential dynamics sampling simulations. We see greater variation in the scores between the ligand-free structures than we do for the ligand-bound structures. This difference could correspond to the range of mobility of the hinge domains between the homologues that results in the fully and partially open sub-states. The outlier 2ak3 displays the same kind of order in the clustering with regards to 4ake, 1dvr and 1ake as seen in the separation of the distribution of LID-CORE and NMP-binding-CORE distances observed in all-atom MD simulations [83].

The BC score provides a measure of the similarity between the intrinsic motions of the proteins, but does not provide information about which parts of the structures accounts for any difference in intrinsic motion. To aid in pinpointing specific differences between them, the web server provides deformation energies and fluctuations calculated for each amino acid. We analysed the deformation energies and the fluctuations of these 8 structures (Figures 3B and 3C). On these plots, the fluctuations of the conserved part are represented using solid lines, and the non-conserved parts in dotted lines. In the comparative fluctuations plot, we find that the CORE domain between the LID and NMP-binding domains has higher relative flexibility in the structures with both domains closed (1ake, 1aky, 2aky, 2eck) than with the others. The structures that have the NMP-binding domain free of any ligand (4ake, 1ak2, 1dvr) display larger peaks within this domain, especially towards the C-terminal end, whereas the structures with ligands in this domain display low levels of fluctuations in contrast. While this trend is not as explicit in the LID domain, we do see that the fully open structure 4ake experiences the most fluctuations in this domain.

In the comparative deformation energies plot, we see that structures that have no ligands or a ligand bound in only one of the domains display a clear opposite trend (a valley instead of a peak) towards the C-terminal end of the NMP-binding domain (positions 60-67 in the alignment index) when compared to structures that have both domains occupied. Upon closer scrutiny on the structure, this region lies in the loop between two helices within the domain that interacts with the ligand. The regions that flank this have been identified to be hinges that facilitate the displacement of the NMP-binding domain that have been described by Henzler-Wildman et al. [75], and fits well with the hinge region predicted by Pontiggia et al. [84]. This region also precedes the part of the NMP-binding domain that accumulates very high strain energy [85]. We observed that this region consists of higher deformation energies when both domains are bound, compared to the structures that have only one or neither of the domains occupied. The contrast in deformation energies in the NMP binding domain, fits well with the clusters obtained from the BC score, and suggests a local structural clue to what causes the difference in conformational degrees of freedom (seen in the BC plot) for these two groups of structures. Based on the BC plot we chose 1aky and 1ak2 as representatives of the closed and open or partially open conformation respectively. Inspection of these structures reveals that the region of the NMP domain with higher deformation energies in the closed state is in close contact with a helix flanking the LID domain. From the comparative fluctuations plot (Figure 3B), we see that both sides of this contact (alignment index 60-67 and 180-187) have increased fluctuations in the open states. Running 1aky and 1ak2 on the single analysis reveals correlation plots with a marked difference in correlation between these two regions. Compare the correlations between the segments with residue index 54-62 and 168-175 in Figure 4A and B marked by solid black lines. These regions in the closed conformation are highly positively correlated (Figure 4B), which is typical for parts of the structure in close contact. While we will be careful in making functional inference from these structural observations, we believe this example serves to illustrate how the comparison of fluctuation and deformation profiles can serve to localize structural features that can explain why the normal modes of related structures are different, as inferred from the BC or RMSIP plots. We have also illustrated how the single analysis on representative structures can aid in such structural inference.

This also shows that despite the lack of ligands represented explicitly in the ENMs, the analysis is still sensitive enough to pick up on changes in flexibility conferred to these points, even when the effects are small. Moreover, the heterogeneity in the conformations of the proteins can be well assessed by the similarity scoring (BC) and further scrutinised using the profile alignment techniques. The implication of comparing multiple structures, instead of the pair of completely open and closed structures, is that it allows us to understand the dynamics of intermediate conformational states of an enzyme in relation to each other, without over-interpreting differences between the two extreme cases.

**CONCLUSION**

WEBnm@ is now a web-tool that provides comparative normal modes analysis on multiple protein structures, in addition to its original functionality as a web-tool for NMA of single protein structures. The computations are performed using the *Calpha* forcefield developed by Hinsen et al[52]. Comparative analyses of aligned structures are to the best of our knowledge not provided by any other web-tools. We have demonstrated that comparative analysis can be used to conduct analysis of evolutionarily related protein families, and study a small set of homologous proteins where the results may provide clues to the differences in flexibility. WEBnm@ is designed to quickly provide simple output, tailored towards teaching and exploratory analysis of proteins while large scale analysis can be accomplished through the web services.

**LIST OF ABBREVIATIONS**

NMA – Normal Mode Analysis

ENM – Elastic Network Model

BC – Bhattacharyya coefficient

RMSIP – Root Mean Squared Inner Product

MD – Molecular Dynamics

MMTK – Molecular Modelling ToolKit

**AUTHORS’ CONTRIBUTIONS**

EF and ST developed, implemented and tested the code for the API for the analyses available in the Single and Comparative Analysis sections and wrote the documentation at the front-end. They served as the main driving force in building webnm@ v2.0 and drafted the manuscript. SMH designed and implemented the webservices. LS and TG contributed code to the API and the front-end (TG). SHG and KT developed the framework and the front-end. NR supervised the project and edited the manuscript. This work is a truly collaborative effort of all eight authors. All authors read and approved the final manuscript.

**ACKNOWLEDGEMENT**

We would like to thank AS Schillinger, C.Grauffel, S.Narawane and A.Abboud, as well as all of our users, for testing and benchmarking of the WEBnm@ 2.0 webserver and webservices.

**FUNDING**

This work was supported by the Bergen Research Foundation (Bergen forskningsstiftelse) and the National Program for Research in Functional Genomics of the Research Council of Norway.

**SUPPLEMENTARY DATA**

**S1.** Flow of the Single Analysis on the webUI. **S2.** Sequence representation of the structural alignment of the homologous Adenylate Kinases. **T1.** TIM barrel proteins fold dataset. **T2.** Sequence similarities of the homologous Adenylate Kinases sequences.

**REFERENCES**

1. Henzler-Wildman K, Kern D: **Dynamic personalities of proteins**. *Nature* 2007, **450**(7172):964-972.

2. Tama F: **Normal mode analysis with simplified models to investigate the global dynamics of biological systems**. *Protein Pept Lett* 2003, **10**(2):119-132.

3. Ma JP: **Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes**. *Structure* 2005, **13**(3):373-380.

4. Bahar I, Rader AJ: **Coarse-grained normal mode analysis in structural biology**. *Current Opinion in Structural Biology* 2005, **15**(5):586-592.

5. Bahar I, Lezon TR, Bakan A, Shrivastava IH: **Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins**. *Chemical reviews* 2010, **110**(3):1463-1497.

6. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M: **Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic**. *Proteins* 2002, **48**(4):682-695.

7. Petrone PP, Vijay S: **Can conformational change be described by only a few normal modes?** *Biophysical Journal* 2006, **90**(5):1583-1593.

8. Dobbins SE, Lesk VI, Sternberg MJE: **Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking.** *Proceedings Of The National Academy Of Sciences Of The United States Of America* 2008, **105**(30):10390-10395.

9. Wako H, Endo S: **Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis.** *Biophysical Chemistry* 2011, **159**(2-3):257-266.

10. Skjaerven L, Martinez A, Reuter N: **Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit**. *Proteins* 2010, **79**(1):232-243.

11. Rueda M, Chacón P, Orozco M: **Thorough validation of protein normal mode analysis: a comparative study with essential dynamics**. *Structure* 2007, **15**(5):565-575.

12. Fuglebakk E, Reuter N, Hinsen K: **Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions**. *Journal Of Chemical Theory And Computation* 2013.

13. Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J: **Exploring the common dynamics of homologous proteins. Application to the globin family**. *Biophys J* 2005, **89**(1):3-13.

14. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR: **An analysis of core deformations in protein superfamilies**. *Biophys J* 2005, **88**(2):1291-1299.

15. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino Dl, Ortiz AR: **Core deformations in protein families: a physical perspective**. *Biophys Chem* 2005, **115**(2-3):125-128.

16. Zen A, Carnevale V, Lesk AM, Micheletti C: **Correspondences between low-energy modes in enzymes: Dynamics-based alignment of enzymatic functional families**. *Protein Science* 2008, **17**(5):918-929.

17. Zen A, de Chiara C, Pastore A, Micheletti C: **Using dynamics-based comparisons to predict nucleic acid binding sites in proteins: an application to OB-fold domains.** *Bioinformatics (Oxford, England)* 2009, **25**(15):1876--1883.

18. Fuglebakk E, Echave J, Reuter N: **Measuring and comparing structural fluctuation patterns in large protein datasets**. *Bioinformatics (Oxford, England)* 2012, **28**(19):2431-2440.

19. Zheng W, Doniach S: **A comparative study of motor-protein motions by using a simple elastic-network model**. *Proc Natl Acad Sci USA* 2003, **100**(23):13253-13258.

20. Zheng W. BBR: **Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin.** *Biophys J* 2005, **89**:167–178.

21. Hollup SM, Fuglebakk E, Taylor WR, Reuter N: **Exploring the factors determining the dynamics of different protein folds**. *Protein Science* 2011, **20**(1):197-209.

22. Zheng W, Brooks BR, Thirumalai D: **Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations**. In: *Biophysical Journal.* vol. 93; 2007: 2289-2299.

23. Echave J, Fernández FM: **A perturbative view of protein structural variation**. *Proteins* 2010, **78**(1):173-180.

24. Echave J: **Why are the low-energy protein normal modes evolutionarily conserved?** *Pure and Applied Chemistry* 2012, **84**(9):1931-1937.

25. Rodgers TL, Townsend PD, Burnell D, Jones ML, Richards SA, McLeish TC, Pohl E, Wilson MR, Cann MJ: **Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors**. *PLoS biology* 2013, **11**(9):e1001651.

26. Marcos E, Crehuet R, Bahar I: **Changes in dynamics upon oligomerization regulate substrate binding and allostery in amino acid kinase family members**. *PLoS Comp Biol* 2011, **7**(9):e1002201.

27. Katebi AR, Jernigan RL: **The critical role of the loops of triosephosphate isomerase for its oligomerization, dynamics, and functionality**. *Protein science : a publication of the Protein Society* 2014, **23**(2):213-228.

28. Kolan D, Fonar G, Samson AO: **Elastic network normal mode dynamics reveal the GPCR activation mechanism**. *Proteins* 2014, **82**(4):579-586.

29. Lukman S, Grant GH: **A network of dynamically conserved residues deciphers the motions of maltose transporter**. *Proteins* 2009, **76**(3):588-597.

30. Seckler JM, Leioatts N, Miao H, Grossfield A: **The interplay of structure and dynamics: insights from a survey of HIV-1 reverse transcriptase crystal structures**. *Proteins* 2013, **81**(10):1792-1801.

31. Yang L, Song G, Carriquiry A, Jernigan RL: **Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes**. *Structure* 2008, **16**(2):321-330.

32. Van Wynsberghe AW, Cui Q: **Conservation and variation of structural flexibility in protein families**. *Structure* 2010, **18**(3):281-283.

33. Micheletti C: **Comparing proteins by their internal dynamics: Exploring structure–function relationships beyond static structural alignments**. *Physics of Life Reviews* 2012.

34. Hollup SM, Salensminde G, Reuter N: **WEBnm@: a web application for normal mode analyses of proteins**. *Bmc Bioinformatics* 2005, **6**:-.

35. Blacklock K, Verkhivker GM: **Experimentally Guided Structural Modeling and Dynamics Analysis of Hsp90-p53 Interactions: Allosteric Regulation of the Hsp90 Chaperone by a Client Protein**. *J Chem Inf Model* 2013, **53**(11):2962-2978.

36. Blacklock K, Verkhivker GM: **Allosteric Regulation of the Hsp90 Dynamics and Stability by Client Recruiter Cochaperones: Protein Structure Network Modeling**. *Plos One* 2014, **9**(1).

37. Bloudoff K, Rodionov D, Schmeing TM: **Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases**. *J Mol Biol* 2013, **425**(17):3137-3150.

38. Mehrbod M, Mofrad MRK: **Localized Lipid Packing of Transmembrane Domains Impedes Integrin Clustering**. *Plos Comput Biol* 2013, **9**(3).

39. Golji J, Collins R, Mofrad MRK: **Molecular Mechanics of the alpha-Actinin Rod Domain: Bending, Torsional, and Extensional Behavior**. *Plos Comput Biol* 2009, **5**(5).

40. Golji J, Mofrad MRK: **A Molecular Dynamics Investigation of Vinculin Activation**. *Biophysical Journal* 2010, **99**(4):1073-1081.

41. Amrute-Nayak M, Diensthuber RP, Steffen W, Kathmann D, Hartmann FK, Fedorov R, Urbanke C, Manstein DJ, Brenner B, Tsiavaliaris G: **Targeted Optimization of a Protein Nanomachine for Operation in Biohybrid Devices**. *Angew Chem Int Edit* 2010, **49**(2):312-316.

42. Stoica I: **Characterization of protein matrix motions in the Rb. sphaeroides photosynthetic reaction center**. *J Mol Model* 2006, **12**(4):468-480.

43. Nicholas Warren AS, Brianna Nicolet, Kristine Albin, Joshua Albrecht, Brenna Bausch, Megan Dobbe, Megan Dudek, Samuel Firgens, Chad Fritsche, Anthony Gunderson, Joseph Heimann, Cheng Her, Jordan Hurt, Dmitri Konorev, Matthew Lively, Stephanie Meacham, Valentina Rodriguez, Stephanie Tadayon, David Trcka, Yer Yang, Sudeep Bhattacharyya, Sanchita Hati: **Comparison of the Intrinsic Dynamics of Aminoacyl-tRNA Synthetases.** *Protein J* 2014, **33**(2):184-198.

44. Gezelter J: **Jmol: an open-source Java viewer for chemical structures in 3D**. In*.*; 2006.

45. **The PyMOL Molecular Graphics System** [<http://www.pymol.org>]

46. **AD-ENM server** [<http://enm.lobos.nih.gov>]

47. Seo S, Kim MK: **KOSMOS: a universal morph server for nucleic acids, proteins and their complexes**. *Nucleic acids research* 2012, **40**(Web Server issue):W531-536.

48. Kruger DM, Ahmed A, Gohlke H: **NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins**. *Nucleic acids research* 2012, **40**(Web Server issue):W310-316.

49. Suhre K, Sanejouand Y: **ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement.** In: *Nucleic acids research.* vol. 32; 2004: 610-614.

50. Lindahl E, Azuara C, Koehl P, Delarue M: **NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis.** *Nucleic acids research* 2006, **34**:W52-56.

51. Eyal E, Yang LW, Bahar I: **Anisotropic network model: systematic evaluation and a new web interface**. In: *Bioinformatics (Oxford, England).* 2006.

52. Hinsen K, Petrescu A, Dellerue S, Bellissent-Funel M, Kneller G: **Harmonicity in slow protein dynamics**. *Chem Phys* 2000, **261**:25-37.

53. Hinsen K: **The Molecular Modeling Toolkit : a new approach to molecular simulations**. *J Comput Chem* 2000, **21**(2):79-85.

54. Hinsen K: **Structural flexibility in proteins: impact of the crystal environment**. *Bioinformatics (Oxford, England)* 2008, **24**(4):521-528.

55. Riccardi D, Cui Q, Phillips J, George N: **Evaluating elastic network models of crystalline biological molecules with temperature factors, correlated motions, and diffuse X-Ray scattering**. *Biophysical Journal* 2010, **99**(8):2616-2625.

56. Soheilifard R, Makarov DE, Rodin GJ: **Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors**. *Physical Biology* 2008, **5**(2):026008.

57. Hinsen K, Kneller G: **Projection Methods for the Analysis of Complex Motions in Macromolecules**. *Molecular Simulation* 2000, **23**(4):275-292.

58. Hinsen K: **Analysis of domain motions by approximate normal mode calculations**. *Proteins* 1998, **33**(3):417-429.

59. **R**. In*.*, Copyright (C) 1989, 1991 Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA edn.

60. Amadei A, Ceruso MA, Di Nola A: **On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations**. *Proteins* 1999, **36**(4):419-424.

61. Amadei A, Linssen ABM, Berendsen HJC: **Essential dynamics of proteins**. *Proteins: Struct Funct Genet* 1993.

62. Bhattacharyya A: **On a measure of divergence between two multinomial populations**. *Sankhyā: The Indian Journal of Statistics* 1943, **7**:401-406.

63. Ichiye T, Karplus M: **Collective motions in proteins - a covariance analysis of atomic fluctuations in molecular-dynamics and normal mode simulations**. *Proteins* 1991, **11**(3):205-217.

64. Reuter N, Hinsen K, Lacapere J-J: **Transconformations of the SERCA1 Ca-ATPase: A Normal Mode Study**. *Biophys J* 2003, **85**(4):99020-.

65. **Zope ToolKit: a set of libraries intended for reuse by projects to develop web applications or web frameworks** [<http://docs.zope.org/zopetoolkit/index.html>]

66. **Grok: a web application framework** [<http://grok.zope.org>]

67. **Hodman: job dispatcher for asynchronous web services.** .

68. **SOAP: Simple Object Access Protocol**.

69. Murzin AG, Brenner SE, Hubbard T, Chothia C: **SCOP: a structural classification of proteins database for the investigation of sequences and structures**. *J Mol Biol* 1995, **247**(4):536-540.

70. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM: **MUSTANG: a multiple structural alignment algorithm**. *Proteins* 2006, **64**(3):559-574.

71. Raimondi F, Orozco M, Fanelli F: **Deciphering the deformation modes associated with function retention and specialization in members of the Ras superfamily**. *Structure* 2010, **18**(3):402-414.

72. Velazquez-Muriel JA, Rueda M, Cuesta I, Pascual-Montano A, Orozco M: **Comparison of molecular dynamics and superfamily spaces of protein domain deformation**. *BMC Struct Biol* 2009, **9**(1):6.

73. Zhang S, Nemutlu, E., Terzic, A., & Dzeja, P.: **Adenylate Kinase Isoform Network: A Major Hub in Cell Energetics and Metabolic Signaling**. In: *Systems Biology of Metabolic and Signaling Networks.* vol. 16; 2014: 145-162.

74. Snow C, Qi, G. and Hayward, S. : **Essential dynamics sampling study of adenylate kinase: Comparison to citrate synthase and implication for the hinge and shear mechanisms of domain motions.** *Proteins* 2007, **67**(2):325–337.

75. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D: **A hierarchy of timescales in protein dynamics is linked to enzyme catalysis**. *Nature* 2007, **450**(7171):913-U927.

76. Katherine A. Henzler-Wildman VT, Ming Lei1 Maria Ott3, Magnus Wolf-Watz, Tim Fenn, Ed Pozharski, Mark A. Wilson, Gregory A. Petsko, Martin Karplus, Christian G. Hübner & Dorothee Kern: **Intrinsic motions along an enzymatic reaction trajectory.** . *Nature* 2007, **450**(7171):838–844.

77. Schulz GE, Müller, C. W. & Diederichs, K.: **Induced-fit movements in adenylate kinases.** *J Mol Biol* 1990, **213**(4):627-630.

78. Gerstein M, Schulz, G. & Chothia, C.: **Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers.** *J Mol Biol* 1993, **229**(2):494-501.

79. Jana B, Adkar, B. V, Biswas, R., & Bagchi, B. : **Dynamic coupling between the LID and NMP domain motions in the catalytic conversion of ATP and AMP to ADP by adenylate kinase**. *The Journal of Chemical Physics* 2011, **134**(3).

80. Miyashita O, Onuchic JN, Wolynes PG: **Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins**. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 2011, **100**(22):12570-12575.

81. Schrank TP, Wrabl JO, Hilser VJ: **Conformational Heterogeneity Within the LID Domain Mediates Substrate Binding to Escherichia coli Adenylate Kinase: Function Follows Fluctuations**. *Top Curr Chem* 2013, **337**:95-121.

82. Daily MD, Phillips, G. N. & Cui, Q.: **Many local motions cooperate to produce the adenylate kinase conformational transition.** *J Mol Biol* 2010, **400**(618–31).

83. Brokaw JBC, J.-W. : **On the roles of substrate binding and hinge unfolding in conformational changes of adenylate kinase.** *Biophysical Journal* 2010, **99**(3420–9).

84. Pontiggia F, Zen A, Micheletti C: **Small-and Large-Scale Conformational Changes of Adenylate Kinase: A Molecular Dynamics Study of the …**. *Biophysical Journal* 2008.

85. Whitford PC, Miyashita O, Levy Y, Onuchic JN: **Conformational transitions of adenylate kinase: Switching by cracking**. *J Mol Biol* 2007, **366**(5):1661-1671.

**FIGURES LEGENDS**

**Figure 1.** Flow of the webUI. (**A**) File upload and job submission, (**B**) modes computation and preliminary analysis, (**C**) list of analyses available, (**D**) profile alignment results, (**E**) covariance similarity analysis

**Figure 2.** Root Mean Squared Inner Product (RMSIP, **A**) and the Bhattacharyya Coefficient (BC, **B**) for all pairs of structures in the TIM beta/alpha-barrel fold. The plots are symmetric with comparisons of a structure to itself on the diagonal and hierarchical clustering represented on the periphery. The colour scale is adapted to visualise contrasts in a data set, hence the same colour might refer to different values in the two plots. Structure names reflect the SCOP classification, with leading characters denoting the SCOP concise classification string.

**Figure 3.** The Bhattacharya Coefficient (BC, **A**) for all pairwise comparison in the alignment of the Adenylate Kinases. The plot shows that the ligand-bound and ligand-free structures cluster in two distinct groups. The comparative normalized fluctuations plot of the Adenylate Kinases (**B**) and deformation energy plot (**C)**. In this plot, the values of the parts that are aligned to all other structures in the alignment in solid lines, and the parts aligned with gaps in other structures are in dotted lines. The NMP and LID domains are annotated by black and grey points, respectively.

**Figure 4.** Inter-residue correlations for selected representatives of the open and closed conformation of the Adenylate Kinases. The open conformation (**A**) is represented by PDB ID 1ak2 and the closed conformation (**B**) by PDB ID 1aky. Solid black lines mark the regions that have greater correlations in the closed conformation than in the open (residue index 54-62 and 168-175).

**TABLES**

**Table 1.** PDB IDs, Chains, Conformational state and Ligands bound of the Adenylate Kinase analysed using Comparative analysis in case study 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PDB ID | Chain | State | Organism | Ligands |
| 2ak3 | A | NMP-binding domain closed | *Bos taurus* | AMP |
| 4ake | A | Both NMP-binding and LID domains open | *E. coli* | None |
| 1ak2 | A | LID domain nearly closed | *Bos taurus* | None |
| 1dvr | B | LID domain closed | *S. cerevisiae* | ATP |
| 1ake | A | Both NMP-binding and LID domains closed | *E. coli* | AP5 |
| 2eck | A | Both NMP-binding and LID domains closed | *E. coli* | AMP and ADP |
| 1aky | A | Both NMP-binding and LID domains closed | *S. cerevisiae* | AP5 and imidazole |
| 2aky | A | Both NMP-binding and LID domains closed | *S. cerevisiae* | AP5 and Mg2+ |