Computers & Graphics (2021)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Visual exploration of large normal mode spaces to study protein flexibility

Pierre Bedoucha[a,c], Nathalie Reuter[b,c], Helwig Hauser[a], Jan Byška[a,d,*]

[a] Department of Informatics, University of Bergen, Bergen Norway
[b] Department of Chemistry, University of Bergen, Bergen, Norway
[c] Computational Biology Unit, University of Bergen, Bergen, Norway
[d] Masaryk University, Brno, Czech Republic

ARTICLE INFO

Article history:
Received January 27, 2021

Keywords: Normal Mode Analysis, Protein Flexibility, Molecular Visualization, Coordinated and Multiple Views

ABSTRACT

When studying the function of proteins, biochemists utilize normal mode decomposition to enable the analysis of structural changes on time scales that are too long for molecular dynamics simulation. Such a decomposition yields a high-dimensional parameter space that is too large to be analyzed exhaustively. We present a novel approach to reducing and exploring this vast space through the means of interactive visualization. Our approach enables the inference of relevant protein function from single structure dynamics through protein tunnel analysis while considering normal mode combinations spanning the whole normal modes space. Our solution, based on multiple linked 2D and 3D views, enables the quick and flexible exploration of individual modes and their effect on the dynamics of tunnels with relevance for the protein function. Once an interesting motion is identified, the exploration of possible normal mode combinations is steered via a visualization-based recommendation system. This helps to quickly identify a narrow, yet relevant set of normal modes that can be investigated in detail. Our solution is the result of close cooperation between visualization and the domain. The versatility and efficiency of our approach are demonstrated in two case studies.

1. Introduction

Proteins come with a large variety of functions, including the catalysis of chemical reactions and the transport of molecules in living organisms. Therefore, studying protein function has been in the scope of biochemists for many decades. Proteins are dynamic structures and it is state-of-the-art understanding that protein motion is often centrally related to protein function [1, 2]. To explore protein dynamics, researchers often utilize in-silico methods, including molecular dynamics (MD) simulations as one popular choice. However, MD simulations are computationally very demanding and simulating large systems for a long time remains overly challenging [3].

A commonly used alternative to MD simulations, which is better suited for studying large molecular systems for longer time intervals, is to decompose the flexibility of a protein into a set of normal modes. This decomposition enables researchers to probe the flexibility of a protein structure, solely based on the 3D organization of its atoms but without providing the resolution along time. One major challenge of normal mode calculation is that it leads to many modes (in the order of the number of atoms) and analyzing them separately or in combination is a practically unbound endeavor. To investigate protein mechanisms, biochemists commonly use multiple conformations of the same structure as hallmarks that allow them to study the transitions between functional states. However, this exploration is limited to studying a few low-frequency normal modes – usually separately, without considering the impact of the possible combinations of these normal modes on the protein

*Corresponding author:
e-mail: jan.byska@gemail.com (Jan Byška)

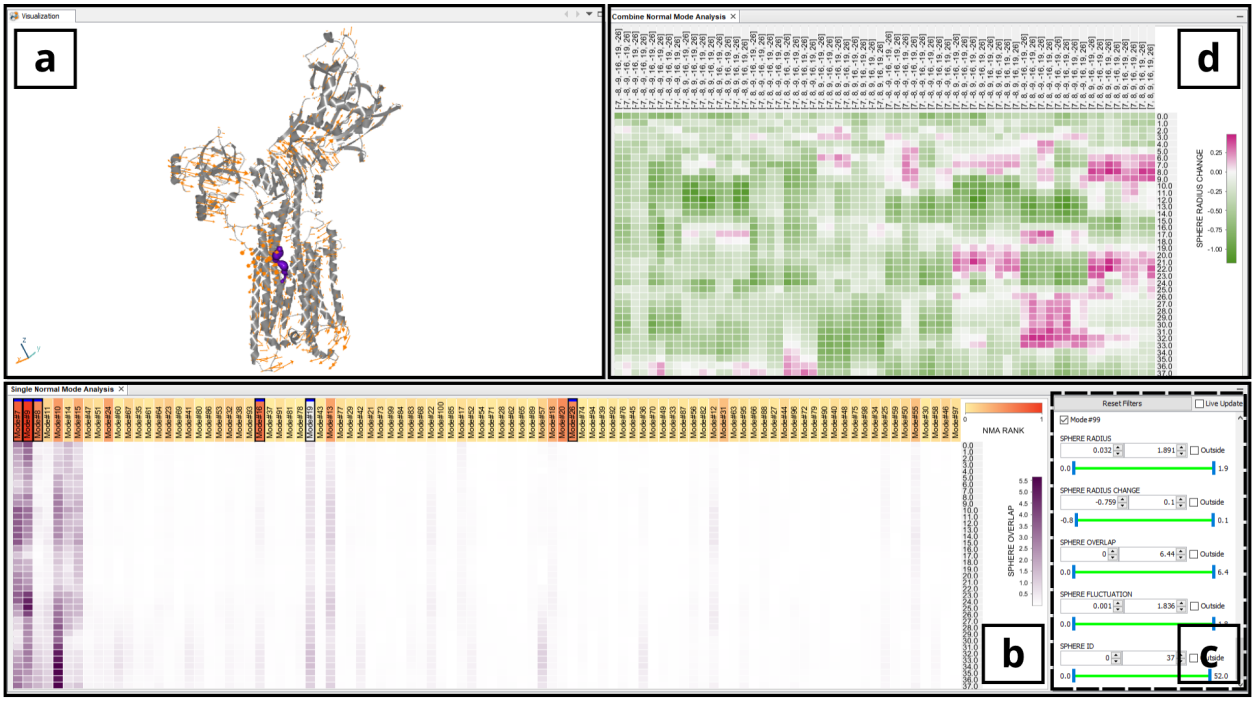

Fig. 1: The overview of our system - a) 3D View depicting the protein, its tunnel (violet), and selected normal mode using vector field representation (orange arrows); b) Single Normal Mode View showing the involvement of individual normal modes in tunnel dynamics; c) filtering and ranking options; d) Combined Normal Mode View providing detailed information about possible combinations of selected normal modes and their influence on tunnel radius.

mechanisms.

In this application paper, we present a new and interactive visual approach used to better exploit the information provided by the normal mode calculation to study links between protein function and dynamics. To do so, we utilize the fact that the protein function is facilitated via interactions with other proteins or smaller molecules (known as ligands). These interactions occur in so-called active sites that are either on the protein surface or deeply buried inside it. In the latter case, protein voids (also known as tunnels or channels) serve as access paths for the ligand to the buried active site. Studying the dynamics of protein tunnels helps domain experts to better understand how proteins interact with ligands – which in turn helps them to understand the function of the protein. Therefore, we use protein tunnel dynamics as a guiding mechanism to identify the most important normal modes.

Our solution is based on a carefully crafted set of 2D and 3D views (see Figure 1). The views enable the quick and extensive exploration of individual normal modes (or their combinations) and their effect on the dynamics of a protein tunnel of interest. Our design revolves around a fast sorting approach, based on computing the similarity of the normal modes to the desired behavior [4], which enables the domain experts to explore the normal modes and their combinations in a more systematic and faster way than what was possible before. To the best of our knowledge, we present one of the first solutions that would not only permit normal mode visualization but also allow exploring combinations throughout the normal mode space. Altogether, the main contributions of this paper are:

- assessment of the tasks for an effective and efficient visual analysis of proteins using normal mode calculation,
- the design of a new interactive tool for the exploratory analysis of a large number of normal modes and their combinations, based on these tasks,
- the design of a visualization-based sorting approach, guiding the user to potentially important normal modes or their combinations, easing the exploration of the vast normal modes space,
- the possibility of performing the analysis only from a single protein conformation

The work presented in this paper was conducted in close collaboration between the visualization and computational biology units at *[removed for the review]* and the authors of this paper belong to both of these groups. To document the value of our approach, we include case studies that show how our solution supports the inference of a relevant protein function from a single protein conformation, where several input structures (representing various functional states) were required so far. This amounts to a significant step forward as it makes studying proteins, where only one conformation is known, much easier.

2. Background

Proteins are dynamic systems, and the way they move has long been regarded as a link between their structure and their function. The intricate motions they can display to perform their goal may be decomposed into principal components. They are known as the normal modes which are used as a mean to help domain experts to study the flexibility of the protein structure.

2.1. The Normal Modes

In general, normal mode calculations are a state-of-the-art method from physics and are defined as an analytic approach to model the flexible motions of mechanical systems. In other words, normal modes are an evaluation of the system interactions. These interactions are characterized as a matrix representing the physical forces between pairs of system particles. Such a matrix is decomposed into independent vectors - here its columns - obtained through eigenvalue decomposition. They form a new orthogonal basis set to describe the particle coordinates and are known as the independent modes of motion. Normal modes classify all available motions a system can undergo given its 3D structure and they are essentially large, non-uniform vector fields (at discrete locations).

In a biological context, normal modes represent all the intrinsic dynamics available to the protein. To transition from one functional conformation (i.e., the particular spatial arrangement of atoms) to another, the protein will deform following a specific subset of modes from its normal mode space. Since using normal mode calculation remains computationally cost-effective with increasing size of the molecule, it can be used routinely also on large molecular structures with thousands of atoms.

The normal mode calculation can be performed using different granularity – the full atom calculation, where all the atoms are considered, and coarse grain (CG) normal mode calculation, where only one atom per amino acid (i.e., the α-carbon of each amino acid) is considered. In this paper, we focus on the CG normal mode calculation, since it enables the study of larger systems, but the presented solution can be easily used for full atom normal mode calculation, as well.

Solving the eigenvalue decomposition of the interaction matrix, describing the forces between the protein atoms, yields $3N$ normal modes (eigenvectors), where N is the number of particles (i.e. α-carbons) in the model [5]. Each normal mode k comprises a set of N 3D vectors $\mathbf{V}_k$ and describes an independent periodic motion of the entire system around the initial position $\mathbf{r}^0$.

From the list of all normal modes, we can reconstruct all possible motions of each protein particle i using the following equation:

$$\mathbf{r}_i(t) = \mathbf{r}_i^0 + \frac{1}{\sqrt{m_i}} \sum_{k=1}^{3N} \mathbf{v}_{ki} a_k \cos(2\pi f_k t + \delta_k) \tag{1}$$

with a_k denoting the thermal amplitude per mode k, f_k being its frequency (linked to the eigenvalue from the original eigenproblem), δ_k the phase shift, and m_i the mass of the considered particle i. It is important to note that thermal amplitude a_k is a function of the given normal mode frequency, highlighting the fact that the lowest-frequency modes are also displaying the largest motions:

$$a_k = \frac{\sqrt{2k_B T}}{2\pi f_k} \tag{2}$$

with, k_B the Boltzmann constant and T the temperature set at $300K$. It describes and normalizes the amplitude of a normal mode k by informing only about the particles' position extrema.

All $\mathbf{V}_k$ are the eigenvectors and consequently form an orthogonal basis set. It is thus possible to project the $3N$ mass-weighted Cartesian coordinates on eigenvector k to get the normal coordinate q_k. Obtaining atomic displacements can be then achieved from one mode, or n normal modes combined, through the following equation:

$$\mathbf{r}_i = \mathbf{r}_i^0 + \frac{1}{\sqrt{m_i}} \sum_{k}^{n} \mathbf{v}_{ki} q_k \tag{3}$$

Equation 3 is linear, and thus, atom movements can be treated as usual vectors and combined accordingly. The phase shifts δ_k from Equation 1 are taken into account by considering the combinations of vector directions (in both positive and negative) when combining k modes.

Also noteworthy, the first six modes (also called trivial modes, $1 \leq k \leq 6$) are rigid-body transformations, and therefore, not of interest. When domain experts analyze normal modes, they usually consider only the lowest frequency modes, also known as the slow modes (e.g., the first 20 non-trivial modes), as they have been shown to experimentally and computationally relate to protein function [6, 7]. Although the higher frequency modes are usually disregarded, due to the smaller variance of motion and the increased complexity of such an analysis, it is known that they can yield useful information [8].

2.2. Normal Mode Analysis

Normal mode analysis (NMA) is one of the molecular modeling methods to complement the experimental procedures which lack dynamical information. NMA represents all the information that can be drawn from the previous fundamental Equations. When exploring and analyzing the results of a normal mode calculation, domain experts might be interested in asking questions including the following:

- Q1 Which amino acids (or domains) of the protein are displaced and which normal modes relate to their motion?
- Q2 Which amino acids (or domains) of the protein are displaced together under the normal modes?
- Q3 Which flexible motions are likely to be relevant for a particular protein function?
- Q4 Which motions are potentially influencing the transport of ligands (or substrates) through a protein tunnel or channel?

Getting answers to these questions is a nontrivial task as the normal mode calculation provides a sizable number of modes without any information about their possible involvement. As drawn from Equation 3 the normal mode space grows with the combination of amplitudes and phases δ_k. Considering all their potential combinations, without proper visualization support, becomes challenging.

Surely, in taking Equations 1 and 2, one can compute the variance per particle i for a normal mode vector set k, defined by the squared fluctuations:

$$F_{ki} = \frac{k_B T \mathbf{v}_{ki}^2}{m_i 4\pi^2 f_k^2} \tag{4}$$

with $\mathbf{v}_{ki}$ being mass-weighted mode vector scaled by its thermal amplitude and weighted by its corresponding frequency. These squared fluctuations describe the extent of flexibility (as in the amount of displacement) of given particles in a structure and can tell the biochemists about the most displaced domains of the protein, inferring intrinsic mechanisms. However, as this information is obtained for every particle, a proper visual representation is needed.

Normal modes being vector fields, the straight-forward method to visualize them involves scaled arrows. During our interviews with domain experts we discovered that it can be useful, especially for non-experts, to reconstruct a trajectory, following a mode in question (or their combinations) and depict it in 3D. Seeing the actual movement enables biochemists to rapidly generate hypotheses about how each normal mode could be involved in protein functions of interest.

One disadvantage of depicting the trajectory is that it remains hard to focus on details and identify the correlation between movements, mainly due to occlusion and visual clutter. Such an analysis, for instance, is important to investigate allostery regulation processes [9] by which the proteins transmit the effects of (un)binding of ligands from one active site to another part of the protein. To investigate this phenomenon, it is common to study which parts of the protein are moving together. Accordingly, any visualization system for normal mode analysis should support the direct analysis of these correlated movements.

To this end, the correlated movements of particles can be obtained from the available non-trivial normal modes as firstly introduced by Ichiye et al. [10] with the cross-correlations:

$$\mathbf{C}_{ij} = \frac{\sum_{k=7}^{3N} \frac{1}{\lambda_k} \mathbf{v}_{ki} \cdot \mathbf{v}_{kj}}{(\sum_{k=7}^{3N} \frac{1}{\lambda_k} \mathbf{v}_{ki} \cdot \mathbf{v}_{ki})^{\frac{1}{2}} \cdot (\sum_{k=7}^{3N} \frac{1}{\lambda_k} \mathbf{v}_{kj} \cdot \mathbf{v}_{kj})^{\frac{1}{2}}} \tag{5}$$

with, $\mathbf{v}_{ki}$ and λ_k being, respectively, the displacement vectors of the eigenvector, and eigenvalue from a non-trivial mode k, and i and j the particles of the system. The correlations indicate which structure particles are coupled in their motions: $C_{ij} = 1$ when the coupled motion of two particles i and j is fully correlated and alternatively $C_{ij} = -1$ if it is fully anti-correlated.

Further, it was also mentioned during the interviews that current workflows for the detection of important normal modes often require multiple conformations of the protein as input. Having two conformations enables the experts to detect normal modes representing the transitions between these experimentally observed functional states (e.g., protein in open and closed conformation). The challenge here is that it is hard and time-consuming to obtain these conformations in the first place, for example, in cases where not much is yet known about the protein function. Hence, the experts would appreciate a solution that enables them to discover potentially functional and relevant normal modes only from the initial protein structure used for the normal mode calculation.

2.3. Existing NMA Workflow

With the presented formalism in mind, we describe the current workflow used by the biochemists to perform normal mode analysis with the available tools. First, the modes are ranked by their frequencies and up to a dozen of the first normal modes (slow modes) are chosen to be studied. Each selected mode is scrutinized on a protein-wide scale; the biochemists mainly answer Q1 to identify the dynamical domains of the protein. To this aim, they reconstruct the trajectories following the normal modes and utilize the fluctuation calculations. Second, the correlated movements are studied in answering Q2 to examine which dynamical domains move together. Generally, this first set of analyses is carried out within a specific tool (e.g., [11, 12]). Then, the slow modes are studied locally as well. The biochemists aim at identifying which flexible motions are important for a particular sub-domain of the protein. This step requires precise geometrical measurements within the protein structure and are carried out with different tools (e.g., [13, 14]). Finally, to narrow down the protein functional motions in answering Q3 and Q4, the biochemists may isolate a tunnel to study its dynamics. Multiple protein structures deformed along each of the slow modes are generated and later loaded in a third tool (e.g., [15, 16]) to compute a tunnel and analyze its dynamical behavior under the normal modes.

This intricate workflow has its shortcomings, for instance, it is very difficult to isolate the exact same tunnel within different protein structures for each of the slow modes across several tools. Further, when exploring the potential functional modes and their combinations the field experts have to switch between all the mentioned tools to confirm or infirm the choice of functional normal modes or their hypotheses on allosteric mechanisms. However, most importantly, due to the current technical limitations, the maximum number of selected modes is usually arbitrary without knowing if any of the disregarded modes may be describing a functional movement.

2.4. Protein Tunnels as a Metric of Interest

As already addressed, protein function is often related to protein tunnels and their dynamics [17, 18]. Tunnels represent a space between the protein amino acids and are usually depicted as a set of contiguous spheres. They connect deeply buried active sites with the environment and enable chemical reactions through which proteins facilitate their function. Thus, when attempting to identify potentially functional modes, domain experts often reported the need for analyzing and quantifying their influence on specific tunnels. To support such an analysis we carefully selected the following measurements that allow us to quantify the effect of a given normal mode on the tunnel:

- The tunnel radius and its changes per sphere to resolve the effect of the normal mode down to amino acid precision.
- The squared fluctuation values of tunnel lining amino acids, describing the variance of motions of the given tunnel under a given normal mode.
- The normal mode overlap that measures the extent to which a normal mode vector field is similar to another vector field representing an atomic displacement. To study the impact of the protein flexible motions on its tunnels, we use a projection method proposed by Hinsen [19]. With a vector field of M displacement vectors, and for mode k,

it is possible to quantify the involvement of the mode in a particular displacement:

$$p_k = \sum_{i=1}^{M} \mathbf{d}_i \cdot \mathbf{v}_{ki} \tag{6}$$

with, $\mathbf{v}_{ki}$ the k^{th} normal mode vector of particle i, and d the corresponding displacement vector to project onto the k^{th} normal mode – p_k^2 then represents the involvement value. To describe the relationship between the tunnel and the normal mode, we set up this general vector field such that it describes some potentially functional movements - in our case the expansion of the tunnel. Widening the tunnel for metabolite pathing requires the surrounding amino acids to be displaced away from the tunnel center. Hence, the displacement vectors **d** are designed such that they are pointing from the center of each sphere and through the surrounding amino acids.

2.5. Data and Task Abstraction

In following the guidelines proposed by Tamara Munzner [20] and to summarize the above subsections, we identified the following crucial tasks, which a system for normal mode analysis must support:

- T1 (Search task) Explore the influence of a single normal mode (or their combinations) on the overall dynamics of the protein and its tunnels (Q1)
- T2 (Query task) Identify the correlated movements of amino acids under the normal modes (Q2)
- T3 (Query task) Identify variance in amino acid motions under the normal modes (Q1)
- T4 (Query task) Summarize different spatial and biochemical properties of a tunnel under the influence of a single normal mode (or their combinations) (Q3, Q4)
- T5 (Query task) Identify the amino acids surrounding the tunnel which are contributing to its dynamics (Q4)
- T6 (Query task) Identify potentially functional and relevant normal modes only from the initial protein structure and its corresponding normal mode calculation (Q3)

Also, as laid out at the beginning of the section, our workflow comprises several input data listed as follows: the initial protein atom positions in 3D (PDB format), the normal mode vector sets (or vector fields) obtained as a list from the normal mode calculation, and a list of computed protein tunnels among which the users can select one to investigate.

3. Related Work

In the past decades, several tools for computing normal modes and their analysis have emerged [21, 12, 22, 23, 11]. While these tools are partially addressing some of the tasks presented above, they do not provide a satisfactory solution for all of them. Normal mode analysis (NMA) refers to a quantitative and visual set of methods. It mainly consists of the analysis of fluctuations, trajectories, vector field representations, and correlations.

All the above tools display the normal mode atomic fluctuations, which highlight the most displaced protein regions, using a 2D chart as a function of the protein amino acid sequence. However, such a representation is insufficient as the spatial information is lost and the representation requires precise knowledge about the relationship between the protein sequence and its structure.

The motion induced by the individual normal modes is commonly depicted either in 3D using animation or as a static visualization using arrows. Noteworthy, the work from López-Blanco [11], implements a more recent visualization of the normal mode vector field with the use of affine arrow glyphs as introduced by Bryden et al. [24]. Unfortunately, these methods do not support the analysis of the protein tunnel dynamics. Despite this, we made sure that our approach supports them as they help to partially address T1 by informing about the protein-wide dynamics which play an important role during the analysis.

Finally, the presented tools allow the analysis of the correlated motions of each structure particle (T2) with correlation matrices showing the coupling of all particle pairs. Additionally, Tiwari et al. [21] provided an external PyMol [25] script to display the 3D links between the significantly correlated structure particle pairs. As such representation preserves the spatial information better than the correlation matrices, we have adapted it in our solution.

To link the intrinsic dynamics of the protein to its function the domain experts often utilize multiple experimentally obtained conformations of the protein. For instance, Tama et al. [4] evaluated the normal modes with respect to their involvement in the transitions between the two known functional states. Several approaches are also using path planning to explore the possible transconformations between the two functional states [26, 27]. Particularly, Al-Bluwi et al. [27] used NMA to steer the conformation path sampling between the start and end protein states. However, precise sampling of the conformational space may involve long computation times depending on the macromolecule size (e.g., 30 hours for a 950 amino acid protein [27]). Additionally, the above methods are implying the inherent need for several protein structures. It limits their use for certain classes of proteins (e.g., transmembrane proteins) that remain, despite the tremendous progress in the past few years [28], still difficult to determine as explained by Cheng [29].

In order to overcome this issue, several methods have been developed from the molecular modeling community. More precisely, the recent work of Mahajan and Sanejouand [30] presents the achievability of predicting the possible end conformation only from a starting structure (T6). The authors suggest identifying the functional normal modes by retaining the modes that are conserved through multiple different normal mode calculations. The disadvantage of this approach is high computational complexity which prevents the interactive normal mode space exploration.

Also, on an exploratory matter from a single protein structure, the work from Mongan [31] introduced a solution to obtain

the principal components of motion for a protein and project its structure on the individual (or combined) eigenvectors. However, the approach does not link the possible motions to any potential protein function and the choice of relevant components is left entirely to the user who has to select the modes (or their combination) one by one.

As can be seen, the exploration and analysis of the vast normal mode space is an immense task. Therefore, our tool provides sorting features that suggest possibly interesting combinations of modes. The topic of guidance is heavily discussed in the visualization community. For instance, Willett et al. [32] described the concept of Scented Widgets that enhances the common navigation elements in the user interface with embedded visual cues. Krause et al. [33] proposed a visual analytic system for guiding users through the process of feature selection for predictive modeling on high-dimensional data. Gladisch et al. proposed a recommendation method for navigation in hierarchical graphs [34]. Finally, an overview of user guidance for visual analytic systems was provided by Ceneda et al. [35]. The authors proposed an extension of the Wijk's [36] model of visualization including the means for user guidance.

In this paper, we propose a method for linking the protein intrinsic dynamics with protein function via analysis of motions influencing protein tunnels or channels. A similar idea was already investigated by Taly et al. [37] and Kurkcuoglu et al. [38]. However, in both cases, the authors either study the normal modes separately or investigate only a limited number of their combinations. Moreover, the approach proposed by Taly et al. [37] also requires a manual selection of atoms involved in the protein channel bottleneck which will then be approximated by a regular pentagon in the plane perpendicular to the main channel.

The methods presented above stress the need for a more exhaustive framework to investigate the link between protein dynamics and function through normal mode analysis and precise tunnel and channel computation. In general, there are several techniques for visualization and visual analysis of protein tunnels and channels, for both static and dynamic cases. In this text we list only a few most relevant methods but a complete overview can be found in the work of Krone et al. [39].

Tools such as Caver Analyst [16] or MOLE 2.0 [15] use Voronoi diagrams to extract and subsequently visualize the evolution of protein tunnels and channels. Here, the most computationally expensive part is to track the inner voids over time. As this step may require hours for a larger dataset, these techniques are unsuitable for analysis of the vast normal mode space. Alternatively, several real-time techniques for the extraction of dynamic cavities inside proteins were proposed [40, 41, 42]. However, these methods only allow studying the evolution of the whole cavity, while for the purposes of normal mode analysis, we require to analyze the impact of a normal mode to different parts of the tunnel.

Finally, also multiple non-spatial techniques were developed over the past decade. For instance, Lindow et al. [43] propose to use rational graphs to explore the migration of a selected inner cavity within protein over time. Byška et al. [44] proposed to use heat maps to investigate the tunnel bottleneck and its dynamics. Also, Byška et al. [45] and Masood et al. [46] developed abstracted representations for depicting the tunnel and its properties including the amino acids over time. Finally, Kolesar et al. [47] proposed a technique for the spatial reformation of the tunnel surface 2D space. While the methods mentioned above can be used to identify amino acids surrounding the tunnel (T5), they are designed only for the analysis of molecular dynamics data and do not support (at least not directly) the normal mode analysis pipeline.

4. Visualisation and Interaction Design

In Section 2.3, we presented the common workflow for biochemical analysis of protein motions and functions with the currently available tools. This workflow has its shortcomings as described above. Therefore, we propose several visual abstractions (see Figure 2) that are enabling quick and flexible exploration of normal modes while addressing the previously introduced tasks (T1-T6). Upon loading the input data, the domain experts have access to two workflow paths for studying movement characteristics on protein-wide (bottom part of Figure 2) or tunnel-wide scales (Figure 2a,b).

On the protein-wide scale, and to focus on allostery and concerted large-amplitude protein motions, the biochemists can compute and visualize normal mode correlations (see Figure 2c, Section 4.3). Alternatively, to get rapid insights about the motion variance, the biochemists have access to the computation of normal mode squared fluctuations with a coloring mapped onto the protein structure (see Figure 2f, Section 4.3). Also, to convey direction when static visualization is required, the normal mode vector field can be directly displayed, either in the form of vector field arrows (see Figure 2g, Section 4.3), or in the form of affine arrows illustrating large-scale motions (see Figure 2h, Section 4.3). Then, for dynamical representation and direct impact of the protein motions on the selected tunnel, the domain expert can compute and display the mode trajectory (see Figure 2d, Section 4.3).

Providing the domain experts want to analyze normal modes on a tunnel-wide scale they can follow an alternative workflow path by exploring the Single Normal Mode View (SNMV, see Figure 2a, Section 4.1) as it displays the impact of the normal modes on each of the tunnel subsections (i.e., spheres) with respect to various geometrical measurements. Upon sorting, filtering, and eventually selecting a sub-set of normal modes in the SNMV, the interactive link with the Combined Normal Mode View (CNMV, see Figure 2b, Section 4.2) allows the domain experts to have access to equivalent geometrical measurements for the combinations of selected normal modes.

In case biochemists need more details on individual normal modes or their combinations, they can choose a column in the SNMV or the CNMV and interact with updated 3D visualizations. Here they can study the corresponding normal mode fluctuations, vector fields and trajectories. To further the local analysis of the motion impact on the tunnel, the biochemists can select a tunnel section (i.e., rows) on either of the SNMV or CNMV and the active geometrical measurement coloring is overlaid onto the tunnel spheres in 3D (see Figure 2e, Section 4.3).

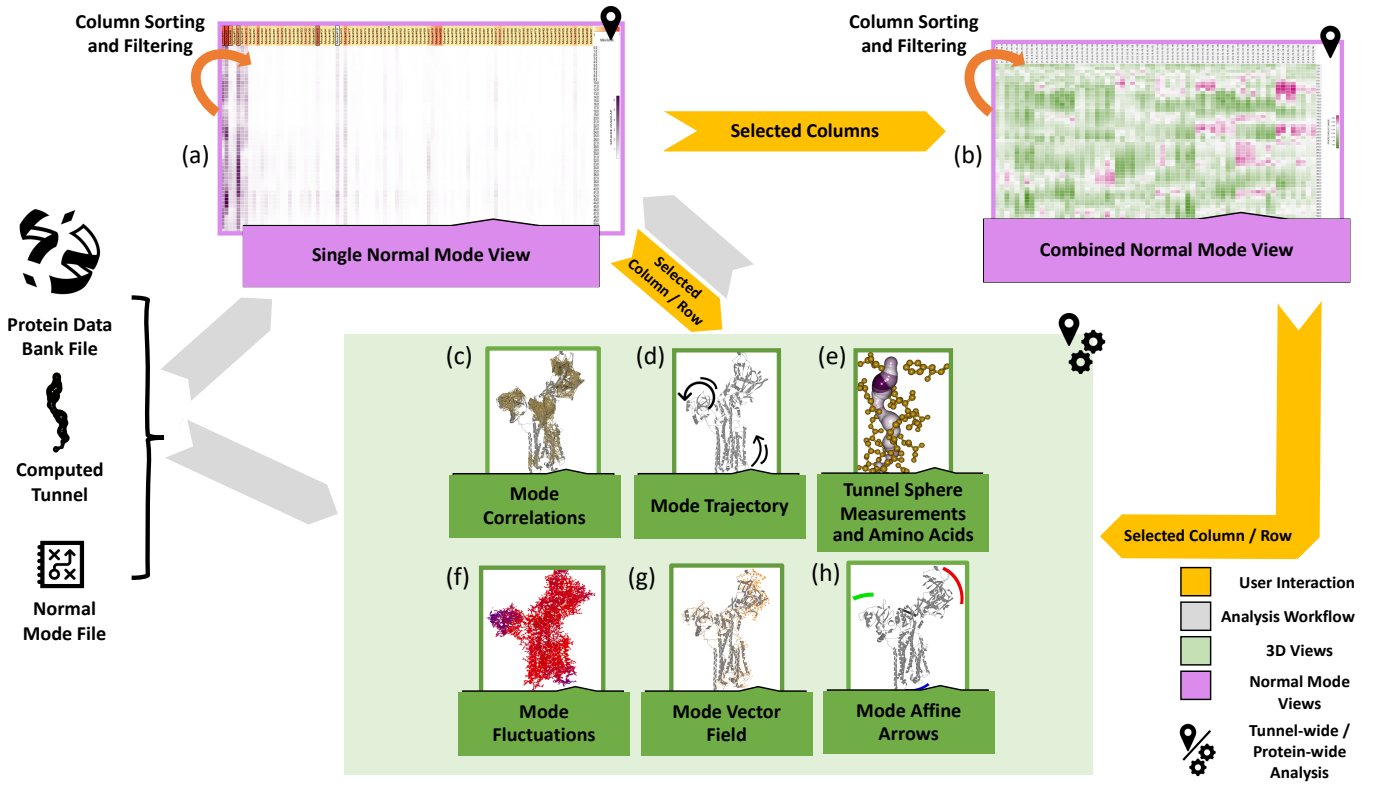

Fig. 2: Implemented workflow of our tool to visualize and quantify the flexible motions of proteins and their tunnels. The data inputs are located on the left-hand side and give the domain expert access to a range of both protein-wide and tunnel-wide dynamics analysis. Yellow arrows represent user interactions with the system which will update some of the visualizations while grey arrows depict the analysis workflow. Letters (a)-(h) indicate the different views in our tool and are detailed throughout Section 4.

This three-step process (i.e., selecting functionally relevant normal modes based on either protein-wide scale or tunnel-wide scale, using the latter to select one of their possible combinations, and visualizing its influence on protein structure) can be easily repeated and results in an interactive feedback loop. It can be used by the biochemists to quickly confirm their hypotheses about the protein functional mechanisms.

In the subsequent sections, we will describe the different views of our approach by following the second presented workflow path, from a tunnel-wide scale and throughout the normal mode space with the SNMV and the CNMV, to a protein-wide scale where a selected mode or a combination of modes can be explored in detail with the 3D View.

4.1. Single Normal Mode View

The independence of the normal modes enables using them as a basis set to describe any of the possible motions of the protein. Within this space, the protein employs an even narrower space of motions to carry out its functional goals. One of the aims of the biochemists is to precisely study this space to infer protein mechanisms and function (T4).

However, selecting relevant normal modes from the decomposition without proper representation is practically unfeasible. In order to pinpoint functionally relevant modes, every mode and every possible mode combination have to be taken into account. Therefore, we design the 2D Single Normal Mode View (SNMV) that allows the biochemists to explore normal modes in a summarised and systematic fashion. It provides an overview of all the normal modes and their influence on the protein spheres composing the tunnel. We choose to focus the analysis on the tunnel spheres instead of tunnel-wide measurements (e.g., tunnel length, curvature, or bottleneck radius) as biochemists need amino acid-wise definition in inferring substrate pathing or functional mechanisms. To solve this task, geometrical measurements of the tunnel under the influence of normal modes are displayed in the SNMV in the form of a matrix (see Figure 2a). The x-axis represents the available normal modes, primarily ranked by their respective frequencies, the slowest one being located on the far left of the matrix chart. The y-axis represents each sphere forming the tunnel, going from inside of the protein (top) to the outer environment (bottom). Each matrix cell represents a value for the selected measurement (e.g., sphere radius) by color. Hovering a given cell, the users will be displayed the exact information about the corresponding mode, the sphere number, and all the geometrical measurements relevant for the particular tunnel part.

Summarizing the geometrical measurements of the tunnel under all possible normal modes requires a fast approach for displacing the tunnel. To this end, our solution proposes to inspect each normal mode with the following method. For each one of them, the initially calculated amino acids which are forming the tunnel boundaries (called tunnel lining amino acids) are displaced following the flexible motions at the user-defined amplitude. Then, the initial tunnel is reconstructed

around the displaced amino acids. The displaced tunnel reconstruction process consists of moving each initial tunnel sphere along the direction of the normal mode displacements. Each initial tunnel sphere center is displaced by the average flexible displacement of all of its surrounding amino acids and the reconstructed sphere radius is adjusted to the shortest distance to the displaced amino acid atoms. Ultimately, the geometrical measurements of the reconstructed tunnel are calculated by comparing each tunnel sphere variations from the initial one.

The above-described method ensures an efficient summarizing of the tunnel spatial properties under the influence of all the individual normal modes, and supports the domain experts in visually isolating probable functional ones (T4). To even further assist the biochemists in exploring and selecting relevant normal modes across their whole space in a quantitative fashion, we design sorting features based on the tunnel and the normal modes (T6). The normal mode sorting heavily utilizes the overlap method described in Section 2 with Equation 6. For each normal mode, the overlap involvement value (or similarity) with another input vector field is calculated.

On the one hand, the sorting approach can be based on the tunnel. Although reconstructing the displaced tunnel with the previously described calculation is fast, it is still too computationally expensive for the interactive sorting of the modes needed here. Consequently, only vibrational displacements of the tunnel lining amino acids are considered. Therefore, we build the input vector field such that it would approximate our desired behavior – representing the largest change in tunnel sphere size.

On the other hand, the users can utilize the sorting approach which is based on a broader analysis of the normal mode space. Upon selecting a normal mode of interest – e.g., a normal mode resulting in biochemically relevant motions of the tunnel -– the biochemists can obtain a fast comparison with all of the other modes in the space. The sorting will run, for all the protein α-carbon atoms, a pairwise overlap comparison between the different mode vectors (see Equation 6, with d_i being here the displacement vectors of another normal mode).

Ultimately, the modes are highlighted based on their overlap values ranking as a recommendation to the end-users. Similarly to the idea of scented widgets [32], we show this information directly on the column headers using color, with the red-most mode being the most influential and the yellow-most the least. Based on this ranking, the normal modes can be subsequently reordered by their average overlap values directly in the SNMV, placing the most influential mode on the left and making it straightforward to select (see Figure 5).

As a result, biochemists can rapidly identify the normal modes that can putatively greatly impact their tunnel (T4) and thus explore a larger number of normal modes than with the already available normal mode visualization tools.

4.2. Combined Normal Mode View

Upon identifying the important normal modes in the SNMV the biochemists require access to combinations of these selected normal modes. In Equation 1 the phases δ_k indicate how the normal mode direction vectors are applied with one another. Indeed, through Equation 3, the phases are taken into account in exploiting the linearity and the vector directions (as in, adding in the same directions, canceling out in opposing directions). Therefore, it is crucial to consider all possible combinations of modes as they are key to allow studying more complex motions (T4).

Therefore, we designed and implemented the Combined Normal Mode View (CNMV) that presents the combinations of selected modes. This view is, similarly to SNMV, based on a heatmap representation (see Figure 3). As discussed in the previous section, the geometrical measurements detailing the tunnel are computed using an identical method but considering the combined mode vectors.

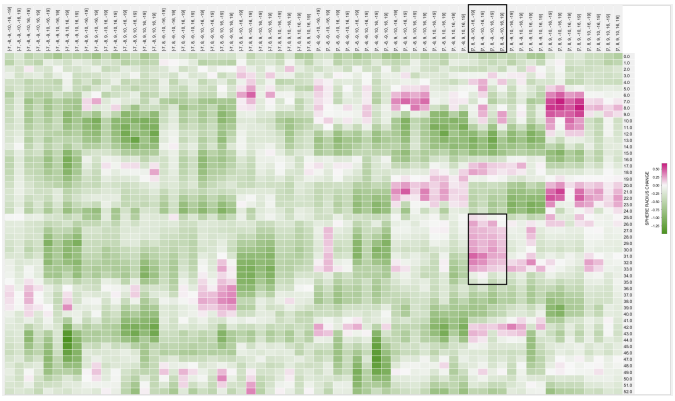

Fig. 3: The selected Combined Normal Mode View. Columns describe the mode phases representing arbitrary directions for each of the normal mode in the dataset, while rows indicate the tunnel sphere numbers from the calculation starting point (in) to tunnel exit (out). Sphere radius difference values are displayed on the chart, with strong violet representing an increase and green a decrease. Opening spheres of the tunnel mid-section and their corresponding mode combinations are highlighted with additional black rectangles.

The x-axis displays each considered combination of the previously selected normal modes while considering their different positive or negative phases. The y-axis represents each sphere forming the tunnel, starting inside of the protein (top) to the outer environment (bottom). Consequently, biochemists have access to an exhaustive view of the reduced space formed by the selected relevant normal modes influencing the tunnel geometry and can quickly verify the putative tunnel-wide dynamics as well as the corresponding protein-wide dynamics on the linked 3D View.

4.3. 3D View

While the SNMV and CNMV, described in previous sections, provides an overview of the normal mode space exploration, they also completely abstract the spatial characteristics of the protein. Therefore, the 3D View can be used to strengthen the understanding of the influence of the normal modes on the tunnel and allows to spatially resolve this information (T1, Figure 2e). We designed our tool to allow the biochemists, by interacting with the other views, to access and depict the amino acids surrounding a individual tunnel spheres. Further, any geometrical measurement is directly mapped onto the tunnel surface, thus unlocking the precise spatial resolution of an influential normal mode on the tunnel (T4, T5). We used color to

represent this quantitative attribute, as position, shape, and size of the tunnel spheres are reserved for tunnel visualization.

Additionally, the 3D View functionalities also focus on protein normal mode features and their analysis. As mentioned in Section 2, the squared fluctuations inform the biochemists about the variance in vibrational motions under a single normal mode or a combination of modes (see Figure 4b and Equation 4). We choose to encode this quantitative attribute with color, as position, shape, and size of structure atoms are reserved for other biochemical attributes. Here, red corresponds to low and blue corresponds to high squared fluctuation values, respectively. The squared fluctuations grants fast and effective access to overall protein flexible motions (T3) and inform biochemists about the prevalent structure mechanisms.

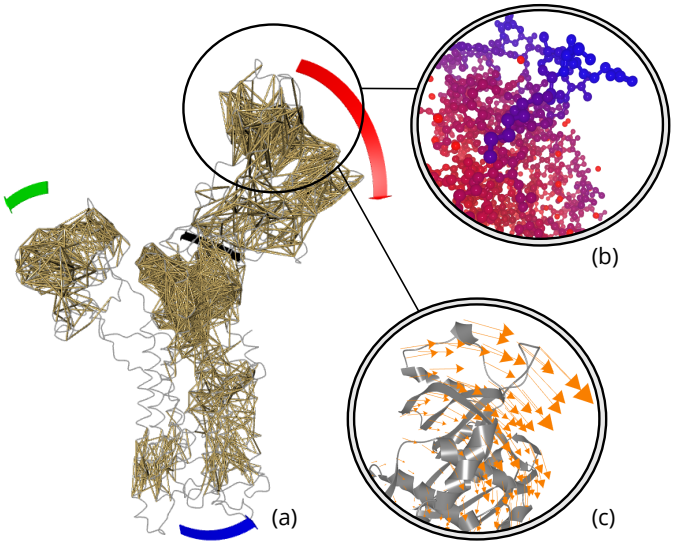

Fig. 4: The protein structure (PDB ID 1SU4) represented as a grey wire trace with the static 3D visualizations of the normal modes. Gray circles depict magnified details for each representation. 2a: The whole loaded normal mode set cross-correlations filtered by 0.97 absolute value. The yellow sticks indicate a strong correlation of the headpieces and that the individual helices are forming correlated clusters throughout the transmembrane region (bottom part). Additional affine arrows by Bryden et al. [24] indicate the large scale dynamics of the subdomains, here for mode 7. 2b: mode 7 squared atomic fluctuation values mapped onto the structure - blue displays high movement variances and red low movement variances. 2c: mode 7 vector field arrow representation. Arrow items are colored in bright orange for visibility and their magnitude describes the motion amplitudes while the protein is represented as gray secondary structure elements. Alternative representations can be interactively switched or combined.

Subsequently, the selected mode vector field representation is available (T1, T3). Even though it is tedious to fully grasp the motions solely from the vector field, this visualization enhances the previously presented fluctuations as it enriches the fluctuation color mapping by providing a notion of the direction and can also be used when an animation is not possible (e.g., paper publication) (see Figure 4c). The displacement vectors are displayed for each amino acid at atoms used during the normal mode calculation (e.g., α-carbon atoms) and are represented by orange arrows in the direction of the vibrational displacements.

Alternatively, and to circumvent the shortcomings of state-of-the-art vector field representation, users have access to affine arrows from the method introduced by Bryden et al. [24], which model the motion by aggregating the arrow vectors of each amino-acid in the different protein subdomains (see Figure 4a).

Also, all the normal mode amino acid cross-correlations are available and can be displayed onto the protein structure (T2) (see Figure 4a). The correlated amino acids are linked in the 3D View by sticks. Users can choose a threshold to display only the highest correlated motion values.

While the biochemists can depict static information regarding the normal modes, by coloring the protein structure or tunnel it the 3D View, they also require more dynamical analyses, which are described in the next subsection, to better understand motions and mechanisms.

4.4. Vibrational Trajectories

As already discussed, the biochemists are often interested in protein motions when studying its function (T1). The state-of-the-art fashion of visualizing atom motions described by the normal modes is through examining their vector fields. However, understanding the protein mechanisms purely from vector arrows is not an easy task. Therefore, we generate protein trajectories to animate the normal mode harmonic motions of the atoms.

The normal modes vectors are obtained from atomic displacements around the initial structure. Thus, the obtained vectors indicate a direction in the 3D space for a vibrational motion rather than an actual displacement. Taking this into account, the trajectory for a single normal mode is generated as follows. Each particle is iteratively displaced in the normal mode direction using an increasingly larger scaling factor in each step to ultimately reach the user-selected scaling amplitude and thus create a trajectory. The number of steps (or snapshots) is a user-defined parameter and represents a simple sampling along the linear vectors. By default, we generate 4 snapshots in the positive directions and 4 in the negative ones. The initial structure is placed in the middle of the resulting trajectory.

To gain insight into the studied protein function, biochemists require analyzing the influence of normal modes on tunnels (T3). As presented in Section 2, normal mode data can be obtained from all-atom or coarse-grained (CG) computations. However, to compute the protein tunnels, we need information about the movement of all atoms. As this information is not present in CG computations, we designed an algorithm to reconstruct the movements of the missing atoms. It is done as follows; for each α-carbon atom along the protein sequence, its two respective neighbors on both sides are selected to compute the transformation between the corresponding atoms in the original and generated trajectory snapshot. The transformation is then applied to all the amino acid atoms around the currently evaluated α-carbon atom.

As explained before, it is not enough to consider normal modes separately but the domain experts also need to investigate their combinations (T1). To generate the vibrational trajectories for combinations of normal modes, we use a similar approach. The only difference is that, before the vibrational trajectory creation, the normal mode directions are combined (see Equation 3) and the resulting direction is used.

5. Case Study

In evaluating the usefulness of the proposed solution, we focused on a total of three case studies – one in the main text and two in the supplementary material. We started with SERCA1 Ca-ATPase and α7 Nicotinic Acetylcholine Receptor. To thoroughly compare our results with the initial work, we made sure that the PDB structures we used in these two studies were the same as the ones in the respective literature. This way, we avoided any potential bias when comparing our results. We used only one initial conformation for the analysis in order to prove the ability of our tool to achieve the same results as conventional tools that could require multiple input conformations (T6). Once validated on the original data, we used our approach also on the GLIC ligand-gated ion channel, with more recent structures, to investigate whether our method captures the modes involved in transitioning from the closed to the open conformation of the protein. The results from the second and third case studies are described in the supplementary material.

The normal modes for all proteins were computed using WEBnm@ [21] tool which uses coarse-grain normal mode calculation and the dataset consists of 200 normal modes and their frequencies, in a format as presented in Section 2.

As described in Section 2 structural biochemists are investigating cues on the motions of protein sub-parts to conclude on their mechanisms of actions. We thus investigated the dynamics of the protein and tried to match them with possible functional displacements. To achieve so, we looked at which of the protein domains were highly influenced by the protein flexibility (T1). Also, in bridging between protein dynamics and function, we were interested in the correlated nature of the displacements to confirm or infer concerted protein mechanisms (T2). To further research any functional link, we computed the relevant tunnel using Caver Analyst [16] and focused on the sorting features to explore the influence of various combinations of normal modes on the known functional tunnel (T3).

In the case studies, the obtained results were systematically compared with the literature in order to assess the tool's correctness. This analysis proved the capability of our solution to support domain experts in identifying possibly functional modes from a single structure (T4, T6). Also, upon different putative results presented below, we were able to detect new hypothesizes regarding the metabolite access mechanisms inside the core of the protein. These results were, due to the limitations of the existing workflows, either not considered or considered merely briefly without any quantification in the previous studies.

5.1. SERCA1 Ca2+ ATPase protein

The transmembrane protein SERCA1 Ca-ATPase (PDB ID 1SU4) is responsible for moving two calcium ions, through its core, from the inside of the cell to the outside. It consists of many different functional domains which can be mainly separated into two groups; the upper part of the structure, inside the cell (called the headpieces, see Figure 4) controlling the molecular machinery, and the transmembrane helices, the lower part of the structure, containing the ion pathways. The main tunnel pathway was chosen and computed based on the work of Møller et al. [48] and Bublitz et al. [49].

Firstly, concerning the link between the protein function and the loaded normal modes, we analyzed how the different dynamics collaborate on the protein-wide scale (T2). Here we used our tool to compute and display the correlations between the movement of individual amino acids considering all normal modes. Through the 3D View, we could easily identify the link between the movements of the headpieces and the motions of the transmembrane helix clusters, as the dynamical domains stand out with the correlations (see Figure 4a). We thus inferred the role of actuators for the headpieces in linking the displacements down through the protein transmembrane core.

As a second step, we needed to investigate the displacements of the headpieces, and thus link them to functional modes to understand the intrinsic mechanisms of the metabolite transfer through the protein (T1). Knowing that the first non-trivial normal modes (sorted by their frequency) are always the most important [6], we visualized their squared atomic fluctuations separately on the 3D View. In comparison to the existing approach where the fluctuation is depicted in line chart as a function of amino acids (e.g., WEBnm@ [21]), our approach using color mapping onto the 3D structure (see Figure 4b) provides information about the parts of the molecule that are influenced by the individual normal modes without the necessity to know their spatial position a priori. Here we identified modes 7, 8 impacting the headpiece of the upper right domain, whilst modes 9, 10, 11, and 12 are influencing the headpiece of the upper left domain (for mode 7, see Figure 4b, blue color part).

As mentioned earlier, the lowest frequency normal modes are well related to functional motions. Nevertheless, in order to understand how the protein works, we needed to isolate the most functional normal modes and by doing so, identify which protein domain motions could play a key role in its function (T4). Since the protein tunnel is directly responsible for the protein function, we used the tunnel dynamics as the guidance for the analysis. We used the tunnel squared overlap measurement presented in Section 2 to identify relevant modes. In Figure 5, several non-trivial modes are identifiable — with strong purple colors and values of localized squared overlap. The heatmap representation strengthens the overall analysis of the normal modes in comparison with existing techniques as it is more straightforward to identify the modes strongly impacting each sphere of the selected functional tunnel (T4) without having to precisely analyze the modes separately. Even more so, in using the filtering and sorting features in the SNMV it was easy to identify the key normal modes. Here, the normal modes 7, 8, 9, 10, 13, 14, 15, and 19 were resulting from a filter threshold of squared overlap values. As the values are depicted for each sphere, we could easily observe which of these modes is influencing the biologically relevant tunnel mid-section susceptible to play a role in the metabolite pathing (spheres 24 to 42). Among these modes, the ones numbered 9 and 10 were the most impactful for the tunnel mid-section, stressing their role in mediating the ion access to the protein core.

The same normal modes were identified in the work of Reuter et al. [50], especially modes 9 and 10, described by the litera-

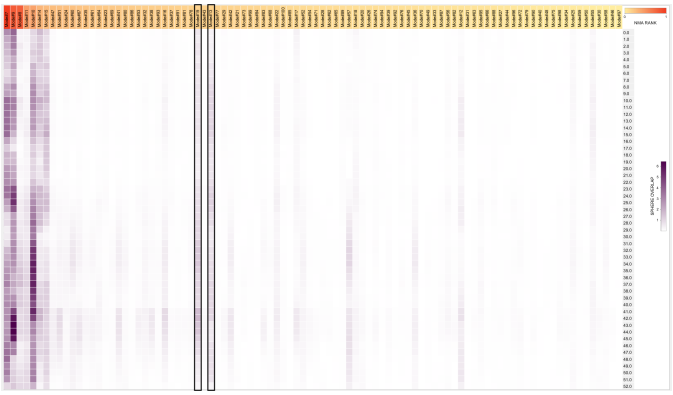

Fig. 5: Single Normal Mode View of the dataset. Columns describe the modes while rows indicate the tunnel sphere numbers, from the calculation starting point (in) to tunnel exit (out). Squared overlap values are displayed on the chart with strong purple representing high mode involvement. The sorting is applied, with modes ranked and colored by their average involvement over the whole tunnel. Modes 13 and 19 are highlighted with additional black rectangles.

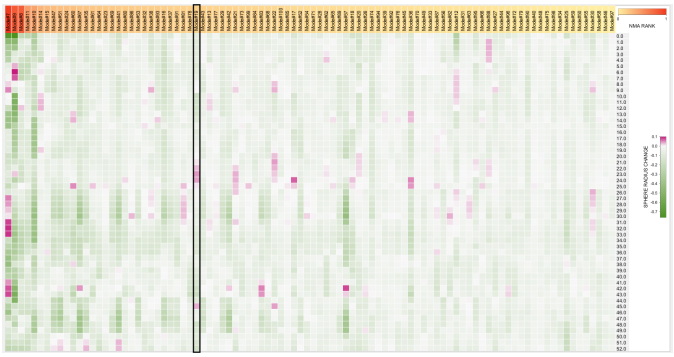

Fig. 6: Single Normal Mode View of the dataset. Columns describe the mode indices while rows indicate the tunnel sphere numbers, from the calculation starting point (in) to tunnel exit (out). Tunnel sphere radius change values are displayed on the chart with strong violet representing an increase and green a decrease. The sorting is applied, with mode indices ranked and colored by their average involvement over the whole tunnel. Opening spheres of the tunnel mid-section and mode 19 are highlighted with an additional black rectangle.

ture as twisting the transmembrane helices for the former, and displacing the headpiece domain for the latter. However, to relate the observed displacements to the protein function, Reuter et al. [50] studied the contribution of each low-frequency mode to the difference between two known active conformations of the studied protein. It resulted only in modes 7, 8, and 15 contributing the most to the functional change of protein conformations. Although precisely describing the observed motions of the aforementioned modes 9 and 10, in linking the modes to the structural protein domains, in their study, the authors could only formulate hypotheses about their quantified involvement with the protein function as they were not highlighted by the contribution method used. In utilizing our tool to isolate the highest values of the squared overlap, we could perform the analysis faster and only from one loaded protein structure and its normal mode calculation. Moreover, we were able to easily localize and quantify the influence of the normal modes on specific parts of the tunnel which was not possible before without tedious manual work.

Finally, to further steer the exploration of the large normal mode space, we utilized the sorting approach of the modes in the SNMV. As already mentioned, Figure 5 presents the ranking of the normal modes based on their impact on the tunnel. Here the modes 7, 9, 8, 11, and 10 have the highest contribution values. Using this view we also isolated modes 19 and 13 exhibiting a high squared overlap per each tunnel spheres but being ranked quite low on the sorting with respect to the overlap value for the whole tunnel. As we wanted to explain this behavior, we switched the coloring of the chart to tunnel sphere radius changes. We could immediately see that the mode 19 is displaying localized positive and negative radius values in the vicinity of the functional tunnel mid-section (see Figure 6). Identifying such mode within the whole normal mode set – let alone describing its possible combination with other modes – would be impossible with the current state-of-the-art normal mode visualization tools.

As such behavior is of possible biological relevance we decided to study the mode 19 further. We thus wanted to scrutinize precisely the most favorable normal mode combinations that would enhance the motion described by mode 19. We utilized the alternative sorting approach (described in Section 2) to highlight the other normal modes based on their calculated similarity with mode 19. The system recommended a subset of modes, with modes 7, 9, 8, and 16 for their probable enhancement of mode 19, and we also selected the mode 10 as it stood out as impacting the tunnel mid-section from the above presented analysis.

To further analyze the narrowed normal mode combinations and their impact on both the overall protein structure and its functional tunnel, we displayed the Combined Normal Mode View (Figure 3) with the tunnel sphere radius change measurement. In comparison to the existing approach where the domain experts generate different protein models with multiple amplitudes and phases using modeling tools (e.g., [13]) to visualize the possible impact of a combination on the protein structure in another tool (e.g., [25]), our solution provides information about the parts of molecule and its tunnel that are influenced by the combination of normal modes in an integrated feedback loop where users can select and modify their normal mode combination choice and directly visualize their impacts on the protein without the necessity of a tedious analysis across several tools. Here, we focused on the first tunnel spheres, close to the ion-bindings site, and the midsection of the tunnel (spheres numbers 24 to 42), representing the biological relevant tunnel mid-section. Ultimately, we could isolate the normal mode combinations displaying the strongest opening of the particular spheres (high violet values in Figure 3) and visualize the putative realistic motions on the linked 3D View to further explore a precise metabolite gating and binding mechanism.

6. Conclusion and future work

Our solution deals with high dimensional data and allows its exploration to gradually build an understanding of field-related concepts. It was developed with biochemists' tasks in mind. We implemented a set of linked 2D and 3D views to ease their work; without the necessity of using multiple tools separately to assess protein functional mechanisms. To demonstrate the usefulness of our tool, field experts conducted several case studies

using a dataset of protein structures and their normal mode calculations. The case studies corroborated that our solution can be used to address the identified tasks (T1-T6).

However, interactive and computationally cost-effective analysis often comes at the expense of limitations. Despite displaying the normal mode properties for fast access to dynamical studies, and comparing them with one another to infer the most functionally relevant ones, we have to stress that our system is oriented on the datasets involving key protein tunnels for metabolite pathing. Also, some preliminary studies are required to investigate the protein atoms involved in the pathing mechanism before being able to probe the protein larger functional motions. Nevertheless, the power of our tool can be used in a two-fold manner. Either the biochemists know the functional tunnel before the analysis and want to draw conclusions on the protein mechanics and large functional motions, as described in Section 5. Or, if the field experts do not know the precise location of any functional tunnel, upon running a tunnel computation (using Caver Analyst or MOLE 2.0 [16, 15]) our tool unlocks the possibility to quickly analyze these tunnels based on the protein flexibility as described by the normal modes. This allows biochemists to further explore individual tunnels and draw conclusions on the metabolite pathing from the generated realistic motions. Also, we believe that several techniques described in this paper are of interest to the broader visualization community as the normal mode decomposition is used in other domains than biology (e.g., physics). Namely, the way we designed the predictors for determining the normal mode influence on the tunnel (i.e., using overlap computation) could be easily extended to other problems as long as the required behavior can be described by vector fields. Alternatively, our fast algorithm for tunnel reconstruction can be of interest to other visualization researchers working on representing tunnels in molecular dynamics simulations as it provides an approximation of the tunnel dynamics.

For the future, we plan to further extend the Combined Normal Mode View. Indeed, in this view, all the selected mode combinations are scrutinized, making it a more computationally challenging task for a larger number of selected modes. We plan to investigate a possible smart sampling, biased with the most prominent mode combinations, that will further narrow the relevant normal mode space which will be strongly linked to the protein function.

References

[1] Hensen, U, Meyer, T, Haas, J, Rex, R, Vriend, G, Grubmüller, H. Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. PloS one 2012;7(5):e33931.

[2] Bahar, I, Lezon, TR, Yang, LW, Eyal, E. Global dynamics of proteins: bridging between structure and function. Annual review of biophysics 2010;39:23–42.

[3] Shaw, DE, Maragakis, P, Lindorff-Larsen, K, Piana, S, Dror, RO, Eastwood, MP, et al. Atomic-level characterization of the structural dynamics of proteins. Science 2010;330(6002):341–346.

[4] Tama, F, Sanejouand, YH. Conformational change of proteins arising from normal mode calculations. Protein engineering 2001;14(1):1–6.

[5] Cui, Q, Bahar, I. Normal mode analysis: theory and applications to biological and chemical systems. CRC press; 2005.

[6] Hinsen, K, Thomas, A, Field, MJ. Analysis of domain motions in large proteins. Proteins: Structure, Function, and Bioinformatics 1999;34(3):369–382.

[7] Lukman, S, Grant, GH. A network of dynamically conserved residues deciphers the motions of maltose transporter. Proteins: Structure, Function, and Bioinformatics 2009;76(3):588–597.

[8] Piazza, F, Sanejouand, YH. Long-range energy transfer in proteins. Physical biology 2009;6(4):046014.

[9] Motlagh, HN, Wrabl, JO, Li, J, Hilser, VJ. The ensemble nature of allostery. Nature 2014;508(7496):331.

[10] Ichiye, T, Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins: Structure, Function, and Bioinformatics 1991;11(3):205–217.

[11] López-Blanco, JR, Aliaga, JI, Quintana-Ortí, ES, Chacón, P. imods: internal coordinates normal mode analysis server. Nucleic acids research 2014;42(W1):W271–W276.

[12] Bakan, A, Meireles, LM, Bahar, I. Prody: protein dynamics inferred from theory and experiments. Bioinformatics 2011;27(11):1575–1577.

[13] Brooks, BR, Brooks III, CL, Mackerell Jr, AD, Nilsson, L, Petrella, RJ, Roux, B, et al. Charmm: the biomolecular simulation program. Journal of computational chemistry 2009;30(10):1545–1614.

[14] Hinsen, K. The molecular modeling toolkit: a new approach to molecular simulations. Journal of Computational Chemistry 2000;21(2):79–85.

[15] Sehnal, D, Vařeková, RS, Berka, K, Pravda, L, Navrátilová, V, Banáš, P, et al. Mole 2.0: advanced approach for analysis of biomacromolecular channels. Journal of cheminformatics 2013;5(1):39.

[16] Jurcik, A, Bednar, D, Byska, J, Marques, SM, Furmanova, K, Daniel, L, et al. Caver analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. Bioinformatics 2018;34(20):3586–3588.

[17] Gouaux, E, MacKinnon, R. Principles of selective ion transport in channels and pumps. science 2005;310(5753):1461–1465.

[18] Pao, SS, Paulsen, IT, Saier, MH. Major facilitator superfamily. Microbiol Mol Biol Rev 1998;62(1):1–34.

[19] Hinsen, K, Kneller, GR. Projection methods for the analysis of complex motions in macromolecules. Molecular Simulation 2000;23(4-5):275–292.

[20] Munzner, T. A nested model for visualization design and validation. IEEE Transactions on Visualization and Computer Graphics 2009;15(6):921–928. URL: http://dx.doi.org/10.1109/TVCG.2009.111. doi:10.1109/TVCG.2009.111.

[21] Tiwari, SP, Fuglebakk, E, Hollup, SM, Skjærven, L, Cragnolini, T, Grindhaug, SH, et al. Webnm@ v2. 0: Web server and services for comparing protein flexibility. BMC bioinformatics 2014;15(1):427.

[22] Suhre, K, Sanejouand, YH. Elnemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic acids research 2004;32(suppl_2):W610–W614.

[23] 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(suppl_2):W52–W56.

[24] Bryden, A, Phillips, G, Gleicher, M. Automated illustration of molecular flexibility. IEEE Transactions on Visualization and Computer Graphics 2010;18(1):132–145.

[25] Schrödinger, LLC, . The PyMOL molecular graphics system, version 1.8. 2015.

[26] Orellana, L, Yoluk, O, Carrillo, O, Orozco, M, Lindahl, E. Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. Nature communications 2016;7:12575.

[27] Al-Bluwi, I, Vaisset, M, Siméon, T, Cortés, J. Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods. BMC structural biology 2013;13(1):S2.

[28] Rubinstein, JL. Cryo-em captures the dynamics of ion channel opening. Cell 2017;168(3):341–343.

[29] Cheng, X, Lu, B, Grant, B, Law, RJ, McCammon, JA. Channel opening motion of $\alpha 7$ nicotinic acetylcholine receptor as suggested by normal mode analysis. Journal of molecular biology 2006;355(2):310–324.

[30] Mahajan, S, Sanejouand, YH. Jumping between protein con-

formers using normal modes. Journal of computational chemistry 2017;38(18):1622–1630.
[31] Mongan, J. Interactive essential dynamics. Journal of computer-aided molecular design 2004;18(6):433–436.
[32] Willett, W, Heer, J, Agrawala, M. Scented widgets: Improving navigation cues with embedded visualizations. IEEE Transactions on Visualization and Computer Graphics 2007;13(6):1129–1136.
[33] Krause, J, Perer, A, Bertini, E. Infuse: interactive feature selection for predictive modeling of high dimensional data. IEEE transactions on visualization and computer graphics 2014;20(12):1614–1623.
[34] Gladisch, S, Schumann, H, Tominski, C. Navigation recommendations for exploring hierarchical graphs. In: International Symposium on Visual Computing. Springer; 2013, p. 36–47.
[35] Ceneda, D, Gschwandtner, T, May, T, Miksch, S, Schulz, HJ, Streit, M, et al. Characterizing guidance in visual analytics. IEEE Transactions on Visualization and Computer Graphics 2017;23(1):111–120.
[36] Van Wijk, JJ. Views on visualization. IEEE transactions on visualization and computer graphics 2006;12(4):421–432.
[37] Taly, A, Delarue, M, Grutter, T, Nilges, M, Le Novere, N, Corringer, PJ, et al. Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. Biophysical journal 2005;88(6):3954–3965.
[38] Kurkcuoglu, O, Kurkcuoglu, Z, Doruker, P, Jernigan, RL. Collective dynamics of the ribosomal tunnel revealed by elastic network modeling. Proteins: Structure, Function, and Bioinformatics 2009;75(4):837–845.
[39] Krone, M, Kozlíková, B, Lindow, N, Baaden, M, Baum, D, Parulek, J, et al. Visual analysis of biomolecular cavities: State of the art. In: Computer Graphics Forum; vol. 35. Wiley Online Library; 2016, p. 527–551.
[40] Krone, M, Falk, M, Rehm, S, Pleiss, J, Ertl, T. Interactive exploration of protein cavities. In: Computer Graphics Forum; vol. 30. Wiley Online Library; 2011, p. 673–682.
[41] Parulek, J, Turkay, C, Reuter, N, Viola, I. Visual cavity analysis in molecular simulations. BMC bioinformatics 2013;14(19):S4.
[42] Jurčík, A, Parulek, J, Sochor, J, Kozlikova, B. Accelerated visualization of transparent molecular surfaces in molecular dynamics. In: 2016 IEEE Pacific Visualization Symposium (PacificVis). IEEE; 2016, p. 112–119.
[43] Lindow, N, Baum, D, Bondar, AN, Hege, HC. Exploring cavity dynamics in biomolecular systems. BMC bioinformatics 2013;14(19):S5.
[44] Byška, J, Jurčík, A, Gröller, ME, Viola, I, Kozlikova, B. Molecollar and tunnel heat map visualizations for conveying spatio-temporo-chemical properties across and along protein voids. In: Computer Graphics Forum; vol. 34. Wiley Online Library; 2015, p. 1–10.
[45] Byška, J, Le Muzic, M, Gröller, ME, Viola, I, Kozlikova, B. Animoaminominer: Exploration of protein tunnels and their properties in molecular dynamics. IEEE transactions on visualization and computer graphics 2016;22(1):747–756.
[46] Masood, TB, Sandhya, S, Chandra, N, Natarajan, V. Chexvis: a tool for molecular channel extraction and visualization. BMC bioinformatics 2015;16(1):119.
[47] Kolesár, I, Byška, J, Parulek, J, Hauser, H, Kozlíková, B. Unfolding and interactive exploration of protein tunnels and their dynamics. In: Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine. Eurographics Association; 2016, p. 1–10.
[48] Møller, JV, Nissen, P, Sørensen, TL, le Maire, M. Transport mechanism of the sarcoplasmic reticulum ca2+-atpase pump. Current opinion in structural biology 2005;15(4):387–393.
[49] Bublitz, M, Musgaard, M, Poulsen, H, Thøgersen, L, Olesen, C, Schiøtt, B, et al. Ion pathways in the sarcoplasmic reticulum ca2+-atpase. Journal of Biological Chemistry 2013;288(15):10759–10765.
[50] Reuter, N, Hinsen, K, Lacapère, JJ. Transconformations of the serca1 ca-atpase: a normal mode study. Biophysical journal 2003;85(4):2186–2197.