Computational 3D imaging to quantify structural components and assembly of protein networks

Pouyan Asgharzadeh a,b,⇑, Bugra Özdemir c, Ralf Reski c,d,e, Oliver Röhrle a,b, Annette I. Birkhold a,b

a Institute of Applied Mechanics, University of Stuttgart, Stuttgart, Germany
b Stuttgart Research Centre for Simulation Technology (SimTech), Stuttgart, Germany
c Plant Biotechnology, Faculty of Biology, University of Freiburg, Freiburg, Germany
d BIOSS – Centre for Biological Signalling Studies, Freiburg, Germany
e FIT – Freiburg Centre for Interactive Materials and Bioinspired Technologies, Freiburg, Germany

ABSTRACT

Traditionally, protein structures have been described by the secondary structure architecture and fold arrangement. However, the relatively novel method of 3D confocal microscopy of fluorescent-protein-tagged networks in living cells allows resolving the detailed spatial organization of these networks. This provides new possibilities to predict network functionality, as structure and function seem to be linked at various scales. Here, we propose a quantitative approach using 3D confocal microscopy image data to describe protein networks based on their nano-structural characteristics. This analysis is constructed in four steps: (i) Segmentation of the microscopic raw data into a volume model and extraction of a spatial graph representing the protein network. (ii) Quantifying protein network gross morphology using the volume model. (iii) Quantifying protein network components using the spatial graph. (iv) Linking these two scales to obtain insights into network assembly. Here, we quantitatively describe the filamentous temperature sensitive Z protein network of the moss Physcomitrella patens and elucidate relations between network size and assembly details. Future applications will link network structure and functionality by tracking dynamic structural changes over time and comparing different states or types of networks, possibly allowing more precise identification of (mal) functions or the design of protein-engineered biomaterials for applications in regenerative medicine.

Statement of Significance

Protein networks are highly complex and dynamic structures that play various roles in biological environments. Analyzing the detailed spatial structure of these networks may lead to new insight into biological functions and malfunctions. Here, we propose a tool set that extracts structural information at two scales of the protein network and allows therefore to address questions such as “how is the network built?” or “how networks grow?”.

1. Introduction

Examining protein-networks in their natural environment is crucial for understanding their roles in cellular processes. Previous studies, which investigated protein networks, mainly focused on biochemical aspects. The basic building blocks of the cytoskeleton have been identified and characterized extensively in vitro [1–6].

However, how protein networks act as a whole is poorly understood. Protein networks consist of filaments formed by polymerization of different types of sub-units. In eukaryotic cells, three main kinds of filaments, i.e., microfilaments, microtubules, and intermediate filaments, constitute the cytoskeleton. However, cytoskeletal structures also exist in bacteria [7] or in plastids [8,9]. These skeletal structures are active polymer gels, whose component polymers and regulatory proteins are in constant flux, constituting a dynamic and adaptive scaffold. It is a highly dynamic 3D network of filamentous proteins linking cellular components. Protein networks are multifunctional, they spatially organize the
content of cells, provides structural support, connect the cell biologically and physically to its environment [10].

Recently, mechanical forces are increasingly recognized as major regulators of cell structure and function [11,12]. Furthermore, the cytoskeleton generates coordinated forces that enable the cell to move and change shape, therefore the filaments constituting the cytoskeleton contribute to cell mechanics [13–16]. Moreover, physical properties of the network have been recently directly linked to cellular physiology [17–19]. Therefore, the organization of the complex links of the polymers and the architecture of the resulting mechanical framework seems to not only play a central role in transmitting stresses and in sensing the mechanical micro-environment [11], but might also directly contribute to the physiology of the cell. Moreover, pathological changes in cells seem in many cases to be linked to alterations in the cytoskeletal network structure, as altered mechanical properties of cancer cells are assumed to be caused by structural changes in the cytoskeleton network [6,20], and structural changes in the cytoskeleton protein tau [21] seem to be linked to Alzheimer’s disease [22].

Imaging of fluorescent-labeled proteins in living cells is a powerful technique for studying protein network overall shape but also its structural details in a spatial and functional perspective [23–25]. Recent advances within the imaging field, e.g., noninvasive multicolor or 3D imaging at the nanometer scale [25,26], enables the imaging of cytoskeletal structures in detail. Three-dimensional imaging of actin has been performed using both stochastic optical reconstruction microscopy (STORM) [26] and photoactivated localization microscopy (PALM) [27]. STORM and PALM further enabled visualization of filamentous temperature sensitive Z (FtsZ), the bacterial homolog of eukaryotic tubulin [28,29]. Other methods such as stimulated emission depletion microscopy (STED) could resolve neurofilaments [30], keratin filaments [31], and primary cilia [32]. Many studies have captured Z-stacks of images using confocal microscopy, while relatively few studies have analyzed the cytoskeleton in 3D. Structured-illumination microscopy (SIM) has been used to resolve actin filament arrays and microtubules in 3D [33,34]. Additionally, the three-dimensional organization of FtsZ in dividing bacteria could be visualized [35,28,29]. The fast advancing imaging technologies allow recently completely new 3D and time-resolved visualizations of physiological processes [36–40] and are therefore advancing our understanding of protein network morphology and physiology. However, extraction of information about morphology and behavior of these networks is to date largely limited to qualitative observations.

The lack of analytical tools for quantifying the structures remains a bottleneck, as manual analysis of large data sets requires a great amount of time and are prone to bias and error. Previous studies on the automated analysis of protein network data focused mainly on segmentation and extraction of the biopolymer network structures [41–45], tracing the shape of individual filaments in 2D [46] or only on curvature and orientation in 3D [47]. A recent study looking at more details of the network is limited to 2D [48]. However, a computerized analysis of the structure of protein networks in 3D would enable to track dynamical processes or identify pathological changes in an automated manner. Additionally, linking the overall shape of a cell (or plastid) to the organization of its internal supporting network structure would give further insights into cell mechanics.

To enable an enhanced investigation of morphological aspects of protein networks, we present a novel automated image processing method allowing for a detailed quantitative spatial network analysis. This method processes high-resolution 3D image data sets of protein networks to investigate the network structure from two different yet strongly connected perspectives. The geometrical characteristics of the network as a continuous body are separated from the properties of the subunits of the structure and their connections. The first point of view investigates the gross morphology of the network. The introduced descriptors provide a quantitative answer to the question “how does the network look like?”. The second one studies the protein network on a smaller scale with the aim of quantification of the organizational characteristics of the network components and their relative positions, connections and distributions. Therefore, a spatial graph representing the network as a set of nodes, segments and connections, is extracted from the 3D geometry. This part of the quantification investigates the design of the network and aims to answer the question “how is the structure built?”. For both perspectives, a number of robust and quantitative descriptors are introduced to enable a reproducible, quantitative characterization of the organization of protein networks.

The method is introduced and tested by applying it to confocal microscopy images of fluorescent-labeled FtsZ proteins of Physcomitrella patens [49] (Fig. 1a), a homolog of the eukaryotic cytoskeleton protein tubulin. Based on these data, we report the first detailed image-based characterization of a protein network structure on a sub-cellular level using measures extracted from the gross morphology as well as the arrangement of a network components.

2. Materials and methods

2.1. Materials

The developed tool set is tested on n = 9 3D confocal microscopy images of FtsZ1–2 protein networks of chloroplasts of Physcomitrella patens. This protein network is due to its relatively simple and dynamic structure [50] an ideal first application to demonstrate and test the method. Furthermore, the similarity to eukaryotic cytoskeleton proteins, like microtubuli, make it easily adjustable to these more complex network structures, as these are, besides a similar molecular structure, also assembled of the same basic structural units (points, nodes, elements, and segments).

Total RNA was isolated from wild type Physcomitrella patens (“Gransden 2004” ecotype) protonema using TRIzol Reagent (Thermo Fisher Scientific, USA) and used for cDNA synthesis using Superscript III reverse transcriptase (Life Technologies, Carlsbad, CA, USA). The coding sequence of PpFtsZ1–2 was PCR-amplified from this cDNA and cloned into the reporter plasmid pAct5::Linker:EGFP-MAV4 (modified from Kircher et al. [51]) to generate the fusion construct pAct5::PpFtsZ1–2:Linker:EGFP-MAV4. Then, 50 µg of this plasmid was used for the transfection process. The moss material was grown in a bioreactor [52,53] and transferred according to the protocol described by Hohe et al. [54]. The transfected protoplasts were incubated for 24 h in the dark, subsequently being returned to normal conditions (25 ± 1 °C, light-dark regime of 16:8 h light flux of 55 µmol s⁻¹ m⁻² from fluorescent tubes, Philips TL ~ 19–65 W/25).

2.2. Confocal imaging

3D microscopy was performed directly on live cells between the 4th and 7th days after the transfection, using a Leica TCS 8T-WS microscope. Images were generated using HCX PL APO 100x/1.40 oil objective with a zoom factor of 10.6. For the excitation, a white laser adjusted to 488 nm was applied. The detection range was set to 503–552 nm for the EGFP channel and 664–725 nm for the chlorophyll channel. The pinhole was adjusted to 106.1 µm. Resulting images have a voxel size of 21 nm in x–y dimensions and 240 nm in z dimension. The image data sets were subsequently
deconvolved using Huygens Professional software (Scientific Volume Imaging B.V) based on the theoretical point spread function and the Classical Maximum Likelihood Estimation (CMLE) algorithm.

2.3. Image processing

To extract quantitative measures of the network structure from 3D microscopic images we designed an image processing framework containing several steps. First, raw images (Fig. 1b) are segmented using a semi-automatic iterative approach (Fig. 1c); second, a 3D geometric volume model is created and its overall shape is analyzed (Fig. 1d). Third, it is converted into a spatial graph, which allows to extract the network topology (nodes, connections, segments; Fig. 2). From these two representatives of the network, descriptors of the overall shape of the network and descriptors of the detailed morphology of the network and the sub-structure are determined.

2.3.1. Segmentation to extract network

The image is segmented using an adaptive local threshold algorithm based on median value in each 3D window (window size = 10x10x10 voxels and constant value = 10). Next, the remaining filament discontinuities inside the network are manually corrected. Segmentation was performed in FEI Amira 6.2.0 (Thermo Fisher Scientific, USA).

2.3.2. Extraction of network gross morphology

The gross morphology of the network is studied as a whole. Therefore, a solid outer surface is defined for the segmented image to find the volume enclosing the network. First, for each slice of the 3D image stack the convex hull, represented by the smallest convex set containing all the voxels, is determined. The combination of all convex hulls of all slices forms a wrapped hull around the whole network (FEI Amira 6.2.0 (Thermo Fisher Scientific, USA)). Second, instead of the detailed network, the solid outer surface of the network represented by its wrapped hull is analyzed. A shape matrix describing the shape and the orientation of the wrapped hull of the network structure is calculated, adapted from the shape analysis for whole cells and pulmonary systems presented by Mc Creadie et al. [55] and Chandran et al. [56], respectively. Therefore, each voxel is represented as \( X(i) = (x, y, z) \), with \( x, y \) and \( z \) as the coordination of the voxel \( i \). Furthermore, for each voxel, the displacement vector from the center of mass is defined as \( M(i) = X(i) - C \), with \( C \) as the center of the mass. The shape matrix representing the solid outer surface is built as:

\[
S = \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} M_x(i)M_y(i) & M_x(i)M_z(i) & M_x(i)M_y(i) \\ M_y(i)M_x(i) & M_y(i)M_z(i) & M_y(i)M_x(i) \\ M_z(i)M_x(i) & M_z(i)M_y(i) & M_z(i)M_x(i) \end{bmatrix},
\]

with \( n \) as the number of voxels in the segmented image. This \( 3 \times 3 \) matrix is created for the covariance of the coordinates of all voxels of the wrapped hull. This part was done in Matlab 2017a (MathWorks, USA).

2.3.3. Calculation of network shape descriptors

All the following steps are performed using an inhouse Matlab code (Matlab 2017a, MathWorks, USA). Shape descriptors are defined and calculated based on the segmented image, its wrapped hull and its shape matrix to characterize the spatial extensions of the network as a whole:

1. The network volume is defined as

\[
V_{nw} = \bar{X} \star \delta_x \star \delta_y \star \delta_z
\]

Fig. 1. (a) Confocal microscopy images of several fluorescent-labeled FtsZ proteins inside chloroplasts of one Physcomitrella patens cell (voxel size: 101 nm in \( x \) – \( y \) dimensions and 300 nm in \( z \) dimension). (b) Raw 3D image of one protein network of FtsZ (voxel size: 21 nm in \( x \) – \( y \) dimensions and 240 nm in \( z \) dimension). (c) Segmented network. (d) Wrapped hull determined from segmented image. Images in (c) and (d) have the same voxel size as (b).

Fig. 2. Transformation of the Segmented Image into a Spatial Graph. (a) Segmented image of a protein network of FtsZ. (b) The spatial graph extracted of the segmented image. Nodes are shown in light green and segments are shown in a yellow → red color code with red representing thicker segments. (c) Zoomed in part of the network showing individual points in white. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
and calculated from the number of the foreground voxels ($n_{PN}$) of the segmented image. $\delta_1$, $\delta_2$, and $\delta_3$ are the extensions of the voxels in $x$, $y$, and $z$ directions, respectively.

2. The enclosed volume of the network is computed as

$$V_{EN} = n_{EN} \ast \delta_1 \ast \delta_2 \ast \delta_3$$

and defined as the space occupied by the network and the empty space inside the network. It is determined by counting the number of voxels ($n_{EN}$) inside the wrapped hull.

3. The network volume density, $\rho_{PN}$, describes how densely the volume inside the network (wrapped hull) is occupied by material. It is determined as the ratio of the enclosed volume and the network volume

$$\rho_{PN} = \frac{V_{EN}}{V_{PN}}$$

4. The greatest and smallest diameters of the network, $d_{PN}^{max}$ and $d_{PN}^{min}$, respectively, are calculated by scaling the respective eigenvalues of the shape matrix, as introduced for whole cell analysis [55]:

$$d_{PN}^{max} = 2 \sqrt{5}\lambda_3,$$

$$d_{PN}^{min} = 2 \sqrt{5}\lambda_1,$$

where $\lambda_3 > \lambda_2 > \lambda_1$ are the eigenvalues of the diagonalized resulted symmetric shape matrix, which define the ellipsoid axes of the wrapped hull.

Spatial anisotropies in the network shape are quantified by analyzing the ratios between the diameters/eigenvalues and the parameters stretch and oblateness of the network are introduced. These descriptors have been previously presented to analyze the shape of bone cells [57].

5. The stretch of the network, $St_{PN}$, describes the elongation of the protein network and is calculated as the difference between the largest ($\lambda_{3}$) and smallest eigenvalue ($\lambda_{1}$) of the shape matrix, normalized by the largest one:

$$St_{PN} = \frac{\lambda_3 - \lambda_1}{\lambda_3}.$$  

$St_{PN} \in [0, 1]$, 0 corresponds to a perfect sphere and 1 refers to an infinitely stretched object (cylinder).

6. The oblateness of the network, $Ob_{PN}$, is defined as

$$Ob_{PN} = 2 \cdot \frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_1} - 1.$$  

$Ob_{PN} \in [-1, 1]$ classifies rod-like and plate-like structures. If the second eigenvalue, $\lambda_2$, is closer to the greatest eigenvalue, $\lambda_3$, then the object is considered to be elongated. An oblateness value equal to $-1$ indicates a perfect rod and a value of $1$ a perfect plate.

2.3.4. Extraction of a spatial graph

To extract information about the network micro-structure, a transformation to a numerical representative, defined by points, nodes and segments, representing the different elements of the complex network, is performed. This transformation process is built upon the concept of tensostructure and spatial trusses introduced by Ingber et al. for the analysis of cytoskeleton and endothelial mechanotransmission [58,59], adapted from an implementation for extracting network geometry of collagen gels by Stein et al. [60]. First the edge voxels are determined using the gradient $\nabla f$ of the image $f$:

$$\nabla f = \frac{\partial f}{\partial x} e_x + \frac{\partial f}{\partial y} e_y + \frac{\partial f}{\partial z} e_z,$$

where $e_x$, $e_y$, and $e_z$ are unit vectors forming an orthogonal basis. Second, the centerlines of the filamentous structures are identified (Fig. 2b) based on calculating a distance map of all voxels from the nearest edge voxel. Afterwards, points are placed at the centerline of each structure entity [61]. A point is placed at any part of the structure at which a change in either the thickness or the direction of the filament occurs (Fig. 2c). Hence, the distances between the points are based on the resolution of the original image and complexity of the structure. Last, all consecutive points are connected by elements. As a result, the following components are determined to numerically represent the network (Fig. 2b, c):

- Points: The basic entity of the extracted spatial graph. Points are connected through elements.
- Elements: Connection between two points.
- Nodes: Points that are connected to more than two other points.
- Segments: A sequence of elements starting from one node and ending at another node.
- Connection: Intersection of segments.

For further analysis, the following information is extracted at each point: an identification (ID) number, coordination, thickness of the filament at that point and the IDs of the neighboring points to which this point is connected. The network extraction steps were performed in FEI Amira (Thermo Fisher Scientific, USA).

2.3.5. Calculation of network element descriptors

The segmented image and the information from the spatial graph are further analyzed together to quantify details of the network structure details (inhouse Matlab code (Matlab 2017a, MathWorks, USA)).

- Node descriptors (Fig. 3):
  1. Number of nodes in the network, $N_n$,
2. Node thickness $t_n$ is determined by the diameter of the filament at the location of the node, $n_i$.
3. Node density, $\rho_n$, is defined by
   \[ \rho_n = \frac{N_n}{V_{En}}, \]
   and determined by the number of nodes normalized to the volume enclosed by the network (1/µm$^3$). It has to be taken in mind, that this in not the same as the network volume density.
4. Node-to-node distance, $d_{n_i\_n_j}$,
   \[ d_{n_i\_n_j} = ||n_i - n_j||_2 = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}, \]
   is calculated as the Euclidean distance between two neighboring nodes, $n_i$ and $n_j$, with coordinates $x_i$, $y_i$ and $z_i$, and $x_j$, $y_j$ and $z_j$, respectively.
5. The node-to-surface distance, $d_{n_i\_Surface}$, represents the closest distance of the node $n_i$ to the surface of the network. This element descriptor represents the local distribution of the nodes within the network.
6. Compactness of the network, $C_{PN}$, is defined by
   \[ C_{PN} = \frac{d_{n_i\_Center} - d_{n_i\_Surface}}{d_{n_i\_Center}}, \]
   with $d_{n_i\_Center} = \frac{1}{n} \sum_{i=1}^{n} d_{n_i\_Center}$, and $d_{n_i\_Surface} = \frac{1}{n} \sum_{i=1}^{n} d_{n_i\_Surface}$.

   where, $d_{n_i\_Center}$, the node-to-center distance, is the distance of each node to the center of gravity of all nodes. It is calculated according to node-to-surface distance (Eq. (10)) by replacing the second node with the center of gravity. We define the compactness of the network as the difference between the mean distance to the center of gravity and the mean distance to the network surface of all nodes, normalized by the mean distance to the center of gravity; $C_{PN} \in [0, 1]$. For $C_{PN} = 1$, all the nodes are placed at the surface of the network. In contrary, $C_{PN}$ converges towards 0 if all nodes are placed near the center of gravity.
7. Node-to-surface to node-to-center distance ratio, $d_{n_i\_Surface}/d_{n_i\_Center}$, provides information on whether the node is located closer to the surface of the network or to the center of gravity.
   - Segment descriptors (Fig. 4a):
     1. The total number of segments is denoted by $N_s$.
     2. Segment length $L_s$ is defined as
   \[ L_s = \sum_{i=1}^{n_p-1} l_i, \]
   where $l_i$ is the Euclidean distance between two consecutively placed points on the segment, $S$, and $n_p$ is the number of the points forming segment $S$.
3. Segment curvature $k_s$ is determined as the menger curvature, i.e.,
   \[ k_s = \frac{2 | \mathbf{v}_{p_1} \times \mathbf{v}_{p_2} |}{| \mathbf{v}_{p_1} || \mathbf{v}_{p_2} || \mathbf{v}_{mn} |}, \]
   where, $n_1$ and $n_2$ are the nodes at the start and end of the segment and $p$ is the point on the segment which has the greatest distance to the straight line from $n_1$ to $n_2$. These three points ($n_1, n_2$ and $p$) form a triangle (shown in blue in Fig. 4a). Segment curvature is calculated as radius of the circle passing through these three points by calculating $\mathbf{v}_{p_1}$, $\mathbf{v}_{p_2}$ and $\mathbf{v}_{mn}$ as vectors from $p$ to $n_1$, from $p$ to $n_2$ and from $n_1$ to $n_2$ respectively.
4. Mean segment thickness $t_{th}$ is given by
   \[ t_{th} = \frac{1}{n_p} \sum_{i=1}^{n_p} t_{th_i}, \]
   where $n_p$ is the number of points on the segment and $t_{th}$ is the thickness of the filament at point $i$ on the segment.
5. Segment inhomogeneity $I$ characterizes how much the segment geometry changes along its length. This can be extracted from the number of points, as these represent the locations where the segment changes in terms of thickness or direction:
   \[ I = n_p, \]
   where $n_p$ is the number of points on a segment and $I$ serves as a quantifier of eccentricity of the points on the segment. A higher number of points on a segment means that the segment exhibits more deviations in thickness and direction.
6. Mean point-to-point distance $d_{p\_p}$ denotes the mean of all distances between two consecutive points on a segment.
   - Connection descriptors (Fig. 4b):
     1. The mean number of connections per node, $n_{nc}$, is the number of segments starting/ending at a specific node. Nodes which connect to only one other node are not included in this measure.

Fig. 4. Segment and connection descriptors. (a) Local thickness of the segments $t_{th}$, are shown in a yellow - red color code with red representing thicker segments and the diameter of the representing line element. The zoomed part of the segment marked by green color shows how the length of the segment is calculated by the summation of the distances between the points on that segment. Curvature of a segment is shown by the purple arrow. The triangle formed by starting and ending nodes and the point on the segment with greatest distance from the straight line crossing the two nodes is shown by blue. (b) Types of connections and evaluation of angles in a sample protein network. Examples of nodes with one, three, four and five connections are marked with their numbers, respectively. The blue arrows show the angles between the 3 vectors of a node with 3 connections. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
2. Open nodes, $n_{oe}$, denotes the percentage of all nodes in the network that are only connected to one other node, i.e., nodes that are end nodes.

3. The mean angles between the segments at a connection, $\theta_i$, is defined as

$$\bar{\theta}_i = \frac{1}{n} \sum_{j=1}^{n} \theta_j,$$

with,

$$\theta_j = \arccos \left( \frac{\mathbf{v}_k \cdot \mathbf{v}_l}{|\mathbf{v}_k| |\mathbf{v}_l|} \right), \quad k, l \in \{1, \ldots, n\},$$

where $\bar{\theta}_i$ is the average of angles ($\theta_j$) between segments meeting in one connection. Angle $\theta_j$ is evaluated by calculating the angle between the vectors $\mathbf{v}_k$ and $\mathbf{v}_l$, which start from node $n_i$ located at the center of the connection and go to the first point on each of the meeting segments. Hence, for a node with three connections, three different angles for each pair of outgoing segments (three pairs) are calculated as $\theta_j$, therefore, $\bar{\theta}_i$ is the mean value of these 3 calculated angles (Fig. 4b, indicated in blue).

Descriptors are calculated using an inhouse Matlab code (Matlab 2017a, MathWorks, USA).

### 2.4. Validation of image processing

We validate the proposed workflow by comparing data determined manually and the computed values using the analytical tools developed. Therefore, manual segmentation and node extraction is performed (Fig. 5a) for all data sets.

#### 2.4.1. Validation of segmentation

Although several segmentation strategies exist [42,44], to date no gold standard for this specific segmentation problem has been developed. However, the focus of this study lies on the quantification of structural characteristics of protein networks and not on advancing the existing segmentation techniques. We chose here a relatively simple segmentation approach based on an adaptive threshold. However, further calculated parameters depend on the segmentation outcome. Therefore, we performed a validation of this part. To do so, the images have been manually segmented (MS) and the segmentation results are compared to the results of the segmentation using the local adaptive threshold (AS) technique. Segmentation results were compared using a Bland Altman diagram [62], where mean values of the two techniques are plotted on the x-axis and the difference between them is plotted on the y-axis.

#### 2.4.2. Validation of node identification and placement

Furthermore, the node identification/placement is a crucial step, as the quantification of the element descriptors depends on it. Therefore, the nodes in all nine networks have been manually extracted (MNE) and are taken as the reference (ground truth)
for the node placement step (ANE) of the proposed method (Fig. 5a). All further descriptors are directly derived from these two intermediate steps and are therefore not validated here.

First, the number of extracted nodes in each network is determined. Second, we calculated the vertex error

\[
E_v = \frac{1}{2} \sum_{m=1}^{N_m} \min_{p \in N_m} \| p_m - p_a \| + \frac{1}{2} \sum_{a=1}^{N_a} \min_{p \in N_a} \| p_a - p_m \|,
\]

and Hausdorff distance

\[
H_D = \max \left\{ \max_{m \in N_m} \min_{p \in N_a} \| p_m - p_a \|, \max_{a \in N_a} \min_{p \in N_m} \| p_a - p_m \| \right\},
\]

between the nodes in the manually and automatically extracted networks. \(N_m\) and \(N_a\) denote manually and automatically extracted networks, \(p_m\) and \(p_a\) are the nodes in the manually and automatically extracted networks, respectively. Third, the distances between the nodes in the manually and the automatically extracted networks have been calculated and a color-code is assigned. The value of the cell \((m, n)\) represents the distance between the \(n\)th manually extracted and \(m\)th automatically extracted node. Last, the relation between the Hausdorff distance and number of the nodes in the network, which is directly related to the size and complexity of the network, has been analyzed.

2.5. Statistical analysis

To identify relationships between calculated descriptors regression analysis were performed and Pearson correlation coefficients reported. Unless otherwise indicated, all results are presented as mean ± standard deviation (SD). All statistical tests were performed in Matlab (Matlab, 2017a, MathWorks, USA).

3. Results

All image data sets were analyzed for validation, then shape and element descriptors were calculated.

3.1. Validation

3.1.1. Segmentation

Manually segmentation resulted in a trend of 5 ± 5\% (1100 voxels) on average greater segmented image (\(p = 0.06\)). For 8 of the 9 images the differences in voxel number lie in an acceptable range of ±1.96 SD (Fig. 5b). Therefore, we consider this error for the purpose of the current study as acceptable.

3.1.2. Node identification and placement

No significant differences between the node numbers determined manually and with the automated method were determined, only a trend of 4.2 more nodes using the proposed automated method (\(p = 0.33\)) was found. For 8 out of 9 networks the differences are in the acceptable range of ±1.96 SD (Fig. 5c). The vertex error is 0.19 ± 0.4 μm and the Hausdorff distance is 0.94 ± 0.20 μm, respectively. Considering the voxel sizes, these values lie in the range of the errors calculated by Xu et al. [47] and Xu et al. [44]: [1.08 ± 1.15] and [1.10] pixels. In all networks, except network 2 (8 out of 9), for each node in the manually

![Fig. 6. Network shape descriptors. (a) Raw image of a representative sample network of FtsZ. (b–d) Wrapped hull of the same protein network shown from three perspectives with the three diameters of the network shown as green arrows. The diameters were extracted from the ellipsoid, defined by the shape matrix of the outer surface. (e) Enclosed volume of the network and network volume. (f) Network volume density. (g) The greatest and smallest diameters of the networks. (h) Stretch and oblateness of the networks. \(n = 9\), data shown as mean ± standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
3.2. What is the shape of the network?

Primary gross morphology features of the network as a whole were evaluated (Fig. 6). A wrapped hull of a network and its three diameters are shown for a sample network in Fig. 6a–d. The extracted network, there exists a node in the automatically extracted network within a distance of 2 µm (dark blue color; Fig. 5d)). This confirms the acceptable error for the method’s ability to place nodes. When the number of nodes increases, the Hausdorff distance tends to merge towards a value less than 1 µm, where the second derivative of the second order polynomial curve \( f(x) = -0.001x^2 + 1.33x + 768 \) and \( R^2 = 0.29 \) fitted to the scattered diagram is \(-0.002 \) (Fig. 5e). We assume to have a similar Hausdorff distance for higher node numbers, which would allow applying this method to more complex protein networks, such as cytoskeletal protein networks. Considering the image resolution and overall size of the networks, the Hausdorff and vertex error values can be considered small enough to allow a quantitative analysis of the images. However, when interpreting the results they have to be taken into account.

### Table 1

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>199 ± 130</td>
</tr>
<tr>
<td>Node density</td>
<td>2.52 ± 0.63 nm²</td>
</tr>
<tr>
<td>Compactness</td>
<td>0.75 ± 0.10</td>
</tr>
<tr>
<td>Node thickness</td>
<td>50 ± 10 nm</td>
</tr>
<tr>
<td>Number of segments</td>
<td>259 ± 168</td>
</tr>
<tr>
<td>Mean segment thickness</td>
<td>20 ± 10 nm</td>
</tr>
<tr>
<td>Mean point-to-point distance</td>
<td>55 ± 10 nm</td>
</tr>
<tr>
<td>Segment inhomogeneity</td>
<td>20.4 ± 3.64</td>
</tr>
<tr>
<td>Number of connections per node</td>
<td>3.11 ± 0.04</td>
</tr>
<tr>
<td>Most observed number of nodes</td>
<td>3 and 4</td>
</tr>
<tr>
<td>Open nodes</td>
<td>8.79 ± 2.38%</td>
</tr>
</tbody>
</table>

3.3. Nodes – basics of the network

Individual nodes are identified and analyzed. For the investigated networks a wide range of number of nodes was found (199 ± 130; Fig. 7a–d), whereas the size of individual nodes (node thickness 50 ± 10 nm) and normalized parameters like the compactness and node density were more homogeneous (0.75 ± 0.10 and 2.52 ± 0.63 node/µm², respectively). Additionally, the relative positions of the nodes in the network give further insight into the construction of the network (Table 1). Connected nodes have been identified and the average node-to-node distance (0.57 ± 0.05 µm) was determined. Analyzing the distribution of distances showed, that 90% of all nodes are closer than 0.93 µm to their closest neighbor. In contrast, nodes have only an average distance of 0.50 ± 0.20 µm to the surface of the network, with 90% of them being closer than 1.16 µm to the surface (Fig. 7e, f). Moreover, for values of node-to-center distance a considerable variation between the networks was observed (Fig. 7g) whereas the node-to-surface to node-to-center distance ratio (Fig. 7h) showed less variation in different networks (2.20 ± 0.74 µm and 0.28 ± 0.09, respectively).

3.4. Segments – filamentous nature of the network

The analyzed networks had 258 ± 168 segments with a mean thickness of 25 ± 5 nm and a mean length of 0.78 ± 0.07 µm. To
obtain further details, distributions of curvatures, lengths and inhomogeneities of segments have been analyzed (Fig. 8a). The most observed curvature was $4.35 \mu m^2/C0^1$ (Fig. 8b), the distribution of lengths peaks at 0.33 $\mu m$ (Fig. 8c), and the segment inhomogeneity, which quantifies the local changes of properties of a segment, peaks at 9.30 $\mu m$ (Fig. 8d).

3.5. Connections – how is the network built?

In the next step, we investigated the structural characteristics of connections between segments. A closer look at the nodes and the connections where filaments meet, is demonstrated in Fig. 8e for a sample network. The most observed number of connections at a node was 3 and the second most observed was 4. Moreover, in average $3.11 \pm 0.04$ connections exist per node and 8.76$ \pm 2.33\%$ of all nodes were open nodes. The distribution of angles differed depending on the amount of connections at a node. At nodes with 3 connections, a high probability of finding an angle around 90° exists (Fig. 8e), whereas at nodes with 4 connections no preferred angle could be identified (Fig. 8f).

3.6. From small to big networks

Combining information of shape and elements allows to shed light on relative changes in the size of the networks. Investigating the relationship between enclosed volume of the networks and calculated shape descriptors such as stretch and oblateness (Fig. 8g) show low correlations ($R^2 = 0.18$ and $R^2 = 0.00$, respectively). In contrast, a higher correlation exists between the number of the nodes in the network and the distance of the nodes from the center.
(Fig. 8h), showing that a higher number of nodes results in nodes being located further away from the center of mass ($R^2 = 0.81$). The size of the network also correlates strongly with the number of nodes in the network ($R^2 = 0.89$; Fig. 8i). When the network volume is twice as great (e.g. 100 $\rightarrow$ 200 $\mu^3$), also the number of nodes is approximately doubled; but with increasing network size the node placement is not homogeneously propagated, as the average distance of the nodes from the center increased at the same time by a factor of 1.5. Furthermore, the combination of node-to-surface and node-to-center distance measurements (Fig. 7e and g) suggests that the average value of compactness for the data is closer to 1 than to 0. This is confirmed by calculating the average compactness (cf. compactness in Table 1). Besides these global relationships between the size of networks and shape and element descriptors, segment growth can be analyzed on a local scale. On one hand, with longer segments the segment inhomogeneity increases (Fig. 8j). On the other hand, the increase in segment inhomogeneity results in smaller distances between the points within segments (Fig. 8k).

4. Discussion

To date, analysis of protein network structures as basis for cytoskeletal behaviour assessment is mainly qualitative. However, with current advanced microscopy and labeling techniques, localized cytoskeletal behavior can be imaged with great detail, allowing for a quantitative network analysis. Here, we analyzed possibilities to quantitatively describe the structure of protein networks and provided a tool set, which contributes to such a quantitative network analysis. To show the capabilities of this tool set, it was validated using confocal microscopy images of FtsZ.

The observed variations in calculated shape descriptors (Fig. 6e–g) of FtsZ show that although there exist great variations between the sizes of the analyzed networks, the network volume density shows lower deviation from the mean value. This conveys that in networks from small to big, the network volume density does not change remarkably. In cases of ellipsoidal-like network shapes, like for FtsZ [64], morphological characteristics such as diameters of the network and relationships between these calculated descriptors can be analyzed (Fig. 6g). Therefore, stretch and oblateness of the network are identified (Fig. 6h). The stretch values indicate a deviation of the shape of the networks from a sphere to a more stretched geometry while the oblateness measurements convey the tendency toward more rod-shaped networks. However, the greater variation in oblateness shows that stretch is, for the analyzed data sets, a better measure of the shape of the network.

Analysis of the element components of protein network allows detailed quantification of the subunits of the network. This analysis enables to quantitatively describe the characteristics of structural units of the network: nodes, connections and segments. Node-to-surface distances are relatively similar for all analyzed FtsZ networks, the distance of the nodes from the center of mass shows a great variation between networks. Quantifying relative positions of the nodes is in future applications useful in terms of studying dynamic settings of networks and to identify changes that occur in the network as a result of external stimuli on the structure of the network or internally motivated changes.

The shape measures allow to relate changes of the protein network size and shape to alterations of the plastid, or cells in the case of other protein networks, morphology, which will allow in future applications to relate protein network alterations to cellular processes. For example, during cell division the cell, and accordingly the cytoskeleton, elongates from its initial form to a more stretched shape [65]. Furthermore, in cancerous cells, cytoskeleton protein transformations lead to changes in mechanical properties of the cancer cells in contraction, stretchability and deformability [66]. For example, epithelial-to-mesenchymal transition in colorectal cancer cells results in transformation of the cytoskeleton into a spindle-shape losing their polarity [67]. Moreover, propagation of tau filaments from cell to cell in Tauopathic diseases have been shown to result in protein network changes [68,69]. All these processes could be investigated quantitatively with the here introduced shape descriptors in future studies.

Analysis of segments have been previously shown by Smith et al. [42] and Alioscha et al. [46] to link actin filaments fragments in segmented images and by Stein et al. [60] for extracting network geometry of collagen gels. Moreover, Xu et al. [47] analyzed different filamentous protein networks. To obtain a complete protein network analysis method, we included these entities in our shape and element analysis and adapted the calculation of these to our structural quantification method. Analyzing the segments of the network does not only shed light on the local attributes of the filaments, but also allows to draw further conclusions on the mechanical properties of the networks, as e.g. the curvature of the segments allows to quantify mechanical forces on the filament [70] or effects of shear stress [46]. Segment length can be used to track changes that will transform the architecture of the network in a more fundamental way. The angles between the segments meeting in one connection give insight into the shape of individual sub-compartments of the network. Also, it might be applicable to track the effects of external stimuli on the network such as shear forces, which elongate the filaments and change the angles between them in the network [46]. In the here analyzed networks a great variation of angles between connections was observed, which might be caused by such external stimuli. Therefore, tracking structural descriptors might enhance the understanding of the dynamic setup of network structures, by allowing to monitor and quantify changes occurring in various processes that alter the shape of the network.

Linking shape and element descriptors gives insights into how the shape of protein networks changes during different processes without being transformed into a more stretched or plate-like shape. More nodes resulted in nodes being located further away from the center of mass. This shows that FtsZ networks might grow in size by adding nodes in radial direction form the center of gravity close to the surface, consequently resulting in networks with higher volume. Moreover, the relation between the size of the network and number of nodes in the network conveys the possibility of a preferred approach of the network to grow. The network seems to grow in volume by adding more nodes to the existing network instead of preserving the nodes and adding more connections to the existing nodes. Finally, we found, that there exists a direct correlation between the length of the segments and the inhomogeneity of the segments.

Our study has several limitations. We study protein networks in living cells. Therefore, assembly and disassembly is a dynamic process that generates inhomogeneities. Furthermore, the spatial resolution, especially in z-resolution, does not allow to resolve all small structures in the images and therefore, impacts the results as well. However, the fast advancing field of microscopy technology will allow to overcome this limitation in future applications. A main limitation of our method is the segmentation, as it is in most imaging studies. Global threshold algorithms such as the Otsu-Algorithm [71] are commonly used to extract the geometry of filaments [48]. However, the presence of imaging-specific artifacts, like filament discontinuities, high signal-to-noise ratio resulting from presence of overexposed pixels and dynamics of the network, decrease the efficiency of global algorithms. Therefore, we used an adaptive threshold algorithm, which accounts for intensity variations. However, variations in the input image quality effects the segmentation outcomes, therefore a manual correction is applied, which introduces user bias. An enhanced segmentation in future studies could include linkage of filament
fragments, as previously presented for microtubule network architecture phenotypes in fibroblasts [48]. Therefore, care must be taken in the analysis to avoid systematic errors. It should also be taken in mind, that the approach used here might not be directly applicable to other network shapes, as we assume an ellipsoidal shape, approximated by 2 vectors and 3 scalar quantities. For other (more complex) shapes, different shape models have to be used.

5. Conclusion and outlook

The regulations of individual proteins and their functions have been investigated in detail in the past [72]. Here, we presented a method to quantify the structure of protein networks. Proteins forming networks in biological environments are many copies of a few key pieces, which can be assembled into a wide range of structures depending on how the pieces are assembled. The method presented here provides a quantitative tool to investigate this meso-scale of the assembly. This analysis can potentially be used to study temporal evolution of the network by batch processing consecutive frames of a time-lapsed life image sequence. Analyzing the structure of the network in each frame and tracking changes of structural features might allow relating internal and external stimuli to the modifications of the structure by utilizing a network-based simulation model and might contribute to determining functionality of the network. Here, we analyzed the FtsZ protein network structure and revealed aspects of this internal organization. In the future, the developed method can be applied to compare networks to identify relationships between structural and functional differences. Furthermore, more complex networks such as cytoskeletons [73,74] can be analyzed, as these are built of the same basic structural units. This analysis would facilitate understanding the links between the interactions of the individual units of the network and the large-scale cellular behaviors depending on them [13]. Linking network structure and functionality by tracking dynamic structural changes over time and comparing different states or types of networks, may allow to more precisely identify (mal) functions or design protein-engineered biomaterials for applications in regenerative medicine.

Contributions

Study concept and design: PA, OR, AB, RR. Molecular biology, imaging and deconvolution: BO. Contributing software: PA. Analysis of data: PA. Interpretation of data: PA, AB. Writing manuscript: PA. Critical revision: All authors. The developed codes in Matlab and the work flow in Amira will be provided by the corresponding author upon request.

Disclosures

The authors state that they have no conflict of interest.

Acknowledgements

This work has been funded by the German Research Foundation (DFG) as part of the Transregional Collaborative Research Centre (SFB/Transregio) 141 ‘Biological Design and Integrative Structures’, A09 project.

References
