

Visualization and Exploration of 3D Toponome Data

S. Oeltze¹ and P. Klemm¹ and R. Hillert² and B. Preim¹ and W. Schubert²

¹Department of Simulation and Graphics, University of Magdeburg, Germany

²Molecular Pattern Recognition Research Group, University of Magdeburg, Germany

Abstract

The toponome of a cell describes the location and topological distribution of proteins across the cell. In toponomics, the toponome is imaged and its inner structure and its semantics are investigated in order to understand how cells encode different functionalities both in health and disease. Toponome imaging results in complex multi-parameter data composed of a 3D volume per protein affinity reagent. After imaging, the data is binarized such that 1 encodes protein present and 0 encodes protein absent. Biologists are particularly interested in the clustering of these binary protein patterns and in the distribution of clusters across the cell.

We present a volume rendering approach for visualizing all unique protein patterns in 3D. A unique color is dynamically assigned to each pattern such that a sufficient perceptual difference between colors in the current view is guaranteed. We further present techniques for interacting with the view in an exploratory analysis. The biologist may for instance “peel off” clusters thereby revealing occluded cell structures. The 3D view is integrated in a multiple coordinated view system. Peeling off clusters or brushing protein patterns in the view updates all other views. We demonstrate the utility of the view with a cell sample containing lymphocytes.

Categories and Subject Descriptors (according to ACM CCS): J.3 [Computer Applications]: Life and Medical Sciences—Biology and genetics

1. Introduction

While the human genome project has revealed the code for all proteins, the next big challenge is to understand how proteins cooperate in cells and tissues in time and space [Sch10]. Although many details on the molecular function and structure of many proteins are known, their corresponding cellular functions cannot simply be derived. This is due to a dependence of the function on the contextual position of a given protein within a protein network. The *toponome* of a cell describes its functional protein pattern, i.e. the location and topological distribution of proteins. In *toponomics*, the toponome is imaged, explored and analyzed for applications in toxicology, drug development and patient-drug-interaction. Toponome imaging results in complex multi-parameter data composed of a 3D volume per protein affinity reagent. In a post-processing step, the data is binarized such that 1 encodes protein present and 0 encodes protein absent. For each data voxel, a binary vector can be constructed over all volumes, which then encodes the local protein co-mapping. Biologists are interested in answering questions such as: Which proteins co-map with which frequency?,

Where across the cell surface do the binary protein patterns cluster?, and How does the clustering differ from cell to cell? These questions guide our research. In previous work, we developed a graph view which encodes co-mapping proteins and co-mapping frequency [OFH*11].

Here, we present a volume rendering approach that generates an integrated 3D visualization of all protein affinity reagent volumes. It is based on the assignment of a unique color to each unique binary protein pattern. Since several thousands of such patterns may exist in a dataset, we offer a dynamic color range distribution for the current view on the data, e.g., a close-up view of a cell. The algorithm then generates a set of unique, perceptually optimized colors for the currently visible patterns. Thus, the visual differentiability of these patterns is improved. To support an exploratory analysis of the toponome, the 3D view is equipped with several interaction techniques, e.g., *similarity brushing*. Furthermore, it is integrated via *linking* in a multiple coordinated view system. Brushing and linking supports the biologist in deciphering the toponome code. We demonstrate the utility of the 3D view with a cell sample containing lymphocytes.

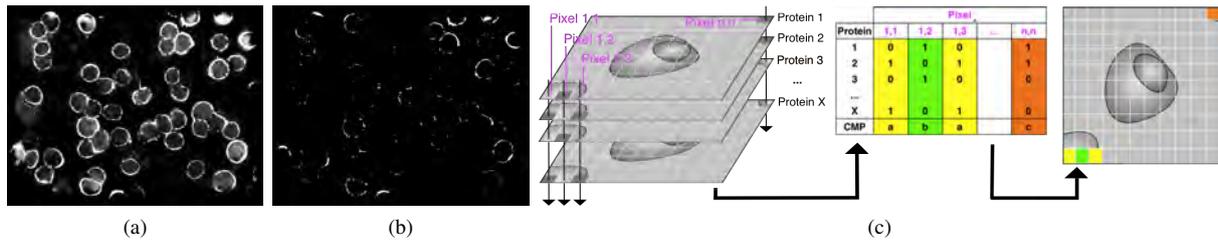


Figure 1: a: Original fluorescence signal in a single slice. b: The same slice as in (a) after thresholding. White pixels indicate protein present. c: Generation of combinatorial molecular phenotypes and display in a 2D toponome map. For each pixel, the binary fluorescence signal of all protein affinity reagents is collected in a combinatorial binary code (table columns). The set of unique binary codes (combinatorial molecular phenotypes) is computed and each code is assigned a color. A toponome map is generated by mapping each pixel to its corresponding color (right subimage).

2. Biological Background

This section explains the toponome and how it is imaged. Further, it introduces *combinatorial molecular phenotypes*, which serve as the input for the 3D visualization.

2.1. The Toponome

The entirety of all protein networks, in which proteins are defined by their protein-to-protein context in any given cell, is defined as the *toponome* [FBKS07, SBP*06, SGK*11]. Its inner structure, its biological code, and its semantics are investigated in *toponomics*. It has been shown that the toponome is hierarchically organized [SBP*06]. It comprises protein clusters which contain *lead proteins* and are interlocked as a network. The lead proteins control the topology of the clusters and their function as a network. However, neither a lead protein nor the protein cluster which is controlled by it can be predicted from molecular data. Thus, the toponome must be imaged in human cells or tissues to understand how cells encode different functionalities both in health and disease. This is essential for finding new drugs in cancer and for detecting protein clusters that can be regarded as a new system of biomarkers in disease [Sch10, SGK*11].

2.2. Imaging the Toponome

The most advanced technique for imaging the toponome is robot-driven multi-parameter fluorescence microscopy TISTM. It is based on a cyclical procedure in which a *tag library* (specific affinity reagents recognizing proteins) is conjugated to one and the same dye. An imaging robot applies these tags to a fixed cell or tissue sample [FBKS07]. It starts with the first dye-conjugated tag and applies it to the sample. The resulting fluorescence image is then captured by a CCD camera, which is connected to an epifluorescence microscope. In the last step of the cycle, the dye is bleached gently to avoid any energy transfer into the remaining proteins [SBP*06]. Then, the second tag is applied and so on.

The labeling of 100 proteins in 100 cycles has been demonstrated [SBP*06, SGK*11]. Imaging can be performed in 3D by modifying the microscope's focal plane. The acquisition of a 3D dataset is costly and may take longer than a day.

2.3. Combinatorial Molecular Phenotypes

After imaging the toponome, a thresholding algorithm is applied to the fluorescence signal (Fig. 1a-1b). This generates a combinatorial binary code for each voxel where 0 indicates protein absent and 1 indicates protein present [BDS10]. The unique binary codes that exist in the data, out of all possible codes, are referred to as *combinatorial molecular phenotypes* (CMPs). A simple technique for visualizing CMPs of a single volume slice is the assignment of a unique color to each CMP and the display of the slice as a colored image also known as *toponome map*. The generation of CMPs and of the toponome map are illustrated by Figure 1c. The binary code that corresponds to a certain CMP very often exists at several pixel/voxel positions. These positions are not randomly spread over the data but clustered at certain locations of the cells. The biologist is interested in these protein clusters since they correspond to functional units of the cells.

3. Related Work

3D microscopy data may consist of a single channel or multiple channels, with each channel showing a different aspect of the data. The most wide-spread visualization technique for 3D microscopy data is direct volume rendering.

3.1. Single-Channel 3D Microscopy Data

Sakas et al. presented a pipeline for visualizing confocal laser scan microscopy (CLSM) data [SVP96]. They integrated Maximum and Minimum Intensity Projection as well as full volume rendering into the rendering step. The volume visualization system VolVis was presented by Kaufman in the context of investigating CLSM datasets of nerve

cells [Kau98]. It supports surface as well as volume rendering and offers global rendering effects such as shadows and reflections. Fang et al. described an approach for microscopy data exploration with a focus on the intuitive design of transfer functions [FDM*00]. They employed an image-based model which defines a transfer function as a sequence of 3D image processing steps, e.g., intensity thresholding and boundary detection. An approach for the reconstruction of cellular structures from optical microscopy data was proposed by Mosaliganti et al. [MCS*08]. They employed sophisticated techniques for extracting cells and separating overlapping cells. The resulting, individually colored cells were then visualized through volume rendering. In recent work, Guo et al. presented a novel volume illustration technique [GYL*12]. It simulates an optical phenomenon in interference microscopy, which accounts light interference over transparent specimens, and thereby enhances the image contrast and structure details.

3.2. Multi-Channel 3D Microscopy Data

Razdan et al. worked on the visualization of multicolor laser confocal microscopy data [RPFC01]. In such data, three lasers providing light at different wavelengths reveal different substances in the same field of view. A RGB composite rendering was generated by means of ray-casting. A volume rendering system for confocal and two-photon fluorescence microscopy data was presented by Clendenon et al. [CPS*02]. They focused on a near real-time visualization of multichannel image stacks on standard PCs by exploiting 3D graphics processors. DeLeeuw et al. presented the ARGOS system for an interactive exploration and batch processing of confocal laser microscopy data [dLVvL06]. They linked a volume rendering of the image data to attribute views, e.g., parallel coordinates, which were equipped with brushing facilities. Bruckner et. al presented a system for an integrated visualization, exploration and annotation of anatomical brain microscopy data and molecular genetic data of fruit flies [BŠG*09]. The microscopy data was visualized through volume rendering with Maximum Intensity Difference Accumulation as a projection method. Wan et al. developed an interactive tool for visualizing multi-channel confocal microscopy data [WOCH09]. Multidimensional transfer functions and several compositing techniques were implemented for an integrated visualization of up to three channels. The approaches creating an integrated visualization of several data channels are closest to our work. However, none of these approaches can handle an arbitrary number of binary data channels (Sec. 2.3).

4. Analysis Framework and Prior Visualization

This section comprises a brief discussion of the in-house toponome analysis framework of our collaborators and their prior way of visualizing the 3D toponome data in an off-line process, i.e., outside the framework.

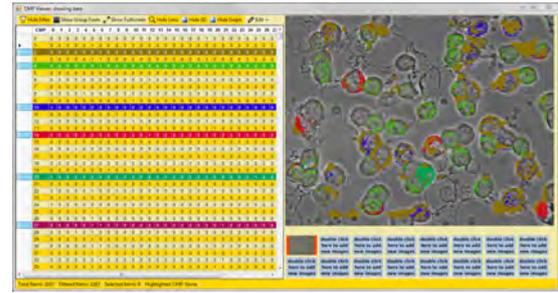


Figure 2: Toponome analysis framework. Multiple views are implemented, such as a table view listing all CMPs as rows (left) and a toponome map showing the location of CMPs selected in the table (right). The toponome map is superimposed on a grayscale phase contrast image. Each ring-shaped structure represents a cell.

4.1. Toponome Analysis Framework

The in-house toponome analysis framework implements multiple coordinated views: a filter view, a table view (Fig. 2, left), a toponome map (Fig. 2, right), and a recently added graph view [OFH*11]. All views are linked and equipped with brushing facilities. If a selection is brushed in one view, it is merged with the selections from all other views and the result is propagated to all views. The filter view facilitates the definition of a template CMP by setting the filter for each affinity reagent to zero, one or no filtering. The table view lists all CMPs as rows. Each selected row is assigned a unique color, which serves as an identifier for this CMP in all views. The table is linked, e.g., to the toponome map, such that the user may select individual CMPs and observe their spatial locations in the cell/tissue sample. The toponome map is often superimposed on a phase contrast image, which serves as a spatial reference (Fig. 2, right). In these images, a relatively clear distinction between cell surface, nucleus, and background is possible. The framework further contains a graph view which encodes the co-mapping of proteins (edges) and the co-mapping frequency (edge width).

4.2. Prior Visualization of 3D Toponome Data

Until now, our collaborators applied the commercial software IMARIS (Bitplane AG, Zurich, Switzerland) for generating volume renderings of their 3D toponome data in an off-line process. The entire procedure for visualizing a 3D toponome dataset took them 1 – 2 workdays. Besides the high expenditure of time, a visualization at interactive frame rates in IMARIS is restricted to ≈ 30 different CMPs on a machine with 4 GB of working memory. Hence, our collaborators focused on the 30 most frequent CMPs out of a set that may comprise several 1000 CMPs. In the following, we describe the time-consuming visualization procedure. Since

the in-house framework could only handle 2D data so far, each volume slice was processed separately. At first, the 30 most frequent CMPs of the volume's middle slice were determined. Although these might not have been the most frequent ones in the remaining slices, they were treated as such in order to generate an inter-slice coherent visualization. In a next step, all n slices were processed sequentially: an image was generated per CMP, showing its locations in this specific slice, the image was screen captured and stored. For a common number of slices $n = 20$, this resulted in 600 screen captures which were then loaded into IMARIS. Each set of images showing the same CMP had to be loaded as a separate channel and perceptually well separated colors had to be defined manually and assigned to the channels.

5. Visualization and Exploration Methods

This section starts with a description of the toponome data, which we employ for demonstration purposes. Then, it elaborates on our volume rendering approach for visualizing the data, which has been integrated in a newly added 3D view of the toponome analysis framework. Finally, it discusses the equipment of the 3D view with interaction facilities.

5.1. Toponome Data

Our methods have been applied to a cell sample containing blood lymphocytes of a healthy subject. Lymphocytes are frequently related to causing chronic inflammatory diseases by entering healthy tissues. In imaging the sample, a large tag library, containing monoclonal antibodies directed against cluster of differentiation (CD) marker proteins, was used. CD marker proteins are expressed, among others, on the surface of immune cells such as lymphocytes. A major challenge is to decipher the cell surface toponome code of these cells in order to detect disease specifics. In this study, 32 CD surface proteins have been co-mapped on the sample (Sec. 2.2). The sample has been imaged at 20 slice locations with a matrix of 658×517 pixels, an in-plane-resolution of 216×216 nm and a slice distance of 200 nm. The data has been binarized and imported into the framework, which detected 2167 CMPs (see [FBKS07] for detailed information on sample preparation, data acquisition, and thresholding).

5.2. Volume Rendering of 3D Toponome Data

The combinatorial molecular phenotypes (CMPs) serve as an input for the volume rendering (recall Sec. 2.3). They are stored per slice in XML files. Each file contains the CMPs of the slice and the pixel positions per CMP. In order to construct a 3D dataset, the XML files are read in and an empty dataset is generated, whose dimensions correspond to those of the measured data. Then, each CMP is assigned a unique RGBA value and this value is stored at those voxels of the new dataset which are associated with the respective CMP (RGBA computation will be discussed in Section 5.3).

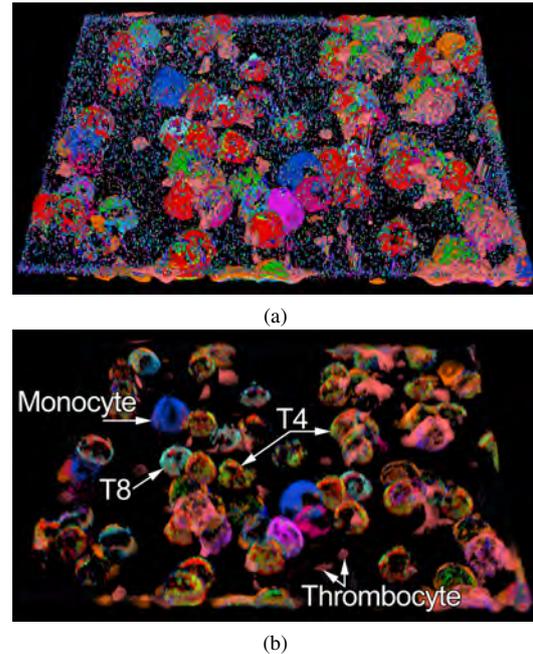


Figure 3: (a) Overview of a cell sample containing lymphocytes (hemisphere-shaped structures). Each of the 2167 CMPs is assigned a unique color. Very small structures are likely to represent noise. (b) A globally assigned transparency visually suppresses these structures. A subset of the major cell types is annotated.

A crucial aspect of volume rendering is the definition of transfer functions for mapping data values to color and transparency. Often, a color scale together with a range of transparency values between zero and one are defined and discretized for that purpose. In our case, a color has already been defined and associated to each data value which renders a mapping unnecessary. The application of transparency to visualizing toponome data is challenging. An increased transparency induces color mixing which hampers the identification of CMPs by their unique color. This scenario is very likely, since a cell or tissue section often contains heterogeneous regions with respect to CMP distribution. At a given pixel position, a different CMP may exist in each slice or all neighboring pixels in the same slice may exhibit different CMPs. Nevertheless, a semi-transparent visualization proved to be valuable in getting an initial overview of the data, in visually suppressing noise and in an interactive data analysis. The latter will be described in Section 5.4. The exploration of 3D toponome data starts with an overview visualization (Fig. 3a). Here, the biologist is particularly interested in the location of cells and in dominant CMP clusters. In case of noise or a high frequency of other small structures that visually overlap with the structures of interest, an increased global transparency simplifies the retrieval of this information (Fig. 3b). It visually suppresses small structures

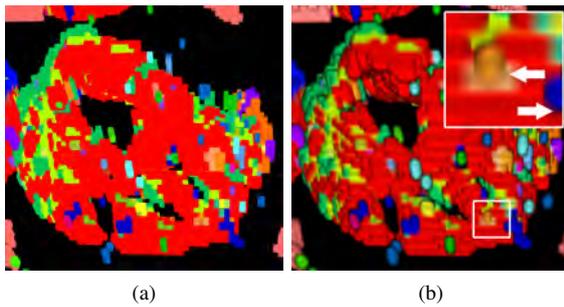


Figure 4: (a) Ray casting using nearest neighbor interpolation at sample points. (b) Trilinear interpolation improves the recognition of depth information and cell surface morphology. However, mixed colors occur at transitions between different CMP clusters and between a cluster and the background (arrows in inset).

while large regions exhibiting identical CMPs remain visible due to a higher accumulated opacity. The global transparency is adjustable and initially set to zero.

The volume rendering has been implemented employing the open-source Visualization Toolkit (VTK, Kitware, Clifton Park, NY, U.S.). VTK offers fast volume rendering techniques exploiting graphics hardware, it implements 3D interaction techniques, e.g., picking, and it is freely available. Furthermore, it integrates well with the existing framework written in C#, by means of the ActiViz software (VTK, Kitware, Clifton Park, NY, U.S.). In VTK, two hardware-accelerated volume rendering techniques are integrated: 3D texture mapping and GPU-based ray casting. We employ the latter, since it is better suited for interactive applications where the input dataset is constantly updated, e.g., through brushing operations. These updates significantly slow down the texture mapper since it always resamples the data to be a power of two in each direction before rendering. In ray casting, the data is sampled along each ray. At each sample position, the corresponding data value is interpolated based on the neighboring values. VTK implements nearest neighbor and trilinear interpolation. We apply the latter since the resulting visualization better conveys depth information and cell surface morphology (Fig. 4). A drawback of trilinear interpolation is the generation of mixed colors along the border between two CMP clusters or a cluster and the background (arrows in inset of Fig. 4b). However, the corresponding regions are very narrow and hence can be distinguished from the real data. The sample distance along the ray as well as the image sampling density have a strong impact on the image quality. If either of them is too low, aliasing artifacts occur. VTK's default values for the sample distance and the sampling density are 1.0 and one ray per pixel, respectively. These settings led to an artifact-free visualization in all our experiments. However, a rendering at interactive frame rates

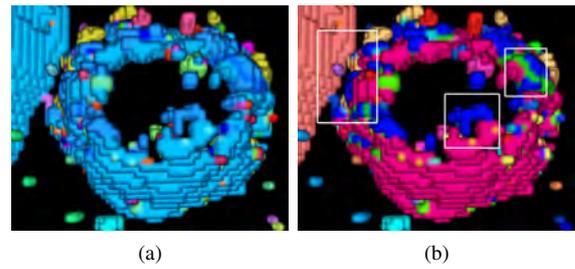


Figure 5: (a) Close-up view of a cell and a neighboring cell (upper left). A large CMP cluster appears turquoise in both cells. (b) A color range redistribution for the same view reveals the true variety of CMPs (insets). It shows that the visible CMP cluster of the neighboring cell differs from the largest cluster of the focused cell.

is hampered for large render windows, e.g., 1600×1200 pixels. As a solution, VTK offers an automatic adjustment of the parameters to a desired update rate, which we set to 15 frames per second. This guarantees a fluent interaction but also causes aliasing artifacts. Hence, we turn the automatic adjustment on when the user is interacting with the scene and turn it off again when interaction stops.

5.3. Perceptually Optimized Coloring

Color serves as CMP identifier across all views of the toponome analysis framework. It is crucial that different CMPs, i.e., their associated colors, may be well discriminated by the biologist. However, a toponome dataset is likely to contain several thousand CMPs. While generating a different color for each CMP is technically possible, visually discriminating these colors by far exceeds the capabilities of the human visual system. Hence, we offer a dynamic color range distribution, that can be activated for the current view on the data, e.g., a close-up view of a particular cell (Fig. 5). It then generates a set of colors with a sufficient perceptual difference for the currently visible CMPs. The colors are transferred to all other views for a coherent visualization. The color range distribution is carried out before the dataset is initially displayed. The user may trigger a redistribution for the current view at any time of the exploration process. In the following, we provide details on the implemented color range distribution algorithm. It strongly improves the former coloring algorithm, which was independent of the number of CMPs and generated 255 shades of color by uniformly sampling the perceptually non-linear RGB color space.

At first, an initial pool of colors is computed in HSV color space by a regular sampling of each component (Alg. 1). From this pool, the n colors, which will be associated to the n visible CMPs, will be drawn. When computing the pool, the hue component is sampled over the full range of 360° with a sample distance of $360^\circ/n$. For the saturation S and the value

Algorithm 1 Compute pool of colors in HSV space.

```

counter ← 0
degree_incr ← 360/n
for hue = 0 → 359 incr : degree_incr do
  for saturation = 60 → 100 incr : 10 do
    for value = 70 → 100 incr : 10 do
      colors[counter] ← [hue, saturation, value]
      counter ← counter + 1
    end for
  end for
end for
end for

```

V component, only a subrange of $[0, 100]$ is considered in order to avoid undersaturated and too dark colors. Both, S and V are iteratively incremented by 10, within the subranges $[60, 100]$ and $[70, 100]$, respectively. With this computation scheme, n hue samples, 5 saturation samples per hue sample and 4 value samples per saturation sample are employed resulting in a color pool size of $n \times 5 \times 4$. Less saturated and darker colors are computed first. They will later be assigned to CMPs with a high frequency of occurrence while highly saturated and bright colors will be assigned to CMPs with a low frequency of occurrence. Thus, large protein clusters remain well visible due to their size and small clusters are easier to perceive due to a striking coloring.

In the next step, the n visible CMPs of the current view are computed by casting a ray per pixel into the scene and determining each intersected voxel and the corresponding CMP. Finally, n colors with a sufficient perceptual difference are drawn from the pool. We compute the difference in the perceptually linear CIELab color space, since a difference between two colors there corresponds well to their perceived difference. As a difference measure, the Euclidean distance is employed [Sha03]. To facilitate distance computations, the color pool is transformed from HSV to RGB [Smi78] and from RGB to CIELab space [JL07]. Before the n colors are drawn, a distance threshold is defined beyond which two colors are considered as sufficiently different. If the threshold is too high, the number of colors that can be drawn is too small. If it is too small, the colors are not perceived as different. Hence, the threshold is determined in an iterative process. The iteration starts with a high value and decreases the value down to the *just noticeable difference* (JND) of two colors in CIELab space which is ≈ 2.3 [Sha03]. The iteration is terminated once n colors could be determined. The initial threshold as well as the step size for decreasing it have been determined empirically and set to $2.3 + 30$ and 0.5 , respectively. A higher initial threshold led to very small color sets, e.g., containing just a single entry. During each step of the iteration, the algorithm tries to draw a set of n colors from the pool. It starts with the first color and adds it to the set. Thereafter, the distance between this color and the second color is computed. If it is higher than the current threshold, this color is also added. If not, the algorithm proceeds with

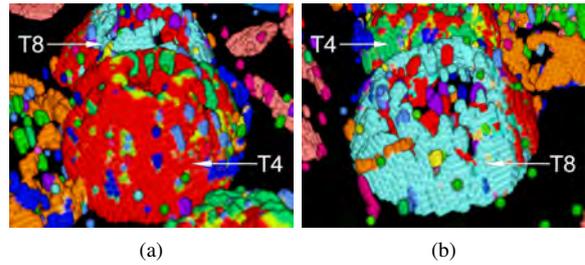


Figure 6: Close-up views of a T4 (a) and a neighboring T8 lymphocyte (b). Note the strikingly different toponomes.

the third color and measures its distance to the first one and so fourth. If more than a single color is contained in the already drawn set, distance tests are carried out between the new color and each of the set members. Only if all tests are passed, the new color is added.

The computation time of the color range distribution depends on the view size and the percentage of background pixels (black image regions). For a view size of 1600×1200 pixels and an overview of the dataset as in Figure 3, the time is 4.4s. For the same view size and a close-up view as in Figure 6a, the time is longer (6.2s) due to a lower percentage of background pixels. A drawback of the current approach is that the colors of the CMPs which are outside the current view are not taken into account. Hence, these colors may also be assigned to CMPs within the view. This poses a problem once the user changes the view, e.g., to an overview representation. In such a case, the color range distribution should be triggered again.

5.4. Interactive Exploration of 3D Toponome Data

In this section, we describe our methods for interacting with the 3D view in an exploratory analysis. After the dataset has been loaded, all CMPs are listed and automatically selected in the table view. Due to linked views, this generates an overview visualization of the cell sample in the 3D view (Fig. 3a). The biologist then modifies the global transparency in order to get an overview of the most prominent CMPs (Fig. 3b). Since the table is initially sorted from top to bottom according to decreasing CMP frequency, these CMPs are listed at its top. This together with the overview helps the biologist in identifying the cell types which are contained in the sample. For example, the cluster of differentiation (CD) marker protein CD4 binds to the surface of T4 lymphocytes. If now CMPs are listed in the table with an entry of 1 in the column representing CD4 and if the corresponding regions cover large parts of a cell's surface in the 3D view, this cell represents a T4 lymphocyte. In the present sample, for instance T4 and T8 lymphocytes as well as monocytes and thrombocytes may be observed. In Figure 3b, T4 lymphocytes, which activate and direct other immune cells, ap-

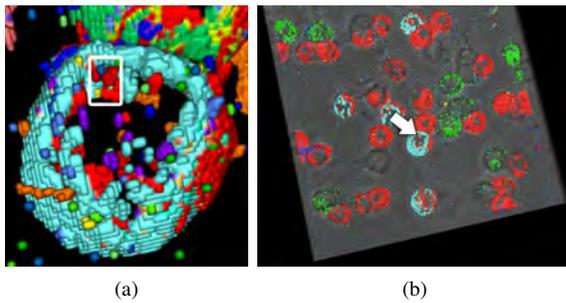


Figure 7: Similarity brushing of a T8 lymphocyte. (a) The biologist brushes an interesting part of the cell surface. (b) The corresponding CMPs are determined across all slices and the 3D visualization is restricted to these CMPs. A phase contrast image stack is volume rendered in grayscale mode and integrated into the scene as a spatial frame of reference.

pear as red and greenish sprinkled hemispherical structures. T8 lymphocytes destroy cells which are effected by a virus. Large parts of their surface are colored in turquoise. Monocytes are another type of white blood cells and appear as mostly blue structures. They are important in immune function and often serve as antigen presenting cells. Thrombocytes are the light red, small, solid structures in between the hemispherical cells. They play a role in blood coagulation. Please note that the coloring of the cells may change across the figures due to color range redistribution (Sec. 5.3).

For our collaborating biologist, the T4 and the T8 lymphocytes are of high interest. Hence, he zooms in to one of these cells and applies a color range redistribution (Fig. 6). Then, the proteins that contribute to the visible CMPs need to be identified. This may be accomplished sequentially by memorizing the color of each CMP and then searching for this color within the selected rows of the table view (Fig. 2). The entries in a detected row which are equal 1 then represent the co-mapped proteins. This approach has several drawbacks. It involves frequently changing the point of gaze between table and 3D view in order to recall the color of interest. This is further complicated as soon as scrolling the table is required. Furthermore, once the row of interest has been detected it needs to be scanned for its entries equal 1. We avoid these problems by implementing a CMP probing through mouse-over interaction. During mouse movement, a ray is cast into the scene starting from the current mouse pointer position. The first hit non-background voxel is reported, the corresponding CMP is extracted, and the names of its co-mapped CD marker proteins are rendered as text at the mouse pointer position. Further, the table row representing this CMP is moved to the top in order to avoid scrolling.

While the biologist samples the cell's surface, he mentally constructs a molecular "face" of the cell which he later compares to the "face" of other cells, e.g., for detecting (patho-

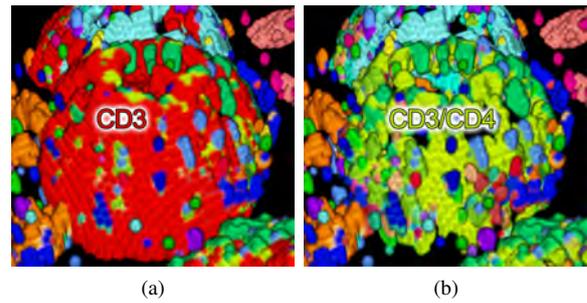


Figure 8: CMP peeling of a T4 lymphocyte. (a) Initial view showing all CMPs. The CMP with a single entry equal 1 for the cluster of differentiation (CD) marker protein CD3 shall be peeled off (CD3 is found in the membrane of T4 and T8 lymphocytes). (b) A CMP with entries equal 1 for CD3 and CD4 is revealed (CD4 is specific to T4 lymphocytes).

logical) variations. We support this comparison by *similarity brushing* (Fig. 7). Here, a cell surface part is brushed in screen space, the corresponding CMPs are determined across all slices and the visualization in all views (including the 3D view) is restricted to these CMPs. Thus, regions containing at least one of the brushed CMPs are revealed. In practice, the biologist defines a rectangular ROI enclosing a cell surface part. Then, rays are cast into the scene for each ROI pixel. Finally, all hit non-background voxels and the corresponding CMPs are determined per ray. In order to augment the 3D visualization, which is restricted to a subset of CMPs and hence, to certain cell parts, a phase contrast image stack is volume rendered and integrated into the scene as a spatial reference (Fig. 7b). The reference is rendered in shades of gray and a global transparency is assigned such that a trade-off between a good visibility of the brushed data and a clear visibility of reference structures is achieved. The reference volume approach has been adopted from superimposing a toponome map on a phase contrast image (Sec. 4.1).

A common problem in 3D data visualization is occlusion. Transferred to toponome data, cell regions corresponding to one CMP occlude others that correspond to another CMP. Hence, we implemented a semi-transparent rendering of the occluding region which is triggered by a point-and-click interaction. After clicking, the desired CMP is determined as described for the CMP probing. This approach is very limited, since overlapping semi-transparent regions are difficult to distinguish and color mixing occurs. Hence, we implemented a CMP peeling interaction (Fig. 8). The desired CMP is selected by means of point-and-click and then, "peeled off" by rendering it fully transparent. Further, it is automatically deselected in all other views of the framework. The peeling step may be repeated any number of times. We enhance the usability of this technique by offering an undo/redo mechanism which is operated via the graphical user interface of the 3D view.

5.5. Anecdotal User Feedback

We gathered anecdotal feedback from our collaborating biologist and a computer scientist who has been working in his laboratory for several years. Both also co-authored this paper. They appreciated the fast generation of the 3D visualization, its completeness with regard to the number of displayed CMPs, and its integration into their framework. They used the color range redistribution extensively (Sec. 5.3). However, they criticized that it ignores the spatial arrangement of the differently colored cell surface parts. Thus, neighboring parts may be assigned rather similar colors as compared to further distant parts. The integration of a reference volume was commended (Fig. 7b). All interaction methods were assessed as useful. Particularly, CMP probing and CMP peeling were frequently used. The computer scientist requested an undo/redo mechanism for the peeling which we added.

6. Summary and Discussion

We presented a volume rendering approach for visualizing 3D toponome data, which significantly reduces the workload of our collaborators. Visualizing a 3D toponome dataset, which took them 1–2 workdays before, now takes \approx 30s. We integrated the 3D visualization into their analysis framework. In this framework, a unique color serves as CMP identifier. We improved the existing coloring scheme by computing a set of perceptually optimized colors. This set may be recomputed for varying views on the data in order to optimize the differentiability of the currently visible CMPs. To support an exploratory analysis, the 3D view was equipped with several interaction techniques. CMP probing supports the biologist in mentally forming the molecular “face” of a cell. CMP peeling provides insight into the composition of a cell’s surface. For instance, a CD3 layer wrapping around a CD3/CD4 layer was frequently observed in T4 lymphocytes (Fig. 8). We further presented the brushing of interesting CMPs in 3D space, which restricts the visualization in all views to these CMPs. In future work, we will conduct a user study to evaluate the perceptually optimized coloring.

Acknowledgements

Technological and biological toponome studies were supported by BMBF grants Biochance, CELLECT, NBL3, NGFN2, NGFNplus, and through DFGschu627/10-1, the Innovationskolleg INK15, and the Klaus Tschira foundation (project toponome atlas).

References

- [BDS10] BARYSENKA A., DRESS A. W. M., SCHUBERT W.: An information theoretic thresholding method for detecting protein colocalizations in stacks of fluorescence images. *J Biotechnol* 149, 3 (2010), 127–131. 2
- [BŠG*09] BRUCKNER S., ŠOLTÉSZOVÁ V., GRÖLLER M. E., HLADŮVKA J., BÜHLER K., YU J., DICKSON B.: Braingazer - visual queries for neurobiology research. *IEEE Trans Visual Comput Graph* 15, 6 (2009), 1497–1504. 3
- [CPS*02] CLENDENON J. L., PHILLIPS C. L., SANDOVAL R. M., FANG S., DUNN K. W.: VoxX: a pc-based, near real-time volume rendering system for biological microscopy. *American Journal of Physiology - Cell Physiology* 282, 1 (2002), C213–C218. 3
- [dLVvL06] DE LEEUW W., VERSCHURE P. J., VAN LIERE R.: Visualization and Analysis of Large Data Collections: a Case Study Applied to Confocal Microscopy Data. *IEEE Trans Visual Comput Graph* 12, 5 (2006), 1251–1258. 3
- [FBKS07] FRIEDENBERGER M., BODE M., KRUSCHE A., SCHUBERT W.: Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system: sample preparation and measuring procedures. *Nat Protoc* 2, 9 (2007), 2285–2294. 2, 4
- [FDM*00] FANG S., DAI Y., MYERS F., TUCERYAN M., DUNN K.: Three-dimensional microscopy data exploration by interactive volume visualization. *Scanning* 22, 4 (2000), 218–226. 3
- [GYL*12] GUO H., YUAN X., LIU J., SHAN G., CHI X., SUN F.: Interference microscopy volume illustration for biomedical data. In *Pacific Vis* (2012), pp. 177–184. 3
- [JL07] JIN L., LI D.: A switching vector median filter based on the cielab color space for color image restoration. *Signal Processing* 87, 6 (2007), 1345–1354. 6
- [Kau98] KAUFMAN A.: Visualizing Nerve Cells with VolVis. In *Pac Symp Biocomput* (1998), pp. 165–176. 3
- [MCS*08] MOSALIGANTI K., COOPER L., SHARP R., MACHIRAJU R., LEONE G., HUANG K., SALTZ J.: Reconstruction of cellular biological structures from optical microscopy data. *IEEE Trans Visual Comput Graph* 14, 4 (2008), 863–876. 3
- [OFH*11] OELTZE S., FREILER W., HILLERT R., DOLEISCH H., PREIM B., SCHUBERT W.: Interactive, Graph-Based Visual Analysis of High-Dimensional, Multi-Parameter Fluorescence Microscopy Data in Toponomics. *IEEE Trans Visual Comput Graph* 17, 12 (2011), 1882–1891. 1, 3
- [RPF01] RAZDAN A., PATEL K., FARIN G. E., CAPCO D. G.: Volume visualization of multicolor laser confocal microscope data. *Computers & Graphics* 25, 3 (2001), 371–382. 3
- [SBP*06] SCHUBERT W., BONNEKOH B., POMMER A. J., PHILIPSEN L., BÖCKELMANN R., MALYKH Y., GOLLNICK H., FRIEDENBERGER M., BODE M., DRESS A. W. M.: Analyzing proteome topology and function by automated multi-dimensional fluorescence microscopy. *Nat Biotechnol* 24, 10 (2006), 1270–1278. 2
- [Sch10] SCHUBERT W.: On the origin of cell functions encoded in the toponome. *J Biotechnol* 149, 4 (2010), 252–259. 1, 2
- [SGK*11] SCHUBERT W., GIESELER A., KRUSCHE A., SE-ROCKA P., HILLERT R.: Next-generation biomarkers based on 100-parameter functional super-resolution microscopy TIS. *New Biotechnology* (2011), In Press, Available Online. 2
- [Sha03] SHARMA G.: *Digital Color Imaging Handbook*. CRC Press, 2003. 6
- [Smi78] SMITH A. R.: Color gamut transform pairs. *SIGGRAPH Comput. Graph.* 12, 3 (1978), 12–19. 6
- [SVP96] SAKAS G., VICKER M. G., PLATH P. J.: Case study: visualization of laser confocal microscopy datasets. In *IEEE Visualization* (1996), pp. 375–379. 2
- [WOCH09] WAN Y., OTSUNA H., CHIEN C.-B., HANSEN C.: An interactive visualization tool for multi-channel confocal microscopy data in neurobiology research. *IEEE Trans Visual Comput Graph* 15, 6 (2009), 1489–1496. 3