Eurographics Workshop on Visual Computing for Biology and Medicine (2016)
S. Bruckner, B. Preim, and A. Vilanova (Editors)

Illustrative Transitions in Molecular Visualization via Forward and Inverse Abstraction Transform

Johannes Sorger[1,2], Peter Mindek[1], Tobias Klein[1], Graham Johnson[3], Ivan Viola[1]

[1]TU Wien, Vienna, Austria [2]VRVis Research Company, Vienna, Austria [3]Allen Institute for Cell Science, Seattle, USA

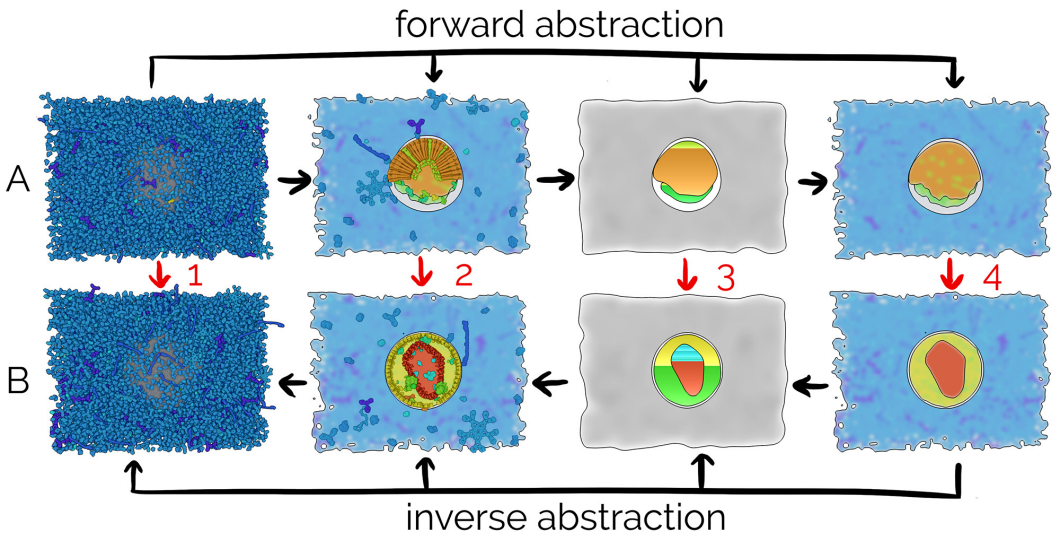


Figure 1: Four levels of visual abstraction used to illustrate a transition between the model of an immature HIV virion (A) and a mature HIV virion (B). The level of information available about the process of the transition between these two development stages of the HIV virion determines, which level of visual abstraction to use to interpolate between the models. We can smoothly interpolate between the levels of visual abstraction (black arrows), in order to visually abstract to a level, at which the interpolation meaningfully illustrates our level of knowledge about the transition (red arrows).

Abstract

A challenging problem in biology is the incompleteness of acquired information when visualizing biological phenomena. Structural biology generates detailed models of viruses or bacteria at different development stages, while the processes that relate one stage to another are often not clear. Similarly, the entire life cycle of a biological entity might be available as a quantitative model, while only one structural model is available. If the relation between two models is specified at a lower level of detail than the actual models themselves, the two models cannot be interpolated correctly. We propose a method that deals with the visualization of incomplete data information in the developmental or evolutionary states of biological mesoscale models, such as viruses or microorganisms. The central tool in our approach is visual abstraction. Instead of directly interpolating between two models that show different states of an organism, we gradually forward *transform the models into a level of visual abstraction that matches the level of detail of the modeled relation between them. At this level, the models can be interpolated without conveying false information. After the interpolation to the new state, we apply the* inverse *transformation to the model's original level of abstraction. To show the flexibility of our approach, we demonstrate our method on the basis of molecular data, in particular data of the HIV virion and the mycoplasma bacterium.*

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms

1. Introduction

Biologists often utilize illustrations and 3D animations to communicate their knowledge to different audiences, such as students or the general public. However, the creation and the maintenance of these animations in regard to new research is a challenging task. When the current knowledge in biology changes due to new scientific discoveries, the illustrations and animations may become out of date and they need to be recreated. This is a time consuming task, since depending on its complexity, it can take weeks or months to create a 3D animation describing a biological process.

To avoid the manual re-creation of illustrations and animations when new knowledge about a phenomenon is gathered, biologists produce complex computational models and simulations, which describe the biological processes and phenomena according to their current knowledge. The communication of these phenomena is then carried out by the visualization of these models. When the biology knowledge changes, the computational models are simply updated, while the visualization pipeline remains the same. This allows for much shorter turnaround times for the communication of science with illustrations and animations.

However, the acquired scientific knowledge about biological phenomena is often sparse and the processes cannot usually be modeled with absolute accuracy. While a lot can be known about a particular biological system, there is usually still even more unknown. Illustrators have to deal with uncertainties when utilizing these biological models. In order to avoid the conveying of false information, they have to respect the information incompleteness when visualizing biological phenomena.

For instance, we have a relatively good understanding of the chemical composition of the HIV at various stages of development, such as the immature and mature stages of the virus. We know which proteins from an earlier stage build up structures in later stages. However, it is not exactly known in which order the chemical reactions are happening. Structural biologists are able to build computational models of the immature and mature HIV, which are relatively accurate. However, they cannot model the process of HIV maturation, which is the transition between these two stages, with the same level of accuracy because of the missing information about this process. This has to be taken into account when these models are used to illustrate the process of the transition of the HIV between its development stages.

When the specifics about certain developments are unknown, scientific illustrators can only show these processes in abstracted ways, e.g., by choosing a viewpoint or representation where the unknown is not visible. The illustrators reduce the level of detail of the available models to match their knowledge about the process, so that the created animation does not imply false information. These abstracted representations are created manually, according to the current knowledge in biology and the illustrator's artistic expressions.

Based on this concept of the visual abstraction, we propose a method for the creation of visual transitions between different molecular models. The method deals with the challenge of conveying incomplete information about the relation between two models. The relation describes the biological process that relates one model to another. Our approach supports the depiction of those processes at multiple levels of visual abstraction, depending on the degree of information that is available.

To enable the interpolation between models at different levels of visual abstraction, we automatically generate several representations of the available models with varying level of detail, which we refer to as *levels of visual abstraction*. It is possible to smoothly interpolate between those representations in order to continuously reduce the level of detail of a visualized model. Before transitioning between two models, we therefore first interpolate to a higher level of visual abstraction, where the interpolation between the models would be meaningful and would correspond to our knowledge about the process. We display the transition between the models in this visual abstraction level, and then inverse transform back to the original representation of the model, i.e., the lowest level of visual abstraction. In this way, we are able to perform a meaningful interpolation between two models which sparsely represent a biological process or a phenomenon. This ensures that the generated animation does not imply false information about the illustrated biological process.

The main contribution of this paper is a method for creating illustrative transitions between molecular models, such as different development stages of a virus, where the actual process of the transition is not exactly known or it is too complex to model. We specifically address the following challenges in our design:

- The visual abstraction of multifaceted molecular data.
- The transition between different levels of visual abstraction.
- The transition of multifaceted molecular data between different models at different levels of abstraction.

2. Related Work

Since our method comprises elements from different visualization areas, the discussion about related work is split into the categories of schematic illustrations of molecular data, and visualization approaches that illustrate or compare change in data.

2.1. Schematic Illustrations of Molecules

Molecular data and corresponding processes on a molecular level are highly complex. Schematizations are therefore often utilized to convey knowledge about these processes on a more abstract level in order to focus on the relevant information for the corresponding context. Goodsell [Goo99] separates molecular representations into atomistic and continuous models. The schematization of atomistic models usually shows either chemical bonds or the surface of a molecule, whereas illustrations of continuous models visualize derived properties of the underlying molecules. Falk et al. [FKRE09] especially emphasize the analysis of signal transduction and therefore propose a schematic visualization that depicts individual molecules and their tracks, or reactions. With the introduction of a cartoon representation for molecules [Ric81] more abstract illustrations [HOF04, WB11] were proposed, especially to schematize the structure of molecules. Cipriano and Gleicher [CG07] propose a simplification of the molecular structure that preserves significant shape features. They utilize symbols placed on the surface to visually encode additional properties. Closely related to our

work is the approach by Zwan et al. [VDZLBI11], which is able to gradually transform the visualization of one molecule between different degrees of visual abstraction. In contrast to their approach, we focus our schematization approach on the quantitative and structural level of complete models of viruses and bacteria, which enables smooth transformations between different states.

2.2. Visualization and Understanding of Change

Comparative Visualization. The comparison and the establishment of relationships between complex objects is often a challenging task. Comparative visualization provides support in fulfilling those tasks. The design of comparative visualizations is traditionally categorized into juxtaposition, superposition or explicit encoding of relationships [GAW*11]. Vivek et. al [VP04] argue that side-by-side comparisons impose the task of finding and quantifying differences on the user instead of on automation. We remedy this drawback by visually encoding quantities directly within the objects. Further, we make use of superposition, but instead of directly overlaying objects, we smoothly transition from one object into another.

Small multiples [Tuf90], coordinated multiple views (CMV), and static images in conjunction with traces and glyphs are typically used to visualize transitions, trends, correspondences or sequences [RFF*08]. Hsu et al. [HMC11] present an automated method for the illustrative visualization of multi-scale phenomena. They generate an image that contains multiple levels of detail of one subject by blending the renderings of multiple pin hole cameras at different zoom levels. VisLink [CC07] establishes relationships between 2D visualizations through placing them in 3D space and drawing edges between different nodes of the visualizations. However, this visual connection of related objects is not feasible for our approach, since our data is 3D, consists of a large number of molecules (more than 20000 in the model of the HIV virion), where the relationship between individual molecules is in many cases not known.

Animated Transitions between Visual Representations. Tversky et al. [TMB02] summarize cognitive studies on the benefits of animation. Although, they conclude that animation alone has not been convincingly demonstrated to be superior to static illustrations, other findings of their study suggest a direction for research on animations in visualization. They report that animation together with basic interaction methods like pausing, partial replaying, zooming and change of perspective might be the key to enhance the effectiveness of animation. Robertson et al. [RFF*08] state that animated transitions of data is more enjoyable and exciting to users. They also found that it was significantly faster, when used for presentation but less exact and less effective for analyzing data.

Kosara et al. [KSH04] use 3D scatter plot matrices to establish a link between the physical layout and the abstract dimensions of the data. Their work is an early attempt to connect information visualization and scientific visualization using points as data primitives. Elmqvist et al. [EDF08] present an exploration technique for multi-dimensional data. They place scatter plots on the sides of a dice and animate the transitions when the dice is rotated. Guilmaine et al. [GVM12] compare different animated transitions of tree structures. They find that hierarchical animation is better suited for the tracking of changes. Basch [Bas11] describes possibilities for animated transitions between volumetric rendering and abstract views like histograms and scatter plots. They use staged animations to reduce occlusion. Hurter et al. [HTCT14] present a more general technique that interpolates data points between different views which are projections of the original data dimensions. They demonstrate applications for volume data, as well as for images. Their approach addresses occlusion with interaction methods like brushing and locking of data points. Heer and Robertson [HR07] show animated transitions between different visual representations of statistical data. Their work makes extensive use of staging to reduce occlusion and clutter. Huron et al. [HVF13] present the visual sedimentation metaphor for the animation of data streams. This metaphor allows to transition from discrete visual elements to a continuous representation. We employ a similar metaphor for the transition of spatial objects to abstract charts.

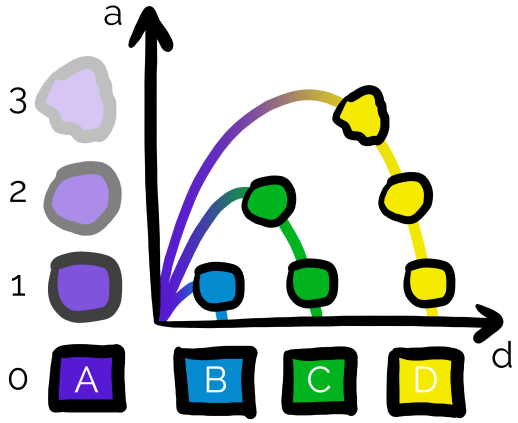


Figure 2: Transition between model *A* and models *B*, *C*, *D*. Axis *d* represents the incompleteness in the mapping between the models. The more incomplete the mapping is, the farther away both models are in mapping space. Axis *a* represents the level of visual abstraction. The higher the distance in mapping space, the higher is the level of visual abstraction that is needed to interpolate between them.

3. Method Overview

While the production of computational models of biological structures already reaches atomic resolution, the highest level of detail at which these biological processes are described is often significantly lower. The knowledge that describes the relations between the stages of development is therefore often incomplete. How should the visual transition between two high-detail models be conveyed when their relations are only modeled in lower detail? Illustrators want to show the highest level of structural detail that is available to them, while at the same time they want to avoid conveying wrong information about the relations between them.

Our method addresses this issue with *forward* and *inverse* visual abstraction. By abstracting a highly detailed structural model down to the level of detail, on which the mapping information between process stages is available, an interpolation from one stage to another is possible without conveying wrong information. After the transition between the two stages of the biological process is completed in the visually abstracted representation, we increase the visual degree of detail of the new stage that we transitioned to again to the highest level by applying the inverse abstraction transformation. With this approach we can display, both, models and relations between models at the highest level of detail available.

This strategy is depicted in Figure 2. Axis *d* describes the degree of incompleteness, in which the relation (or mapping) between the two models is described.The further away two models are on this axis, the less detail of how one model relates to the other one is available. We denote this as the distance in *mapping space*. A distance of zero in mapping space corresponds to a complete description of the relation between two models (independent of how detailed the models are). If the models are described on a higher level of detail than the relations between them, the distance in mapping space increases. Axis *a* describes the levels of visual abstraction of a model. The further two models are apart in mapping space, the higher we have to abstract both models before we can interpolate them. We have to abstract the models to the level of visual detail that corresponds to the level of detail, at which the relation between the models is described. The number of abstraction levels therefore has to be symmetrical to the maximal degree of incompleteness in the relation between the models, i.e., the maximal distance in mapping space. We depict this symmetry by the curves that are drawn between two models in Figure 2.

How many levels of abstraction there should be, and how the distance between two models is measured, depends on the type and complexity of the data that should be presented. In the case of molecular models at the scale of complex viruses and simple bacteria, we define four levels of abstraction on Axis *a*, as well as four discrete distances in mapping space on Axis *d*.

The molecular data, on which we demonstrate our method, describes biological organisms at atomic resolution. Biologists model their mesoscale data according to so called recipes [JAAA*15] that describe the molecular and structural composition of a model. These recipes are executed by a packing algorithm that populates the space with the specified macro molecules. The result is a 3D molecular model of an organism that consists of tens of thousands of molecular instances, each comprised of thousands of atoms. This comparatively large number of molecular instances originates from a relatively small range of protein types. In the case of the HIV virion, 20.000 molecular instances are distributed across 42 different types of molecules. The molecules are densely packed within their respective compartments of the model and also constitute the walls of the compartments. Figure 1 displays two models of the HIV virion at the four different levels of visual abstraction.

Figure 3 displays the four degrees of incompleteness, i.e., the four discrete distances in the mapping space between molecular models that we identified. In the following, we describe the levels of visual abstraction that are necessary to match two models at different degrees of incompleteness.

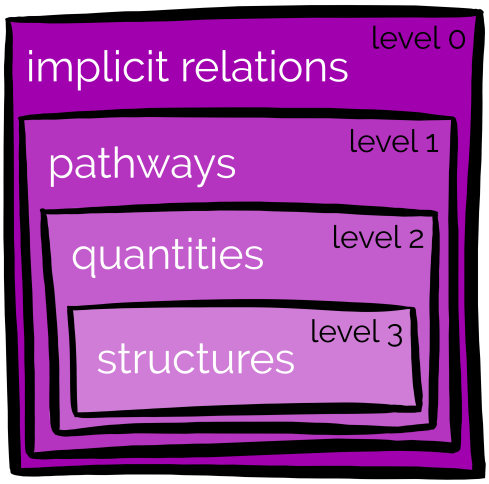
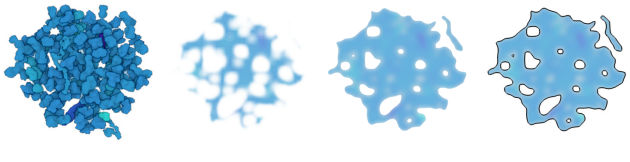


Figure 3: Degrees of incompleteness: Depending on the amount of available mapping information, the corresponding level of visual abstraction (Levels 0-3) has to be used for a transition between two models. If only information about high level structure correspondence is available, we have to perform the transition at visual abstraction Level 3. If we also have information about the relation of protein quantities per structure across models, we can perform the transition at Level 2. If we also know about the protein pathways that connect the chemical composition of one model to the composition of the other, we only have to abstract to Level 1 before the transition. If the information about the development of a biological model is so detailed that it can be described by a simulation, we do not need to abstract the model any further to avoid conveying false information.

3.1. Level 0 - Implicit Relations

Abstraction of data and representation: The lowest level of visual abstraction corresponds to the highest level of detail that a molecular model is represented in. The cellular data is depicted at atomic resolution. Even though it features the highest level of detail, this level is still an abstraction from reality, as the data is not measured but computationally modeled and simulated. Typically, only macro molecules are represented while smaller, more common molecules, such as water molecules, are omitted from the presentation.

Matching of models: For a proper matching of relations between models at any given level of detail, the entities that the respective level of detail is composed of need to be related to each other. At this level of abstraction, the data is displayed in its highest level of detail, at atomic resolution. A transition between two models at this level therefore requires a definition of relations between the individual protein molecule instances in each model. This corresponds to a distance of zero in mapping space at the lowest level of abstraction. Processes at this level of detail describe, for instance, how individual proteins split up and merge to form new proteins and structures, e.g., in the life cycle stages of a cell. The presentation of processes that transform a biological data set from one state to another at this level of detail is typically handled by simula-



(a) (b) (c) (d)

Figure 4: Simplified rendering of molecular data in consecutive steps where the 3D information about the individual molecules is merged into simple geometries. The transition from (a) to (b) shows the blurring with a Gaussian filter. (c) depicts the result after a steep ramp function is applied to the alpha channel to achieve hard edges. Finally, a Sobel operator is applied on the alpha channel (d) to output the display contours of the generated shape.

tions, i.e., at an implicitly described or procedural level, as manual matching for up to hundreds of thousands of molecules is virtually impossible.

3.2. Level 1 - Pathways

Matching of models: If the knowledge about the detailed processes that transform a model from one developmental or evolutionary state to another is not available, the relations between two models have to be defined at a lower level of detail, i.e., at a higher level of abstraction. At this level, the relation between individual molecule instances is not known anymore. However, knowledge about the relation between protein types in each model, i.e., *protein pathways*, is still available. This requires the matching of protein types across models in terms of how they split up in one model and/or merge into new protein types in the other model.

Further, *protein quantity matching information* is still available. The quantities of each protein type in a model are known, as well as how the quantities relate to the quantities of the other model, e.g., 50% of protein A combine with protein B to result in protein C in the other model. *Protein shape information* is still available, as well as the shape of *high level structures* in the biological organism, such as cell membranes. This degree of incompleteness of mapping information corresponds to a distance of 1 in mapping space in respect to the lowest level of abstraction (Level 0). By visually abstracting the model representation to this level, the distance in mapping space becomes zero and a smooth transition between the two models that avoids false information is possible.

Abstraction of data and representation: In order to visually abstract a model from Level 0 to Level 1, we reduce the number of displayed molecules, thus abstracting the model to a level of detail where we do not have to display the missing information of individual molecule instance relations anymore. Now we can match and smoothly interpolate between the shapes of molecule/protein types, the quantities of each type, the shapes of high level structures, e.g., compartments, and the localizations of protein types across compartments.

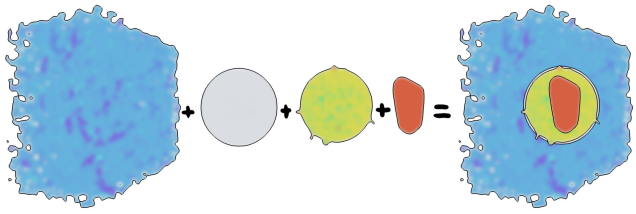


Figure 5: The simplification process of the molecular model is applied to all compartments individually, which are then rendered in layers and subsequently composed through alpha-compositing.

3.3. Level 2 - Quantities

Matching of models: Matching at the second level of visual abstraction is necessary, when we do not possess information about the relation between the proteins across models, i.e., about the protein pathways. The relation of protein type quantities and high level structures across models is still available though.

This degree of incompleteness of mapping information corresponds to a distance of two in mapping space in respect to the lowest level of abstraction. We therefore have to transform the models to the second visual abstraction level before we can interpolate them properly.

Abstraction of data and representation: Since we cannot show the relation between protein types due to the missing pathway information, we abstract the protein type representatives away. What is left is the high level structures, i.e., compartments. We therefore have to substitute the quantitative information that was represented in the protein representative count in the previous level of abstraction. This can be achieved by applying abstract visual encodings, for instance, by filling the compartment structures with colored bars that represent the protein type quantity in the respective compartment.

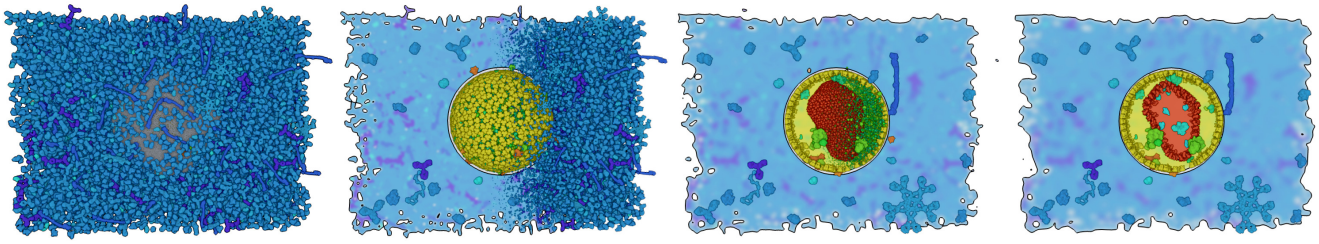
3.4. Level 3 - Structures

Matching of models: At this degree of incompleteness of mapping information, we only have information about which high level structures in model A correspond to which high level structures in model B. Information about protein type and quantity relations is not known anymore. We therefore have to further increase level of visual abstraction of our models in order to interpolate them properly.

Abstraction of data and representation: To further increase the level of visual abstraction, we remove the quantitative per-compartment information of protein types. This leaves us with just the high level structures that now allow a proper matching of the models.

4. Method Implementation

In this Section, we describe the implementation of our method for molecular data. All four proposed levels of visual abstraction, applied to two distinct models, can be seen in Figure 1 marked with the respective numbers.



(a) (b) (c) (d)

Figure 6: This sequence of images shows the transition between the visual abstraction levels one and two of the mature HIV virion. The noisy and dense high resolution molecule information is reduced in a delayed peel-away fashion from left to right and for each compartment. The inner structures of the model are revealed one by one. The dense structure makes way for the blurred background that represents the structure of each compartment. Representatives of each protein type are rotated toward the viewer and enlarged.

4.1. Simplified Rendering

Since the visual abstraction levels 1-3 are based on the display of high level structures, i.e., molecule compartments, we apply a simplified rendering approach to create these compartment shapes. This method merges the 3D information about the individual molecules into aggregated geometrical shapes representing individual compartments within the model, together with their inherent hierarchy.

To create these aggregated geometrical shapes, we take the rendering of the molecular model and blur it with a Gaussian filter including the alpha channel in a post processing step. The radius of the filter specifies, to which degree the compartment shapes will be simplified. The blurring serves as a fast and easy way to create a smooth structure from the noisy patterns of the densely packed macro molecules.

Subsequently, a steep ramp function is applied to the alpha channel of the blurred rendering. This results in hard edges, which are however smoothed by the Gaussian filter. Finally, we apply a Sobel operator on the alpha channel and use the output to display contours of the generated shape. These contours allow us to distinguish the blurred compartment shapes from each other. This process is illustrated in Figure 4.

The simplification process is applied to all compartments individually, i.e., only the molecules of a given compartment are rendered in the same pass. This results in a rendering of the individual compartments within separate layers. The layers are subsequently composed through alpha-blending, as illustrated in Figure 5. During the composition, the layers representing the most enclosed structures are added last. This ensures that the hierarchy of the compartments is visible in the simplified rendering from any viewing direction. However, this requires that the underlying model has an onion-like structure of the individual compartments. If there is too much overlapping of the compartments, the clarity of the simplified rendering is reduced.

4.2. Level 0 - Implicit Relations

On this level, the molecular visualization framework *cellVIEW* [LAPV15] is used to display the molecular data. Simplified rendering is not applied on this level. We use the cellVIEW output as a starting point for the simplified rendering and subsequently for all the other levels of abstraction. In principle, other molecular visualization techniques could be used as a starting point, as well.

4.3. Level 1 - Pathways

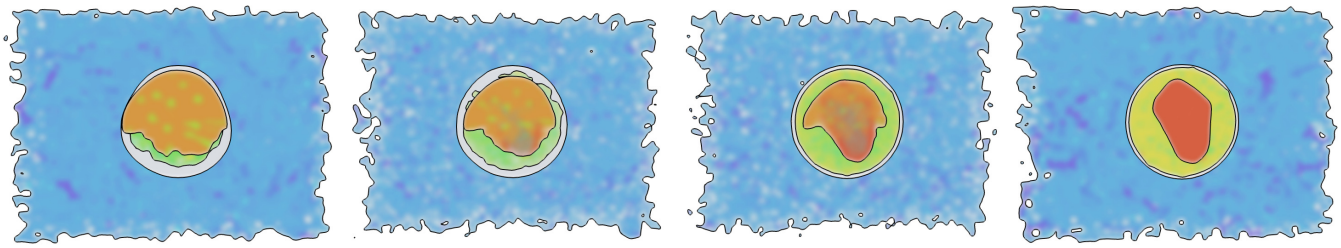
On the second level of visual abstraction, we display the structural information of the model, including the hierarchy of its compartments, individual membranes defining the compartments, as well as representative specimen of matrix proteins in the model. These representatives serve as examples of the chemical composition of the individual compartments within the model. This level of abstraction features a cross-section view of the entire model, while the number of individual molecules is significantly reduced.

By presenting representatives of proteins in a cross-section view, we free up screen space to enable the display of pathway information. The remaining protein representatives can then be used to demonstrate chemical processes, such as the splitting and merging of proteins across model states.

We use the simplified rendering of the model as background on top of which we display selected protein representatives. The representatives are split into two categories: matrix proteins and membrane proteins.

For matrix proteins, the amount of selected specimen of each protein type is proportional to the amounts of the molecules of this protein type in the original model, in order to convey the relative quantities of the proteins per type. They are chosen randomly from all the protein instances, approximating uniform distribution within the data.

The orientation of the selected specimen is modified so that they face the current viewpoint. We find the smallest dimension of the minimum bounding box of each protein type (x, y, or z). Each selected specimen is then rotated so that the smallest dimension of its bounding box is parallel to the viewing direction. This way, the shapes of the selected specimen are better conveyed. Additionally, the representative specimen are scaled up by a constant factor (in our implementation, we use factor of *2*) so that their shapes are better visible.



(a) (b) (c) (d)

Figure 7: This sequence of images shows the transition between the immature and mature HIV virion at visual abstraction Level 3. We abstract both models to this level, since we only have information about the compartment to compartment relations between them for the matching. The sequence shows how the orange, green, and gray structures in (a) correspond to the red, yellow and gray compartments in (d).

Representatives of the membrane proteins are neither rotated to face the current viewpoint nor scaled up, since they also convey structural information about the individual membranes. For the membrane proteins, all those are selected whose principal direction is perpendicular to the viewing direction. Therefore, the displayed membrane proteins form boundaries between individual compartments within the simplified rendering and show the shape and molecular composition of these compartments.

Since all the selected specimen are located on their original positions, and only their rotation and scale changed, it is possible to smoothly interpolate from the first level of abstraction to the second one without introducing unnecessary visual clutter. This is achieved by displaying the original molecular rendering of the model on top of the simplified rendering with a blurring radius of 0. Subsequently, the rotations and the scales of the selected specimen are interpolated to the desired values, while the scales of all the other molecules are interpolated towards zero. We use spherical linear interpolation (SLERP) for interpolating rotations which are specified by quaternions. For interpolating the scales, linear interpolation is used. We achieve an incremental abstraction of the compartment shapes by continuously increasing the blurring radius of the simplified rendering, concurrently with the rotation and scale interpolation.

The time offset for starting the interpolation of the scales and rotations of the individual molecules can be made dependent on a distance field, so that not all the molecules are interpolated at the same time. Similarly, this interpolation can be performed sequentially for the individual compartments. This transition is depicted in Figure 6.

4.4. Level 2 - Quantities

In this level of visual abstraction, all the spatial information of the individual protein molecules is replaced by more abstract quantitative information. The simplified rendering is used as context. We overlay selected compartments from the simplified rendering by stacked bars, which show the quantities of the individual protein types within those compartments.

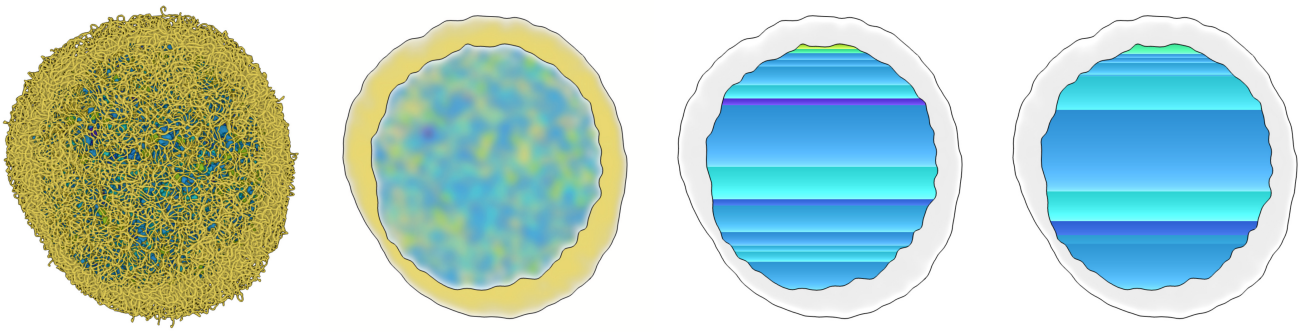
In our implementation, we encode the protein quantities in the height of the bars. This is achieved by finding the topmost and bottommost pixels of the overlaid compartments and taking the difference of their y coordinates to calculate the total height of the compartment. This height is then proportionally divided to create the individual stacked bars. Alternatively, the protein quantities could be encoded in the bar areas, or by other information visualization means, such as a tree-map within a compartment.

The protein colors are used for the respective bars. Additionally, a color gradient is applied to each bar, so that the color luminance slightly increases with decreasing y coordinates. This ensures that there is a visual separation between consecutive bars, which might be of similar colors. Currently, the protein coloring is randomized, as there is yet no standardized color scheme for these macro molecules. The coloring of the simplified shapes representing the compartments results from the blurring of colors of the proteins within each compartment.

To transition to this level of visual abstraction from either of the previous levels, all the molecules displayed on top of the simplified rendering are continuously scaled down until they disappear. Simultaneously, the stacked bars are faded-in through alpha blending. The color saturation of the compartments where the stacked bars are not shown is continuously decreased to steer the viewers' attention towards the stacked bars, while the shapes of other compartments provide the context (see Figure 8 (c) and (d)).

4.5. Level 3 - Structures

In the highest level of abstraction, only the structural information of the model is shown. This is essentially the simplified rendering of the model without any overlays, as described in Section 4.1. To transition to this level of abstraction, the molecules displayed on top of the simplified rendering are removed by continuously scaling them down. Stacked bars from the previous level of abstraction are faded away through alpha blending.



(a) (b) (c) (d)

Figure 8: The transition of mycoplasma. (a), (b) and (c) show the transition from visual abstraction Level 0 to Level 2. (c) and (d) show a transition at visual abstraction Level 2 between two different model states of the mycoplasma (indicated by the change in the distribution of different protein quantities).

5. Results

In this Section, we demonstrate the results of our method by applying it to two molecular data sets. The first example depicts the transition of an HIV virion from its immature to its mature state. The second example depicts the transition of a mycoplasma bacterium between two simulation states.

5.1. HIV Data Set

In this example, we show the relation between two data sets of an HIV virion at different development stages: immature and mature.

We know in theory that the large proteins in the immature model split up into multiple smaller proteins that, in the mature virus, correspond to three different compartments: the virus membrane, the capsid, and the capsid interior. However, we do not have the actual pathway information that describes the relations between proteins in the different maturity states. We only know, which compartments correspond to which ones in each model.

We know about the quantities of protein types in each model, since we have detailed models of both states. However, we do not have a direct relation between the quantities in both models, since many intermediate development steps between these two states are unknown. We therefore also cannot depict the relation of quantities accurately. We have to abstract the immature HIV model to the highest level of visual abstraction (Level 3) before we can show it's relation to the mature HIV model without conveying false information.

The transition between both models at visual abstraction Level 3 is shown in Figure 7. The orange, green, and gray structures in (a) correspond to the red, yellow and gray compartments in (d). We achieve the transition between both models at this level of visual abstraction simply by scaling the blurred structures in the simplified rendering of the immature HIV virion to zero, while simultaneously scaling the structures of the mature HIV virion from zero to their original size. After the transition, we apply the inverse abstraction on the mature data set back to the lowest level of visual abstraction, as can be seen in the lower half of Figure 1.

5.2. Mycoplasma Data Set

We applied our method to a model of mycoplasma mycoides (Fig. 8). With this data set, we have the opposite situation as in the previous example. Many high resolution simulations of the mycoplasma life cycle stages are available [KPC14], while the mycoplasma 3D model is a work in progress. Biologists are developing it since more than one year at the time of writing. We therefore only have access to one incomplete model but to many detailed quantitative simulation results.

In contrast to the previous example, this quantitative information between development states has been sampled at a high frequency. This means, that interpolating the quantities between two time steps of the simulated model would not convey false information about the relation of quantities.

This means that we do not have to abstract all the way to the highest level of abstraction (Level 3). However, even if the detailed protein pathway information between models would be known, we do not have the complete 3D model information required for an interpolation at Level 1. Therefore we can only show the development of protein type quantities in visual abstraction Level 2.

The transition between time steps in the development cycle of the mycoplasma is achieved by simply interpolating the quantities of proteins. This results in the growing and shrinking of the corresponding bars that fill the respective compartments. Annotations could be used to explain the meaning of the individual bar colors.

Further, we do not necessarily have to inverse the abstraction of the model after we finished the transition of protein quantities between different model states, since we only have one 3D model of the mycoplasma.

6. Discussion and Expert Feedback

In this Section, we first discuss the design choices that lead to our method and subsequently present feedback, which we received from two domain experts.

6.1. Design Choices

The choice of organizing the visualization in four different levels of visual abstraction results from the characteristics of our data. Our molecular data comprises molecular structures that are organized in different compartments that are hierarchically ordered. In order to apply our method to data that features different characteristics, the number of abstraction levels may have to be adjusted. The applicability of our approach is not conceptually limited to molecular data. The approach is applicable to any domain that exhibits data with the same characteristics, namely a possible discrepancy between the modeled or measured detail in data points and the relation between these data points.

Using animated transitions to demonstrate the relationship between two different model states provides an implicit way of showing the link between related structures. In some cases, explicit encoding of these relations might be more desirable, e.g., by drawing edges that link related objects [WPL*10, CC07]. In our case however, a direct linking is not feasible for our data characteristics, since the concrete relation between entities is not always known.

6.2. Domain Expert Feedback

We received feedback from two domain-experts, a researcher that specializes in molecular illustrations and animations, and a researcher with the focus on molecular modeling. Both were very excited about the concept of automatically transitioning between two molecular data sets, as they had to create such results manually in Maya until now. In fact, the idea and inspiration for our method came from discussions with one of the researchers during another project. The researchers were already positively surprised when we showed them a prototype of our method where we interpolated between the HIV models. Also the levels of abstraction that we designed received positive attention. One domain expert stated that the abstraction successfully conveyed the compartmentalizations and compartment-to-compartment relationships in the virus. Also the "painting style" that the image space blending achieved was complemented. The illustrator stated, *"This transition does a great job of taking a densely packed model and simplifying it to reveal the composition of molecules that are isolated to each compartment, while still retaining a simplified rendition of the compartment."* The experts appreciated that the transition conveyed additional information that would not have been visible with a simple depiction of the original and target state. *"The transition lets the viewer see the complexity of the structure, where the final image is simplified."*

7. Conclusion and Future Work

In this paper, we propose a method that deals with the creation of illustrative transitions between molecular models, where the actual process of the transition is not or only partially known. Our method utilizes visual abstraction to achieve smooth transitions between different models while respecting the known relational information. Instead of directly interpolating between two different models, we gradually forward transform the models into a level of visual abstraction that matches the preciseness of information about the relationship. These transitions provide intuitive connections between the structural and quantitative characteristics of the two dissimilar models. We exemplary demonstrate the flexibility of our approach on the basis of data sets of the HIV virion and the mycoplasma bacterium. We have received positive feedback from domain experts in the field of biological modeling, and animation. The concept of our approach is applicable to all types of data that exhibits the same characteristics, i.e., where relations between individual (hierarchical) data elements across different data samples can be described on multiple levels of detail and where the knowledge about the model states might not match the knowledge about the relation between those states.

With the proposed method we have started to explore how to relate different states of molecular data with varying levels of relationship knowledge. A natural advancement for this method is the design of transitions for cases where the level of relationship knowledge varies between regions or categories of connected objects. For instance, if only the relationship between certain kinds of molecules is known but not between others, different levels of abstractions could be blended together.

We hope to see further advancements, inspired by our approach, in dealing with missing information in bio-medical data visualization, and in uncertainty visualization in general.

Acknowledgments

Johannes Sorger and Peter Mindek contributed equally to this work. This project has been funded by the Vienna Science and Technology Fund (WWTF) through project VRG11-010 and also supported by EC Marie Curie Career Integration Grant through project PCIG13-GA-2013-618680. Johannes Sorger has been partially supported in the scope of the FWF-funded project P24597-N23 (VISAR) and the COMET K1 program of the Austrian Funding Agency (FFG). We would like to thank Ludovic Autin and Mathieu Le Muzic for insightful comments.

References

[Bas11] BASCH C.: *Animated Transitions Across Multiple Dimensions for Volumetric Data*. Master's thesis, Institute of Computer Graphics and Algorithms, TU Wien, 2011. 3

[CC07] COLLINS C. M., CARPENDALE S.: Vislink: Revealing relationships amongst visualizations. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1192–1199. 3, 9

[CG07] CIPRIANO G., GLEICHER M.: Molecular surface abstraction. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1608–1615. 2

[EDF08] ELMQVIST N., DRAGICEVIC P., FEKETE J. D.: Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1539–1148. 3

[FKRE09] FALK M., KLANN M., REUSS M., ERTL T.: Visualization of signal transduction processes in the crowded environment of the cell. In *IEEE Pacific Visualization Symposium* (2009), IEEE, pp. 169–176. 2

[GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization 10*, 4 (2011), 289–309. 3

[Goo99] GOODSELL D. S.: Atomistic vs. continuous representations in molecular biology. In *Visual Representations and Interpretations*. Springer, 1999, pp. 146–155. 2

[GVM12] GUILMAINE D., VIAU C., MCGUFFIN M. J.: Hierarchically animated transitions in visualizations of tree structures. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2012), ACM, pp. 514–521. 3

[HMC11] HSU W.-H., MA K.-L., CORREA C.: A rendering framework for multi-scale views of 3d models. *ACM Transactions on Graphics 30*, 6 (2011), 131. 3

[HOF04] HALM A., OFFEN L., FELLNER D.: Visualization of complex molecular ribbon structures at interactive rates. In *Proceedings of IV 2004 - 8th International Conference on Information Visualisation* (2004), IEEE, pp. 737–744. 2

[HR07] HEER J., ROBERTSON G.: Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1240–1247. 3

[HTCT14] HURTER C., TAYLOR R., CARPENDALE S., TELEA A.: Color tunneling: Interactive exploration and selection in volumetric datasets. In *Proceedings of IEEE Pacific Visualization Symposium* (2014), pp. 225–232. 3

[HVF13] HURON S., VUILLEMOT R., FEKETE J.-D.: Visual Sedimentation. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2446–2455. 3

[JAAA*15] JOHNSON G. T., AUTIN L., AL-ALUSI M., GOODSELL D. S., SANNER M. F., OLSON A. J.: cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nature methods 12*, 1 (2015), 85–91. 4

[KM13] KOSARA R., MACKINLAY J.: Storytelling: The next step for visualization. *IEEE Computer (Special Issue on Cutting-Edge Research in Visualization) 46* (2013), 44–50.

[KPC14] KARR J. R., PHILLIPS N. C., COVERT M. W.: WholeCellSimDB: a hybrid relational/hdf database for whole-cell model predictions. *Database* (2014). 8

[KSH04] KOSARA R., SAHLING G. N., HAUSER H.: Linking scientific and information visualization with interactive 3d scatterplots. In *Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision* (2004), pp. 133–140. 3

[LAPV15] LE MUZIC M., AUTIN L., PARULEK J., VIOLA I.: cellVIEW: a tool for illustrative and multi-scale rendering of large biomolecular datasets. In *Eurographics Workshop on Visual Computing for Biology and Medicine* (2015), pp. 61–70. 6

[MLF*12] MA K. L., LIAO I., FRAZIER J., HAUSER H., KOSTIS H. N.: Scientific storytelling using visualization. *IEEE Computer Graphics and Applications 32*, 1 (2012), 12–19.

[RFF*08] ROBERTSON G., FERNANDEZ R., FISHER D., LEE B., STASKO J.: Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1325–1332. 3

[Ric81] RICHARDSON J. S.: The anatomy and taxonomy of protein structure. *Advances in protein chemistry 34* (1981), 167–339. 2

[SH10] SEGEL E., HEER J.: Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1139–1148.

[TMB02] TVERSKY B., MORRISON J. B., BETRANCOURT M.: Animation: Can it facilitate? *International Journal of Human-Computer Studies 57*, 4 (2002), 247–262. 3

[Tuf90] TUFTE E.: *Envisioning Information*. Graphics Press, Cheshire, CT, USA, 1990. 3

[VDZLBI11] VAN DER ZWAN M., LUEKS W., BEKKER H., ISENBERG T.: Illustrative molecular visualization with continuous abstraction. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 683–690. 3

[VP04] VERMA V., PANG A.: Comparative flow visualization. *Visualization and Computer Graphics, IEEE Transactions on 10*, 6 (2004), 609–624. 3

[WB11] WAHLE M., BIRMANNS S.: Gpu-accelerated visualization of protein dynamics in ribbon mode. In *IS&T/SPIE Electronic Imaging* (2011), International Society for Optics and Photonics, pp. 786805–786805. 2

[WH07] WOHLFART M., HAUSER H.: Story telling for presentation in volume visualization. In *Proceedings of the 9th Joint Eurographics / IEEE VGTC Conference on Visualization* (2007), EUROVIS'07, Eurographics Association, pp. 91–98.

[WPL*10] WALDNER M., PUFF W., LEX A., STREIT M., SCHMALSTIEG D.: Visual links across applications. In *Proceedings of Graphics Interface 2010* (2010), Canadian Information Processing Society, pp. 129–136. 9