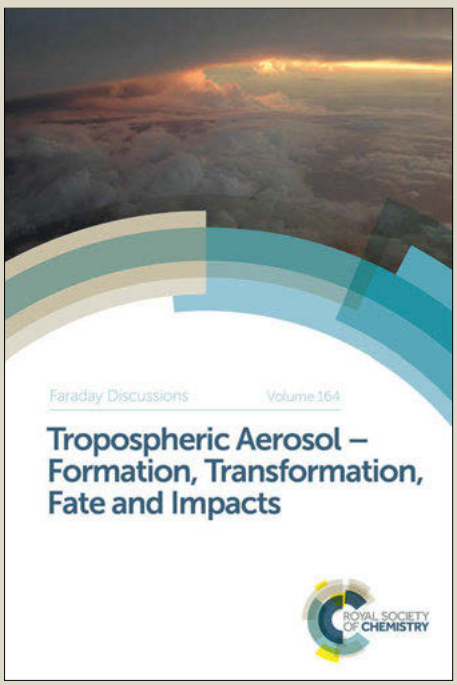
Faraday Discussions

Accepted Manuscript

This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting. All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: http://rsc.li/fd-upcoming-meetings



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
PAPER www. rsc.org/faraday_d | Faraday Discussions

Visual Analysis for Space-Time Aggregation of Biomolecular Simulations

Thomas Ertl,*[a] Michael Krone,[a] Stefan Kesselheim,[b] Katrin Scharnowski,[a] Guido Reina,[a] and Christian Holm[b]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/c000000x

Conducting a current through a nanopore allows for the analysis of molecules inside the pore because a current modulation caused by the electrostatic properties of the passing molecules can be measured. This mechanism shows great potential for DNA sequencing, as the four different nucleotide bases induce different current modulations. We present a visualisation approach to investigate this phenomenon in our simulations of DNA within a nanopore by combining state-of-the-art molecular visualization with vector field illustration. By spatial and temporal aggregation of the ions transported through the pore we construct a velocity field which exhibits the induced current modulations caused by ion flux. In our interactive analysis using parametrisable three-dimensional visualisations we encountered regions where the ion motion unexpectedly opposes the direction of the applied electric field.

1 Introduction

Visualisation has become widely-accepted for improving the understanding of complex data sets from real-world measurements as well as simulations. This is often referred to as 'generating insight' [1]. However, with the proliferation of data sets of rapidly increasing size, even the most concise and abstract visualisation techniques are facing scalability problems. Although this data scalability issue can be mitigated somewhat with algorithmic optimisation as well as with parallel hardware, the human employing a visualisation can hardly be scaled arbitrarily from a perceptual and cognitive point of view. These problems are analysed in detail by Thomas and Cook [2], paving the way for a new approach that tightly integrates automated analysis steps with highly interactive visualisations. In this way the human is put in the loop, steering the analysis process as insight is generated. This approach has been termed *visual analytics* and can be applied to generic abstract data from arbitrary domains. Keim et al. [3] established the *visual analytics mantra* as a simple guideline: "analyse first – show the important – zoom, filter and analyse further – details on demand". Lately, *visual analysis* is seen as a specific variant that applies these concepts to spatial data specifically prevalent in

[a] *VISUS, University of Stuttgart, Germany. E-mail: {ertl | kroneml | scharnkn}@vis.uni-stuttgart.de*
[b] *ICP, University of Stuttgart, Germany. E-mail: {kessel | holm}@icp.uni-stuttgart.de*

the life sciences. According to Kehrer et al.[4], this requires a tight integration of the visual mapping, i.e. the representation of the data, visual interaction concepts and computational analysis.

We have already applied this concept successfully to different scenarios related to biophysics and biochemistry. One example is the time-dependent analysis of the interaction between polymer and solvent[5]. Here, the solvent in the relevant hydration shell is visualised, while the remaining solvent molecules can be filtered out in real time. Different properties of the solvent can be projected onto the solvent-excluded surface of the polymer, for example the hydrogen bond loci, which are automatically computed for the data set. Additionally, the varying conformations over time can be accumulated to show the spatial extents covered during the whole simulation run and to illustrate the potential clustering of hydrogen bonds at the related surface area. Another example is the visualisation of solvent pathlines relevant to protein cavities[6]. The solvent movement is aggregated over time and can be clustered according to user-defined parameters as well as filtered with respect to proximity to an area of interest. The resulting visualisation concisely depicts paths on which water can access, for example, the active site of a protein. In both cases, temporal or spatial aggregation is used to provide a meaningful visualisation that supports the analysis process. A further example in which aggregation is employed to give a clearer picture is the work of Lindow et al.[7]. Here, a partially open channel in a protein is extracted and tracked over time. The extracted paths through the channel are aggregated over time. This results in an aggregated path that shows whether a small molecule would be able to traverse the channel during the simulation, despite the channel never being fully open.

In the work presented in this paper, we use aggregation to illustrate the behaviour of ions near DNA in a nanopore. Our visualizations allow researchers from the field of biophysics to interactively analyse their simulations, which is an important step in proposing or rejecting hypotheses about the characteristics of the investigated systems. Nanopores are isolated, nanometre-sized holes in thin membranes. When a voltage is applied between the two water compartments separated by the membrane, the pores turn into a sensitive biosensor. A small, but measurable electric current goes through the pore, and its magnitude depends largely of the conductance of the nanopore. The conductance of the nanopore interior changes whenever a biomolecule enters the pore. Detecting these modulations yields information about the molecule. The greater goal of nanopore scientists is to develop a read-out device for the sequence of DNA[8,9]. The electric conductance is caused by ions dissolved in the surrounding water. Typically, potassium chloride (KCl) is used as an electrolyte. The modulation of the current, thus, depends on the interaction of ions with the molecule that is studied. Understanding the origin of the current modulations is of crucial importance to interpret the observed electric current signal. In simulations presented before[10] it was shown that a complex interplay of different interactions between DNA and ions, including water flow, constitute the observed current modulations. The complexity of the three-dimensional geometry, however, obscured the nature of the current modulation as well as the mechanisms creating it. Therefore, we aggregate the ions and their velocities spatially and temporally in order to analyse the ion flux in three dimensions.

We present an integrated visualisation application that allows to analyse the aforementioned simulation of DNA within a nanopore. More specifically, we focus on the visual analysis of ion distribution and flux near the DNA. Our tool combines visualisation techniques for particle data and field data (3D scalar and vector fields). An analyst can choose between several methods to represent different characteristics of the data. The visualisations can be parametrised (e.g. using specific filters or thresholds) to reduce visual clutter and to refine data exploration. This allows experts from the field of biophysics to gain insight into their simulations and, thereby, supports visual discovery of interesting features. We employ GPU-accelerated algorithms for our application to maintain interactive frame rates even during parameter adjustment. This facilitates the intuitive exploration of the data further, as it enables to see the effect of a parameter on the visualisation immediately.

Publicly available programs for visualisation of molecular data sets like Visual Molecular Dynamics (VMD)[11] or Chimera[12] offer a lot of general functionality to illustrate and analyse biomolecules, but often lack some specific features to visualise other data types like vector fields. General-purpose visualisation software like VTK/Paraview[13], however, is usually not tailored to biomolecular data and does not provide sufficient functionality in this regard. Although all of the aforementioned frameworks can be extended using plugins, we opted for our own MegaMol[14] framework, as it supports efficient rapid prototyping of visualisations and is consistently designed for GPU-accelerated algorithms and fast rendering, which are crucial parts of our interactive visual analysis application.

2 Biophysical Background

Fig. 1 (right) shows a typical nanopore setup as used in experiments in the group of Dekker[15,16]: A small hole is drilled into a thin, free-standing silicon nitrite (SiN) membrane with a transmission electron microscope. The wafer is installed in a microfluidic setup such that the pore is the only connection between two water compartments. In one of the compartments a double-stranded DNA is dissolved. Electrodes are inserted into both compartments and a voltage is applied between them. As the DNA is charged, molecules in the vicinity of the pore are dragged into the pore and can be detected as they alter the electric current inside the pore. Our study was intended to gain a quantitative understanding of the modulation of the current, following the experiments of Smeets et al.[15]. They observed an interesting behaviour: If the concentration of the surrounding electrolyte is low, the presence of DNA increases the conductivity. In contrast, it is decreased for high electrolyte concentrations. This indicates a competition between different physical effects.

Our simulation approach to understanding the conduction is similar to that of Luan et al.[17] and sketched in Fig. 1. We concentrate on the central region of a longer nanopore and simulate a cylindrical pore with a DNA molecule. A DNA strand consisting of 20 base pairs is fixed at the centre in the box. Ions and water molecules are added to the system and an electric field is applied along the pore axis. Periodic boundary conditions along the axis assure that whenever a molecule leaves the simulation box on one side, it enters again on the other side. With this simulation setup we sacrifice some realism in favour of a qualitative

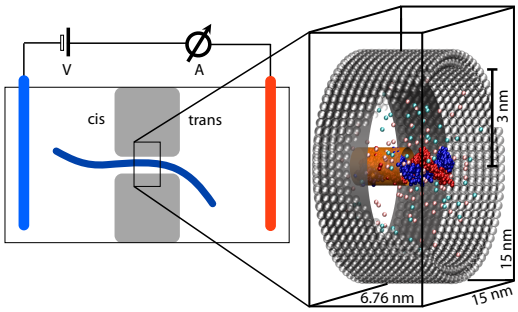


Fig. 1 Schematic setup of the simulated system. We concentrate on a small central region of a larger nanopore (right). The region includes a double-stranded DNA helix containing 20 base pairs surrounded by ions and water molecules (not shown here).

analysis, but we still can find good agreement with experiments.

The total current is the product of the ion density and their velocity integrated over an arbitrary cross section. The magnitude of the current can hence be decomposed into two aspects: the total number of ions that are present in the pore and the question how easily they can move in the applied field. For the first aspect we investigate the distribution of ions around a DNA. Two competing mechanisms are present. First, the DNA occupies a certain volume in space, which is not available to ions and, therefore, the number of ions is reduced, compared to a DNA-free system. Second, the DNA is highly negatively charged and, therefore, surrounded by a cloud of positive K^+ ions with an extent of a few nanometres. The aspect of ion mobility also contains of two contributions. The water movement is induced by the K^+ ions. This *electroosmotic flow* enhances the conductance as it facilitates to motion of the majority of ions. However, in the vicinity of the rough DNA surface the ions are less mobile than in free solution, which reduces the conductance. The interplay of these four effects is responsible for the signal that is measured. Our simulations show excellent agreement with experiments regarding the change of conductivity with respect to the electrolyte concentration. We were able to show that all effects are of similar magnitude, but the surface friction effect is necessary for the observed current reduction.

The next challenge on the route towards a DNA sensor would be to reliably discriminate between different DNA sequences with this method. Originally, we used a DNA consisting only of Cytosine-Guanine base pairs (CG). We performed additional simulations with a DNA containing only Adenine-Thymine (AT) base pairs. The simulation box is of length 6.76 nm in the direction of the pore axis and in total 52 K^+ ions and 12 Cl^- ions were added so that the overall system is electrically neutral. Spring forces are attached to the P-atoms of the DNA backbone to keep the DNA in a fixed position. An electric field of 0.2 V/nm parallel to the pore axis is applied. This causes the ions to move, which in turn also sets the water in motion. In total we performed 100 simulations of 3.5 million steps, corresponding to 7 ns each.

We are interested in the ion current density, which can be decomposed into the product of the ion density and the mean local ion velocity. Since the DNA is fixed in space and can only vibrate weakly, the ion density can be calculated

from aggregating the particle data on a uniform 3D grid in time. Every 0.4 picoseconds, corresponding to 200 simulation steps, the ion positions are written to disc for further analysis. The data aggregation is performed afterwards, using a trilinear interpolation scheme. By construction the system is invariant under simultaneous translation along the axis by 3.38Å and simultaneous rotation by 36 degrees. This invariance can be used to increase the statistical accuracy of the data. During aggregation, each ion is counted 20 times for each simulation time step, once for each of the 20 base pairs of the DNA. Every periodic instance of the ion is rotated and translated according to the orientation of the base pair. That is, we exploit not only the periodic boundary conditions of the simulation but also the periodicity of the DNA strand within the simulation box. The velocity is calculated by taking the ratio of flux density and ion density. For regions with low densities, the statistical accuracy of this quantity is very low, so that below a density threshold the velocity is set to zero.

3 Visualisation Techniques

We use a combination of established visualisation techniques to represent particle data as well as scalar field data and vector field data. A general problem in dense 3D visualisation is visual clutter, especially when showing multiple features of a data set. Our application allows the analyst to use the different visualisation techniques selectively in order to remedy this issue. Furthermore, it offers specific filters for the individual techniques to reduce the amount of visualised data points and ranges. All our visualisations are fully interactive and use GPU-acceleration wherever possible in order to enable smooth, uninterrupted visual exploration of the data. In the following subsections, we summarise the employed visualisation methods and briefly describe their implementation.

3.1 Isosurface Extraction

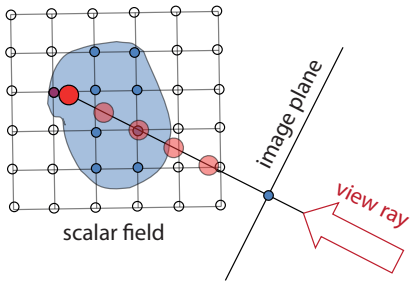


Fig. 2 Volume ray marching. For each pixel in the image plane, the scalar field is traversed along a view ray. Scalar values are sampled in an uniform interval using trilinear interpolation (red dots).

Isosurfaces, which extract features defined by a level-set of a scalar function, are widely used to analyse three-dimensional, spatial scalar data. They are often visualised using either a mesh generated by method like Marching Cubes[19] or

direct volume rendering. In our application, we use GPU-based ray marching [20], a direct volume rendering method that can be used to render a smooth isosurface. The basic idea of ray marching is to traverse the scene along a ray that starts at the camera, thereby sampling the scalar field in a given interval (see Fig. 2). If the scalar value crosses the given isovalue between two sample points (which are obtained using trilinear interpolation), the location of the isosurface along the ray is approximated using linear interpolation. Isosurface ray marching yields high visual quality and runs very fast on current graphics hardware [20]. It has the additional inherent benefit that it renders semi-transparent surfaces correctly, whereas the triangles of mesh-based surfaces have to be sorted for correct transparency. In our OpenGL GLSL [21] implementation, we use multiple isovalues in order to capture the corresponding nested surfaces of the ion density. This allows us to encircle several density ranges at the same time (cf. Fig. 4). While transparency helps to see objects within the isosurface, it can also add to the visual clutter of the scene. Therefore, our application also offers the possibility of using clipping planes to partially remove individual isosurfaces.

3.2 Streamlines

Streamlines are often used in visualisation to represent flows, since they illustrate its dynamics intuitively. Given a time-varying vector field $v(x,t)$, a streamline $L(s)$ is a curve that is tangent to v at any fixed time t. Hence, streamlines can be seen as the solution to the initial value problem of an ordinary differential equation

$$\frac{dL(u)}{du} = v(L(u),t), \tag{1}$$

where $L(0) = x_0$ is the location of the initial value or *seed point*. The most common method to numerically integrate streamlines is fourth order Runge-Kutta integration.

One of the biggest challenges in streamline integration is to choose the proper seed points. The goal is to render as few streamlines as possible in order to avoid visual clutter caused by occlusion and visual overload while still representing all relevant characteristics of the underlying vector field. Various methods that address this problem have been presented. In our case, we simply seed the streamlines randomly within a certain range of isovalues on a user-defined clipping plane. This is sufficient since it allows examining the ion flux around the DNA. The user can also define the number of streamlines and adjust the number of iteration steps (i.e. the length of the streamlines). The streamlines are computed interactively on the GPU using CUDA. We use forward and backward integration in order to extend the streamline in both directions. A GLSL geometry shader implementation that generates a tube around a line is used to render the aforementioned streamtubes. The streamtubes can be colour-coded according to flux or velocity. The clipping plane and isovalue that is used for streamline seeding can also be passed to the isosurface extraction. Hence, the streamlines are seeded inside of the isosurface, which can be clipped at the seeding plane.

3.3 Glyph Rendering

Particles are typically visualised using a geometric object that represents the properties of the corresponding data point, so-called glyphs. To add further information, glyphs can be colour-coded according to a specific property of the underlying data, for example colouring spheres representing atoms by element, charge or index of the polynucleotide chain.

The glyphs in our application are rendered using GPU-based ray casting [22]. In principle, this method is analogous to ray tracing using only primary rays (i.e. no shadow rays, reflection rays etc.). A proxy geometry is rendered that covers the whole glyph with respect to the view and a fragment shader is used to compute the intersections between the view ray passing through every fragment and the glyph. If a ray does not hit the glyph, it is discarded, otherwise, the lighting at the hit point is computed (based on the surface normal, the colour and the light properties). For objects with an implicit surface that can be described by a combination of few polynomial functions of low degree (e.g. spheres, cylinders, tori etc.), this technique is significantly faster than rendering classical triangle meshes.

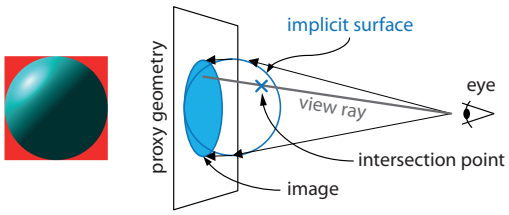


Fig. 3 Glyph ray casting for a sphere. Left: The final, pixel-accurate image of the sphere. Right: For each pixel of the proxy geometry, a view ray is intersected with the implicit surface of the sphere. If the ray hits the sphere, the colour of the corresponding pixel is determined based on the sphere colour and lighting (Phong shading).

The vector field describing the ion flux can also be rendered using glyphs. For each grid cell, we show an arrow glyph (formed by a cylinder and a cone) that points in the direction of the vector field. The size of the arrow can be scaled according to the ion density or flux at this point. Additionally, the arrow glyphs can be filtered by arrow length (i.e. ion flux) or ion density. Different colour mappings can be applied to show more information about the vector field at this point, e.g. density, flux, or direction. While the arrow glyphs are, in some ways, redundant with respect to the streamlines, they can give a more detailed representation of the field, while the streamlines show the general characteristics of the field more intuitively. The arrow glyphs, however, should be applied only sparingly, because they are prone to create occlusion and visual clutter.

4 Results & Discussion

The visual analysis tool described above was applied to explore the simulation results for the nanopore conductance. It follows the visual analysis paradigm for exploratory data analysis and has led to these findings.

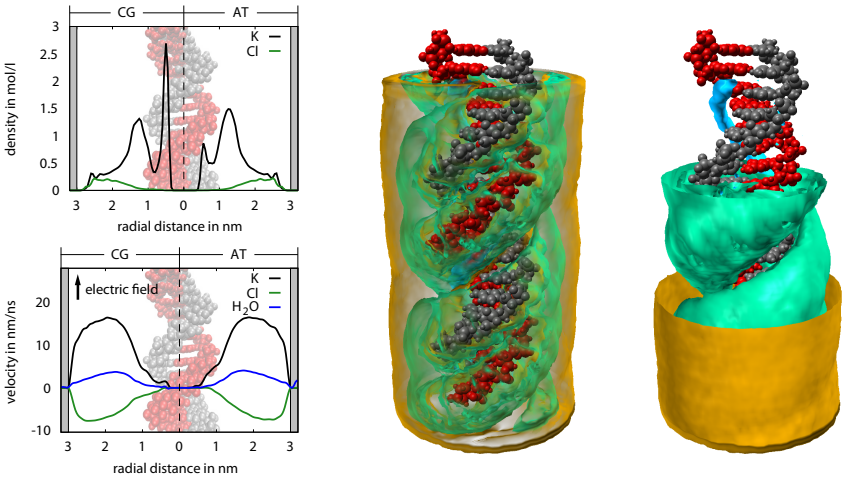


Fig. 4 Left: One-dimensional projection of the ion density (top) and the ion velocity (bottom) for both CG- and AT-DNA. The right parts of the diagrams show the curves for CG-DNA, while the left parts show the curves for AT-DNA. The DNA strand and the pore walls are given for reference. Right: Two screenshots taken from our application showing three ion density isosurfaces for the CG-DNA molecule. The innermost, blue isosurface encapsulate ions inside the major groove of the DNA and strictly follows the helix shape. The outermost, yellow isosurface exhibits an almost perfectly circular cross section with a radius of $\sim$2 nm, indicating that for larger distances a rod-like model of DNA is absolutely adequate.

First we investigate features of the ion density. In the left Panel of Fig. 4, we display the ions' density and their velocity, including the water velocity, in a one-dimensional projection of the data. The mean ion density is calculated from the trajectories by binning the data based on the radial distance. For both datasets, the ion density exhibits two peaks of different magnitude at around r =0.6 nm and $r = 1.2$ nm. By comparison to the DNA structure, it is apparent that the innermost peak corresponds to ions entering the grooves of the DNA molecule. The magnitude of these peaks is different for the CG- and the AT-DNA, which will be investigated later. At $r > 1.2$ nm, due to the charge of the DNA, a cloud of K^+ ions forms which decays to around 0.3 mol/l at the pore boundary. This diffuse layer of counter ions is similar for both datasets. The velocities will be discussed below. In the rightmost panel of Fig. 4, we display three representative isosurfaces of the DNA density for the CG dataset. The blue surface corresponding to the isovalue of the highest density winds tube-like around the DNA helix inside the major groove of the DNA molecule. This structure corresponds to the first peak in the projected data. The intermediate, green isosurface has a complicated shape but is mostly cylindrical, with a gap in the major groove of the DNA. The third, yellow surface is almost indistinguishable from a cylinder, reflecting the fact that around $r > 1.5$ nm, almost all three-dimensional influence of the DNA double helix on the ion density has vanished.

In Fig. 5 we display a comparison of the ion densities closest to the DNA surface. We combine a transparent isosurface rendering with a stick representation

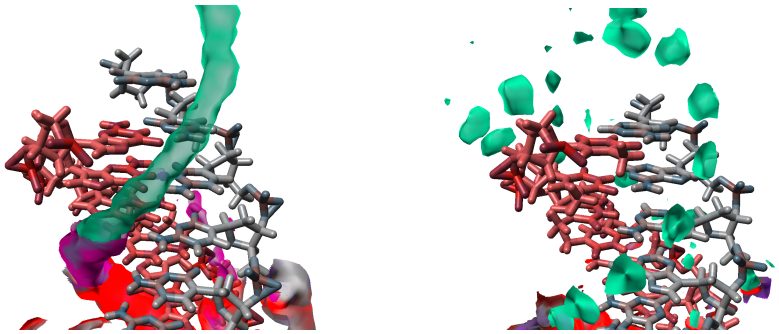


Fig. 5 Binding of K^+-ions to a CG-DNA (left) and an AT-DNA (right). For the CG-DNA the innermost layer of ions binds very tightly to the major groove. The ions prefer close binding to the oxygen and nitrogen groups of the bases. In case of AT-DNA the major groove is far less attractive. Several binding sites can still be identified, most importantly the strongly negatively charged oxygen atoms at the backbone.

of the DNA molecule, which is cropped to a region of interest at integral base pairs. It can be observed that the tube-like structure of the CG dataset is located directly at the N and O groups of the C or G base. The vicinity of two such atoms appears to create a strongly attractive site for K^+ ions. At positions between two base pairs, the O atom of the next pair is also close, and this effect makes the tube structure continuous. For the AT-DNA the situation is entirely different. There is no single attractive site in the groove, but the K^+ ions are close to the strongly negatively charged phosphate group of the backbone. There they can be in the vicinity of two oxygen atoms. No clear binding sites in the grooves can be found. This is reflected in the first peak in the upper diagram Fig. 4 of the AT dataset being considerably lower than for the CG dataset.

Next, we focused on the vector field of ion velocities for the interesting regions with high ion density found in the last step. As described in Section 3.2, the streamlines (or -tubes) give a good impression of the general behaviour of the vector field. Fig. 6 shows a combination of streamtubes and arrow glyphs for regions with high ion density. Upon closer investigation of the velocity field for the CG data set, we noticed that there are regions where the majority of ions does not move in conformity with the electric current, as there are regions in the velocity field where the vectors point in the opposite direction. This effect is still visible if we use a lower temporal sampling rate and is, therefore, not due to short-time motion of the ions (cf. Fig. 6).

We adapted our visualisations to investigate this phenomenon in more detail. Fig. 7 shows the arrow glyph visualisation of the ion velocity vector field inside the grooves (regions with high ion density, green isosurface). The arrows are colour-coded according to their direction: Red arrows indicate velocity vectors that point in the direction of the electric current. This is the anticipated direction of ion motion. There is, however, a region in the major groove where the ions move in the opposite direction. Intuitively, this should not happen, since the force of the applied current is stronger than other competing effects (e.g. negatively charged atoms in the DNA). Again, this is no effect of short-time motion, as it is also clearly visible on a longer time scale (cf. Fig. 7 right). This phenomenon is

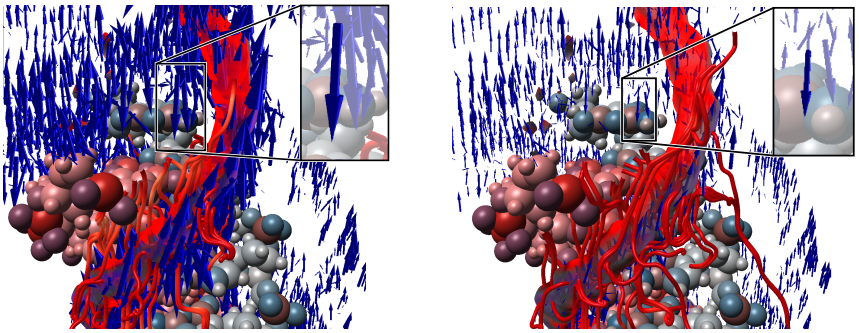


Fig. 6 Ion motion near the CG-DNA taken with two different sampling rates: 5/ps (left) and 0.1/ps (right). The streamlines are coloured according to the normal component of the local ion velocity. They give an overview of the vector field in interesting regions, while the arrows support a more detailed analysis of the velocity. On short time scales a rapid, ballistic ion motion is apparent. If the trajectories are smoothed by using a lower sampling rate, the short-time motion is averaged out. On long time scales, the motion of the ions is largely suppressed, reflecting the friction effect near the surface. The current of the applied electrostatic field should pull all ions upwards; however, there are regions where the ion flow is directed downwards (see highlighted arrows in the inset cutouts). This effect is not due to short-time motion (left), as it is also observable on longer time scales (right).

not fully understood as it does not comply with anticipated behaviour.

This discovery shows the importance of visual analysis of simulation results. In particular, parametrisable, flexible three-dimensional visualisations can be of great importance to finding unanticipated effects or features in the data. While our initial goal was to verify and illustrate the ion flux in the simulated nanopore, it led to the detection of regions where ion motion opposes the electric current. This phenomenon was not observable in the one-dimensional projection of ion velocities shown in Fig. 4. In the future, we want to further investigate the motion of the ions within the grooves more closely in order to determine their effect on the overall conductivity within the nanopore.

5 Future Work

In the future, we want to extend our visualisation application in order to enable a more detailed analysis. As mentioned above, the regions within the grooves where the ion motion contradicts the electric current are of special interest. One possibility would be to apply the concept of solvent pathlines[6] to the trajectories of the ions. The paths of the ions, however, are likely to be much noisier than the water paths that Bidmon et al.[6] investigated. Therefore, we would need alternative methods for smoothing, filtering, and clustering. An example is the clustering of trajectories not only by their absolute spatial position but also by their position in relation to the nucleotide bases.

The visualisations used in this work to analyse the ion flux could also be applied to the water molecules in the simulations, since water molecules are brought into motion by the ions. Visualisations of the velocity of the ions and the velocity

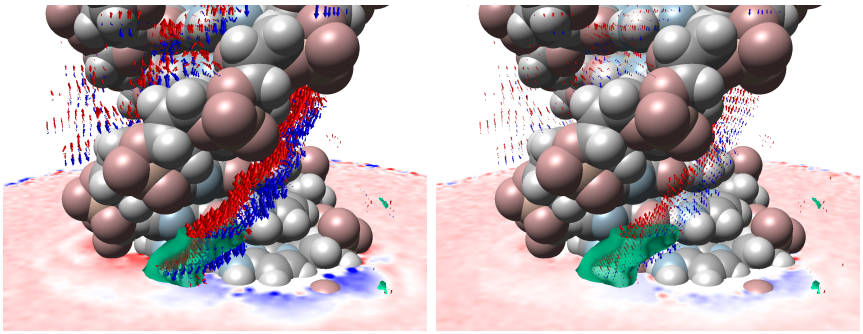


Fig. 7 Ion motion inside the major and minor groove of CG-DNA taken with two different sampling rates: 5/ps (left) and 0.1/ps (right). The green isosurface encircles the region with high ion density, which is following the major groove. In this region, ion motion is partially in compliance with the electric current (upwards, red arrow glyphs). There is, however, a distinct region where the direction of the ion motion opposes the direction of the current (blue arrow glyphs). The atoms of the DNA are colour-coded by their chemical element using the common CPK colouring scheme.

of the water molecules would show their mobility with respect to the pore and the DNA and to what extent they comply with each other.

Another interesting direction for future work would be to compare the all-atom simulations analysed in this paper with simulations using the Poisson-Boltzmann equation (implicit solvent and ions). Here, a specialized comparative visualisation could help to identify whether this would lead to differences in the resulting ion density and flux.

6 Summary & Conclusions

In this paper, we presented a visualisation application for simulations of DNA within a nanopore. The density and motion of the ions is of special interest, since they play an important role in the resulting electric current. As this current depends on the molecule within the pore, it allows to use the pore as a biosensor. The greater goal is to use nanopores as read-out devices for DNA sequencing. In our work, we investigated the current modulation caused by two different types of DNA, namely a DNA strand consisting of AT base pairs and one consisting of CG base pairs. Here, visual analysis turned out to be a driving force for making unexpected discoveries in the simulation data. Using our visualisations, we were able to identify regions in the CG-DNA where the ion motion contradicts the applied electric current. This might have an effect on the overall current through the nanopore, which would be important for the application of DNA sequencing. We also showed that dimensionality reduction, although helpful in many cases, can conceal interesting features that are only observable in 3D. As discussed in the introduction, visual analysis is a powerful tool to generate insight. During our work, the close collaboration between experts from the fields of visualisation and biochemistry greatly facilitated the process, since it allowed us to create a visualisation tool that is tailored to a specific task but still flexible enough to support making novel discoveries. Adapting and refining the visualisations during the

project in was an important building block for the analysis leading to the results presented in this paper.

Acknowledgements

This work was partially funded by the German Research Foundation (DFG) as part of the Collaborative Research Centre SFB 716 (projects D.3, D.4 and C.5).

References

1 L. Rosenblum, R. Earnshaw, J. Encarnacao, H. Hagen, A. Kaufman, S. Klimenko, G. Nielson, F. Post and D. Thalmann, *Scientific Visualization: Advances and Challenges*, Academic Press, 1994.

2 *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, ed. J. J. Thomas and K. A. Cook, 2005.

3 D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas and H. Ziegler, *Visual Data Mining*, Springer, 2008, pp. 76–90.

4 J. Kehrer and H. Hauser, *IEEE Transactions on Visualization and Computer Graphics*, 2013, **19**, 495–513.

5 B. Thomaß, J. Walter, M. Krone, H. Hasse and T. Ertl, International Workshop on Vision, Modeling and Visualization, 2011, pp. 301–308.

6 K. Bidmon, S. Grottel, F. Bs, J. Pleiss and T. Ertl, *Computer Graphics Forum (Proceedings of EUROVIS 2008)*, 2008, **27**, 935 – 942.

7 N. Lindow, D. Baum, A.-N. Bondar and H.-C. Hege, *Proceedings of IEEE Symposium on Biological Data Visualization (biovis12)*, 2012, 99 – 106.

8 J. J. Kasianowicz, E. Brandin, D. Branton and D. W. Deamer, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 13770–13773.

9 A. Meller, L. Nivon, E. Brandin, J. Golovchenko and D. Branton, *PNAS USA*, 2000, **97**, 1079–1084.

10 S. Kesselheim, W. Müller and C. Holm, *Physical Review Letters*, Accepted Tuesday Nov 5, 2013.

11 W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

12 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of Computational Chemistry*, 2004, **25**, 1605–1612.

13 W. J. Schroeder, K. M. Martin and W. E. Lorensen, *The Visualization Toolkit: An Object-Oriented Approach to 3-D Graphics*, Pearson, 1997.

14 *MegaMol Project Website*, http://www.vis.uni-stuttgart.de/megamol (last accessed: 2013-12-20).

15 R. M. M. Smeets, U. F. Keyser, D. Krapf, M.-Y. Wu, N. H. Dekker and C. Dekker, *Nano Letters*, 2006, **6**, 89–95.

16 S. van Dorp, U. F. Keyser, N. H. Dekker, C. Dekker and S. G. Lemay, *Nat Phys*, 2009, **5**, 347–351.

17 B. Luan and A. Aksimentiev, *Physical Review E*, 2008, **78**, 021912.

18 T. Can, C.-I. Chen and Y.-F. Wang, *Journal of Molecular Graphics and Modelling*, 2006, **25**, 442–454.

19 W. E. Lorensen and H. E. Cline, Proceedings of ACM SIGGRAPH Computer Graphics and Interactive Techniques, 1987, pp. 163–169.

20 M. Hadwiger, P. Ljung, C. R. Salama and T. Ropinski, ACM SIGGRAPH Asia 2008 courses, 2008, pp. 1–166.

21 R. J. Rost, B. Licea-Kane, D. Ginsburg, J. M. Kessenich, B. Lichtenbelt, H. Malan and M. Weiblen, *OpenGL Shading Language*, Addison-Wesley Longman, Amsterdam, 3rd edn, 2009.

22 G. Reina and T. Ertl, Eurographics/IEEE VGTC Symposium on Visualization, 2005, pp. 177–182.

