doi:10.2533/chimia.2022.748 Chimia 76 (2022) 748–753

Visual Anatomy of an Article – Lessons Learned and Taught in Five Figures

Michel Rickhaus*

Dedicated to Professor Michal Juríček on the occasion of his fortieth birthday

Abstract: To grow is to teach. In that spirit this article aims to teach best practice for visualizations and at the same time to highlight some of the harvest we could reap in our lab. Following figures recently published in one of our manuscripts, we demonstrate how creativity in figure-making goes hand in hand with storytelling, a deepened understanding of the scientific work and an increase in lateral thinking inside the research avenue. We hope it will help to promote scientific visualizations as a useful, even if sometimes harsh, mirror to give new angles to the laboratory work.

Keywords: Data visualization · Shape-assisted self-assembly · Tutorial · Visual literacy

Michel Rickhaus started his research group at the University of Zurich in August 2019 as an SNSF Ambizione fellow. His main themes are synthesis of polyaromatic species and supramolecular assemblies. Besides the lab, he spends hours making figures, thinks about visual tools and how to pass knowledge on to the next generation, drinks lots of tea in the process and generally loves to bring talent together.

The Design Process

How do you approach figures? Raise your hand if it goes something like this: you perform experiments, you collate the data. You write up the results, make graphs or figures and drop them in the manuscript, caption it something like "NMR of structure **1a**", then move to spend weeks optimizing your surrounding text. For many of us, it took a lot longer to compile and analyze the data and write the manuscript than it did to showcase the data. Default setting of your data treatment program is usually the way to go. Click, export, done.

Writing a manuscript is a solitary thing. It is hard, especially the first time, but generally accepted to be essential for a scientist. Making figures on the other hand, is a lot like cooking. Everyone has to eat, so we learn how to feed ourselves. Graphics are the meat in a manuscript so we all make them, but there is a difference between being fed and eating well. With everything that is going on, there is little time to learn visual design. And hence the relief that most software can generate some sort of graph automatically. What seems almost forgotten in the process is that it is *people* that will read the paper. And people are a visually driven species. It is estimated that about 10 Mb/s of the sensory influx in our brain is dedicated to visual processing, roughly eight times more than all other senses combined.[1] Yet, when it comes to visualizing our data, we often stick to the default approach. Why? Maybe because visualizations seem to require complex programs, or maybe your peers do not see it as time well spent. Why make good figures if the audience already knows what is going on? And there lies the danger: you may have seen your data hundreds of times, spent some good few years with it in fact. But most people haven't (and won't). And if they don't get it, they won't endorse it. Thinking of how to present your data and how to make it accessible *quickly* is as important as the data itself – if you want to make a lasting impact that is. *Have pity on the reader*,[2] as Kurt Vonnegut puts it, aim to make others understand – and therefore be understood.

Missing experience in visualizations can serve as Occam's razor – the struggle with *mise-en-scène*, having limited knowhow and available abilities for visualizations can help in simplifying the process. Conversely, most often one eventually finds it is the underlying story one is struggling with – not the technical aspects of figure making. Making good figures is merciless in revealing the limit of one's understanding. But the research is, at least in my experience, the better for it. At the University of Zurich, we have begun to teach data visualization as a graduate course (*Visualize your Science – make others see what you mean, CHE 915*). Together with my colleague and dear friend Prof. Michal Juríček (to whom this article is dedicated), we have taken the first steps to help our students expand their graphical literacy. And in turn we were found to be strongly influenced in our own research.

This tutorial highlights a few lessons learned that helped us to design approachable figures – we showcase those in the form of a commentary of selected figures from our latest article. Each of these figures stands representative for a distinct role within a research paper: the first figure of the article (the underappreciated gateway), the data figure (how we can guide the reader through our treated datasets) with a little excursion on presenting larger datasets, the table (how to navigate numbers & indicators) and the TOC figure (the one-shot marketing chance).

Shape-assisted Self-assembly

Central to good (visual) communication is to know what you intend to communicate. In our laboratory, one of the fundamental questions we ask is: what governs intermolecular order? Having worked as a materials chemist, I have experienced how one can shape functional structures by careful molecular modifications. As a synthetic chemist, I have also learned to appreciate the challenges to scale these systems to the size regime of the semiconductor industry. By sticking to what we know best – that is, how to make small molecules – we can bridge the persistent size gap. We introduce structural information in monomers that sorts them into large arrays with function as an emergent consequence of their

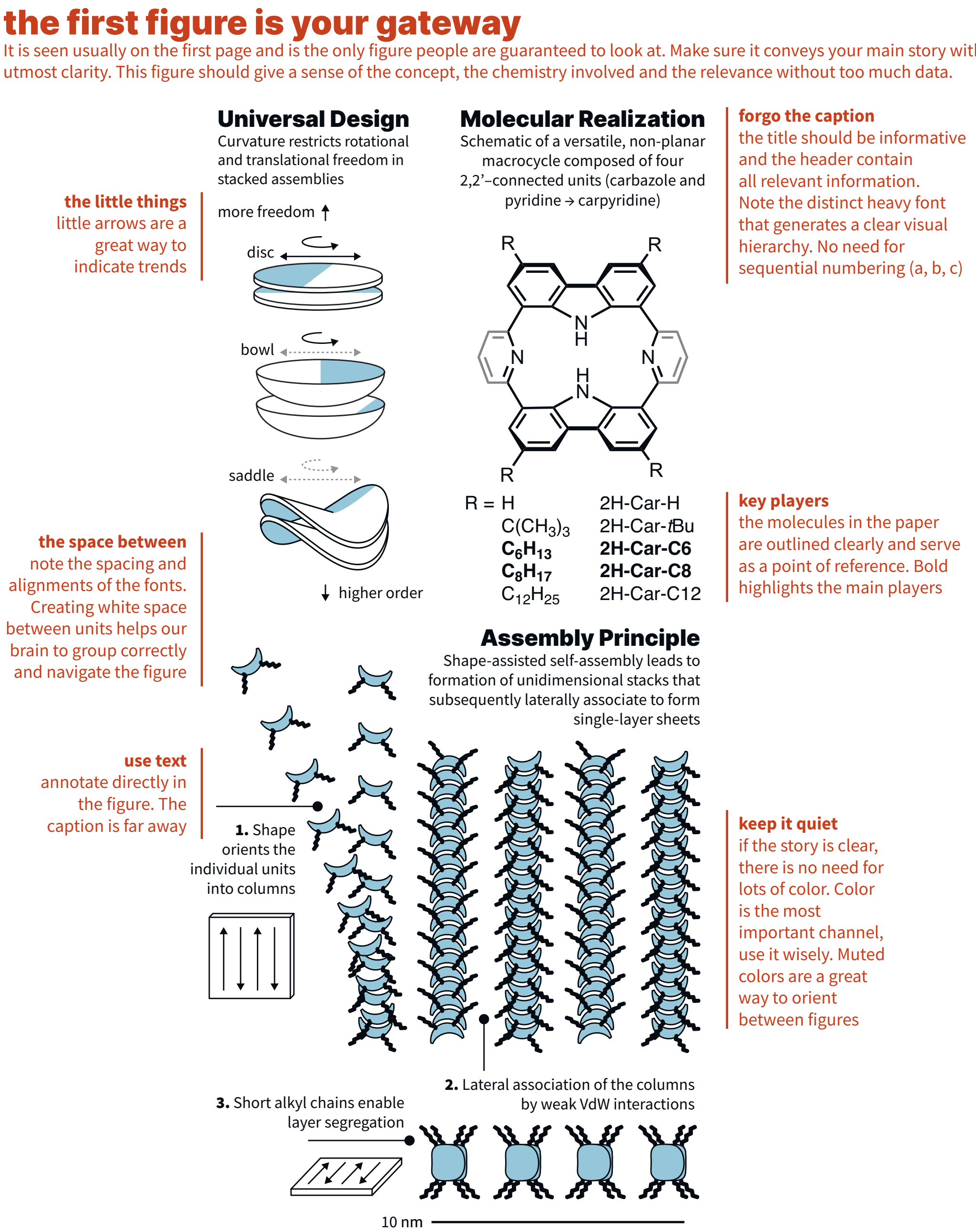
Correspondence: Dr. M. Rickhaus, E-mail: michel.rickhaus@chem.uzh.ch, Department of Chemistry, University of Zurich, CH-8057 Zurich

order. But to structure assemblies, one needs to direct the intermolecular interaction. Besides the covalent bond, which is prone to error propagation when performed sequentially for large numbers of times, it is soft interactions that are particularly suited to instigate intermolecular order. Most prominently used are salt bridges or hydrogen bonds; appending those to a sufficiently rigid core that holds them in a favorable arrangement can lead to astonishing polymeric structures. The strength of such a 'soft' assembly compared to a covalent polymer is that the monomers usually stay interchangeable, leading to self-healing, stimuli-responsive systems. From a molecular design point, the core of such monomers is typically seen as a placeholder, or with sufficiently large π-systems, as a chromophoric or electronic platform but not necessarily as a structuring element. This is where our shape-based approach comes in – what if we could get the *core to assist in the assembly*?[3] That might liberate the sidechains for function and maybe even allow to dispose of strong interactions altogether, enabling a finer gain of responsiveness. As polyaromatic chemists, we settled – surprise – on π–π interactions as our weak monodirectional interaction of choice. The simplest electronic analogue we intend to mimic are wires, which translates to discotic stacks in the lingo of supramolecular chemists. Polyaromatics are known for their ability to twist and bend when defects are introduced into their lattice or bulk is attached to their rim. It seemed logical that monomers composed of a saddle-shaped core would approach each other such that their π–π interactions would be maximized. They could do so only by stacking eclipsed onto each other, the saddle acting as a directional bias. That is in a nutshell our concept of shape-assisted self-assembly. And that is the story we tell in our manuscript.[3]

The First Figure

In my experience, the first manuscript figure is underused and underappreciated. The reader has already opened your article, so why bother with an explanatory figure? I would argue that often the abstract figure (or TOC figure) will not be shown in the PDF and hence it falls to this figure to convince the reader to stay on.

As chemists, we are not just curators of data. Behind our work stands data creation and with it a concept, an idea. The first figure of the manuscript should reflect that and set the mood. Think of it as a map that localizes your reader. It contains information about the type of chemistry, the molecules or systems, and entices with the outlines of the chosen approach. In our case, the idea that shape will guide an assembly by restricting the possibilities of successful interaction. The rest? We let the figure talk.



The Data Figure

Noticed how we did not need a lengthy paragraph explaining the concept? The first figure basically worked on its own. And there was not even a caption. Now, the reader has understood the concept and the broad sense of the paper, he/she has obtained a sense of orientation and is hungry for more. Excellent. Time to pitch data. Data comes in wide variety and there are excellent guides on how to present data using a variety of methods, plot types, and so on (see the reference section and for some suggestions, see the end of the article). Instead, let us now consider the intention of data figures. Mostly, they are there to build confidence in the presented work. Data figures in a manuscript are *highlight* figures—the full dataset will be stored in the supporting information (SI). Summarizing data is difficult in itself but especially hard is striking the right balance between accuracy without overload and simplification without imprecision.

Set the right tone. Generally, a subdued figure is better. The reader at this point is already invested. Using color sparingly but to the maximal effect, highlighting areas of interest while forgoing effects like drop shadows and blurs helps to strike a tone of quiet competence.

Use labels. It is a common practice to add all information in the caption and try to minimize text in a figure. But consider a map, arguably a highly visual representation of spatial data. Mentally move all annotations of that map to captions and legends. Now consider describing a road on that map. Hard, right? Adding text to data figures helps with navigation. If you have a data outlier, why not annotate it with 'increased background scattering'. Or a kink in a progressive line: adding a little arrow and writing 'addition of catalyst'. Treating data figures as maps can help to substantially improve accessibility.

Dare to leave out. The supporting information is the place for complete, detailed displays of datasets. The manuscript is not. You need to select, and not only in the overall number of figures. There are two strategies that are often used in the world of data visualization: highlight or exemplify.

Highlight. When incorporating lots of traces in a data figure, it soon becomes crowded. A common choice is to strike a balance between included data traces and those left out. Usually, that results in too many graphs in the same plot because the more data you can show, the more convincing your trends and thus your conclusions become. A trick that data visualists play, especially with large datasets, is to show all of them, but without color or annotation, almost like a background. Only a few key traces are highlighted with full annotation. In this way, one profits from the combined dataset without visually overwhelming the reader. The data is shown fully annotated as single graphs in the SI. This approach is particularly good for line graphs with changes over time.

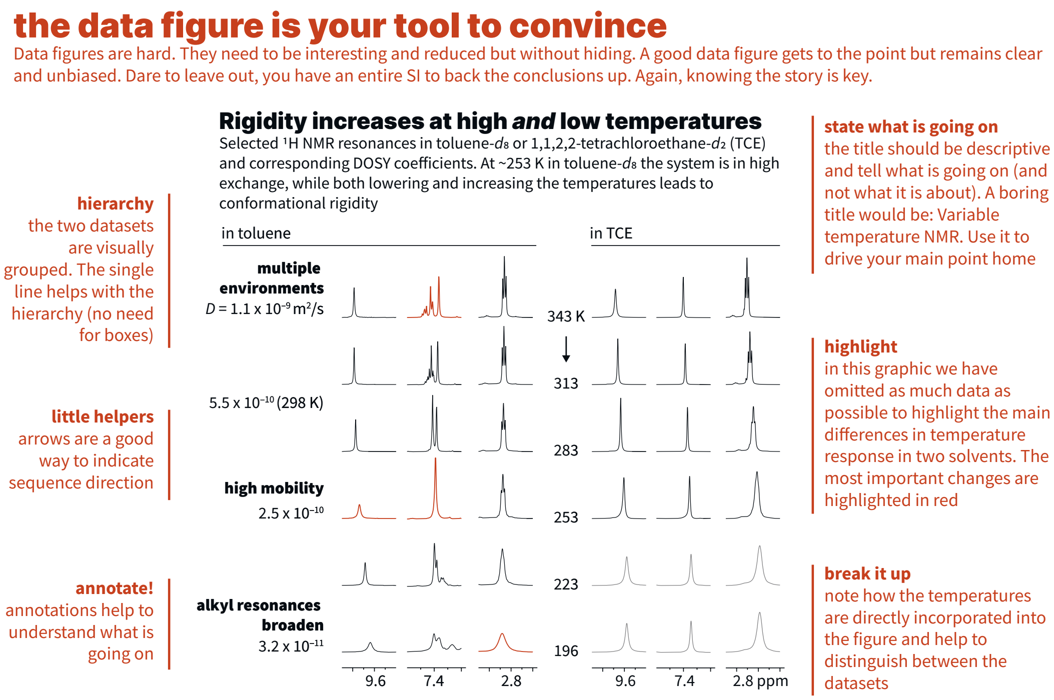
Exemplify. If the dataset contains many similar examples that strongly overlap, an exemplifying approach can be more beneficial: a single data trace is selected (for instance an NMR spectrum). The subsequent datasets are not shown but treated and parameters derived. An appended table just underneath the figure shows changes in shifts (or association constants, or …), which facilitates comparisons while maintaining visual interest. An additional bonus of this strategy is that the spectra shown can accommodate more annotations.

The Table Figure

Most of us never really consider tables as being a visual tool. But tables are everywhere. Websites are essentially nothing else than sophisticated versions of tables. In the scientific context, a good table can help to indicate trends and give overviews over the derived parameters of the presented work. Compared to other figures, there are good rules for constructing solid tables, of which we show a few in the corresponding figure. An exemplary guide is found in ref. [4].

The Table of Contents (TOC) Figure

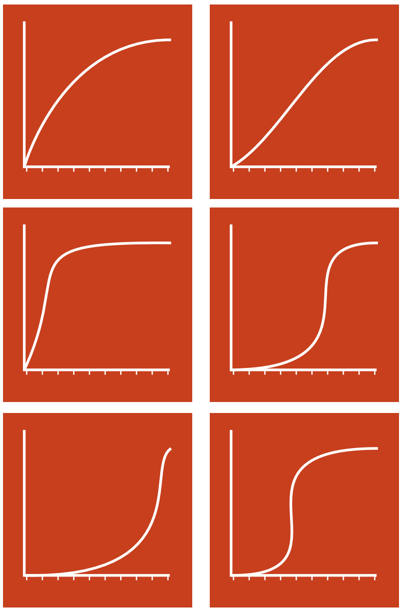
Finally, the TOC figure. It is meant as a graphical summary of the manuscript and featured on the website of the publisher. From my experience, this figure is usually almost an afterthought. After all, your story has come together, the manuscript is ready to be submitted (or already has been accepted), so the TOC figure seems not that important. Right? It is true that for you, this figure usually comes last in the design process. But for your public it comes first. The TOC figure is the point of contact when browsing social media, the ASAP pages of the journals, *etc*. We are visual beings and drawn to figures, and especially color. The TOC figure is thus often studied before the title. A good summary figure is key to getting noticed. It should be seen as the cumulation of a manuscript more than a necessary evil. And nothing states that you have understood the presented science more than when you are able to tell the entire story in a single image.

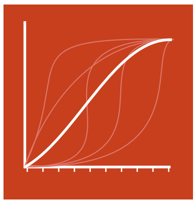


excursion: presenting data differently

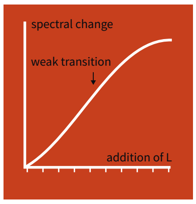
To show multiple datasets, we often tend to show them all at once, either by showing them as individual graphs side-by-side or as a composite in a single graph. Two strategies that create much better visual interest while maintaining integrity are either to highlight and label only the most important graphs, while the others are shown but without emphasis or color and not labeled. Or by picking an exemplary graph and show the derived data in a table underneath for comparison. The full datasets are shown in the Supporting Information (SI).

use the SI
don't show too much in the main manuscript - six times the same graph gets boring fast. To show all data is what the SI is for. Instead, consider the two strategies on the right





highlighting
all six data traces are shown, but only a representative trace is emphasized. The others remain unlabeled and colorless



#	K_D	#	K_D
1	K_1	**4**	K_4
2	K_2	**5**	K_5
3	K_3	**6**	K_6

exemplify and derive
pick a highlight or a representative example and show the full graph, including annotations. *Derive* the other dataset (calculating K_D or shift changes, etc.). Show this data in an associated table directly underneath

the table as a visual tool

Tables are great! The ruleset of making good tables is usually similar and exemplified below going from the raw table (top-left) to the final product (bottom right). A more comprehensive ruleset can be found in *Journal of Benefit-Cost Analysis*, **2020**, 11, 151-178.

T/K	H_a	H_b	H_c
293	9.78	8.02	7.42
283	9.77	8.02	7.41
273	9.77	8.02	7.40
263	9.76	8.03	7.40
253	9.74	8.03	7.39
243	9.72	8.04	7.39
233	9.70	8.04	7.35
223	9.69	8.04	7.33
213	9.67	8.04	7.31
203	9.65	8.04	7.29

stage 0: Raw table

T/K	**H_a**	**H_b**	**H_c**
293	9.78	8.02	7.42
283	9.77	8.02	7.41
273	9.77	8.02	7.40
263	9.76	8.03	7.40
253	9.74	8.03	7.39
243	9.72	8.04	7.39
233	9.70	8.04	7.35
223	9.69	8.04	7.33
213	9.67	8.04	7.31
203	9.65	8.04	7.29

stage 1: hierarchy
separate the header and the bottom. Don't do gridlines. Align text to left numbers to right

T/K	**H_a**	**H_b**	**H_c**
↓293	↓9.78	8.02	7.42
283	9.77	8.02	7.41
273	9.77	8.02	7.40
263	9.76	8.03	7.40
253	9.74	8.03	7.39
243	9.72	8.04	7.39
233	9.70	8.04	7.35
223	9.69	8.04	7.33
213	9.67	8.04	7.31
203	9.65	8.04	7.29

T/K	**H_a**	**H_b**	**H_c**
→ 293	9.78	8.02	7.42
→ 283	9.77	8.02	7.41
273	9.77	8.02	7.40
263	9.76	8.03	7.40
253	9.74	8.03	7.39
243	9.72	8.04	7.39
233	9.70	8.04	7.35
223	9.69	8.04	7.33
213	9.67	8.04	7.31
203	9.65	8.04	7.29

stage 2: reading direction
spacing guides the reading direction. Left to right: add spacers between rows. Top to bottom: add spaces between columns

stage 3: highlight outliers
indicate regions of interest in your data. Here we are looking at shift changes in the data, so we include them and mark substantial outliers

T/K	**H_a**	**H_b**	**H_c**
293	9.78, +/−0.00	8.02, +/−0.00	7.42, +/−0.00
283	9.77, −0.01	8.02, +/−0.00	7.41, −0.01
273	9.77, −0.01	8.02, +/−0.00	7.40, −0.02
263	9.76, −0.02	8.03, +0.01	7.40, −0.02
253	9.74, −0.04	8.03, +0.01	7.39, −0.03
243	9.72, **−0.06**	8.04, +0.02	7.39, −0.03
233	9.70, **−0.08**	8.04, +0.02	7.35, −0.05
223	9.69, **−0.09**	8.04, +0.02	7.33, **−0.09**
213	9.67, **−0.11**	8.04, +0.02	7.31, **−0.11**
203	9.65, **−0.14**	8.04, +0.02	7.29, **−0.13**

stage 4: declutter and annotate
decluttering is key. The newly added shift trends are only needed when new changes happen, so we can safely delete them. Same goes for repeated units etc. Indexes are great for adding annotations below the bottom line of the table. Titles typically go on top

NMR shifts (center of signal) of the **2H-Car-C6** aromatic region in toluene-d_8 (δ; Δδ) in ppm

T/K	**H_a**	**H_b**	**H_c**
293	9.78	8.02	7.42
283	9.77, −0.01	8.02	7.41[a], −0.01
273	9.77	8.02	7.40[a], −0.02
263	9.76, −0.02	8.03, +0.01	7.40[a]
253	9.74, −0.04	8.03	7.39[a], −0.03
243	9.72, **−0.06**	8.04, +0.02	7.39[a]
233	9.70, **−0.08**	8.04	7.35[a], −0.05
223	9.69, **−0.09**	8.04	7.33[a], **−0.09**
213	9.67, **−0.11**	8.04	7.31[a], **−0.11**
203	9.65, **−0.14**	8.04	7.29[a], **−0.13**

[a]broad, assignment unclear

Design for the medium. The TOC figure is different for the journal you publish in than for Twitter (or any other social media outlet) and different again for newspaper outlets. It needs to be adaptable enough for any media.

Reduce to the max. Simplicity is key and nowhere as important as in a TOC figure. Microsoft used eye-tracking software to analyze how long people look at tweets. On average: 2.92 seconds.[5] There is very (!) little time to make someone understand what is going on in the manuscript.

Beware of color. Color is probably the most powerful visual channel we have. If you plan to use it in a TOC figure, use it to the maximal effect, highlighting the single most important feature of your figure. Multiple colors tend to blend together when scrolling, so usually one or maybe two is more than enough.

A common theme. If you choose to have multiple images published, say for a highlight article somewhere besides your manuscript, make sure they tell a common story. A cornerstone of our story is the role of saddle-shapes, so we made sure to have it featured anywhere we can.

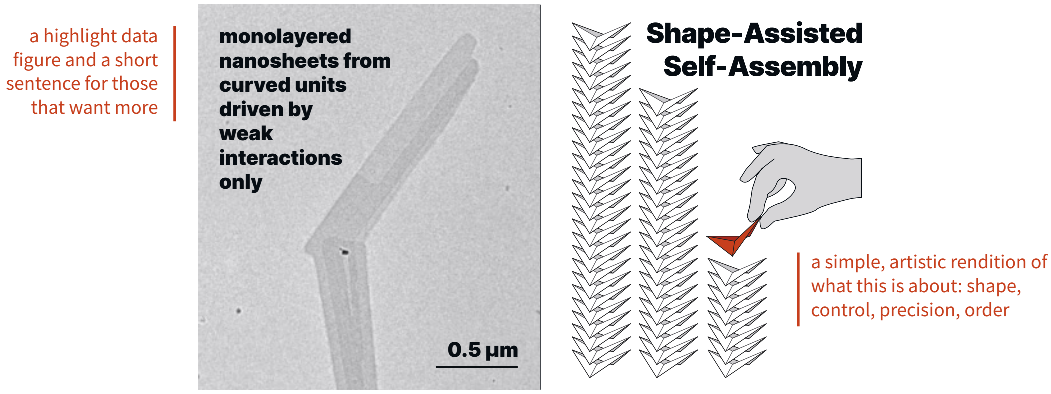
Closing Remarks

Making good figures is hard, but not necessarily for the technical aspects. A good figure can easily come out of Excel or PowerPoint as much as a dedicated high-end software solution. The key is almost always not the artistic ability in data-driven visualizations, but in knowing the story one wants to tell. The question should not be what color to use but what aspects are important, what is the conceptual novelty of the work, whether to convince or represent. Figures are the cumulation of an elevator pitch. You have spent years obtaining your data and your figures should reflect that in the best possible way. Work by identifying the main arch, then decide in how many acts the story will be told. Sketch things, no need for anything fancy, paper and pen will do nicely. It will help you to keep things simple and approachable. Discuss the figures with your peers, you are designing for others after all. Again, making figures can be an excellent self-check to identify knowledge gaps.

Most importantly, do not be afraid to experiment, to iterate and to throw away. It takes a while to get it right even for the most seasoned designer and it is often this process that uncovers additional aspects and viewpoints. Done right, drawing figures can lead to a deepened understanding of one's own work. And by consequence, finally to be understood.

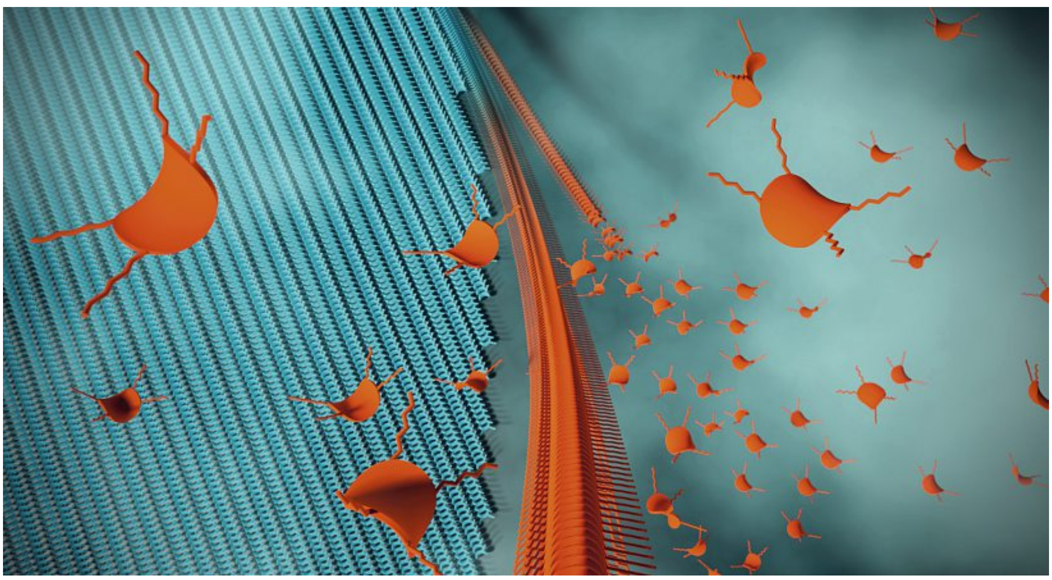
in *medias res* with the TOC figure

The TOC figure is the doorway for your readers. Keep it simple, but not plain. It should maximize clarity and help the readers orient - remember they haven't read the paper yet. Use color sparingly and to maximal effect.



advertise beyond the TOC figure

If you design for a news outlet or a cover, you can really let your artistic side shine. Note how thematically it links up with the TOC figure and the theme of the paper. We have the shape-assisted aspect, we see the material top-down (like in the TOC figure)



Acknowledgements

This tutorial essay contains many lessons learned, and was influenced by many of my peers but maybe by none more than Prof. Michal Juríček. He has been a driving force and a muse for my approach to scientific visualizations.

I thank the funding agencies, in particular the Swiss National Science Foundation (PZ00P2_180101) and the Fonds zur Förderung des akademischen Nachwuchses (FAN) to have the liberty to work creatively, every day.

Received: July 12, 2022

[1] M. Zimmermann, 'The nervous system in the context of information theory', in 'Human Physiology', Eds. R. F. Schmidt, G. Thews, Springer-Verlag, **1989**.

[2] K. Vonnegut, S. McConnell, 'Pity the reader: On writing with style', Seven Stories Press, **2019**.

[3] J.F. Woods, L. Gallego, P. Pfister, M. Maaloum, A. Vargas Jentzsch, M. Rickhaus, *Nat. Commun*. **2022**, *13*, 3681, https://doi.org/10.1038/s41467-022-31482-2.

[4] J. A. Swabish, *J. Benefit-Cost Analysis* **2020**, *11*, 151.

[5] S. Counts, K. Fisher, 'Taking It All In? Visual Attention in Microblog Consumption', The AAAI Press, **2011**.

Our literature on soft materials and shape-assisted self-assembly

J. F. Woods, L. Gallego, P. Pfister, M. Maaloum, A. Vargas Jentzsch, M. Rickhaus, *Nat. Commun*. **2022**, *13*, 3681, https://doi.org/10.1038/s41467-022-31482-2

For a recent review, see: L. Gallego, J. F. Woods, M. Rickhaus, *ChemRxiv* 10.26434/chemrxiv-2022-9vgk6.

Recommended literature on data visualization

J. A Schwabish, 'Better Data Visualizations', Columbia University Press, **2021**.

J. Koponen, J. Hildén, 'Data Visualization Handbook, Aalto University', **2019**.

A Cairo, 'The Functional Art: An introduction to information graphics and visualization', New Riders, **2012**.

A. Cairo, 'The Truthful Art: Data, Charts, and Maps for Communication', New Riders, **2016**.

A. Cairo, 'How Charts Lie: Getting Smarter about Visual Information', W. W. Norton & Company, **2019**.

For table design, see: J.A. Swabish, *J. Benefit-Cost Analysis*, **2020**, *11*, 151.

For color selections, see: L. C. Muth, 'Your Friendly Guide to Colors in Data Visualisation', blog.datawrapper.de/colorguide, retrieved 12.07.22 and L. C. Muth, 'How to pick more beautiful colors for your data visualizations', blog.datawrapper.de/beautifulcolors, retrieved 12.07.22.