From Tradition to Insight Support
What we can demand from our Chemistry Visualization tools

Mario Valle*
Swiss National Supercomputing Centre (CSCS)

September 25, 2006

Abstract

There has always been a strong relationship between chemists' understanding of chemical phenomena and the associated external representations. It is thus common wisdom that visualization tools should play a prominent role in the chemistry discovery cycle, and they do indeed. But looking at them with a non-chemist's eye reveals that they implement some interesting visualization ideas, but fall short on the perceptual support side and, more important, on the implementation of human-computer interaction (HCI) ideas. In this work we substantiate those findings with a tool survey and an inventory of chemistry representation methods. We hope those two results will be the basis for engaging a discussion with the chemistry research community to clarify, understand and define the characteristics that an *effective* chemistry visualization tool should have.

1 Introduction

Chemistry is a visual science, and data representation plays a central role in its discovery process. There has always been a strong relationship between chemists' understanding of chemical phenomena and the associated external representations. For example Kekulé attributed his success to his "more eclectic approach to theoretical schools, and his *irresistible need for graphic imagery*" [1, p. 366–377]. Sure the existing visualization tools serve and support well the research community; sure they contribute to solve the specific problems that often initiated their development. Nonetheless we started looking at them with a non-chemist, visualization-expert eye to see how the visualization principles, techniques and methods are applied in the well-defined area of chemistry structures visualization and, as a consequence, how we could suggest an evolution path.

We analyzed 20 chemistry visualization tools (the list is in section 4.1) to see how they fulfill the goal of being an important piece of the chemist's discovery cycle. This activity has been carried on as part of the support offered by the *Data Management, Analysis and Visualization* group to the *Swiss National Supercomputing Centre (CSCS)* users.

We found that chemistry visualization tools lag behind. The main complains are in the area of perceptual support given by the usual structural representations and in concrete application of human-computer interaction (HCI) principles. After substantiating this claim, we report our

*e-mail: mvalle@cscs.ch – web: http://www.cscs.ch/~mvalle/

finding about the surveyed tools (sect. 4.1) and provide an inventory of the common representation methods those tools implement (sect. 4.2).

We wish that those two results become the basis for engaging a discussion with the chemistry research community to clarify, understand and define the characteristics that an *effective* chemistry visualization tool should have. For this reason this report is not a finished product, but instead it is more a research plan that will be detailed and modified as the project progresses.

2 The visualization scientist's contribution

Visualization scientists and chemistry researchers have distinct goals and distinct priorities related to visualization tools. But the objective should always be to foster scientific insight. To approach this goal, which is the contribution the visualization scientist, and this work in particular, could offers to the research community?

An obvious contribution could be the creation of a "super-tool" that implements all the aspects overlooked in the current ones. This obvious response has indeed various drawbacks: first and foremost, it creates simply the tool $N + 1$ disregarding the loyalty of current users to their preferred tools; second, it lessens the major contribution this work could provide.

We think that this work's contribution could be the creation of a fruitful dialog between domain experts of both fields. This discussion could help generate creative ideas in both camps. We think that the lack of specific chemistry knowledge by visualization scientist and therefore the expected misunderstandings that could arise will be beneficial. First of all we could propose solutions and representations that are not constrained by *prior knowledge* and as such this solution could steer the chemistry researcher to an entirely unexpected thinking path. Second, the interaction between the two fields helps clarify assumptions and surface latent ideas. Third, the real life application of visualization ideas put them through a beneficial reality test that is much needed by any visualization technique.

We can summarize the benefits of this work for the chemistry researcher in a better understanding of the potentialities of visualization tools and techniques, less restrains in asking and discussing new and unusual functionalities with the visualization scientists and, last, in producing ideas that could concretely enhance existing tools.

A final comment for the tools developers: please do not take as a personal offense the results of the survey. Every tool has given a valuable contribution to the research community. Simply we want to learn more and advance the chemistry visualization field.

3 Tool survey findings

The most important observations about the current portfolio of chemistry visualization tools could be condensed in the following items:

Representational conformism. The analyzed tools provide almost the same set of representations. Is this choice conclusive, or are we overlooking more cognitive opportunities?

Undifferentiated set of functionalities. Often tools seem more *me-too* tools and not platforms to implement and propose new methods.

Tools are not invisible. The majority of tools surveyed have intrusive user interfaces (GUI) that almost always distracts from the visualization and analysis work.

Visualization equated to post-processing. We found more emphasis on rendering than on exploration support, on producing images for publication than on using images for understanding.

Tools snob mental processes. The mental processes of chemists are often overlooked by the visualization tools developers. Moreover tools provide no support to associate thinking to working.

Tools taken as given. Available functionalities often dictate the contribution of the tool to the discovery process. In a certain sense, discovery process adapts to the tool functionalities and not vice versa.

"New is better" syndrome. It is taken for granted that computer visualization is better than old technology like paper diagrams and tangible models. But what have we lost in the transition?

3.1 Representation conformism

The analyzed tools provide almost the same set of representations (see sect. 4.2 for more details). Citing Murray-Rust: "The key point is that these *electronic chemical stencils* did not change, and in fact reinforce, conventional (and often ambiguous) symbolism. In some cases the programs were designed to display excellently on paper, only adequately on the screen and much less so within formal data-structures. The creators of such programs often constrain the chemical author to a finite set of chemical glyphs with the consequence that semiotic innovation is thereby not encouraged" [2].

This imposed limitation carries a risk: "Working with a stagnant set of representations creates a more subtle and profound problem. By restricting the set of available visualizations, we limit the ways that scientists think about their model and thereby limit potential insights" [3].

Then we ask: is the proposed set of representations the only conceivable one, or simply does everyone proposes the same set by tradition? We think that studying and implementing new representations could change this trend and help insight, since models used always shape the research questions posed [4] because signs are not passive media for expressing preconceived ideas, but actively contribute to meaning. For example bonds represented as lines have made the "bond" a real entity on the same level as the atoms.

To be fair, more variety of representations is found in tools focused on biochemistry. But for other fields the representations used are always the traditional ones, forgetting that there are "non-representable" materials like foams that are gaining importance or metallic and ionic compounds for which the molecular visual metaphor is not correct.

In any case there is no single *perfect representation*, only representations that help solve specific problems. This idea could motivate us to move from thinking "how to (technically) represent" to a "what should be represented" approach. This shift of perspective puts also into focus the perception and cognition characteristics of a good representation that guides and focuses attention by removing unneeded details and by integrating multiple levels of information.

Even if this refocusing could not be done immediately, the effectiveness of the current set of

representations could be enhanced by providing guidelines on how to choose a representation over another considering the problem under analysis.

3.2 Undifferentiated set of functionalities

Looking at the current tools the first impression we have is that they are more *me-too* tools than something built to provide a very unique functionality not offered by existing tools. The reasoning path of tools developers seems to move from existing techniques to their implementations, instead of going from the chemistry problem and researcher goal to the implementation of "what works".

This design attitude compounds with the limited extensibility of the tools making the integration of different or unusual visualization techniques beside the ones offered by the tools difficult. We do not know if this is the only reason, but we detected a general difficulty by chemists to ask or discuss new functionalities they want integrated in the existing tools.

On a related note, a visualization tool should couple visualization and analysis; otherwise visualizations could be interesting, but not informative. In this area we found more variety, mostly in the interfacing to popular computational codes. But this does not represent yet full integration of visualization and analysis.

Instead there are two examples of functionalities that could help analysis: representation simplification methods and multiple views on the same data. Both of them are into reach of visualization developers and both of them use our perception and cognitive capabilities to enhance understanding of the data. So why very few tools provide them?

3.3 Tools are not invisible

The majority of the tools analyzed have the classical WIMP (Windows, Icons, Mouse, Pointer) user interface and we have found only one of them with a different one. Often the limitations of this traditional approach are exacerbated by intrusive GUIs that almost always distract from the visualization and analysis work. In addition the tools have frequently a rigid usage and interaction model that forces the discovery process to adapt to them and not vice versa.

We know that scientists are not HCI experts and visualization scientists have difficulty to understand fully the scientists' needs. But something should be done if we want to move to the next effectiveness level.

To understand what could be done to alleviate the problem, we can look to the area of video games. Here the immersion, called *playability*, is of supreme importance. The gamers do not want to be distracted by peripheral interaction issues; they want to reach the game goal. They want *fluid interaction* with the tool [5]. This is exactly the same goal pursued by a researcher looking to an unknown structure. Think about options selection: currently the user should move from the graphical scene, reach the menu bar, find the right menu and push the right option. Instead something similar to a 3D menu widget or a semi-transparent display positioned near the attention point distracts a lot less. To work on this theme, we will start from some studies [6, 7] that cover specifically innovative user interfaces for chemistry visualization tools.

Another aspect of the tool non-invisibility is the integration of visualization in the discovery

loop: the visualization tool is often considered an addition to an existing discovery loop, not something integral to it. For example connecting the computational code output to the visualization program still requires a consistent manual effort.

3.4 Visualization equated to post-processing

When looking at the tools feature set and usage patterns seems that the visualization tool is still perceived as an ancillary tool, not as a mainstream research one. In other words, visualization tools are classified mostly in the category of post-processing tools.

The effectiveness of visualization tools is compromised by this categorization: calling them post-processing tools means that they visualize what is already known. Furthermore, concentrating on rendering issues —aesthetics, lights and colors— tends to induce a passive attitude in the researcher. Both mind-sets reduce the discovery potential offered by exploratory visualization.

Presenting and communicating results is obviously an important part of the researchers' work, but they can communicate only something that has been already understood, so exploratory visualization, in our opinion, should play a more prominent role in the discovery cycle.

3.5 Tools snob mental processes

Looking at the current visualization tools we feel like chemists missed the opportunity of thinking about the real role visualization could play in the discovery process.

A complaint, which is shared by the whole world of scientific visualization tools, is that tools do not provide support to the thinking process. They provide no assistance to associate thinking with working. The support could range from knowledge recording during exploration, to annotation, to backtracking support and up to the creativity support offered, for example, by multiple views of the same data or by suggesting unexpected connection and thinking paths.

Another approach could be to look at the users' mental model of the tool behavior, at the expectations they have about the tool, and at the differences between those tools and the generic scientific visualization ones. This questioning could focus attention on the part of the tool that provides added value to a chemist.

Furthermore we should consider that creativity and representation often go together: "we can hardly be amazed by the fact that metaphor is a prime suspect in the search for an agent that can help us think creatively. It is also relatively little disputed that metaphors somehow fulfill the function of triggering or guiding cognitive inference" [8, pp. 183-208].

3.6 Tools taken as given

"Most researchers are a slave to the program and to the coordinates" said the molecular illustrator Irving Geis [9] meaning that, at least sometime, chemists rely more on tools than on their mental mechanisms. Worse, they do what the program can do without pretending more support for their specific research. But remember that tools shape the discovery process, as much as the models and representation do.

We found that there is a mismatch between "work to be done" and "tool". What could happen

instead if we could access a "perfect" tool? What will we ask to it? This exercise could stimulate us to surface the unverified assumptions that we made and to see which opportunities we are missing.

3.7 "New is better" syndrome

It is taken for granted that computer based 3D visualization is better than physical ball-and-stick construction kits or paper diagrams. But we can ask ourselves: are computers making chemistry more visual? Are we aware of the constraints that have been introduced moving to computer-supported 3D visualizations?

For example a tangible physical model provides understanding by building and touching and gives automatic steric constrains enforcement, capabilities normally missing in current tools. And more, mechanical models let the chemists "think with their hands". With them the researcher is free from the single point of view constrain, can manipulate the model, can touch it, measure test, dissect or assemble. In other words the material model acts as a material analog. Again, can we say the same of our current tools?

Another example is given by the 2D reaction diagrams: they contain structures, annotations, highlight of important parts, simplified representations to focus attention. From one side 3D tools strive to offer more "faithful" representations, but from the other one they forget to take advantage of the work done in the past to overcome limitations of the available media, like paper.

So which opportunities do we miss? Could we look at those previous technologies and extract ideas applicable to our computerized tools?

4 Current project deliverables

This project is *work in progress*; there are some partial outputs that we details in the following sections, but the most useful parts and the most useful results will came only with the users involvement and input.

The first project output is a survey of existing visualization tools (sect. 4.1). The initial goal of this survey was to have ready a pondered suggestion for our users that want to start using visualization in their work.

The second output (sect. 4.2) is a report summarizing the common representation methods currently offered by the surveyed tools.

Both works were started also to collect nice and interesting ideas to be included in the tools we are developing at CSCS.

4.1 Tools surveyed

The list of surveyed tools with a brief list of the main points found is contained in the report "Existing chemistry visualization tools survey". The testing that started this whole project is available on the Web (see sect. 7). To be considered for this survey a tool must be:

1. Non-commercial

2. Multi-platform
3. Not too old
4. With active development and user communities

In those tools we started looking at good ideas implemented and things to avoid. Then we added a look to original representations and HCI principles application.

As per today (October 11, 2006) the report is not yet available, but the starting material is available on-line (see sect. 7).

4.2 Representation methods survey

The companion report "Representations in Chemistry – A survey" contains a list of the logical data type that are normally considered as input for chemistry visualization tools and a summary of the common representation methods currently offered. A final section lists some unusual method sometime proposed that could be of interest for normal research work.

As per today (October 11, 2006) the report is not yet available, but the starting material is available on-line (see sect. 7).

5 Future project proposal

The current chemistry visualization tools portfolio is serving the researcher community very well. On the contrary you can think we forgot here those contributions and depicted a very depressing situation. Do not take it at face value; instead think that, to help the chemistry community, we need to move to another level of effectiveness in our tools.

Here are some ideas that could become future projects that we will strive to develop in the near future:

1. New rendering methods that enhance the perception of the structures.
2. Simplification methods to show only the essential parts of a structure.
3. Integration inside the discovery loop (e.g. integration with computation and computational steering).
4. Imagine a "super tool". No limitations, no problems in adding functionalities. What will you ask to such a tool?
5. Innovative interaction methods to focus user attention.
6. Collecting researchers' feedback by using "viz laboratories": small, focused applications that highlight a specific chemistry visualization technique to collect practitioners' feedback in real life situations.
7. Multiple representations and views usage.

6 Conclusion

This work has been carried on as part of the support offered by the *Data Management, Analysis and Visualization* group to the *Swiss National Supercomputing Centre (CSCS)* users. As a group we want first and foremost to foster scientific insight by the researchers that use the CSCS facilities in their work. To reach this goal we provide, for example, advice in selecting tools, development to enhance existing tools and education about fruitful visualization usage.

The short term result of this work is the preparation of material to support those activities by providing testing of existing tools and guidelines about chemistry representations to use. But the most interesting outcome would be the start of a dialog with the chemist research community to clarify, understand and define the characteristics that an *effective* chemistry visualization tool should have.

7 Working materials

To support discussion the following on-line resources are available:

- The ChemViz Blog to comment on those ideas:
  `http://chemviz.blogspot.com/`
- The discussion Wiki accessible (after registration) from:
  `https://twiki.cscs.ch/bin/view/TWiki/ChemViz`
- The initial *representation in chemistry* survey on:
  `http://www.cscs.ch/~mvalle/ChemViz/representations/`
- The test of chemistry visualization tools:
  `http://www.cscs.ch/~mvalle/ChemViz/tools.html`

References

[1] Alan J. Rocke. Hypothesis and Experiment in the Early Development of Kekulé's Benzene Theory. *Annals of Science*, 42:355–381, 1985.

[2] Peter Murray-Rust, Henry S. Rzepa, Simon M. Tyrrella, and Yong Zhang. Representation and use of Chemistry in the Global Electronic Age. *Org. Biomol. Chem.*, 2:3192–3203, 2004.

[3] Thomas C. Palmer. A Language for Molecular Visualization. *IEEE Computer Graphics and Applications*, 12(3):23–32, May/June 1992.

[4] Roald Hoffmann and Pierre Laszlo. Representation in Chemistry. *Angewandte Chemie (int.l edition)*, 30(1):1–16, January 1991.

[5] Jeff Dyck, David Pinelle, Barry Brown, and Carl Gutwin. Learning from Games: HCI Design Innovations in Entertainment Software. In *Proc. Graphics Interface 2003*, February 2003.

[6] Joachim Almgren et al. Tangible User Interface for Chemistry Education: Visualization, Portability, and Database. In *SIGRAD 2005 The Annual SIGRAD Conference Special Theme - Mobile Graphics*, pages 19–24, Lund, Sweden, November 2005.

[7] Robert van Liere, Arjan Kok, Jean-Bernard Martens, and Martijn van Tienen. Interacting with Molecular Structures: User Performance versus System Complexity. In *Eurographics Symposium on Virtual Environments*, pages 147–156, Aalborg University, Denmark, October 2005.

[8] E. Steinhart. *The Logic of Metaphor: Analogous Parts of Possible Worlds*, volume 299. Dordrecht: Kluwer Academic, 2001.

[9] Bruce Paul Gaber and David S. Goodsell. The art of molecular graphics Irving Geis: Dean of molecular illustration. *Journal of Molecular Graphics and Modeling*, 15(1):57–59, February 1997.