Illustrating Publication@Source: The Behaviour of Benzo-15-crown-5 at the Solution/Air Interface by Experiment and Molecular Dynamics Simulations.

Esther R Rousay, Hongchen Fu, Jamie M Robinson, Stuart E. Murdock, Jonathan W Essex, Jeremy G Frey
School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ

Abstract

Using the concepts of publication@source technology, the data obtained from a range of spectroscopy and surface tension experiments, and that obtained from a number of related large computational simulations, have been stored in such a way as to enable a viewer to trace back through the analysis methods to the raw data. The storage of the simulation data is viable using the BioSimGrid at Southampton. In order to trace back to the raw data via chains (or trees), which include assumptions made and models used, a resource description framework (RDF) representation and an ontology including descriptive terms are employed.

1. Combining Experiment and Simulation

Second Harmonic Generation (SHG) is a non-linear laser technique that allows us to investigate the behaviour of adsorbed molecules at the air/water interface in the presence of the same molecules in the bulk water phase. The experiment is sensitive to certain averages of over the molecular orientation distribution of the adsorbate molecules; in this case the crown ether molecules, and typically assumptions have to be made about the nature of this distribution (e.g. highly peaked) to interpret the experiments. We have used SHG to probe a number of different interfaces however combining the experimental studies with molecular simulations allows us to estimate the nature of the distribution with much greater certainty and thus improve the interpretation of the experiment.

Figure 1: 3D structure of the benzo-crown ether. See http://ecrystals.chem.soton.ac.uk/145/ for detailed solid state crystal structure information.

This combined experimental and simulation workflow imposes problems when it comes to publication, since the two components are interrelated in terms of their interpretation and importance. Moreover, the large amounts of raw data, simulation and experiment, require extensive manipulation to yield a small number of significant physical observations. This project therefore represents an ideal case study for the publication@source paradigm, in which all data and associated mathematical operations are made available to a reader.

Figure 2: Area per molecule at each concentration derived from the surface tension isotherm (filled diamonds). The open circles indicate the points along this curve at which simulations have been performed.

2. Experiments

The partially soluble surfactant benzo-15-crown-5 (figure 1) has been studied at the aqueous solution/air interface has been studied using surface tension measurements, second harmonic generation (SHG) spectroscopy. The
Surface tension measurements showed that the minimum area occupied per benzocrown molecule is about 88 Å² (see figure 2). In the SHG experiment, the polarisation dependence and concentration dependence SHG studies showed only a slight change in effective orientation occurred over the concentration range studied. The effective tilt angle determined assuming a peaked distribution, is close to 36 deg. This is close to the SHG magic angle consistent with a very broad weakly anisotropic distribution. The molecular dynamics simulations studies were undertaken in an attempt to resolve these issues.

Figure 3:
Two different conformations of the crown ether differing in the relation orientation of the benzene and crown rings.

3. Molecular Dynamics at the air/water interface

To study this system in greater depth, a set of atom-based molecular dynamics (MD) simulations were run with different numbers of benzocrown molecules on the water surface, consistent with the surface concentrations present in the experimental studies. These simulations were run as parallel calculations on the new AMD Opteron-based Iridis 2 cluster at Southampton. From these simulations, the way in which the crown ether behaves (i.e. any packing or orientation preference) may be determined. These results may in turn be linked to the experimental studies, whereby parameters required for the interpretation of experiment are derived directly from the simulation. Two distinct benzocrown conformations seen in a low surface concentration molecular dynamics simulation (see figure 3).

The simulation trajectories obtained using NAMD⁴, are large, and require complicated tools for their interpretation. The BioSimGrid storage and analysis software is therefore being used for this work, in which raw data is stored in a distributed fashion using flat-files, while the associated metadata is in a replicated relational database. The analysis scripts are written in python, and may be stored to allow a full understanding of how the data have been treated.

![Simulation trajectories obtained using NAMD](image)

**Figure 4**: The average tilt angle derived from the simulations as a function of the density of crowns adsorbed on the surface

4. The Interface

Average tilt angle between the chromphore (the benzene ring) molecular axis and surface normal, to which the SHG experiments are sensitive, was calculated directly from MD simulations. The results, shown in figure 4, are different from the value derived directly from the SHG experiment. As indicated above, several assumptions are made in the analysis of the SHG polarisation data, both about the nature of the surface orientational distribution and the dominant microscopic molecular properties, that contribute to the macroscopic response. Clearly these assumptions need to be reviewed.
Our attention focussed first on the orientational distribution, which is shown in figure 5 for different densities of the crown ether molecules at the interface. At low concentrations a broad distribution is observed with the most probably angle around 90 deg, that is lying in the surface.

Figure 5: Tilt angle distribution for different densities of the crown ether at the interface

As the concentration of the crowns increases the distribution maximum shifts slightly to larger angle, but a further increase in concentration gives rise to an increasing bimodal distribution with a peak ca. 45 deg and at ca 150 deg, as if at higher concentration some crown molecules are being forced up or down into the interface.

5. Publication@Source

To achieve the aims of the publication@source project, two problems need to be solved. First, how the experimental and analysis processes are described, and second, how the processes applied to the data may be displayed. Semantic technologies, namely Resource Description Framework (RDF) and an ontology to record the processes are used. The ontology defines the vocabularies of the experimental analysis processes to facilitate the sharing of data amongst collaborators in different organisations.

The RDF representation captures the fact that each step in the analysis process is performed according to a model, and allows for additional assumptions (parameters) to be included and comments to be made about each step. These comments can subsequently be rendered to provide the explanation of the model used at each step.

The simplified data analysis tracking represented in this work considers largely linear data flows. The TriScape browser can display branching links but the current menu system does not provide an ideal view in these cases; this restriction will be lifted in the next version of the TriScape Browser. The raw RDF is viewable in a separate window.

The relationships between the steps in the analysis are described using RDF expressions. At a higher level the predicates and vocabularies used to describe the analysis process are defined in the ontology. The process, which is referred to as the OntoProcess, is implemented using OWL (W3C Web ontology Language), with a hierarchy structure.

6. The RDF Technology

By presenting the data in RDF form with the ability to trace back to the raw data, the publication is complete, and a viewer is able to see the entire process by which the data have been treated. Such dynamic linking also means that should any new data be obtained on the system, it may be added to the database, and linked to the original publication to allow any interested parties to be kept up-to-date with the information. It also means that a viewer can determine the quality of the data obtained.

The need to represent the wide range of chemical objects can be achieved by using XML-based technologies. Figures can be plotted in the scaleable vector graphics (SVG) by the interactive-SVG utility, thereby simplifying the provision of hyperlinks to each data point on a graph. The raw data is dynamically pulled from a database using a URI stored in the data description, which directs a web service to resolve a request as an SQL query to the relational database. The points delivered may then be operated on by some mathematical function, for example, averaged.
7. Details of the TriScape RDF Browser & Model

The Triscape browser is a generic RDF browser and it can be used to browse any valid RDF resources. It has the following features:

Server Side application: The Triscape RDF browser is implemented as a web application (Servlet) running on the Tomcat server and making use of the Jena API from HP laboratories. On the client side a call, such as a hyperlink or XLink (in SVG) sends an HTTP request to the server, which then calls the relevant RDF model. The RDF browser renders the relevant content of the RDF model as HTML.

Display Objects: For each statement in the resource model, the browser locates its predicate (property) and the relevant object, as well as the corresponding vocabulary entry defined in the ontology. If the object is not a resource then the description from the vocabulary is displayed as the definition, together with the value of the object. If however, the object is another resource, then the vocabulary label is displayed together with a link to its definition, and the URI of that resource. The latter includes the HTTP request to that resource so that the client can follow the chain further. The resource may be a call to a web service, such as a statistical calculation, which can be invoked via the request from the client.

Display parent: The Triscape browser also supplies a link from an object to its parent resource. The parent resource is located by searching the RDF to find a resource that points to the current item. In the scenario of a chemical process, the parent resource is usually the previous step. As above a link is provided to enable the chain to be followed in this direction.

Display Menu: For its use as a process tracking system, a plug-in component is added which supplies a menu of all the steps in the process and is displayed on the left-hand side of the page. The menu is currently limited to displaying linear analysis chains.

8. Conclusions

The active versions of the papers and presentations implied by this paper allow the complex data and data analysis pathways used to interpret both the experimental and simulation data to be explored and the data from any and every step to be made available for further exploration and re-use. The information is presented to the user in HTML or SVG and thus generally renderable without too much difficulty. The linkages in the analysis steps are represented using RDF, which allows the computer to explore these relationships as well as present them for human comprehension. This presents a significant advance on the provision of supplementary data to a publication, but does now place an additional burden on the ‘publisher’ to ensure that the material can indeed be made available in this form. Adoption of more semantic standards will be required before this can be assumed to be as easy as browsing the current web – not itself always that easy in practice!