SUMO
A SONIFICATION UTILITY FOR MOLECULES

Florian Grond

ZKM - Center for Art and Media
Institute for Music and Acoustics
Lorenzstr. 19, D-76135
Karlsruhe, Germany
grond@zkm.de

Fabio Dall'Antonia

Karl-Franzens-University Graz
ZMB - Center for Molecular Biosciences
Dept. Structural Biology Humboldtstr. 50/III
8010 Graz, Austria
fabio.dallantonia@uni-graz.at

ABSTRACT

In this paper we present SUMO, an open source software environment, which is designed to facilitate the open development of molecular sonifications for everyday research in chemistry and structural biology. Sonifications of chemical data are developed since more than 25 years but surprisingly auditory display is not yet a scientifically established mode to interact and explore molecular data. Before presenting SUMO we introduce the implications of presenting molecular data to the sonification community. For chemists and structural biologists, we briefly review different sonification approaches made so far and discus their potential. Within this broader scope we situate SUMO, the *lab proof* sonification framework. We describe the software environment in detail and present two implementations of methods for sonifying conformations of amino acids and B factors.

[Keywords: sonification framework, macromolecules, PyMOL plug-in, Open Sound Control, SuperCollider]

1. INTRODUCTION

In this paper we will present our sonification framework SUMO for the molecular viewing program PyMOL. Since this interdisciplinary enterprise between chemistry/structural biology and sonification involves the background of different fields we want to line out in the introduction some general thoughts on the structure and representation of molecules. These general considerations are part of our conviction, why auditory displays can make a valuable contribution to the well established field of molecular visualization. This section wants to introduce the auditory display community to the implications of representing molecules from a chemical perspective.

1.1.The amount of available molecular structures

During the last decades, researchers used various techniques to determine a vast amount of macromolecular structures in the field of structural biology. The general aim of macromolecular structure determination is the explanation of its function in its natural environment. The result of a characterized macromolecule is roughly speaking a data set that contains the positions in xyz of all atoms belonging to the investigated molecule plus further relevant information. These data are saved in a standardized file format known as PDB. The number of resolved structures is constantly growing. A repository of all the so far resolved structures can be found in the protein data bank [1]. As of December 2007 there are 48091 Structures available. For the field of sonification this is interesting in so far as this huge amount of structurally complex and interesting data is publicly accessible in a standardized format. This is a key requisite to develop established auditory displays.

1.2.The complexity of macromolecules focusing on proteins

To briefly describe the complexity of an example class of macromolecules we need to know that a medium sized protein may consist of approximately 3000 non-hydrogen atoms. These atoms are organized in a well ordered 3-dimensional hierarchical manner: The primary structure, a chain of amino-acids which are covalently linked. The secondary structure describes the next neighbor arrangement of amino-acids such as α-helices and ß-sheets. The tertiary structure describes the arrangement of secondary structure elements known as the fold of a protein. The quaternary structure describes the arrangement of domains of one protein chain or the assembly of different macromolecules into one functional complex.

1.3.Representing the unseen

Since the complex structure of molecules cannot be seen with the naked eye, chemistry has from early on developed systematic methods to represent its research object. The history of visual representations of molecules goes from symbolic representation of chemical agents in alchemy and leads to modern chemistry and its contemporary images of molecules. Throughout this development we can observe in many steps how the insight into the structures of molecules has coevolved with their representations[1].This coevolution oscillated between concrete and abstract approaches always negotiating between the need to come close to what was imagined through theoretical considerations and the restrictions of the medium of representation at hand[2].

Modern computer graphics have given us the possibility to view molecules as interactive 3 dimensional objects on a computer screen. This remarkable possibility lets us all too often forget that we are dealing here with well established yet constructed representations and not the molecule itself. The molecule itself is not an object of defined shape and has no colors, yet the depiction with well formulated boundaries

[1] Alchemistic Symbols were 2 dimensional and underwent an interesting metamorphosis to become the 3D representation of molecules as we know them today.

[2] One well known example is the Fischer projection, which was designed by Emil Fischer in order to depict stereochemistry two dimensionally in an unambiguous way on paper.

and additional coloration adds tremendously to our understanding.

In 1981 Richardson [2] [3] invented well known cartoons as one of the most useful simplification for the representation of features like α-helices and ß-sheets (see Figure 1). This creative intervention in drawing molecules shows that with the increasing size of information on the display and the therefore growing visual complexity, a well designed abstraction is sometimes superior to a detailed realistic picture, which we do not posses anyway. It further shows that creative and innovative forms of representations constantly need to be invented. The balance between theoretical constructs and its creative representation has to be constantly renegotiated. This process, which might remind one of artistic practice, is particularly true in times of increasing complexity of the represented research object, as biochemist Schatz describes in [4]

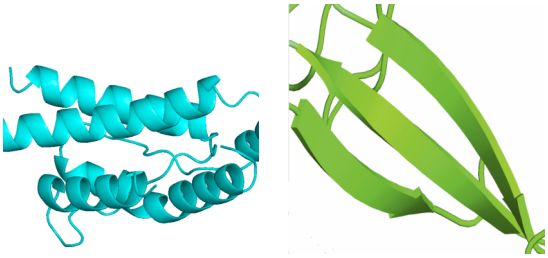

Figure 1: The structures show from left to right several α-helices and a ß-sheet as invented by Richardson in 1981

1.4.Why additional modes of representation?

There are many different molecular viewing programs available; which allow to manipulate molecules in various ways like highlighting certain features and choosing from different display modes. Molecules can be grabbed with the mouse pointer and rotated into any position. The main purpose of these programs is to assist in the analysis and functional structure interpretation which can only partly be automated and still needs manual intervention and visual inspection.

Despite all these advanced display methods for macromolecules the interpretation of their functions and the detection of functional significant deviations from one molecule to another is still a very intricate procedure with obstacles like missing the atomic detail in an overview setting, missing the context when zooming in, or simple obstruction through presenting the 3D data with the means of central perspective. Although there are multiple possibilities for presets, the user's interaction with the representation of the molecule in the viewing program remains a singular experience requiring expertise, which is by its nature difficult to turn into an automated process.

In this paper we describe, how sonification or more general auditory display can be a valuable complementary contribution to display macromolecules in combination with the already established molecular viewing programs. The tight integration of our acoustic and visual senses results in a multimodal experience for the user and can therefore substantially contribute to the process of gaining and eventually communicating expertise.

In the next paragraph we briefly recapitulate sonifications of data from chemistry, then we describe the software framework, and later we demonstrate with two implemented use cases how sonification can support the interpretation of macromolecules.

2. DEFINITION OF TECHNICAL TERMS, STATE OF THE ART IN SONIFICATION

In the interdisciplinary cross-section of auditory display and chemistry, it must be mentioned that the term sonification is already occasionally used instead of ultrasonification in the field of biochemistry and microbiology. Ultrasonification is a technique where ultrasonic waves are applied to chemical mixtures or tissue cultures with varying intensity in order to facilitate certain reactions or to break membranes open to access organelles or molecules. We mention this term in passing to demonstrate that applied sound of sufficient energy can have an impact on a cellular/molecular level. The other direction, which is the unmediated emission of continuous sounds from molecules themselves[3], has not been observed [5].

2.1.Existing sonifications in chemistry

In the following part of this section, we line out already existing work that uses in one way or another sound in combination with chemistry; this is mainly meant as a brief introduction for chemists with an interest in auditory display. In order to situate our project we give an overview about scientific publications and related websites. The selection is certainly not complete but highlights different motivations and technical approaches.

2.1.1.Artistic approaches

One motivation that is often encountered is artistic. For an overview over related projects, I refer to [6], [7], [8] and well annotated collections of links in the WWW [9], [10]. Here, molecular or genetic data serve as an abstract score in order to generate musical patterns. The core assumption is that the complexity of protein chains or DNA strands results in equally interesting sound/music. In this conceptual approach, many sonification choices are driven by aesthetic considerations and not by scientific research questions. Data sonifications for scientific purposes should also sound pleasant, but in this case, sound parameters must not only be tuned to serve specific aesthetic needs as it is the case in sonification for musical purposes. This leads to choices, which dramatically increase perceptual contrast where they are from a scientific point of view less significant. At the same time attention is dragged away from potentially interesting details in the data.

2.1.2.Didactic projects

As the review of Garcia-Ruiz et al. [11] shows**,** an abundant motivation is didactics. Here sonification has the big advantage to address our often neglected auditory senses and to integrate them in the process of acquiring knowledge. SUMO can also be used for didactic purposes, but we aim to develop a framework and principles that are easily applicable to all molecular structures and not only to some demonstrative examples. Ideally in the future, sonification as a purely didactic tool becomes obsolete because learning the structures of macromolecules also means getting accustomed to their well designed and established sonic representations.

[3] This definition does not include explosions or bubbling noises from evaporating gas, which could be taken as an early form of sonification which allowed to monitor fermentation for instance. There are examples of audible amplified sounds mediated through the atomic force microscopes from yeast cell cultures[5]

2.1.3.Improving accessibility

Projects such as Brown et all [12] and Lunney [13] aim at making molecular structures accessible for the visually impaired. These projects are re-representing already existing visual information for our auditory senses. With our project, we want to introduce sonification to mainstream scientific practice in chemistry, therefore our approach is slightly different; for typical research situations, we assume the scientists to have an abstract spatial/visual concept of the molecule, that's why we want to use sonification mainly, but not exclusively, for non obvious and complementary information.

2.1.4.Conversions to MIDI sounds

Many popular sonification applications convert data into MIDI [14] notes, in order to communicate with sound synthesis programs. There is for instance *Proteinmusic* by King and Angus [15], or the well known and versatile *ArtWonk* [16] by Dunn. This program offers the *BioEditor* as an extension, a program that helps to convert DNA and protein data sets. *ArtWonk* was used in the collaboration with biologist Clark for the project *Life Music*: *The Sonification of Proteins* [17]. Another more recent project which is worth mentioning is *gene2music.* The project website [18] offers a systematic and linked overview over related research. PROMUSE [19], a further MIDI based project, has a focus on protein structural alignment, which is clearly a research related motivation. There is also the commercialized project *Molecular Music®* [20] the sound of which suggests that the MIDI format was used. Sound renderings of more than one minute for a single macromolecule show that most of these programs focus on the aspect of algorithmic complexity in music. Beyond that fact, the old MIDI data format reduces the choice for sonification methods tremendously; it gives these approaches a more musical impetus.

2.2.Reception in the scientific communities

Looking at most of the publications in this field, we can find with Yeung (1980) [21] and Ohno et all (1986) [22] only two who published their original works in journals related to natural science. Most of the other works were published in either chemical education related journals [23] or in the field of human computer interfaces.

Two different review articles shed an interesting light on the reception of the research attempts so far. In *The Biochemist* in 2002 Sansom [24] reviews several projects and lists URL links that combine sound and chemistry. Because of the chosen projects, which had a rather artistic focus, Sansom describes sonifications as something between catchy science news and a curious small side approach of mainstream research; the potential of sonification for the understanding of underlying data structures is mentioned but not systematically evaluated.

In *Interacting with Computers* we find the highly recommendable review about sonification in chemistry by Garcia-Ruiz et al. [11]**.** In their paper the authors trace attempts to combine sonification with chemistry back to the early 80ies; they further report many different research projects in the 90ies and some after 2000. To pick just a few interesting points from this review, Hofstatter (1980) [25] seems to be the first to associate mRNA with sound. Most of the reviewed research projects had a didactic focus, except for Yeung in 1980 [21**]** also professional chemists were asked to evaluate the sonifications. Byrne [26] finds in her PhD thesis that for the understanding of atomic structures interaction plays a more dominant role than immersion. The implementations of the reviewed projects ranged from desktop applications to immersive virtual environments.

2.3.Recent efforts and ongoing projects

CORSAIR is an ongoing project for multimodal virtual environments. The focus is the development of innovative tools for data exploration, such as the CAVE-like set-up at LIMSI [27]. Additional modalities, like sound and haptics are believed to make data exploration more intuitive, faster and possibly reveal new phenomena. The chemistry related application is protein docking.

Recent efforts to apply sonification for discrimination tasks in chemistry were undertaken by DeCampo and Huber during the workshop *Science by Ear* [28] at the IEM in Graz, where one workgroup focused on developing acoustic representations for data from chromatographically separated polysaccharides.

2.4.SUMO in context

After this overview we want to situate our project with respect to all the previous work so far. With regard to music and composition we emphasize that our focus is first to develop a scientific display. Nonetheless we pay attention to the acoustic design and we welcome everybody interested in SUMO to use and abuse it creatively. For our project relevant definitions of sonification and auditory display, can be found in Kramer [29] and Hermann [30]. We are aware that translations from data to sound might contain subjective elements. In order to follow a systematic method we refer to the sonification design strategies by DeCampo [31]. For the specific question, on how to tune sonification parameters for a well balanced perceptual contrast to discriminate scientific data features we plan to apply the method suggested by Hermann in [32].

3. MOTIVATING THE SOFTWARE DESIGN AND ITS SPECIFICATION

3.1.Obstacles for auditory displays in chemistry

Given all the interesting and promising records about early and contemporary sonification attempts related to chemistry, one might wonder why auditory display is not yet a common interaction mode in scientific programs like molecular viewers in everyday research praxis.

3.1.1.Lack of standards

Garcia-Ruiz et al. [23] conclude that beyond the well known cultural bias against sonification there is also a lack of standards and difficulties in implementing some auditory display techniques. This fact can be addressed with open standards and contemporary open source software environments. This is the main reason, why we will make the code of our plug-in available as open source. SUMO as an open interface should nurture discussion and evaluation of results in order to work towards an integrated, well established and accepted auditory display for molecules.

3.1.2.The need of a pragmatic approach

From a chemical perspective, the following considerations seem important to us: big virtual environments exhibit certainly tremendous possibilities, but they are unlikely to become an abundant tool in the research process because of their sheer size and costs. As we see in Byrne, interactivity, and not immersion, plays the major role. Therefore we want to stick to the principle *small is beautiful*. For the chemist in the lab, there are no specific hardware requirements to use SUMO, except for a pair of headphones. Yet SUMO is still easily scalable as far as the audio part is concerned.

3.1.3.Compatibility to established visualization programs

Although there are desktop applications such as ArtWonk and ProteinMusic, they lack graphical representations of their sonified data, which are meaningful from a chemical standpoint. While scientists are listening to sonifications and learn to appreciate their potential, it is crucial that they continue to work effectively by using the familiar visual representation. To this end we decided to start out with an already existing, mature and modular molecular viewing environment.

3.1.4.Real-time

As far as the amount of data allows we try to setup a system that renders sonifications in real time. This is important because the acceptance in the research community depends highly on how smooth the new display integrates into the work flow.

3.1.5.Flexibility, modularity

All established sonification techniques like earcons, auditory icons, parameter mapping [29] and model based sonification [33] should be easy to implement. This is certainly the main difference to the already existing MIDI based PROMUSE program. The visual and the audio renderings take place in two different programs, which are communicating by network. This allows to distribute the computational load over more than one computer. All the components of the software environment are further cross platform (Windows, Linux and Mac) compatible.

3.2.PyMOL

Due to the aforementioned pragmatic considerations, we chose as starting point the molecular viewing program PyMOL [34]**.** PyMOL is written in the programming language Python and uses OpenGL for the graphical rendering of the molecular structures. It is widely used within the scientific community because of its manifold features and the advanced plug-in architecture. The biggest advantage of using this well established molecular viewing program is that all data handling routines like parsing the PDB files, loading and displaying multiple files and selecting and deselecting substructures, are already well implemented and tested.

For future work PyMOL can further load and handle other relevant data files, which hold complementary information to PDB files. PyMOL exhibits a further advantage: there are many numerical libraries available in python like (numpy [35] and scipy [36]) which allow an efficient preprocessing of the data before they are sent to the sonification program.

3.3.SuperCollider

As suggested in [37], we rely on the mature high-level real time audio synthesis programming language SuperCollider [38] (SC) as the sound rendering software. SC was originally developed for Mac and ported successfully to Windows and Linux within the last years. SC allows for the implementation of all existing sonification techniques as described above. The biggest advantages of SC over many other sound rendering programs are the real-time capacities together with its scalability.

3.4.OSC

The communication between both programs is implemented with the Open Sound Control protocol OSC [39] [40] [41]. This protocol allows the efficient transfer of standard data types over the network. In our software framework PyMOL is the OSC client, which transfers over UDP all the data to SC. OSC is an integral part in SC, and an implementation for Python is available [42]. Many other programs like PureData or Max/MSP have OSC packages too. This modular approach allows therefore that sonifications are also implemented with alternative sound applications.

3.5.SUMO, the PyMOL plug-in

The software interface was implemented as a plug-in for PyMOL. This plug-in can be opened from the PyMOL menu giving an extra window (see Figure 2). In this window a button allows to boot SC. Through the modification of the SC startup file, a SC script is loaded and interpreted automatically. This

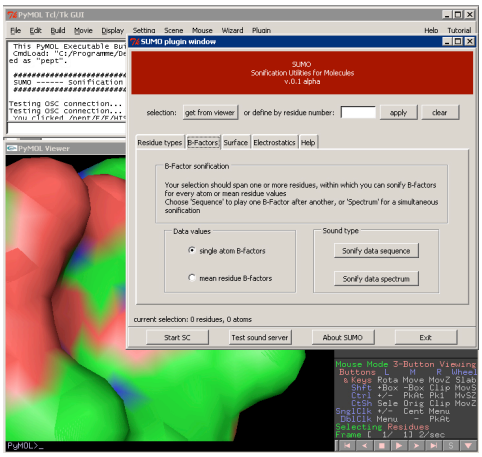

Figure 2: Surface rendering of a Protein in PyMOL together with the sonification plug-in in the foreground.

creates the OSCresponders, turning SClang into an OSC server. The OSC connection can be tested from the plug-in right away, returning a test sound.

All the interface elements are programmed in Python utilizing the Tkinter library. We separated interface from data handling routines, to develop a well structured and reusable code base.

Molecular structures can be selected from the viewer in various ways. Selections can be made for instance from the amino acid sequence, or from the viewer window with dragging the mouse. Then the selection is imported into the plugging. Buttons for different sonification methods send the selected data to SC via OSC. Every sonification method matches a specific OSCresponder.

The first implemented data structure, Table 1, which we send via OSC to SC, is derived from the information contained by the PDB file. It is a concatenated one dimensional list containing 8 items for each atom, (see Table 1). In SC we provide routines so that this list can be easily converted into a list of sub lists, each sub list representing one atom. Further hierarchies of sub lists can be created based on entry number 3, the amino acid type. Reorganizing the data in SC is necessary if we want to process all of them within one OSC responder. But depending on the sonification method data can be send in OSC bundles from the SUMO plug-in as well.

List item	Data type	Description
1	`integer`	Atom, a consecutive number through all atoms of the PDB file.
2	`string`	Atom type describing its position in the Amino acid, like C for carbon, N for nitrogen, CA for alpha carbon, etc.
3	`integer`	Number of the amino acid residue
4	`string`	Amino acid type, the atom belongs to. This string represents the amino acids standard abbreviations (Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val)
5	`float`	X position of the atom
6	`float`	Y position of the atom
7	`float`	Z position of the atom
8	`float`	B factor. This number indicates the certainty with which the position of this atom could be determined

Table 1: List items for an atom as sent from PyMOL via OSC to SC.

4. TWO SPECIFIC SONIFICATIONS OF PROTEIN STRUCTURES

We implemented two sample sonifications with our setup. We consider these sonifications not as the only possible solutions, how these molecular structures should be sonified. They are rather meant as demonstration, how this sonification environment can be used. With these examples we also want to show that, beyond the mainstream sonifications of sequences (peptide chains or DNA), there are other interesting data features in molecules, waiting for an auditory display, to be discovered. You can find sound samples to this section in: http://sonification.kommerz.at/chemie/sonifications.html

4.1.Amino acids

The first example is the sonification of Amino acids, the building blocks of proteins. As described in the introduction, they form the peptide chain, which turns into the characteristic fold of a protein. There are 20 standard proteinogenic amino acids. Their common feature is their ability to form peptide bonds; their residues exhibit different chemical properties, in terms of acidity and polarity. Within one type of amino acids, particularly those with long side chains like arginine, the spatial arrangement of each side chain (chemists speak of *conformation*, structural biologists use the term *rotamer*) varies from instance to instance due to the rotational degrees of freedom of side chain single-bonds. These variations are difficult to identify, by just looking at the structure, see Figure 3.

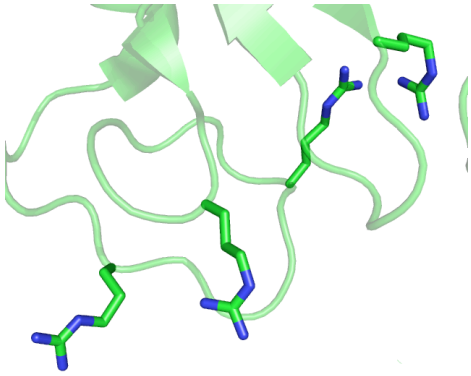

Figure 3: Four arginine residues with different side chain rotamers (example from a lysozyme molecule, PDB 1YQV).

Our task was to design an acoustic signature, which would allow to distinguish the conformational differences. Therefore we designed prototypes of parameterized earcons.

For the sound design we tried to negotiate between two juxtapositions: on the one side, the limited number of 20 amino acids makes it in principle possible to learn to distinguish individual sounds for each amino acid. Once all these acoustic representations are learned, conformational differences could be discovered. On the other side absolute acoustic representations of data structures will always remain difficult to identify for the musically untrained ear. Similarly, DeCampo describes in [31] two opposite sonification applications, which is either displaying well known information or rendering unknown data in order to discover hidden relations. In the case of rotamers of amino acids we encounter the gray zone between those two cases. We therefore tried to find an approach, where different amino acids have different sounds. At the same time we thought it could be useful to express the conformational difference directly in the earcons by comparing it with a corresponding reference amino acid. This reference takes on the functions of the origin in a graph. As an example, for all arginines, this reference can be for instance an arginine in a conformation representing a minimum of potential energy for arginine.

In order to convert the 3D data points of the atom positions into sound we computed first the pair wise distances of all atoms. The first earcons consisted of two short sounds. We mapped the pair wise distances to the frequency list of the *Klank* filter bank in SC, a bank of fixed frequency resonators which can be used to simulate the resonant modes of an object. We excited the filter with single sample impulses. The first attack, which was played back on channel one was filtered through the pair wise distances from the reference amino acid. The second attack, which came shortly afterwards, was played back on channel two and was filtered with the current conformation of the chosen amino acid. The sound examples feature proline and glutamic acid. The resulting sound was pleasant and different amino acid types gave an acoustically different impression. However conformational differences within one amino acid type were difficult to distinguish.

In further approaches we used the Formlet, a resonant filter whose impulse response is like that of a sine wave with a Decay2 envelope. Each pair wise distance was represented with one Formlet. All the distances were played with a certain time offset and they were distributed from left to right, depending on how much they deviated from the reference amino acid. We

tuned the parameters manually and finally arrived at a point where conformational identity with the reference amino acid was a short blob, and any difference in conformation would make this blob spread out in time and differentiate in space. We kept parameters for two different time scales. For the first time scale you can hear alanin, valin and glutamine. For the second time scale we recorded arginine and tyrosine. As it can be clearly heard, with this solution conformational deviations and similarities could be easily perceived.

4.2.B-Factors:

The sonification of B-factors is the second example prototype. B-factors or *Debye-Waller factors* are used in macromolecular crystallography to describe the displacement of atoms in experimentally determined structures. B-factors depend on the amplitude of the thermal (vibrational) displacement (u) according to:

$$B = 8\pi^2 u^2$$

and are hence often called "temperature factors". More generally however, the atomic displacement in crystals is not only due to thermal motion, but also due to a statistical dislocation resulting from the slightly deviating atom positions in the ensemble of 10^{15}-10^{16} crystal unit cells. Protein B-factors typically range from 10 to 80 Å^2 and the lower this value; the better defined are the atom positions. Errors in the refinement of crystallographic models lead to high B-factors as well. Thus, they are important indicators for the structural accuracy of a macromolecule model. Existing display methods for these factors make it usually difficult to extract information as it can be seen in Figure 4.

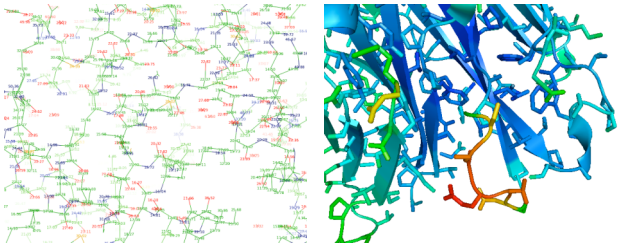

Figure 4: standard methods of displaying B-factors as numbers attached to the atom position on the left and as color coded carton and sticks on the right. In both situations occlusion is the main problem to retrieve information from the image.

In the given example we sonified average values for each amino acid residue. In the PyMOL window, where the molecule is displayed, a region of interest could be selected. Then the plug-in computed the mean values for each amino acid residue and sent it to the OSCresponder. In SC we sorted these values, and mapped them exponentially to frequencies from 100 to 2000 Hz to account for the nonlinear pitch/frequency relationship. The values were played in succession using the Formlet filter, excited by an impulse. The playback was also distributed between the channels, starting with low B-factors on the left moving towards higher ones on the right. The difference between the values defined the time offset, so that clustering in the data showed in the sonification.

The result gave an acoustically interesting arpeggio like overview over the B-factors of the selected region. 6 sonification samples show selected regions of different size ranging from 28 to 221 mean values of B-factors from amino acids. Without zooming into the details of the structure, it could be clearly acoustically identified if a region had generally low or medium B-factors. Singular residues with high factors always stood out towards the end of the arpeggio as short high tones.

5. FUTURE WORK

Since we present the sonification examples as prototypes only, we do not point here at specific possible future improvements of the sonifications of amino acid conformations and B-factors. Of course all the proposed sonifications have to be evaluated before they can enter everyday research. Looking more generally at the Data Sonification Design Space Map in [31], we found that one strategy for improvements is to represent rotamers, for instance, with less data by focusing on chemically informed, preprocessing schemes in order to create more sonifications strategies in the future.

The data structure which we have made available so far is certainly the core data related to molecules but it is not the only one of interest. As the next steps we will develop data interfaces for the electrostatic grid. Secondary structure information is also an interesting target for sonification. The mid term perspective is to make data from molecular dynamics available for sonification, as well as data from the sculpting process in PyMOL, which is a real time interactive modification and minimization process for small range molecules.

6. CONCLUSIONS

With SUMO we have introduced a framework which has a high potential to bring auditory display into everyday research praxis in chemistry. Our framework exhibits all the necessary features for scientific method development. Most important for the future acceptance of SUMO is the fact that it is an open source environment which allows for inspection, evaluation and participation of researchers from both the structural chemistry/biology and the sonification community. From the structural scientists' perspective it is further based on an already existing tool familiar to them, allowing for a smooth integration into their everyday work flow. In the near future, when the code is mature, sufficiently documented and made open source, we hope that this framework attracts programmers from natural sciences as well as from the sonification community.

7. AKNOWLEDGMENTS

We would like to thank Anton Huber for hosting Florian Grond at the Institute of Chemistry at the Karl Franzens University. Further we like to thank Walter Keller and Karl Gruber for fruitful discussions as well as Ludger Brümmer for his support during the initial phase of this project. Christopher Frauenberger, Thomas Hermann, Till Boverman and Alberto de Campo were helpful with SC and Georg Steinkellner with python related questions.

8. REFERENCES

1 PDB: *An information portal to biological macromolecular structures*, http://www.pdb.org

2 Richardson, J. S.. *The anatomy and taxonomy of protein structure*. Adv Protein Chem. 1981, 34:167–339.

3 Richardson, J. S.. *Early ribbon drawings of proteins*. Nat Struct Biol. 2000, 7:624–625

4 Schatz, G *Porträt eines Proteins* February 7 2007, NZZ http://www.nzz.ch/2007/02/07/fe/articleEQ18Y.print.html

5 Pelling, A. E. Sehati, S Gralla, E B Valentine, J S. Gimzewski J K. *Local nano mechanical Motion of the Cell Wall of Saccharomyces cerevisiae* Science 20 August 2004:Vol. 305. no. 5687, pp. 1147 – 1150 DOI: 10.1126/science.1097640

6 Quinn, M. *The climate symphony and other sonifications of Ice Core, Radar, DNA, Seismic and Solar Wind Data*, Proceedings of the 2001 International Conference on Auditory Display, Helsinki Finnland.

7 Quinn, M. Quinn, W. Hatcher, B. *For Those Who Died: A 9/11 Tribute* Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, 6-9 July 2003

8 Acosta, A. Camacho, D. Burbano, A. *Open(ing) Source* http://open-ing-source.net

9 Clark, M. A., *Genetic Music: An Annotated Source List* Texas Wesleyan University http://whozoo.org/mac/Music/Sources.htm

10 DNA & Protein Music www.nslij-genetics.org/dnamusic

11 Garcia-Ruiz, M.A. Gutierrez-Pulido, J.R. *An overview of auditory display to assist comprehension of molecular information* Interacting with Computers 2006 18 853–868 doi:10.1016/j.intcom.2005.12.001

12 Brown, A. Pettifer, S. Stevens, R. *Evaluation of a Non-Visual Molecular Browser*. In Proc. of the 6th International ACM Conference on Computer and Accessibility, 2004, 40-47

13 Lunney, D. *Development of a data acquisition and data analysis system for visually impaired chemistry students*. Journal of Chemical Education 1994 71, 308.

14 MIDI *Musical Instrument Digital Interface* 1983 www.midi.org

15 King, R.D. & Angus, C.G *PM: Protein Music*. CABIOS (1996) 12, pages 251-252 www.aber.ac.uk/~phiwww/pm

16 http://algoart.com/artwonk.htm

17 Dunn, J. Clark M.A. *Life Music: The Sonification of Proteins* Leonardo On-Line © 2004 ISAST www.leonardo.info/isast/articles/lifemusic.html

18 gene2music: www.mimg.ucla.edu/faculty/miller_jh/gene2music/home.html

19 Hansen, M D. Charp, E Lodha, S Meads, D Pang A, *PROMUSE: A System For Multi-Media data presentation of protein structural alignments* Pacific Symposium on Biocomputing 1999 4:368-379

20 www.molecularmusic.com

21 Yeung E. S, *Pattern recognition by audio representation of multivariate analytical data* Anal. Chem. 1980**;** *52*(7) pp 1120 - 1123; DOI:10.1021/ac50057a028

22 Ohno, S., Ohno, M.,. *The all pervasive principle of repetitious recurrence governs not only coding sequence construction but also human endeavor in musical composition*. Immunogenetics 198624, 71–78.

23 Many sonification projects were published in the Journal of Chemical Education http://jchemed.chem.wisc.edu

24 Sansom, C. *DNA makes protein —makes music?* The Biochemist December 2002. 40 www.biochemist.org/bio/02406/0039/024060039.pdf .

25 .Hofstadter, D.R, *Godel, Escher, Bach: an Eternal Golden Braid*, Vintage Books, New York 1980.

26 Byrne, C.M. *Water on tap: the use of virtual reality as an educational tool.* PhD Thesis, 1996 Department of Industrial. Engineering, University of Washington.

27 Brian FG Katz, Olivier Warusfel, Patrick Bourdot, and Jean-Marc Vezien, *CoRSAIRe Combination of Sensori-motor Rendering for the Immersive Analysis of Results* Proceedings of the 2nd International Workshop on Interactive Sonification, York, UK, 3 February 3 2007.

28 DeCampo, A. Frauenberger, C et al, *Science by Ear Workshop* 2007: http://sonenvir.at/workshop,.

29 *Auditory Display, Sonification, Audification, and Auditory Interfaces*, Gregory Kramer (Ed.), Addison Wesley, 1994

30 http://sonification.de/main-def.shtml

31 DeCampo, A. *Toward a data sonification design space map*, Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada, June 26 - 29, 2007

32 Hermann, T. Bunte, K. ,Ritter H. *Relevance-based interactive optimization of sonification* Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada, June 26 - 29, 2007

33 Hermann, T *Sonification for Exploratory Data Analysis*, Ph.D. thesis, Bielefeld University, Bielefeld, Germany, Feb 2002.

34 Pymol, a molecular visualization system: http://pymol.sourceforge.net

35 scientific computing with Python: http://numpy.scipy.org

36 Python for mathematics, science, and engineering: http://www.scipy.org

37 DeCampo, A. Frauenberger, C. Höldrich, R. *Designing a Generalizted Sonification Environment* Proceedings of Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004

38 SuperCollider: http://sourceforge.net/projects/SuperCollider

39 Wright, M. Freed, A. *Open Sound Control: A New Protocol for Communicating with Sound Synthesizers* International Computer Music Conference ICMC 97

40 Wright, M. Freed, Momeni, A. *Open Sound Control: State of the Art* Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada.

41 http://opensoundcontrol.org

42 SimpleOSC an OSC implementation for python: http://www.ixi-audio.net/content/body_backyard_python.htm