Using the velocity auto-correlation function to characterize functional "noise" in bio-molecules

M. Sidore*, Y. Meriguet*, A. Kudashova*, M. Lechelon†, M. Gori†, M. Pettini†, J. Torres*, L. Varani*
*Institute of Electronics and Systems (CNRS UMR 5214),
University of Montpellier, 860 rue St. Priest, 34095 Montpellier, France
†Center of Theoretical Physics (CNRS UMR 7332),
Aix-Marseille University, Marseille, France

***Abstract*—Characterization of current noise in simulated electrical systems relies on the Fourier transform of the auto-correlation of electrons velocity. A similar characterization can also be done in the simulation of biological systems, which relies on a statistical mechanics description through the calculation of the trajectories of the atoms over time. In such a system, the quantity calculated in the same way as electrical noise can be defining, whether it has functional significance or can be compared to experiments in the Mega to TeraHertz range. To exemplify this, we carried out the simulation of a model protein, the Bovine Serum Albumin, and outline the necessary precautions needed to obtain biologically relevant information.**

I. Introduction

Characterization of noise in simulated electrical systems often relies on the Fourier transform of the auto-correlation function of the velocity, or the fluctuations of the velocity, of electrons.

A similar approach can also be used in a biological context, with macromolecules such as proteins or nucleic acids, albeit the "noise" calculated through the auto-correlation function of the velocity of the particles (here, atoms), translates to the vibrational spectrum of the molecule [1], [2]. Using this approach through the simulation of a protein, which relies on a statistical mechanics description through the calculation of the trajectories of the atoms over time, we can uncover dynamical sub-millisecond movements that are often required for an adequate apprehension of biological function [3]–[6].

In the simulation of biomolecules using molecular dynamics, the position and velocity of each atom is recorded every Δt. This allows us to characterize what we can then call molecular "noise", that is in fact how the molecules vibrate. When calculating the Fourier transform of the auto-correlation function of the velocities of the atoms of a bio-molecule, we gain access to its vibrational density of states which can then directly be compared to experimental spectra such as the THz spectra of the same bio-molecule [2]. Through this approach, we can couple simulation and experiment to gain insights into this molecule's signature molecular vibrations, which can be important from a biological and physical point of view.

To exemplify the vibrations we can uncover through this method, we have simulated a system containing the model protein Bovine Serum Albumin (BSA) and obtained its vibrational density of states. BSA is a 583 amino-acid protein (Figure 1) involved in the maintain of oncotic pressure in blood vessels that also acts as a carrier for steroids, fatty acids and hormones. Its main use in biochemistry is as a standard in many experiments.

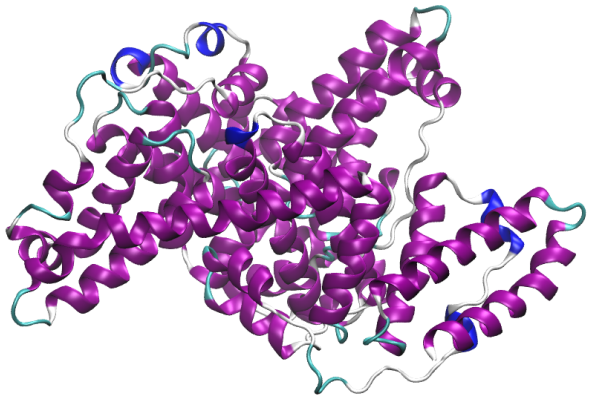

Fig. 1. Cartoon representation of the Bovine Serum Albumin (PDB ID: 3V03). The alpha helices are in purple, the turns are in cyan, the coils are in white and the π-helices are in blue.

After outlining the conditions which need to be met to simulate a system that is physical enough to obtain relevant vibrations, we compare the obtained spectra with previously published data.

II. Methods

Construction of the system

The all atom coordinates of the Bovine Serum Albumin were taken from the Protein Data Bank (PDB ID: 3V03 [7]). BSA is a model protein extensively used in biophysics which is an ideal candidate to be simulated given its lack of cofactors and can moreover be compared to experiments. Missing atoms were built with a homemade script which used the MODELLER [8] software, available at https://framagit.org/msidore/ieee_vdos. The system was then built using the GROMACS tools. During this procedure, a solvation shell of 20 Å was used to ensure that the protein wouldn't see itself through periodic boundary conditions; indeed, THz experiments have

uncovered solvation shells in the nanometer scale [2]. The resulting system has a size of 13 nm^3. The system was then ionized with 0.15 M Na^+ and Cl^-. The final system has 75106 water molecules, 230 Na^+ and 214 Cl^-.

The CHARMM36 forcefield [9], [10] coupled to the TIP3P water model was used. This rigid 3 points water model was chosen because it is sufficient to reproduce long range solvation sphere perturbations [2].

A control system, containing only a 3 Å solvation sphere in a vacuum, has also been built as a control and similarly equilibrated and simulated.

Simulation parameters

All Molecular Dynamics simulations were performed using the GROMACS 2018 simulation package [11]–[15]. A two-step minimization comprised of a steepest gradient step and a conjugate gradient step up to an energy tolerance of 500 kJ/mol/nm. The Van der Waals and Coulomb cutoffs were set at all times to 1.2 nm. The system was then carefully equilibrated.

The first step was a 50 ps NVT run during which the Cα of the protein were position restrained, with a 1 ps Δt and a velocity rescale [16] thermostat set at 300 K with a 0.2 ps time constant. The second step was a 50 ps NPT run with the same parameters plus a Berendsen barostat [17] set at 1 bar with a 2.0 ps time constant.

After these first steps of equilibration, 40 ps of NPT were carried out with constraints set to none and Δt set to 0.5 fs. While this parameter is usually set to all bonds, which is reasonable in most situations [18], constraining the bonds could negatively affect the vibrational density of states by removing small degrees of freedom. A subsequent 160 ps NPT run was then used to remove the position restraints. A return to the NVT ensemble was then carried out for 400 ps with a Nose-Hoover thermostat [19] set at 300 K with a 0.2 ps time constant.

The production run was then carried out in NVT for 1 ns using Particle Mesh Ewald [20] for long-range electrostatics with a PME order of 4 and a grid spacing of 0.12 nm. During the production run, the coordinates and velocities were saved every 8 steps (4 ps) to ensure sufficient sampling to obtain the vibrational density of states.

Analysis

The VMD software [21] was used to render the protein and the assignment of the secondary structure was predicted with the STRIDE software.

Graphical output has been produced with the Grace software.

The vibrational density of states is estimated from the Fourier transform of the autocorrelation function of atomic velocities:

$$VDOS(f) = \int \frac{\langle \vec{v}(0)\vec{v}(t)\rangle}{\langle \vec{v}(0)\vec{v}(0)\rangle} exp(i2\pi ft)dt \qquad (1)$$

The analyses were performed using the GROMACS post-processing tools *dos* and *vacf* on the protein atoms. The vibrational density of states characterizes the oscillations of the molecule in the frequency domain because the motion of the atoms is of an oscillatory nature (atomic velocities self-correlate in a periodic manner). More precisely, all vibrational bands of covalently bonded atoms (such as bond stretching and bond bending) are shown in the vibrational density of states spectrum. For these analyses, water molecules were discarded from the trajectory because the saving of the velocities every 4 fs led to storage issues.

III. Results

The aim here is to produce and outline how the auto-correlation of velocities, which can be conceptualized as "noise", can also be used within biological molecules to obtain signature vibrations. Figure 2 shows the calculated velocity autocorrelation function according to equation 1.

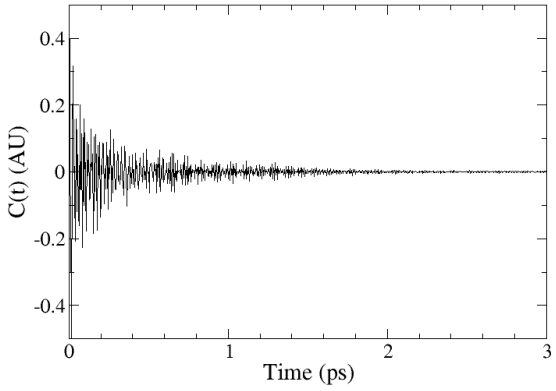

Fig. 2. The velocity autocorrelation function using the velocities of the atoms of the protein in Arbitrary Units (AU) as a function of time. The velocities of the atoms self-correlate to 0 on the scale of a few picoseconds.

The velocity autocorrelation function exhibits a fast oscillating behaviour superimposed to an exponential decay and vanishes on a time scale of few picoseconds. The Fourier transform of this autocorrelation function then turns the periodicities into frequency bands and Figure 3 shows the whole normalized vibrational density of states.

The general shape of the density of states is rather complicated showing several resonances in a wide frequency domain. The frequency peaks found at frequencies higher than 50 THz correspond to the fastest motions in our system. Most importantly, their highest amplitudes are of the same level as the peaks below 50 THz, which indicate that the system contains enough water molecules. Indeed, the 90 THz band has a high amplitude if the protein doesn't have enough water molecules for a complete solvation (Data not shown) and it has been shown that enough water is needed to allow its full range of vibrations.

However, since we are in a solvated environment and water absorbs THz radiations heavily, only the frequencies in the

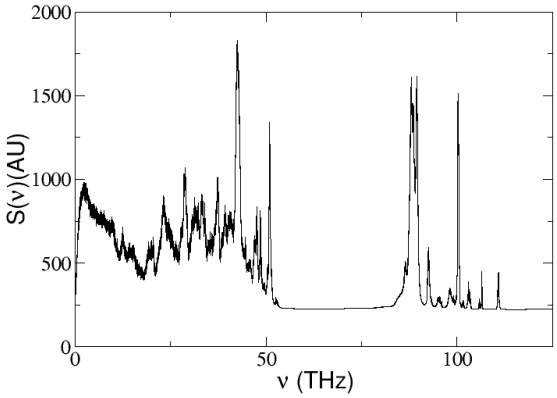

Fig. 3. Whole normalized vibrational density of state of the model protein Bovine Serum Albumin simulated for 1 ns.

low-frequency THz range can be compared with available experiments obtained using THz spectroscopy techniques. Figure 4 shows a zoom of the vibrational density of states in this range.

In this frequency range, we can then compare with experiments. In solution [22], BSA displays a monotonous increase between 0 and 2 THz, which is in accordance with our calculations (Figure 4). After 2 THz, the frequency slowly decreases and reaches a plateau, which is not found in experiments but can be explained by the effects of water on the measurements.

IV. Discussion and Conclusion

In this paper, we have outlined the relevance of the autocorrelation of the velocities of particles in biology to investigate molecular "noise", translated to specific molecular vibrations. In this protocol, the preparation of the system takes the most considerations, from a careful equilibration to the absence of bond constraints and the size of the solvation sphere.

Indeed, the solvation sphere is measured and hypothesized to be several nanometers wide. In our simulations, this matter is entangled with the choice of the water model: can a rigid 3 point water model reproduce a perturbation that extends this far and if it does, what is the minimal amount of water we can put in our system ? Since it has been shown that the TIP3P model can indeed propagate a perturbation far enough from the protein surface [2], the use of this model seems to be reasonable, but that doesn't tell if the links between this perturbation and protein dynamics reproduce what happens in experiments. Nonetheless, with it comes a practical matter: if a simple water model is perturbed far from the protein surface, it means that we need a big amount of water molecules to prevent the protein from "seeing itself" through the periodic boundary conditions, which in turn limits the length of the simulation.

Another practical problem is the storage space necessary to save velocities of a big system every 4 fs. Indeed, several dozens of To would be needed for the whole trajectory and

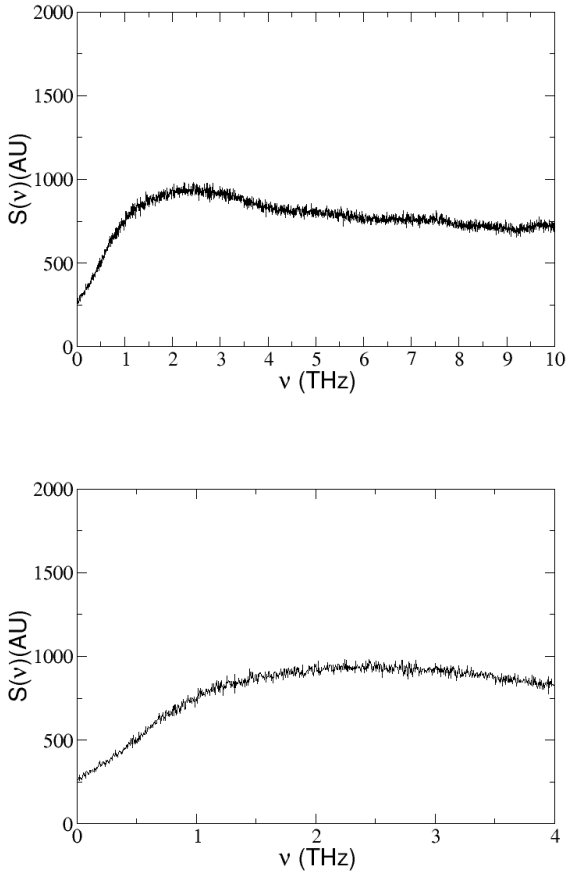

Fig. 4. Normalized vibrationnal density of state, zoomed at 10 THz (up) and 4 THz (down). This frequency range displays the maximum range that can be compared with experiments in a solvated environment.

we had to discard water molecules in the saved trajectory to meet our available storage.

This approach can also realistically be compared to the experiment in the low THz range and be used to identify frequency shifts associated with changes in the atoms dynamics when the molecule is driven far from equilibrium due, for instance, to energy absorption from the external medium.

We stress also that a limit of these simulations concerns the absolute amplitude of the calculated frequencies: molecular dynamics simulations are based on a classical model while an accurate estimation of these amplitudes would require a quantum treatment. Moreover, experiments of THz spectroscopy in a protein solution is also tricky and can lead to artefacts [4], [6], [22], [23], and especially with this protein which can be considered "sticky". As a consequence, the comparison between simulations and experiments is still problematic and additional work is needed to achieve reliable physical information.

However, we want to emphasize that the calculation of the autocorrelation of the atoms' velocities can nonetheless provide useful insights into protein dynamics when measured

and calculated frequency bands are shifted, and the protocol here presented would allow this possibility.

Moreover, the protein here studied provides a working example that may lead to more calculations of molecular "noise" and comparison with experiments.

V. Acknowledgments

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 765426 (TeraApps). This project has also received funding from the LabEx NUMEV.

References

[1] S. Goncalves and H. Bonadeo, "Vivrational Densities of State From Molecular Calculations," *Physical review B*, vol. 46, no. 18, pp. 19–21, 1992.

[2] O. Sushko, R. Dubrovka, and R. S. Donnan, "Terahertz spectral domain computational analysis of hydration shell of proteins with increasingly complex tertiary structure," *Journal of Physical Chemistry B*, vol. 117, no. 51, pp. 16 486–16 492, 2013.

[3] L. Meinhold, J. C. Smith, A. Kitao, and A. H. Zewail, "Picosecond fluctuating protein energy landscape mapped by pressure temperature molecular dynamics simulation." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 44, pp. 17 261–5, 10 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17956984http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2077243

[4] G. Acbas, K. A. Niessen, E. H. Snell, and A. G. Markelz, "Optical measurements of long-range protein vibrations," *Nature Communications*, vol. 5, pp. 1–7, 2014.

[5] Y. He, J.-Y. Chen, J. R. Knab, W. Zheng, and A. G. Markelz, "Evidence of protein collective motions on the picosecond timescale." *Biophysical journal*, vol. 100, no. 4, pp. 1058–65, 2 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21320451http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3037599

[6] F. Lipps, S. Levy, and A. G. Markelz, "Hydration and temperature interdependence of protein picosecond dynamics," *Physical Chemistry Chemical Physics*, vol. 14, no. 18, p. 6375, 4 2012. [Online]. Available: http://xlink.rsc.org/?DOI=c2cp23760a

[7] K. A. Majorek, P. J. Porebski, A. Dayal, M. D. Zimmerman, K. Jablonska, A. J. Stewart, M. Chruszcz, and W. Minor, "Structural and immunologic characterization of bovine, horse, and rabbit serum albumins," *Molecular Immunology*, vol. 52, no. 3-4, pp. 174–182, 10 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0161589012002842?via%3Dihub

[8] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling using Modeller." *Current protocols in bioinformatics*, vol. Chapter 5, pp. Unit–5.6, 10 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18428767http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4186674

[9] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, "CHARMM36m: an improved force field for folded and intrinsically disordered proteins," *Nature Methods*, vol. 14, no. 1, pp. 71–73, 1 2017. [Online]. Available: http://www.nature.com/articles/nmeth.4067

[10] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, "Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models," *Journal of Chemical Theory and Computation*, vol. 6, no. 2, pp. 459–466, 2 2010. [Online]. Available: http://pubs.acs.org/doi/10.1021/ct900549r

[11] E. Lindahl, B. Hess, and D. van der Spoel, "GROMACS 3.0: A package for molecular simulation and trajectory analysis," pp. 306–317, 2001.

[12] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 43–56, 1995.

[13] M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess, and E. Lindah, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, pp. 19–25, 2015.

[14] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, "GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.

[15] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.

[16] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *Journal of Chemical Physics*, vol. 126, no. 1, 2007.

[17] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 10 1984. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.448118

[18] W. F. van Gunsteren, M. Karplus, and W. F. van Gunsteren, "Effect of Constraints on the Dynamics of Macromolecules," *Macromolecules*, vol. 15, no. 6, pp. 1528–1544, 1982.

[19] W. G. Hoover and B. L. Holian, "Kinetic moments method for the canonical ensemble distribution," *Physics Letters A*, vol. 211, no. 5, pp. 253–257, 2 1996. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/0375960195009736

[20] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *The Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.470117

[21] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 2 1996. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0263785596000185

[22] J. Xu, K. W. Plaxco, and S. J. Allen, "Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy," *Protein Science*, vol. 15, no. 5, pp. 1175–1181, 2006. [Online]. Available: http://doi.wiley.com/10.1110/ps.062073506

[23] S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, M. Gruebele, D. M. Leitner, and M. Havenith, "Protein Sequence- and pH-Dependent Hydration Probed by Terahertz Spectroscopy," *Journal of the American Chemical Society*, vol. 130, no. 8, pp. 2374–2375, 2 2008. [Online]. Available: https://pubs.acs.org/doi/10.1021/ja0746520