Project PN-III-P4-ID-PCE-2020-2783-P
Scientific Report 2022

Scientific summary and results

The present research project has two main goals: 1) the development and kick-start of an online database of Vibrational Circular Dichroism spectra, and 2) the validation of the GA-VCD protocol using the VCD spectra collected in the newly created database. The GA-VCD protocol, was developed in house and uses artificial intelligence algorithms to assign the absolute configuration of chiral compounds using VCD spectroscopy.

In July 2022 we launched online the VCD-Machine database. The database can be accessed online at the following address:

https://vcd-machine.provitam.ro/database

The database was advertised at two international conferences, which has led to two important international collaborations. Furthermore, the GA-VCD protocol has been tested successfully on 20 flexible molecules.

Progress summary

There were three objectives scheduled for 2022. All of them have been achieved.

1) Simulation of VCD spectra (Act. 1.1 in the "Project Implementation Plan")

Together with Mr. Szabolcs Jako we have computed IR and VCD spectra for 21 flexible molecules. In total, more than twenty two thousand conformations have been computed at DFT level of theory this year. 13 molecules have not yet been uploaded online because they are the subject of a few scientific articles which have not been published yet.

Currently, the VCD-Machine database contains the experimental and simulated IR and VCD spectra of 33 molecules. It is important to note here the contributions of Dr. Rina Dukor (Biotools Inc. Florida, USA) and Dr. Joao M. Batista Jr. (Federal University of Sao Paulo, Brasil) who shared with us a large number of experimental spectra for

the vcd-machine database, i.e. 120 and 30, respectively. 25 of these spectra have been uploaded online, while many of the remaining ones will be simulated and uploaded online next year.

2) Validation of the GA-VCD protocol (Act. 1.2 in the "Project Implementation Plan")

The GA-VCD protocol was tested extensively on 20 flexible molecules using the IR and VCD spectra simulated so far. These preliminary tests have shown that this newly proposed protocol for assigning the AC of chiral compounds is very effective and that in difficult situations it yields results that are significantly more reliable than those predicted with the standard chiroptical spectroscopy protocol.

3) Testing and optimisation of the developed analysis methods for VCD (Act. 1.3 in the "Project Implementation Plan")

The results of the GA-VCD tests performed so far help us optimise this protocol further. Under my supervision, Mr. Gabriel Marton has developed a new Python GA-VCD code, which is faster and much more effective at optimising Boltzmann factors of the low energy conformers. Furthermore, the GA-VCD code was combined with a hierarchical clustering analysis, which provides a simple and intuitive physical interpretation of the results predicted by the GA-VCD protocol. This allows one to assess on the fly the reliably of the predictions made by the GA-VCD protocol, which in turn makes this protocol even more effective. In addition, I have written a series of script programs, which automates the process of simulating and organising the IR and VCD spectra of hundred/thousands of conformers. This reduces the process of assigning the AC of a chiral compound using the GA-VCD protocol to a very simple procedure, which requires one to type a few very simple commands on a computer.

Results dissemination

The results obtained so far in this research project have been presented at two international conferences (i.e. a 20 minutes oral presentation at the CD2022 conference in New York City, USA and a 30-minutes invited talk at the VOA7 conference in Edmonton, Canada). We are currently working on two publications, one of which will be submitted by the end of this year. Additionally, a collaboration with Dr. Prashant Kumar and Prof. N. Kotov has resulted in an article, which has been recently accepted in the very prestigious Nature journal.

Executive summary of the preformed activities

01.2022 – 12.2022: Simulation of IR and VCD spectra.
In February 2022, the Computational Chemistry lab built by dr. V.P. Nicu at University Lucian Blaga Sibiu was moved at the Provitam Research Foundation. This lab was used from January 2022 till December 2022 almost continuously to compute infrared absorption and circular dichroism spectra for 21 molecules (more than 21k conformers in total).

04.2022 – 12.2022: Development of GA-VCD python code.
This large number of computed spectra was used in combination with experimental spectra (digitised from scientific publications or provided by international collaborators) to thoroughly investigate the performance of the GA-VCD protocol. A critical analysis of the obtained results has revealed that the original FORTRAN GA-VCD code was relatively inefficient and slow. As a result, a more efficient, Python code was developed by mr. Gabriel Marton under the supervision of dr. Nicu. In addition, the original genetic algorithm (GA) was complemented with a hierarchical clustering algorithm. By assessing the trends in the conformers spectra, this clustering algorithm allows one to assess the reliability of the prediction made by the GA-VCD protocol.

07.2022 – 12.2022: Results dissemination.
The preliminary results of the GA-VCD tests performed in 2022 have been presented at two international conferences. That is, on July 25 at the CD2022 conference in New York City, USA (https://www.cd2022.org/day2) and on August 10 (as invited speaker) at the VOA7 conference in Edmonton, Canada (http://voa7.voaconference.com/program).

In September 2022 I started writing two articles which will report the first conclusive results about the performance of the GA-VCD protocol. The first publication will be submitted before the end of this year. In addition, a collaboration with Dr. Prashant Kumar and Prof. N. Kotov has resulted in an article, which has been recently accepted in the very prestigious Nature journal.

Valentin Paul Nicu
December 7, 2022