ERC Starting Grant for HITS group leader

Computational chemist **Ganna Gryn'ova**, head of the Computational Carbon Chemistry (CCC) group at the Heidelberg Institute for Theoretical

Studies (HITS), has been selected to receive a prestigious scientific grant: The European Research Council (ERC) awarded her an ERC Starting Grant of approximately €1.5 million for her project PATTERNCHEM. Dr. Gryn'ova's proposal is one of only 10% of projects that are selected for funding. "We are very happy for Ganna and also proud of this success story," stated HITS Scientific Director Frauke Gräter, "because these grants show the high quality of our research, and now, researchers in six of thirteen groups are working on or participating as a beneficiary in an ERC Grant."

Hailing from Ukraine, Ganna Gryn'ova earned her PhD in computational chemistry at the Australian National University, Canberra, in 2014, followed by a postdoctoral fellowship at École polytechnique fédérale de Lausanne, Switzerland. In April 2019, she established the CCC research group at HITS. Among her other accolades, Dr. Gryn'ova has received a Marie Skłodowska-Curie Actions Individual Fellowship (2016) from the European Commission and is a member of the Elisabeth Schiemann Kolleg at the Max Planck Society (as of 2021). She is also a co-investigator for the German Research Council's Sonderforschungsbereich (Collaborative Research Center) "SFB1249: N-heteropolycycles as functional materials" as well as a member of the Interdisciplinary Center for Scientific Computing (IWR) at Heidelberg University.

Harnessing patterns in functional materials

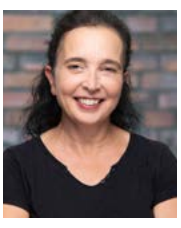
Functional organic materials possess myriad chemical and physical properties that are valuable for environmental remediation, green catalysis, drug delivery, and more. The CCC group uses state-of-the-art computational chemistry to explore and develop these materials. Thus far, the team has laid the groundwork for accurately and efficiently simulating two-dimensional organic materials and the interactions between these materials using small molecular targets. Building on this work, the new ERC-funded PATTERNCHEM project will exploit the topological fingerprints of novel and tantalizing materials, such as graphene derivatives, covalent organic frameworks, and hyperbranched polymers. Ultimately, this line of research aims to establish unconventional ways of designing and optimizing functional organic materials while targeting a given scientific or industrial application.

Via Data
The HITS blog can be found on the "Scilogs" portal at https://scilogs.spektrum.de/via-data/.

HITS

A new lead in FAIR data

ELIXIR – an intergovernmental organization that brings together life-science resources from across Europe – has named new leaders for several platforms. Among them is HITS researcher **Ulrike Wittig** from the Scientific Database and Visualization (SDBV) group, who will serve a two-year term in the data platform together with Patrick Ruch (Switzerland) und Silvio Tosatto (Italy). The researchers' term began in January. ELIXIR makes it easier for scientists to find and share data, exchange expertise, and agree on best practices.



HITS group leader selected as member of the International Space Science Institute

Kai Polsterer, head of the Astroinformatics (AIN) group at HITS, was selected as a member of an international team at the International Space Science Institute (ISSI). The team comprises experts in star formation, survey data, big data analysis, and machine learning and uses multi-wavelength data to derive a new evolutionary scheme for young stellar evolution. In the ISSI, scientists from all over the world meet in a multi- and interdisciplinary setting to reach for new scientific horizons.

New Position: Acting group leader in the MBM group

As of this year, HITS Scientific Director and MBM group leader **Frauke Gräter** is being supported by **Camilo Aponte-Santamaría** as acting group leader of the Molecular Biomechanics group (MBM). As the HITS Scientific Director has to spend a lot of time managing the Institute, the role of acting group leader has been established to help lead the group in scientific matters in accordance with the Scientific Director.

New employees and visiting scientists

Master's students:	Evgeni Ulanov (HITS Lab, CST, & MBM)
Pre-docs:	Jonathan Teuffel & Aysecan Ünal (both MBM)
Post-docs and research associates:	Michael Bazot (TOS), Susan Eckerle (SDBV), & Antony Noll (TOS)
Visiting scientists:	Matheus Vitor Ferreira Ferraz (MCM)

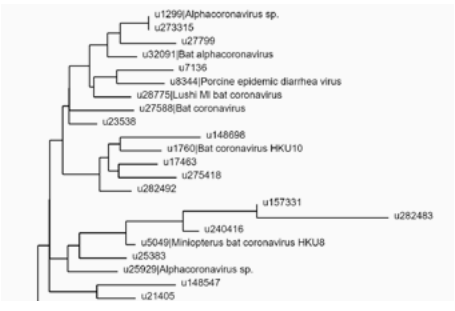
HITS groups 03/2022): *Astroinformatics (AIN), Computational Carbon Chemistry (CCC), Computational Molecular Evolution (CME), Computational Statistics (CST), Data Mining and Uncertainty Quantification (DMQ), Groups and Geometry (GRG), Molecular Biomechanics (MBM), Molecular and Cellular Modeling (MCM), Natural Language Processing (NLP), Physics of Stellar Objects (PSO), Scientific Databases and Visualization (SDBV), Stellar Evolution Theory (SET), Theory and Observations of Stars (TOS).*

HITSters

Digital tools for effective virus research

The diversity of viruses on our planet remains almost entirely unclear because only a small fraction of existing viruses are known to science. The current SARS-CoV-2 pandemic has revealed the devastating consequences that emerging viral diseases can have for humankind. It is therefore critical to categorize the global diversity of viruses with the aid of methods from computer science and to make the data usable for science.

Public sequence databases have become a vast repository of genetic data, with contributions being made by researchers around the world. These data stem from biological research groups that generate sequence data, be it to study the soil microbiome of the Amazon rainforest or to investigate the spread of diseases such as the SARS-CoV-2 virus.



Typically, such studies obtain genetic-sequence data not only from the intended target organism, but also from other organisms whose genetic material happens to be included in the sample. Such incidental data may be of particular interest to other researchers because these data are not the focus of the original study and are therefore usually ignored. However, they are still deposited in the public databases.

Unearthing this hidden treasure trove would require researchers to search through immensely large and distributed datasets because the freely accessible public databases contain sequence data in the order of petabytes (i.e., one million gigabytes). Researchers in the international Serratus project have developed Serratus, an open-source cloud computing infrastructure that is able to perform petabyte-scale sequence alignment.

Number of newly discovered viruses increases tenfold

"Our infrastructure enables efficient searches to be made in the Sequence Read Archive, one of the most popular public sequence repositories," explained co-author **Pierre Barbera** (Computational Molecular Evolution group) at HITS. Pierre developed software that calculates and analyzes the phylogenetic trees of all the studied species. Using the developed tools, researchers were able to identify more than 130,000 new RNA viruses – a tenfold increase in the number of known virus species. These viruses include previously unknown members of the coronavirus family related to the SARS-CoV-2 virus, novel viruses related to the hepatitis D virus, and novel bacteriophages – that is, viruses that specifically target bacteria.

The international research team has published the results of its study in the journal Nature. The data from the project are open source and can also be found on the website www.serratus.io.

Edgar, R.C., Taylor, J., Lin, V. et al. Petabase-scale sequence alignment catalyses viral discovery. Nature, 26 January 2022. DOI: 10.1038/s41586-021-04332-2 / https://www.nature.com/articles/s41586-021-04332-2

Research

Beyond the limits: SIMPLAIX

Traditionally, the investigation of molecular mechanisms and the rational design of molecules and materials for targeted applications have been guided by physics-based modeling and simulation. While this process has revolutionized modern science and technology, it remains necessary to map, explore, and analyze the infinitely complex and variable structure-property space of molecular systems across scales. This is where data-driven and machine-learning methodologies offer a promising approach. But the major obstacles to applying data-driven methods to atomistic systems – for example, dealing with complex three-dimensional structures and integration within multiscale simulation algorithms, to name just two – must first be resolved.

SIMPLAIX is a new 3-way inter-institutional collaboration between HITS, Heidelberg University, and the Karlsruhe Institute of Technology (KIT). It aims to pool the expertise of the three partners to address the challenge of bridging scales from molecules to molecular materials using multiscale simulation and machine learning.

In SIMPLAIX, these methods are being developed and employed to study a set of challenging problems in biomolecules and molecular materials within 8 multidisciplinary, inter-institutional research projects. Examples of the topics studied include discovering the connection between collagen-related diseases and defenses against radical damage, introducing machine learning for supported organic electrode materials, speeding up classical simulations or quantum chemical computations via predictions from machine learning, and predicting where inorganic and organic materials will break under force.

SIMPLAIX is coordinated by HITS researchers **Rebecca Wade** and **Frauke Gräter**, **Ganna Gryn'ova** is one of the 8 Principal Investigators. The initiative was launched in October 2021. It is funded by the Klaus Tschira Foundation and supported by in-kind contributions from KIT and Heidelberg University. During the initiative, 8 young researchers will be hired to work on the projects. Some positions have already been filled. The official SIMPLAIX inauguration event will take place on 12 April 2022 in the Studio Villa Bosch, Heidelberg.

Why is SIMPLAIX called SIMPLAIX?

A "simplex" is a generalized triangle, and this shape reflects the triangular (3-way, trilateral) collaboration between the 3 participating institutions.

- **SIM** stands for SIMulation.
- **AI** stands for Artificial Intelligence.
- **X** represents the interdisciplinary eXchange between simulation, AI, molecules, and materials.
- SIMPLAIX is pronounced /'sɪm.pleɪ̯ks/ (IPA)

Imprint | Dr. Peter Saueressig (V.i.S.d.P.), saueressig@h-its.org, Tel. +49 6221 533 245 | Pictures: HITS, Annette Mück, Gülay Keskin, Serratus Project, L.M.F. Junco (UNIANDES) | www.h-its.org

Beyond the limits