RECENT DEVELOPMENTS IN ATOMISTIC MODELING: MACHINE LEARNING MODELS AND DATASETS, METHODS, SOFTWARE RELEASES, AND SCIENTIFIC EVENTS

Stevan Armaković[1] and Sanja J. Armaković[2,*]

[1] *University of Novi Sad, Faculty of Sciences, Department of Physics, Novi Sad, Serbia;* stevan.armakovic@df.uns.ac.rs

[2] *University of Novi Sad, Faculty of Sciences, Department of Chemistry, Biochemistry and Environmental Protection, Novi Sad, Serbia;* sanja.armakovic@dh.uns.ac.rs

*Correspondence: sanja.armakovic@dh.uns.ac.rs

Abstract: In this review, we summarize a remarkable series of developments in atomistic modeling that unfolded over just six weeks, from mid-May to the end of June 2025. This extraordinary sequence began on May 13, when Meta's Fundamental AI Research team released the OMol25 dataset, a large-scale, high-accuracy quantum chemistry dataset, along with an accompanying paper on arXiv and the model on Hugging Face. On the same day, they also released universal models for atoms (UMA) on Hugging Face, with the related paper published on arXiv on June 30, 2025. Around the same period, the widely used package for atomistic calculations, ORCA, was updated to version 6.1 and officially released on June 17, introducing substantial methodological and performance improvements. In the area of semiempirical methods, the group of Prof. Grimme released a new and powerful method, g-xTB, on June 24. As the successor to the GFN family, g-xTB is a general-purpose extended tight-binding method that offers significantly improved accuracy at a modest computational cost. These advances coincided with WATOC 2025, the triennial meeting of the theoretical and computational chemistry community, held in Oslo from June 22 to 27, with g-xTB presented in session on June 26. Taken together, these developments represent one of the most dynamic and impactful periods in the recent history of atomistic modeling. During this time, we also released a new version of our platform, Atomistica.online 2025, our contribution aimed at enhancing accessibility and usability in molecular modeling.

Keywords: atomistic modeling, machine learning potentials, AI-driven atomistic simulations, OMol25, UMA, ORCA 6.1, g- xTB, semiempirical methods, molecular datasets, WATOC 2025, atomistica.online

1. Introduction

Atomistic modeling plays a central role in modern chemistry and materials science. Methods such as, for example, density functional theory (DFT), molecular dynamics (MD)

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

simulations based on force fields, and others provide fundamental insights into molecular structure, reactivity, and dynamics across a broad range of scientific and technological applications [1–3]. These techniques are crucial for making advancements in areas such as drug discovery, catalysis, materials design, and energy storage [4–9].

Despite their importance and usefulness, high-accuracy atomistic methods, particularly those based on quantum mechanics, are often computationally very demanding. The cost of performing atomistic calculations on large molecular systems or exploring extensive conformational spaces can quickly become unavailable, therefore, even if one owns modern and expensive hardware, the system of interest cannot be studied. To address this challenge, several strategies have been developed. Several approaches exist, while one standard involves introducing additional layers of approximation, leading to the development of, for example, semiempirical methods. These approaches, such as the GFN family of methods [10–14], reduce computational cost by parameterizing quantum mechanical interactions, although a certain price is paid in terms of generality and accuracy.

In recent years, the application of machine learning has gained increasing attention, aiming to complement, or even sometimes, replace or approximate quantum chemical calculations [15–17]. In this approach, mathematical models are trained to reproduce the results of expensive quantum mechanical methods, such as DFT, using large datasets of reference calculations. Once ML models are obtained, scientists can use them to predict molecular properties orders of magnitude faster than conventional quantum methods, while often retaining comparable accuracy within their domain of applicability. This approach denotes a shift from physics-based modeling to data-driven prediction.

The success of the machine learning approach depends critically on the availability of high-quality datasets. Development of traditional software for atomistic calculations is also significant, as it enables researchers to perform calculations and thus generate datasets that can be cross-correlated with experimental data. Building and maintaining these codes is a challenge in every aspect, especially as new theoretical methods and parallel computing architectures emerge. ORCA [18–25], for example, a widely used code for performing a number of types of calculations, stands out as a notable example of sustained innovation and usability.

Finally, scientific conferences continue to play a vital role in fostering collaboration, disseminating results, and setting new directions for the field. The triennial World Association of Theoretical and Computational Chemists (WATOC) congress [26] is one such event that enables connections between researchers from around the globe to exchange ideas, present new work, and shape the future of theoretical and computational chemistry.

Building upon the ongoing evolution of data-driven methods, software development, and community engagement outlined above, this review provides an overview of several significant developments that unfolded in the short span of less than five weeks in 2025, from May 13 to June 27 (Figure 1).

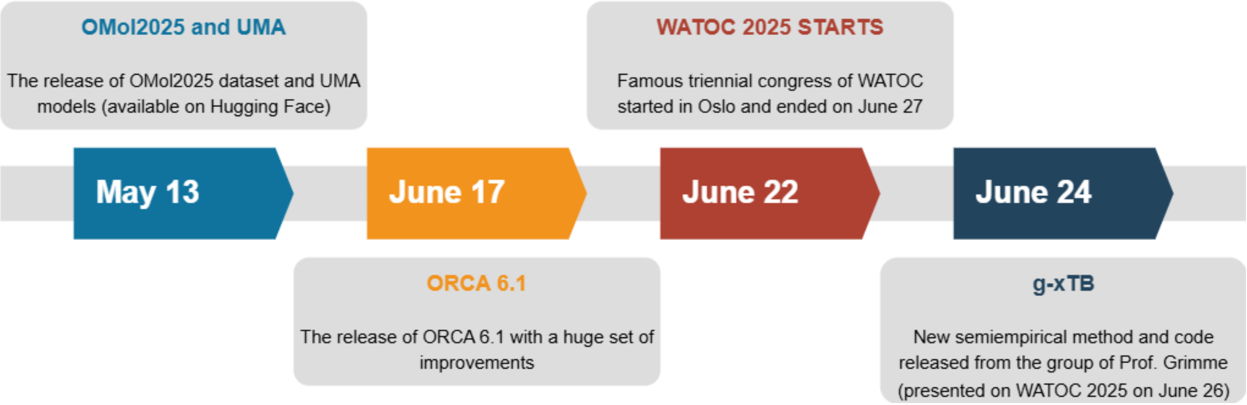
OMol2025 and UMA
The release of OMol2025 dataset and UMA models (available on Hugging Face)
WATOC 2025 STARTS
Famous triennial congress of WATOC started in Oslo and ended on June 27
May 13
June 17
June 22
June 24
ORCA 6.1
The release of ORCA 6.1 with a huge set of improvements
g-xTB
New semiempirical method and code released from the group of Prof. Grimme (presented on WATOC 2025 on June 26)

Figure 1. Timeline of key developments in atomistic modeling during six weeks in 2025

During this brief window, the Meta's Fundamental AI Research (FAIR) team [27] released OMol25 [28], the largest and most diverse dataset of DFT-calculated molecules to date, along with a Universal Models for Atoms (UMA) [29] trained on over 30 billion atoms. Simultaneously, the ORCA software suite advanced to version 6.1, bringing powerful new features and improved efficiency to a widely used code for performing various types of atomistic calculations. In the same period, Grimme's group introduced g-xTB [30], a successor to the popular GFN family of semiempirical methods, pushing the limits of fast, accurate simulations even further. Concluding this dynamic period, the WATOC 2025 congress in Oslo [31] gathered hundreds of researchers. We also briefly note our contribution to this vibrant ecosystem through updates to the Atomistica.online platform [32,33].

2. OMol25 and UMA: Advancing Molecular Modeling with Data and Models

One may think of Meta as a technology company primarily known for its role in social networking and virtual platforms. However, in recent years, Meta has also become a significant factor in fundamental scientific research through its FAIR initiative. While FAIR has been recognized for contributions to language modeling and AI hardware, its parallel efforts in computational chemistry and materials science have grown steadily, now culminating in a series of releases in 2025 that position it as a major contributor to the field of atomistic modeling.

The release of OMol25 and UMA represents huge contributions. These resources address two important aspects in machine learning for molecular simulation: access to large, high-quality datasets and the availability of generalizable and high-accuracy ML potentials. OMol25 resulted from hundreds of millions of quantum mechanical calculations spanning a wide diversity of chemical space, while UMA offers a trained, ready-to-use potential applicable across molecules and materials. Importantly, both were made openly available to the community, setting a new standard for transparency, scale, and accessibility in the area of atomistic calculations.

2.1. OMol25 Dataset: Scale, Quality, and Scope

The OMol25 dataset is a dataset of unprecedented scale. It was developed after more than 100 million quantum-mechanical calculations, each computed using DFT at the ωB97M-V/def2-TZVPD level of theory with a high-resolution 99590-point integration grid. In total, the dataset required over 6 billion core-hours of compute, making it one of the largest and most computationally demanding initiatives in the history of molecular simulation.

OMol25 was designed to overcome key limitations of earlier molecular datasets, such as QM9 [34] or SPICE [35,36], which were constrained by relatively narrow chemical coverage, inconsistent theory levels, or small system sizes. By contrast, OMol25 includes:

- 83 elements, far beyond the typical C-H-O-N-based molecules in prior datasets.
- System sizes up to 10× larger than in earlier efforts.
- Consistent DFT-level accuracy across all entries, enabling clean and transferable machine learning model training.

The dataset is structured around three core chemical domains:

- Biomolecules, including protein-ligand and nucleic acid complexes, were sampled through docking, restrained MD, and tautomer enumeration.
- Electrolytes, such as ILs, redox-active clusters, and solvent-gas interfaces, with structures extracted from MD trajectories and quantum cluster sampling.
- Metal complexes were generated using the GFN2-xTB method combined with the AFIR protocol to explore reactive pathways and diverse spin states.

Moreover, OMol25 includes recalculations of several existing datasets, such as ANI-2X [37], SPICE [35,36], and Transition-1x [38], at the same level of theory, ensuring consistency and extending its relevance. This approach positions OMol25 not only as a

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

standalone resource but also as a unifying layer for previously fragmented datasets in the field.

In terms of coverage and accuracy, OMol25 is widely seen as the molecular equivalent of what ImageNet [39–41] represented for computer vision: a large-scale, standardized dataset that catalyzes the development of general-purpose, high-performing models. It shifts the focus in the field from data acquisition to model innovation and fine-tuning, especially for academic and industrial groups who now have access to a shared high-quality training foundation.

2.2. UMA Models: Toward General-Purpose Neural Potentials

Accompanying the OMol25 dataset, Meta FAIR released a family of machine learning potentials known as UMA. These models are designed to provide general-purpose predictions of atomic interactions across a vast chemical space, bridging molecules and materials, and reflecting a new level of ambition in the development of neural network potentials (NNPs).

The UMA models are trained not only on OMol25 but also on a broader family of FAIR datasets, including OC20 [42], ODAC23 [43], and OMat24 [44]. This multi-dataset training was made possible through a novel Mixture of Linear Experts (MoLE) architecture. Inspired by mixture-of-experts strategies in natural language processing, MoLE enables UMA to learn from chemically heterogeneous datasets without significant performance degradation or increased inference cost.

In technical terms, UMA builds on Meta's earlier equivariant Smooth Energy Network (eSEN) framework [45], incorporating additional structure to support multi-domain generalization. The training process follows a two-stage strategy: first, a direct-force model is trained, followed by fine-tuning into a conservative-force model, which guarantees that the learned force field derives from a potential energy surface. This conservative fine-tuning not only improves physical ground but also enhances performance in tasks like MD, where energy conservation and smooth gradients are essential.

UMA models are released in different sizes (small, medium, large), though only the small model is currently openly available. Despite its modest size, the UMA-small model achieves mean absolute errors (MAEs) below 1 kcal/mol on representative benchmark tasks. Notably, the model matches or exceeds the accuracy of established DFT methods across several tasks while being orders of magnitude faster.

Perhaps most importantly, UMA is positioned as a general-purpose model, intended to serve as a pre-trained backbone for a wide range of downstream tasks. Much like large

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

language models are fine-tuned for translation or summarization, UMA can be fine-tuned for applications such as reactivity prediction, conformer generation, or solvation energy estimation. Its training scale and architectural design suggest a future where pre-trained atomistic models become a standard starting point in computational chemistry workflows.

In summary, the UMA models mark a shift from task-specific potentials to foundation models for atomistic simulation. When paired with the OMol25 dataset, they enable new levels of speed and accuracy in modeling systems that were previously out of reach, including large biomolecular complexes and catalytically relevant metal systems.

3. ORCA 6.1

While machine learning-based approaches have garnered increasing attention, traditional software for atomistic calculations remains indispensable for high-accuracy modeling, method development, and benchmarking. Among the most widely used electronic structure programs in both academia and industry, ORCA continues to evolve with each release, balancing cutting-edge functionality with usability and computational efficiency. The release of ORCA 6.1 in 2025 marks another critical step forward for the community relying on quantum chemical methods.

Prof. Frank Neese leads the ORCA project [18–25] at the Max-Planck-Institut für Kohlenforschung in Mülheim, Germany. Over the years, Prof. Neese has guided the development of ORCA into a full-featured, high-performance electronic structure platform that supports both foundational research and practical applications across a wide range of chemical disciplines. We dedicate space to ORCA not only because it is among the most powerful atomistic codes, if not the most powerful, but also because it remains completely free for academic use, a policy that greatly supports education and research worldwide.

ORCA 6.1 introduces a range of improvements spanning performance, accuracy, and user experience. One of the most notable developments is enhanced GPU acceleration, which significantly reduces computation times for hybrid DFT and post-Hartree-Fock methods. This makes the package highly competitive in high-throughput applications and large-scale simulations, particularly relevant as system sizes and data demands continue to grow in fields like catalysis, materials science, and bioinorganic chemistry.

The update also strengthens support for multi-node parallelism, enabling more efficient scaling across modern high-performance computing clusters. These improvements reflect a broader trend in computational chemistry software: adapting to evolving hardware architectures while preserving precision, flexibility, and methodological rigor.

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

On the theoretical side, ORCA 6.1 brings expanded support for modern dispersion corrections, new basis sets, and enhanced continuum solvation models, broadening its applicability to diverse molecular environments. The software also maintains strong support for multireference methods, excited-state calculations, EPR/IR/Raman spectroscopy, and relativistic effects, reinforcing its relevance for both fundamental and applied research.

It is worth noting that ORCA also played a central role in the creation of the OMol25 dataset: the FAIR team used ORCA 6.0.1 to perform the 6 billion core-hours of DFT calculations required to generate the dataset. This underscores ORCA's robustness, scalability, and reliability even under the most demanding computational conditions.

Finally, the ORCA development team places strong emphasis on user accessibility and educational support. Extensive and well-organized documentation is available, including a detailed user manual and a growing collection of tutorials that help guide users through common tasks, advanced methods, and best practices. This commitment to supporting the community, especially students and early-career researchers, further solidifies ORCA's role as a central tool in the field of computational molecular modeling.

4. g-xTB: a new generation of semiempirical methods

Semiempirical quantum mechanical (SQM) methods continue to play a crucial role in atomistic modeling, offering a balance between computational efficiency and chemical accuracy. These methods are especially valuable for tasks requiring extensive sampling, rapid geometry optimizations, or simulations of large molecular systems. Just recently, the group of Prof. Stefan Grimme introduced g-xTB, a next-generation SQM method that marks a significant advancement over its widely used predecessor, GFN2-xTB.

g-xTB is derived from extended tight-binding (TB) approximations to Kohn-Sham DFT and is designed to close the gap between traditional SQM methods and full DFT calculations in terms of accuracy, robustness, and transferability. The method is minimally empirical and has been parameterized on a highly diverse training set that includes not only standard molecular systems, but also challenging species such as transition metal complexes, actinides, and so-called "mindless molecules".

Several key innovations distinguish g-xTB from earlier xTB models: an atom-in-molecule adaptive atomic orbital basis, Hamiltonian that incorporates range-separated approximate Fock exchange, higher-order charge-fluctuation terms, atomic correction potentials, and a charge-dependent semi-classical repulsion function.

These methodological enhancements result in a model that consistently outperforms GFN2-xTB, often reducing mean absolute errors by half. In benchmarking across ~32000 relative energies, including datasets for thermochemistry, conformational energetics, non-covalent interactions, and reaction barriers, g-xTB achieved a weighted total mean absolute deviation (WTMAD-2) of 9.3 kcal/mol on the general main-group thermochemistry, kinetics, and non-covalent interactions (GMTKN55) database [46] benchmark, placing it on par with low-cost DFT methods. The model also shows substantial gains in domains where SQM and even DFT often struggle, such as transition metal systems, relative spin-state energies, and orbital energy gaps.

Despite these accuracy improvements, g-xTB maintains a computational profile that is only moderately more expensive than GFN2-xTB, typically incurring a 30-50% overhead [30]. This modest cost increase is more than justified by the gains in performance, making g-xTB the strongest candidate for large-scale screening, pre-optimization, and property prediction workflows.

Importantly, g-xTB is open source and designed to be a general-purpose replacement for the GFNn-xTB family. In many practical cases, it also serves as a viable alternative to low- and mid-level DFT methods, particularly when computational resources are limited or large systems are involved.

In summary, g-xTB represents a robust, transferable, and efficient SQM method that offers DFT-like performance at TB speeds. Its release marks an important evolution in semiempirical modeling and complements the broader landscape of atomistic simulation methods, including both traditional ab initio and machine learning-based approaches.

5. WATOC 2025

While methods, models, and software define the technical landscape of atomistic modeling, the role of scientific meetings in shaping the field's direction cannot be overstated. Conferences offer a platform for presenting new results, exchanging ideas, and building collaborations that often seed the next generation of scientific progress. In this regard, the WATOC conference series holds a unique and prestigious position.

The 13th Triennial WATOC Congress, held in Oslo, Norway, from June 22-27, 2025, brought together hundreds of participants from around the world. Hosted at the Oslo Congress Centre and organized by the Hylleraas Centre for Quantum Molecular Sciences at the University of Oslo, WATOC 2025 featured 12 plenary lectures, 150 invited talks, and three poster sessions, spanning 75 thematic areas that reflect the various aspects of the field, from electronic structure theory and MD to catalysis, spectroscopy, materials modeling, machine learning, and quantum computing.

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

A notable addition to this year's program was Young WATOC, held on June 21, which provided a stage for early-career researchers who defended their PhDs in 2020 or later. This satellite event showcased the work of a new generation of theorists and simulators, fostering dialogue between junior and senior scientists in an inclusive and forward-looking environment.

WATOC 2025 arrived at a moment of exceptional activity in the field. Several of the central developments discussed in this review, OMol25, UMA, g-xTB, and ORCA 6.1, were prominently featured in presentations, informal discussions, and poster sessions. These contributions not only highlighted the technological momentum in atomistic modeling but also sparked community-wide conversations on the future role of machine learning, software sustainability, and open science.

In summary, WATOC 2025 reaffirmed its place as a cornerstone of the theoretical and computational chemistry community. Its timing, coinciding with a period of rapid and far-reaching advances in atomistic modeling, made the congress especially impactful, serving both as a showcase of progress and a catalyst for future directions in the field.

6. Atomistica.online 2025: a platform for accessible molecular modeling

In parallel with several large-scale initiatives from major research institutions and industrial laboratories, smaller academic efforts continue to play a meaningful role in the availability of tools for atomistic modeling. As part of this broader movement, we released a significantly upgraded version of our platform, Atomistica.online 2025 (https://atomistica.online) [32,33], during this particularly active period in the field. While modest in scope compared to flagship machine learning models or production-grade electronic structure packages, Atomistica.online aims to provide an accessible and practical environment for conducting a broad range of atomistic modeling tasks, with particular emphasis on education, rapid prototyping, and early-stage research.

The platform has been completely redesigned with a new interface based on modern material design principles, offering a cleaner layout, faster loading times, and improved responsiveness across devices. Molecular visualization is now fully integrated into the browser using 3Dmol.js, allowing users to interactively view and manipulate molecular structures without installing additional software. These features support immediate visual feedback in applications such as geometry optimization, molecular descriptor validation, and structure generation, especially useful in teaching environments and quick-turnaround modeling scenarios.

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

Functionally, Atomistica.online 2025 offers access to a growing set of tools related to atomistic calculations. Notably, support has been added for the newly released g-xTB method [30] developed by Prof. Stefan Grimme and collaborators. Calculations using the GFN family of methods [10–14] and g-xTB can be executed directly through the browser interface, streamlining the process for users who may not have access to local infrastructure.

In step with emerging trends, Atomistica.online also integrates machine-learned potentials, including OMol25 and UMA, two high-impact models recently released by the Meta FAIR team. These models, initially designed for large-scale molecular energy prediction and catalyst–adsorbate optimization, can now be accessed via a simplified online interface. Tools based on the OC20 framework are also included, enabling the generation of relaxed surface-adsorbate systems without requiring users to handle complex installations or command-line workflows.

A significant area of focus in this release has been the modeling and design of ionic liquids (ILs). We developed and deployed a series of machine learning models, collectively referred to as IonIL-IM, for the prediction of density and viscosity of imidazolium-based ILs. These models are trained on curated datasets, constructed using global minimum geometries optimized via the GOAT protocol at the GFN-FF level. Two models are available for density prediction: IonIL-IM-D1, which uses only three simple descriptors for fast screening, and IonIL-IM-D2, which includes seven features for improved accuracy. For viscosity, the IonIL-IM-V model employs nine descriptors, including DFT-derived features such as ESP metrics and local ionization energy. Each model is available as an interactive tool requiring no programming skills: users input descriptors and instantly receive predictions, all within the browser.

To complement these models, an ILs Generator enables users to combine over 1300 cations and 450 anions to generate pre-optimized ion pairs suitable for further modeling or machine learning analysis. Structures are downloadable for seamless integration into quantum mechanical or MD workflows. Together, the models and generator support streamlined design and exploration of ILs for applications in green chemistry, energy storage, and solvent development.

The platform also includes several auxiliary tools aimed at accelerating typical computational workflows. These include a batch input file generator for ORCA, a surface property analyzer based on Multiwfn [47–50].

Importantly, in 2025, Atomistica.online introduced a live usage analytics dashboard, providing visibility into how the platform is used over time. Users and contributors can now explore real-time statistics on app visits, descriptor usage, and tool engagement by accessing the "Detailed usage stats" section. This feature not only enhances platform transparency but also offers a foundation for evaluating outreach impact and exploring potential academic collaborations or sponsorship opportunities based on real-world engagement.

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

Atomistica.online 2025 continues to be offered entirely free for academic and educational purposes, in line with our goal of supporting open science and inclusive access to computational tools. While it does not seek to compete with the scale of major institutional projects, it aims to complement them by enabling rapid, browser-based modeling for those who need accessible, interactive tools in real time. We hope that this platform, with its combination of quantum methods, machine learning models, curated datasets, and real-time usage insight, contributes a small but valuable part to the broader ecosystem of modern molecular simulation.

7. Conclusions

In the six-week period between May 13 and June 27, 2025, the scientific community engaged in atomistic modeling witnessed a remarkable wave of advancements. Meta FAIR's open release of OMol25 and UMA set new standards in dataset scale and model generality. At the same time, the methodological improvements embedded in ORCA 6.1 and the next-generation g-xTB semiempirical method provided even more resources for computational scientists. WATOC 2025 also played a valuable role by offering a timely venue for researchers to share insights, present new tools, and discuss ongoing developments in the field during this active period.

Taken together, the developments highlighted in this review reflect not only rapid progress in specific tools and methods but also a broader shift in how atomistic modeling is practiced and shared. They suggest a future in which machine learning, methodological robustness, open datasets, and accessible software work in tandem to support both established and emerging researchers in molecular science.

Amid these developments, we launched a new version of Atomistica.online, an academic, browser-based platform designed to broaden access to modern modeling tools. Though limited in scale, it brings together semiempirical methods, machine learning models for ILs, input and structure generators, and live usage analytics, illustrating how focused, accessible tools can support progress in the application of atomistic calculations.

References

[1] R. Thomas, Y.S. Mary, K.S. Resmi, B. Narayana, S.B.K. Sarojini, S. Armaković, S.J. Armaković, G. Vijayakumar, C.V. Alsenoy, B.J. Mohan, Synthesis and spectroscopic study of two new pyrazole derivatives with detailed computational evaluation of their reactivity and pharmaceutical potential, J. Mol. Struct. 1181 (2019) 599–612. https://doi.org/10.1016/j.molstruc.2019.01.014.

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

[2] R. Thomas, Y.S. Mary, K.S. Resmi, B. Narayana, S.B.K. Sarojini, S. Armaković, S.J. Armaković, G. Vijayakumar, C.V. Alsenoy, B.J. Mohan, Synthesis and spectroscopic study of two new pyrazole derivatives with detailed computational evaluation of their reactivity and pharmaceutical potential, J. Mol. Struct. 1181 (2019). https://doi.org/10.1016/j.molstruc.2019.01.014.

[3] Y.S. Mary, Y.S. Mary, R. Thomas, B. Narayana, S. Samshuddin, B.K. Sarojini, S. Armaković, S.J. Armaković, G.G. Pillai, Theoretical Studies on the Structure and Various Physico-Chemical and Biological Properties of a Terphenyl Derivative with Immense Anti-Protozoan Activity, Polycycl. Aromat. Compd. 41 (2021) 825–840. https://doi.org/10.1080/10406638.2019.1624974.

[4] G. Zheng, M. Xu, Q. Zhang, L. Mao, Z. Liu, M. Song, DFT investigation on sulfur-induced linear defects in MnCo2S4 for enhanced energy storage capabilities, J. Energy Storage 126 (2025) 117095. https://doi.org/10.1016/j.est.2025.117095.

[5] D. Adekoya, S. Qian, X. Gu, W. Wen, D. Li, J. Ma, S. Zhang, DFT-Guided Design and Fabrication of Carbon-Nitride-Based Materials for Energy Storage Devices: A Review, Nano-Micro Lett. 13 (2020) 13. https://doi.org/10.1007/s40820-020-00522-1.

[6] M. Andersen, K. Reuter, Adsorption Enthalpies for Catalysis Modeling through Machine-Learned Descriptors, Acc. Chem. Res. 54 (2021) 2741–2749. https://doi.org/10.1021/acs.accounts.1c00153.

[7] Y.-K. Lee, Density Functional Theory (DFT) Calculations and Catalysis, Catalysts 11 (2021) 454. https://doi.org/10.3390/catal11040454.

[8] N. Ye, Z. Yang, Y. Liu, Applications of density functional theory in COVID-19 drug modeling, Drug Discov. Today 27 (2022) 1411–1419. https://doi.org/10.1016/j.drudis.2021.12.017.

[9] V.T. Sabe, T. Ntombela, L.A. Jhamba, G.E.M. Maguire, T. Govender, T. Naicker, H.G. Kruger, Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review, Eur. J. Med. Chem. 224 (2021) 113705. https://doi.org/10.1016/j.ejmech.2021.113705.

[10] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, Extended tight-binding quantum chemistry methods, WIREs Comput. Mol. Sci. 11 (2021) e1493. https://doi.org/10.1002/wcms.1493.

[11] C. Bannwarth, S. Ehlert, S. Grimme, GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions, J. Chem. Theory Comput. 15 (2019) 1652–1671. https://doi.org/10.1021/acs.jctc.8b01176.

[12] S. Ehlert, M. Stahn, S. Spicher, S. Grimme, Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods, J. Chem. Theory Comput. 17 (2021) 4250–4261. https://doi.org/10.1021/acs.jctc.1c00471.

[13] C. Bannwarth, S. Ehlert, S. Grimme, GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions, J. Chem. Theory Comput. 15 (2019) 1652–1671. https://doi.org/10.1021/acs.jctc.8b01176.

[14] S. Grimme, C. Bannwarth, P. Shushkov, A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86), J. Chem. Theory Comput. 13 (2017) 1989–2009. https://doi.org/10.1021/acs.jctc.7b00118.

[15] M. Chen, Z. Yin, Z. Shan, X. Zheng, L. Liu, Z. Dai, J. Zhang, S. (Frank) Liu, Z. Xu, Application of machine learning in perovskite materials and devices: A review, J. Energy Chem. 94 (2024) 254–272. https://doi.org/10.1016/j.jechem.2024.02.035.

[16] L. Fiedler, K. Shah, M. Bussmann, A. Cangi, Deep dive into machine learning density functional theory for materials science and chemistry, Phys. Rev. Mater. 6 (2022) 040301. https://doi.org/10.1103/PhysRevMaterials.6.040301.

[17] R. Pederson, B. Kalita, K. Burke, Machine learning and density functional theory, Nat. Rev. Phys. 4 (2022) 357–358. https://doi.org/10.1038/s42254-022-00470-2.

[18] F. Neese, An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix, J. Comput. Chem. 24 (2003) 1740–1747. https://doi.org/10.1002/jcc.10318.

[19] F. Neese, F. Wennmohs, A. Hansen, U. Becker, Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange, Chem. Phys. 356 (2009) 98–109. https://doi.org/10.1016/j.chemphys.2008.10.036.

[20] F. Neese, F. Wennmohs, U. Becker, C. Riplinger, The ORCA quantum chemistry program package, J Chem Phys 152 (2020) Art. No. L224108. https://doi.org/10.1063/5.0004608.

[21] F. Neese, The SHARK Integral Generation and Digestion System, J Comp Chem (2022) 1–16. https://doi.org/10.1002/jcc.26942.

[22] F. Neese, Software update: the ORCA program system, version 4.0, WIRES Comput Molec Sci 8 (2018) 1–6. https://doi.org/10.1002/wcms.1327.

[23] F. Neese, The ORCA program system, WIRES Comput Molec Sci 2 (2012) 73–78. https://doi.org/10.1002/wcms.81.

[24] F. Neese, Approximate second-order SCF convergence for spin unrestricted wavefunctions, Chem Phys Lett 325 (2000) 93–98. https://doi.org/10.1016/s0009-2614(00)00662-x.

[25] F. Neese, Software update: the ORCA program system, version 5.0, WIRES Comput Molec Sci 12 (2022) e1606. https://doi.org/10.1002/wcms.1606.

[26] WATOC - Congress, (n.d.). https://www.watoc.net/watoc.congress.html (accessed July 27, 2025).

[27] Research, AI Meta (n.d.). https://ai.meta.com/research/ (accessed July 27, 2025).

[28] D.S. Levine, M. Shuaibi, E.W.C. Spotte-Smith, M.G. Taylor, M.R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, N.C. Frey, X. Fu, V. Gharakhanyan, A.S. Krishnapriyan, J.A. Rackers, S. Raja, A. Rizvi, A.S. Rosen, Z. Ulissi, S. Vargas, C.L. Zitnick, S.M. Blau, B.M. Wood, The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models, (2025). https://doi.org/10.48550/arXiv.2505.08762.

[29] B.M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J.R. Kitchin, D.S. Levine, K. Michel, A. Sriram, T. Cohen, A. Das, A. Rizvi, S.J. Sahoo, Z.W. Ulissi, C.L. Zitnick, UMA: A Family of Universal Models for Atoms, (2025). https://doi.org/10.48550/arXiv.2506.23971.

[30] T. Froitzheim, M. Müller, A. Hansen, S. Grimme, g-xTB: A General-Purpose Extended Tight-Binding Electronic Structure Method For the Elements H to Lr (Z=1–103), (2025). https://doi.org/10.26434/chemrxiv-2025-bjxvt.

[31] WATOC 2025, Home - WATOC 2025 - Oslo Kongressenter, Home (n.d.). https://www.watoc2025.no/gyroconference.eventsair.com (accessed July 27, 2025).

[32] S. Armaković, S.J. Armaković, Atomistica. online–web application for generating input files for ORCA molecular modelling package made with the Anvil platform, Mol. Simul. 49 (2023) 117–123. https://doi.org/10.1080/08927022.2022.2126865.

[33] S. Armaković, S.J. Armaković, Online and desktop graphical user interfaces for xtb programme from atomistica.online platform, Mol. Simul. 50 (2024) 560–570. https://doi.org/10.1080/08927022.2024.2329736.

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79

[34] R. Ramakrishnan, P.O. Dral, M. Rupp, O.A. von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Sci. Data 1 (2014) 140022. https://doi.org/10.1038/sdata.2014.22.

[35] P. Eastman, B.P. Pritchard, J.D. Chodera, T.E. Markland, Nutmeg and SPICE: Models and Data for Biomolecular Machine Learning, J. Chem. Theory Comput. 20 (2024) 8583–8593. https://doi.org/10.1021/acs.jctc.4c00794.

[36] P. Eastman, P.K. Behara, D.L. Dotson, R. Galvelis, J.E. Herr, J.T. Horton, Y. Mao, J.D. Chodera, B.P. Pritchard, Y. Wang, G.D. Fabritiis, T.E. Markland, SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials, (2022). https://doi.org/10.48550/arXiv.2209.10702.

[37] C. Devereux, J.S. Smith, K.K. Huddleston, K. Barros, R. Zubatyuk, O. Isayev, A.E. Roitberg, Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens, J. Chem. Theory Comput. 16 (2020) 4192–4202. https://doi.org/10.1021/acs.jctc.0c00121.

[38] M. Schreiner, A. Bhowmik, T. Vegge, J. Busk, O. Winther, Transition1x - a dataset for building generalizable reactive machine learning potentials, Sci. Data 9 (2022) 779. https://doi.org/10.1038/s41597-022-01870-w.

[39] ImageNet Challenge 2012 Analysis, (n.d.). https://www.image-net.org/challenges/LSVRC/2012/analysis/ (accessed July 27, 2025).

[40] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, Int. J. Comput. Vis. 115 (2015) 211–252. https://doi.org/10.1007/s11263-015-0816-y.

[41] K. Yang, K. Qinami, L. Fei-Fei, J. Deng, O. Russakovsky, Towards fairer datasets: filtering and balancing the distribution of the people subtree in the ImageNet hierarchy, in: Proc. 2020 Conf. Fairness Account. Transpar., Association for Computing Machinery, New York, NY, USA, 2020: pp. 547–558. https://doi.org/10.1145/3351095.3375709.

[42] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C.L. Zitnick, Z. Ulissi, Open Catalyst 2020 (OC20) Dataset and Community Challenges, ACS Catal. 11 (2021) 6059–6072. https://doi.org/10.1021/acscatal.0c04525.

[43] The Open DAC 2023 Dataset and Challenges for Sorbent Discovery in Direct Air Capture | ACS Central Science, (n.d.). https://pubs.acs.org/doi/10.1021/acscentsci.3c01629 (accessed July 27, 2025).

[44] L. Barroso-Luque, M. Shuaibi, X. Fu, B.M. Wood, M. Dzamba, M. Gao, A. Rizvi, C.L. Zitnick, Z.W. Ulissi, Open Materials 2024 (OMat24) Inorganic Materials Dataset and Models, (2024). https://doi.org/10.48550/arXiv.2410.12771.

[45] X. Fu, B.M. Wood, L. Barroso-Luque, D.S. Levine, M. Gao, M. Dzamba, C.L. Zitnick, Learning Smooth and Expressive Interatomic Potentials for Physical Property Prediction, (2025). https://doi.org/10.48550/arXiv.2502.12147.

[46] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions, Phys. Chem. Chem. Phys. 19 (2017) 32184–32215. https://doi.org/10.1039/C7CP04913G.

[47] T. Lu, A comprehensive electron wavefunction analysis toolbox for chemists, Multiwfn, J. Chem. Phys. 161 (2024) 082503. https://doi.org/10.1063/5.0216272.

[48] T. Lu, Q. Chen, van der Waals potential: an important complement to molecular electrostatic potential in studying intermolecular interactions, J. Mol. Model. 26 (2020) 315. https://doi.org/10.1007/s00894-020-04577-0.

[49] T. Lu, S. Manzetti, Wavefunction and reactivity study of benzo[a]pyrene diol epoxide and its enantiomeric forms, Struct. Chem. 25 (2014) 1521–1533. https://doi.org/10.1007/s11224-014-0430-6.

[50] T. Lu, F. Chen, Multiwfn: A multifunctional wavefunction analyzer, J. Comput. Chem. 33 (2012) 580–592. https://doi.org/10.1002/jcc.22885.

Citation: S. Armaković and S.J. Armaković, AIDASCO Reviews 3, 1 (2025) 21-35. DOI: https://doi.org/10.59783/aire.2025.79