Integrated Database of Force-Field Parameters, Experimental Measurements and Molecular Dynamics Simulations as a Tool for Force Field Development

Molecular Dynamics (MD) simulation is a crucial computational tool for investigating nucleic acids (RNA, DNA), proteins, and other (bio)molecular systems, generating large volumes of data daily. However, the interoperability and reusability of these datasets remain significant challenges.[1] We present IDA (Integrated DAtabase of force fields and datasets from experiments and MD simulations), an innovative platform designed to support a broad range of MD simulation types.[2] IDA accommodates outputs from standard simulations, importance sampling techniques, simulated annealing, and enhanced sampling methods such as replica-exchange approaches. Upon upload, datasets are converted into an internal, standardized format optimized for reuse, enabling users to perform diverse analyses and access results instantly. In addition to simulation data, IDA stores experimental datasets – primarily NMR data – and retrieves relevant information from public databases. A key feature of IDA is its integration of widely used force fields (FFs) in the Universal Molecular Force Field Format (UMFFF). This format includes dedicated files for each molecular type (e.g., RNA, DNA, proteins, solvents, counterions). FF libraries from major MD engines (e.g., AMBER, GROMACS) are automatically converted into UMFFF upon upload and can be exported back into their original formats, ensuring compatibility across platforms. This allows users to initiate new simulations in their preferred MD engine, compare uploaded FF libraries with existing ones, and identify FFs associated with submitted topologies. Unlike existing MD databases,[3] IDA is specifically designed to support force field development by interlinking simulation datasets, experimental data, and FF libraries. It offers a framework for evaluating the performance and accuracy of both established classical non-polarizable FFs and newly developed ones. This integrated environment enables data mining from large, heterogeneous datasets and enhances the predictive power and reliability of MD simulations. Ultimately, IDA aims to facilitate the parameterization of new force fields by providing advanced analytical tools that leverage reweighting techniques and machine learning – both of which rely on accurate, consistent, and well-structured data.

[1] Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R. A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurečka, P., Walter, N. G., Otyepka, M., Chem. Rev. 2018, 118, 4177. [2] Banáš, P., Mlýnský, V., Číž, D., Furmánek, R., Pilat, N., Pauw, V., Hachinger, S., Šponer, J., Martinovič, J., Otyepka, M., bioRxiv 2024 (https://doi.org/10.1101/2024.12.03.626554) [3] Beltrán, D., Hospital, A., Gelpí, J.L., Orozco, M., Nucleic Acids Res. 2024, 52(D1):D393–D403.

Primary author(s) :   Prof. BANÁŠ, Pavel (Czech Advanced Technology and Research Institute, CATRIN, Olomouc, Czech Republic); Dr. MLÝNSKÝ, Vojtěch (Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic); Dr. ČÍŽ, David (IT4Innovations, VSB – Technical University of Ostrava, Ostrava-Poruba, Czech Republic); Dr. FURMÁNEK, Radek (IT4Innovations, VSB – Technical University of Ostrava, Ostrava-Poruba, Czech Republic); Dr. PILAT, Nestor (IT4Innovations, VSB – Technical University of Ostrava, Ostrava-Poruba, Czech Republic); Mrs. ŠPAČKOVÁ, Anna (IT4Innovations, VSB – Technical University of Ostrava, Ostrava-Poruba, Czech Republic); Dr. PAUW, Viktoria (Leibniz-Rechenzentrum (LRZ), Garching bei München, Germany); Prof. HACHINGER, Stephan (Leibniz-Rechenzentrum (LRZ), Garching bei München, Germany); Prof. ŠPONER, Jiří (Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic); Prof. MARTINOVIČ, Jan (IT4Innovations, VSB – Technical University of Ostrava, Ostrava-Poruba, Czech Republic); Prof. OTYEPKA, Michal (Czech Advanced Technology and Research Institute, CATRIN, Olomouc, Czech Republic)

Presenter(s) : Dr. MLÝNSKÝ, Vojtěch (Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic)