[BIOINFO 2024]

Rapid and Sensitive Protein Complex Alignment with Foldseek-Multimer

Woosub Kim[1], Milot Mirdita[2], Eli Levy Karin[3], Cameron L.M. Gilchrist[2], Hugo Schweke[4], Johannes Söding[5,6], Emmanuel Levy[4,*], and Martin Steinegger[1,2,7,8*]

[1]*Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea.*
[2]*School of Biological Sciences, Seoul National University, Seoul, Republic of Korea.*
[3]*ELKMO, Copenhagen, Denmark.*
[4]*Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel.*
[5]*Quantitative and Computational Biology, Max-Planck Institute for Multidisciplinary Sciences, Göttingen, Germany.*
[6]*Campus Institute Data Science (CIDAS), University of Göttingen, Germany.*
[7]*Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Republic of Korea.*
[8]*Artificial Intelligence Institute, Seoul National University, Seoul, Republic of Korea.*
**Contact: emmanuel.levy@gmail.com, martin.steinegger@snu.ac.kr*

The ability to understand protein interactions is critical for unraveling the mechanisms behind their functions. Advances in computational structure prediction are vastly augmenting the hundreds of thousands of currently-available protein complex structures. Translating this avalanche of quaternary structures into discoveries requires efficient alignment and comparison, which is computationally prohibitive with state-of-the-art methods. To tackle this challenge, we present Foldseek-Multimer, a protein complex alignment method, that computes complex alignments from compatible chain-to-chain alignments, identified by efficiently clustering their superposition vectors. Foldseek-Multimer is 3-4 orders of magnitudes faster than the gold standard method US-align, while producing comparable alignments to it; allowing Foldseek-Multimer to compare billions of complex-pairs in a day. Foldseek-Multimer is free and open-source software available at github.com/steineggerlab/foldseek and is accompanied by a webserver available at search.foldseek.com and the BFMD database.