BhageerathH-Pro: Protein Tertiary Structure Prediction Server

Akshata Hegde[1], Smriti Pranjal[1], Devendra Prajapat[1], Shashank Shekhar[1] and B. Jayaram[1,2,3]

1– Supercomputing Facility for Bioinformatics & Computational Biology, 2 – Department of Chemistry, 3 – Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India - 110016

akshata@scfbio-iitd.res.in, bjayaram@scfbio-iitd.res.in

Key: *Auto:Y; CASP_serv:N; MSA:N; Fragm:Y.v; EMA:Y; MD:Y*

BhageerathH-Pro protein tertiary structure prediction server is based on an *ab initio*/homology hybrid methodology. It integrates several methodological innovations designed over the years in SCFBio, such as grid sampling, empirical energy-based scoring, physicochemical filters for screening decoys, RM2TS and NCL methodologies for structure generation, ProTSAV for structure selection, together with template-based homology modeling and molecular dynamics refinement.

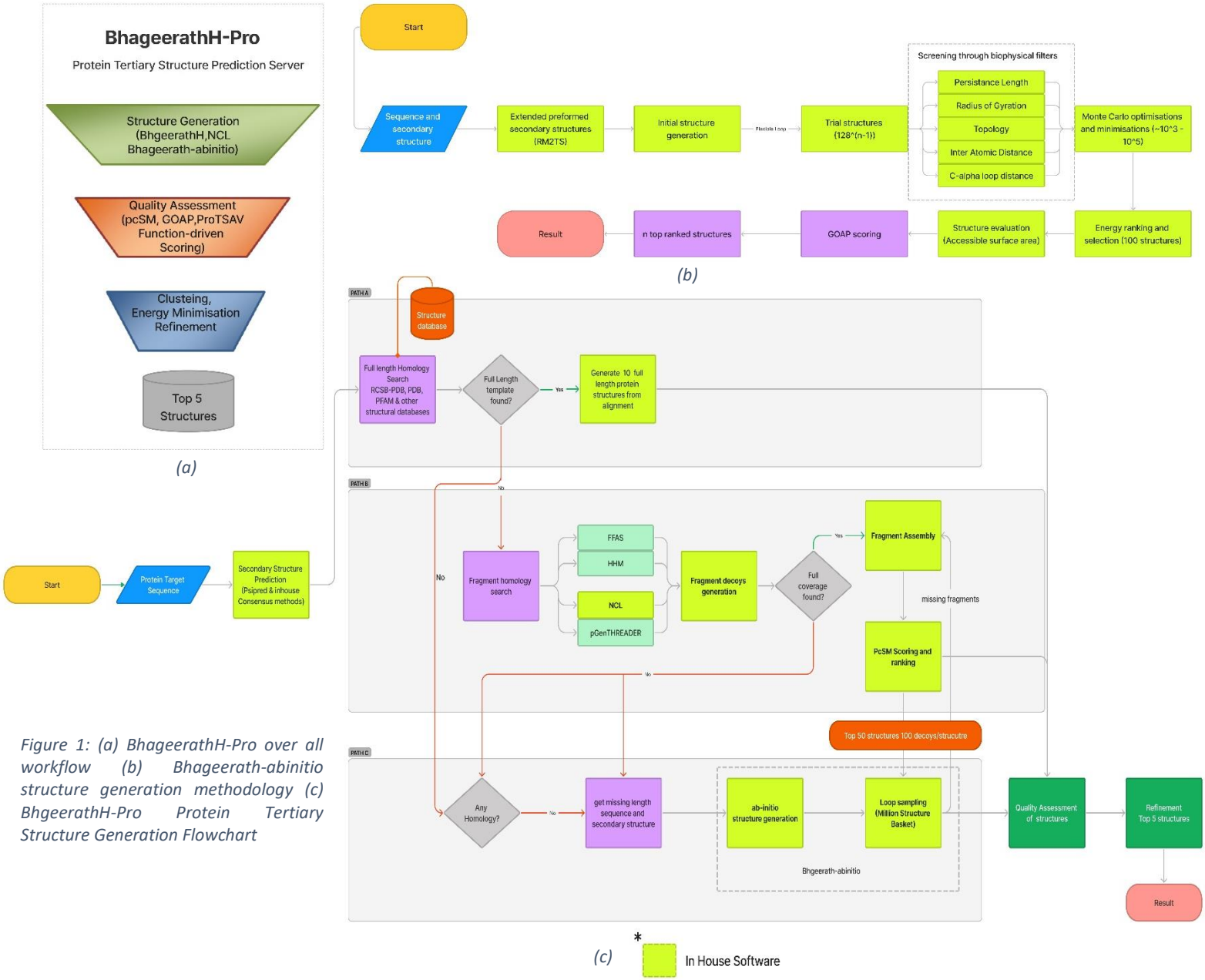

Figure 1: (a) BhageerathH-Pro over all workflow (b) Bhageerath-abinitio structure generation methodology (c) BhgeerathH-Pro Protein Tertiary Structure Generation Flowchart

Despite the now well-known successes of Alphafold2, we refrained from inserting alphafold2 codes into our methodology and intended to benchmark our server against released experimental structures, and Alphafold predicted structures. The expectation, as we analyze the shortcomings of our server, from CASP to CASP, is that someday soon, physics-based methods can yield reliable tertiary structures.

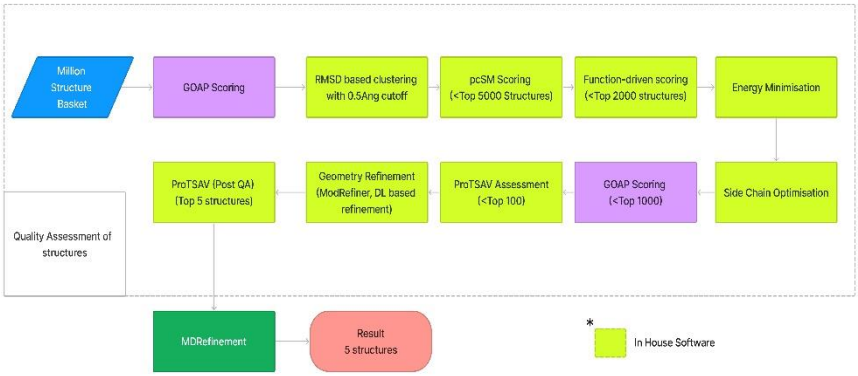

Figure 2: Quality Assessment pipeline for BhageerathH-Pro

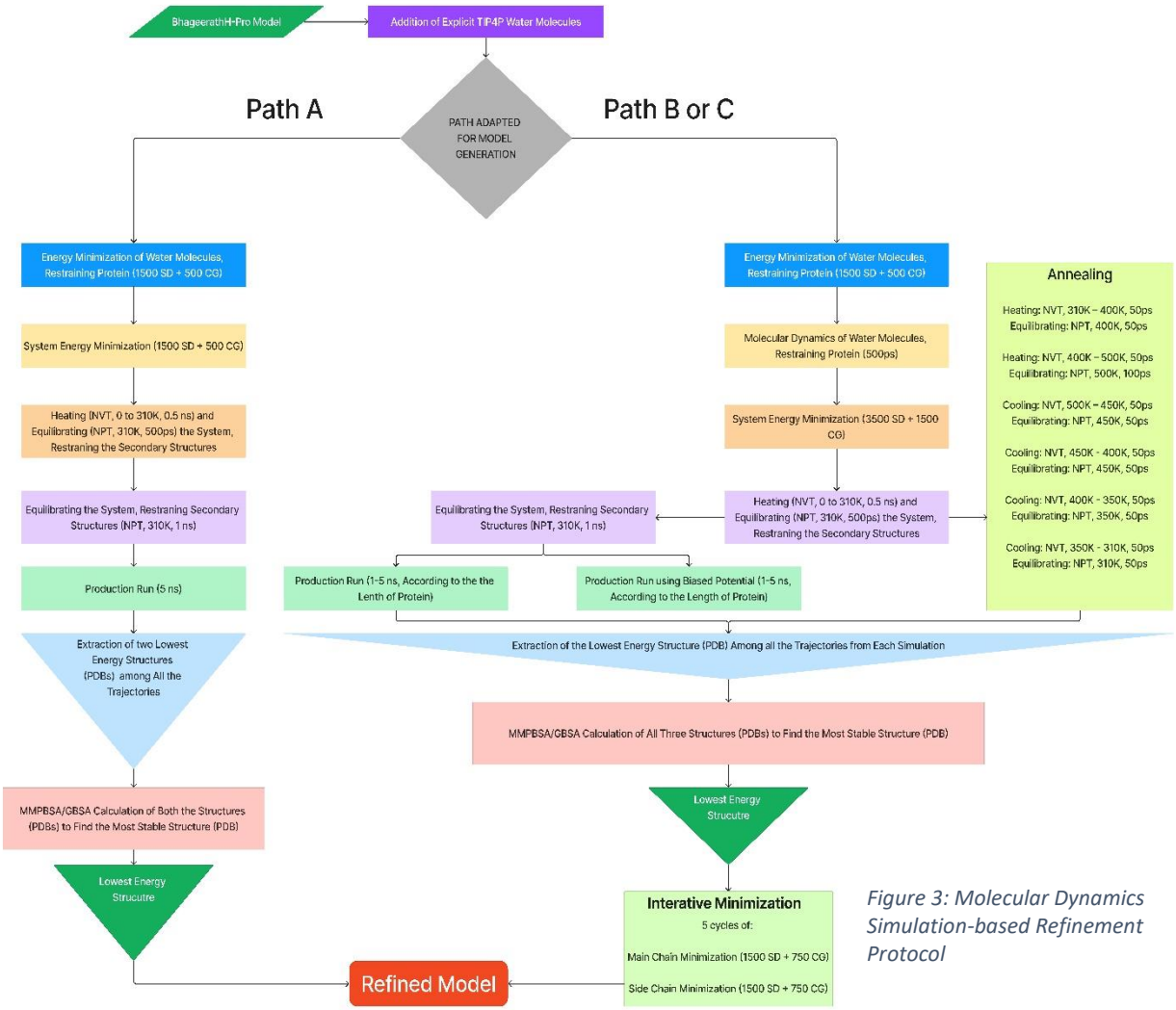

Figure 3: Molecular Dynamics Simulation-based Refinement Protocol

Methodology

BhageerathH-Pro is an advanced version of Bhageerath-H[1] which consists of three major steps of structure generation, quality assessment, and refinement respectively [Figure 1(a)]. The structure generator methodology follows three paths namely, Path A, Path B, and Path C [Figure 1(c)]. Path A caters to the full-length template modeling, while Path B includes the New Chemical Logic of amino acids driven protein alignment and decoys generation[2] and fragment assembly using the previously developed StrGen[3] algorithm. The Bhageerath-*ab initio* module [Figure 1(b)] is implemented as Path C for *ab initio* structure prediction in the absence of homology with an advanced version of RM2TS[4] and an updated smart Bhageerath[5], that works on loop sampling. The quality assessment pipeline is designed to pull out the best possible structure from the basket of almost a million conformations generated [Figure 2]. The sample conformations are clustered and filtered to retain mutually exclusive topologies, that are scored and ranked using iterative scoring modules like pcSM[6], Function-driven scoring[7], and ProTSAV[8] to extract the top 100 structures. The selected structures are energy minimized using PROSEE[9], and geometry optimized using the deep encoder-based refiner.

The top 5 models thus generated are subjected to molecular dynamics for final refinement. The models resulting from path A are refined using conventional molecular dynamics (5ns). [Figure:3 Path A], whereas the models obtained from path B or C are refined using Conventional MD, Accelerated MD, and Annealing [Figure:3 Path B or C]. Once the simulations are finished, the lowest energy conformer from each simulation is subjected to MMPBSA/GBSA calculations to find the most stable conformer, which is further refined using iterative main chain and side chain minimizations.

Results

The pipeline outlined above is fully automated and fielded in the recently concluded CASP15 under the TS category as the BhageerathH-Pro server. BhageerathH-Pro has performed reasonably well so far for the targets whose native structures are released in RCSB till now. Out of 7 targets (whose native information is released), BhageerathH-Pro succeeded in predicting 4 of them under low-resolution structures and 3 of them under 5Å of RMSD. Structures submitted from our server are benchmarked with the now famous Alphafold2 structures, and it is observed that for more than 50% of the targets (i.e., 49 out of 94), BhageerathH-Pro structures are within 3Å of RMSD. We have further noticed that mutational effects and functions are well distinguishable in BhageerathH-Pro structures. Most of BhageerathH-Pro modules and tools are freely available in the public domain for the user community.

Availability

BhageerathH-Pro is an open-source web server and is available at the SCFBio website: http://www.scfbio-iitd.res.in/bhageerathH+/.

Acknowledgments
This research is supported by the CoE project of the Department of Biotechnology and NSM project administered by Meity and CDAC, Govt. of India to SCFBio. The authors are thankful to Prof. Aditya Mittal for helpful discussions and suggestions.

1. Jayaram, B., Dhingra, P., Mishra, A. et al. Bhageerath-H: A homology/ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. BMC Bioinformatics,**2014**, **15**(Suppl 16), S7.
2. Rahul Kaushik, Ankita Singh and B. Jayaram. "Where informatics lags, chemistry leads," *Biochemistry*, **2018**, 55(5), 503-505.
3. Dhingra, P. and Jayaram, B. A homology/ab initio hybrid algorithm for sampling near-native protein conformations. J. Comput. Chem., **2013**, 34, 1925-36
4. Debarati DasGupta, Rahul Kaushik, and B.Jayaram "From Ramachandran Maps to Tertiary Structures of Proteins," *J. Phys. Chem. B*, **2015**. 119 (34), 11136 - 11145.
5. B.Jayaram, K.Bhushan, et al., "Bhageerath: An energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins" Nucl. Acids Res., **2006**, 34, 6195-6204.
6. Avinash Mishra *et al*, "Capturing Native/Native like Structures with a Physico-Chemical Metric (pcSM) in Protein Folding" BBA-Proteins and Proteomics.**2013**,1834, 1520-31.
7. P.Amita, *et al*; "Mask Blast with a new chemical logic of amino acids for improved protein function prediction" Proteins: Structure, Function, and Bioinformatics, **2021**,89(8),922-924.
8. Ankita Singh, Rahul Kaushik, Avinash Mishra, Asheesh Shanker, and B.Jayaram "ProTSAV: A Protein Tertiary Structure Analysis and Validation Server," BBA-Proteins and Proteomics, **2016**,1864(1),11–19.
9. P. Narang, K. Bhushan, S. Bose, and B. Jayaram, "Protein structure evaluation using an all-atom energy based empirical scoring function", J. Biomol. Str. Dyn., **2006**, 23, 385-406.