Accepted for publication in GTHTM-2024 conference proceedings, Editors: Dr. Anil Saxena & Dr. Sisir Nandi, Springer, 2024

A Hands-on Perspective on Physico-Chemical Versus AI/ML Methods along the Genome to Drug Pathway

B. Jayaram[0000-0002-5495-2213]

Supercomputing Facility for Bioinformatics & Computational Biology (SCFBio), IIT Delhi,

Department of Chemistry, IIT Delhi &

Kusuma School of Biological Sciences, IIT Delhi,

Hauz Khas, New Delhi -110016, India

Email: bjayaram@chemistry.iitd.ac.in; Website: www.scfbio-iitd.res.in

Abstract

Availability of genomic sequences in public domain has spurred intense algorithmic developments for identifying protein coding regions (gene finding) and for establishing their functions. This in turn accelerated protein structure prediction efforts, particularly relevant for proteins crucial to pathogens and those involved in disease conditions, to further enable structure-based drug design endeavours. This modern drug discovery pathway (Genome -> Gene -> Protein -> Drug) in our hands became a set of software suites collectively called '*Dhanvantari*' which embodies *Chemgenome*, *Bhageerath* and *Sanjeevini* web-suites to traverse through the genome to drug pathway with entry along any point. While developing the above science and software suites, our focus has been on energy, forcefield and molecular simulation based methods which are referred to here as physico-chemical methods. AI/ML methods have literally stormed in during the last few years into these research areas making it almost difficult to ignore their strengths. The result is the emergence of integrated AI/ML and physico-chemical methods for improved accuracies and greater success rates in new molecule predictions against drug targets. This brief report sketches how these methods have evolved in our hands to help accelerate drug discovery.

Keywords: Genome Annotation, Protein Structure Prediction (PSP), Computer Aided Drug Design (CADD), Energy Based Physico-Chemical Methods (PCM), Artificial Intelligence (AI) / Machine Learning (ML) Methods

1 The Genome to Drug Pathway

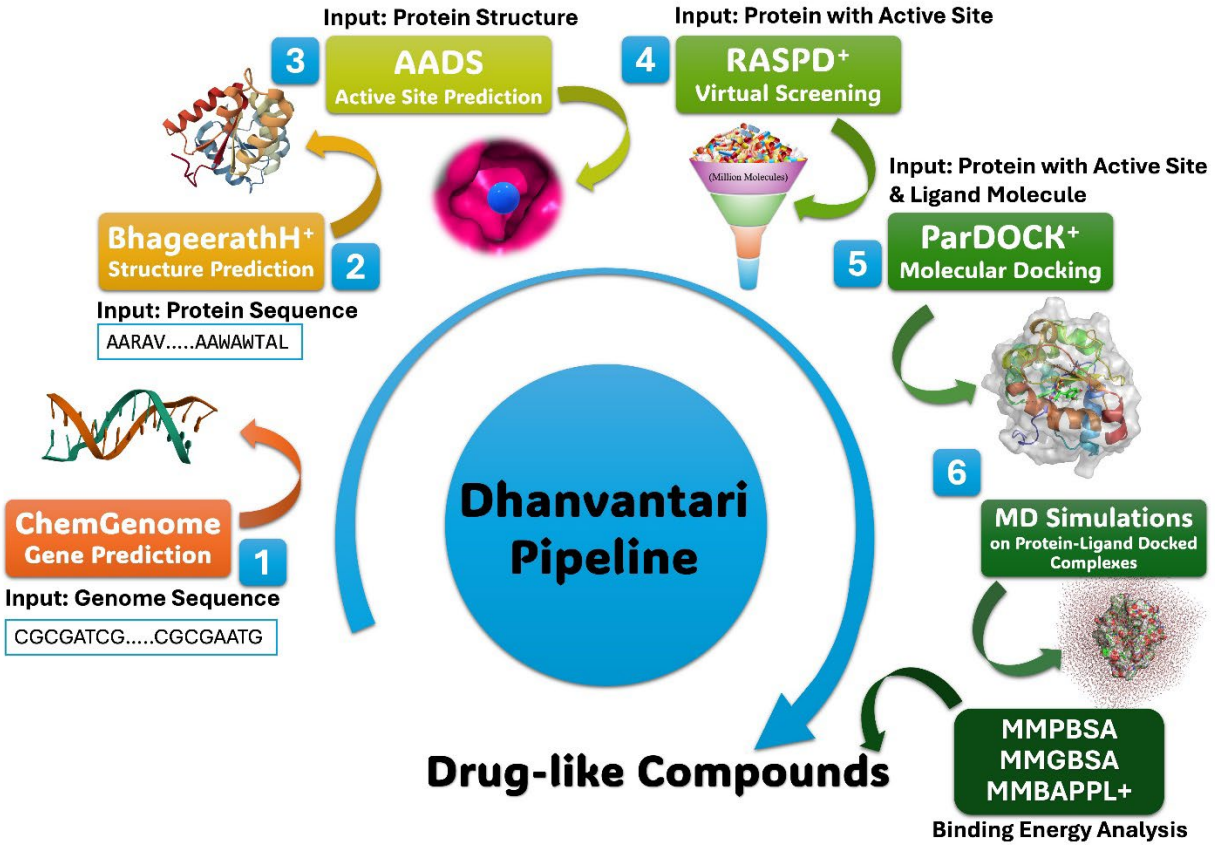


Fig. 1. *In silico* drug discovery assembly line developed at SCFBio. Step 1 is for genome annotation, finding protein coding genes here, step 2 for protein tertiary structure prediction, steps 3,4 and 5 which are part of *Sanjeevini* suite are for identifying lead compounds which are further processed through simulations and *post facto* binding free energy analyses to suggest candidate molecules for experimental testing. User can enter at any step in the above *Dhanvantari* pipeline.

Over the past score and more years, we were focused on developing physico-chemical methods from Genome to Drug discovery [1-3] with entry at any step along the pathway as shown in Fig. 1. While this work was finding strong roots, remarkable progresses were made in the applications of AI/ML in drug discovery in recent years. We trace the advancements in computer aided drug discovery with these distinct approaches in our Laboratory and how integration of these methods are yielding reliable predictions of drug-like candidate molecules for biomolecular targets.

2 Gene Prediction

During the first few years of the century, several methods were proposed for gene finding. GLIMMER, GENSCAN, GENESCAN, GenomeThreader, GRAIL, Augustus, are a few to name [4-9]. An exhaustive list of annotation software is available in Ref. 10. These are typically based on higher order Markov models, hidden Markov models, pattern recognition, discriminant analysis, neural networks, Fourier transforms and several other sophisticated statistical and mathematical techniques mostly in nucleotide sequence space. A common problem encountered with these early methods was the lack of universality and heavy dependence on sparse experimental datasets for training [11]. It was clear that we needed to go beyond sequences into property space for achieving universality. This led us to develop *Chemgenome* based on forcefields [12] and molecular dynamics simulation derived DNA energetics, focusing on universality [13,14]. The method was able to capture sensitivity and specificity upto 90% independent of the species. While this was the case during early part of the century, most of the above mentioned gene finding methods, augmented with newer ML techniques and expanded experimental training sets, now deliver > 95% sensitivity for any specific species, although universality is lacking. Our inspiration to continue with physico-chemical methods for deciphering the language of DNA and its apparent success hinges on two important discoveries. One was the discovery of the conjugate rule [15] and the second was the observation that structural and energetic features of DNA conveyed their functional destiny [16-19]. While physico-chemical methods (PCM) with both energetics and structural information superposed on sequence information, promise broader applicability, the accuracies are still hovering around 90%. Couple of areas where PCMs are leading are in promoter prediction, and in intron-exon boundary detection wherein sequence level promiscuity is high. The ChemEXIN [20] sketched in Fig. 2 shows unprecedented accuracies in intron-exon boundary detection. We expect that PCM and ML combinations will get us to a Genome Reader very soon.

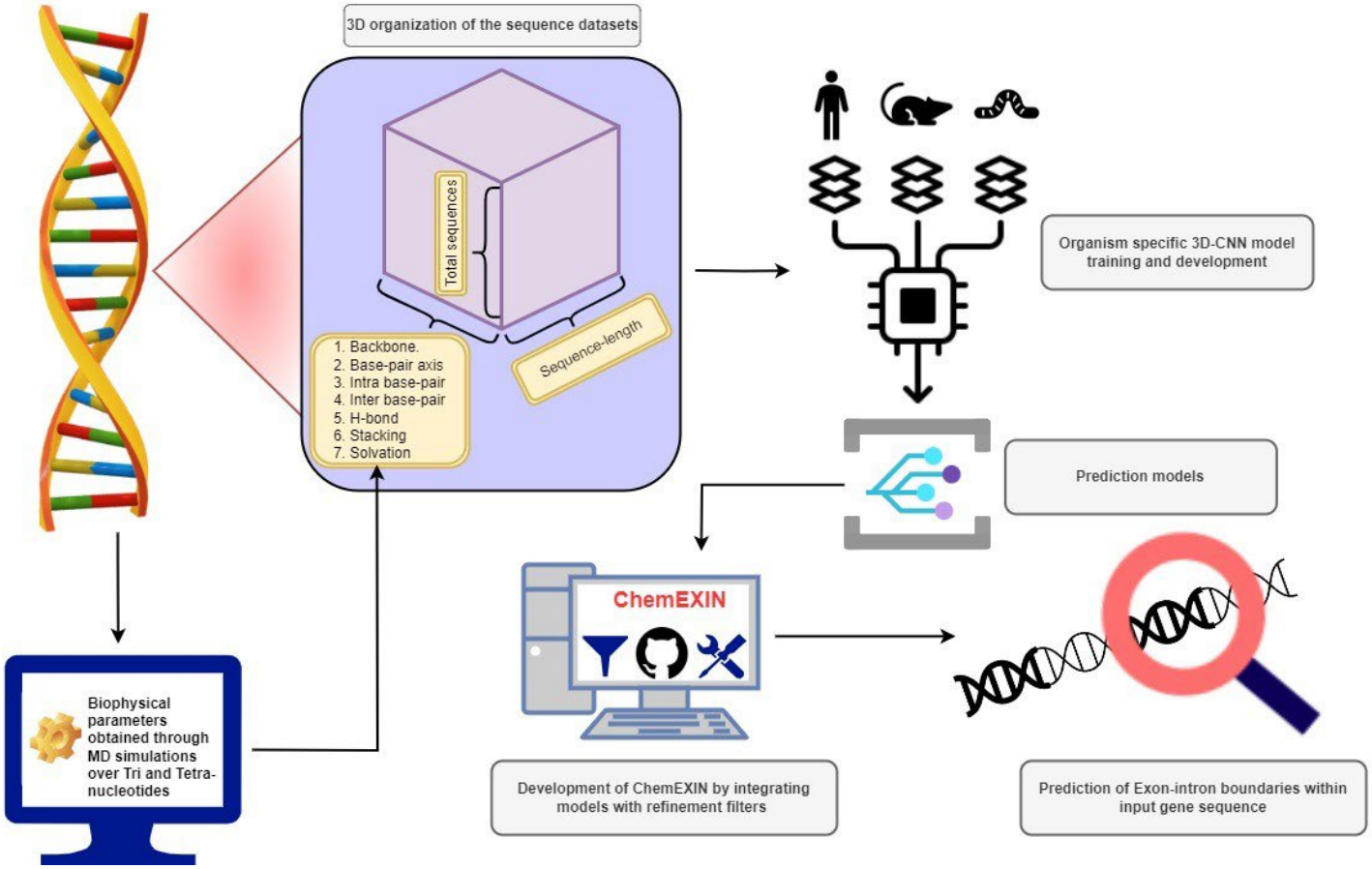
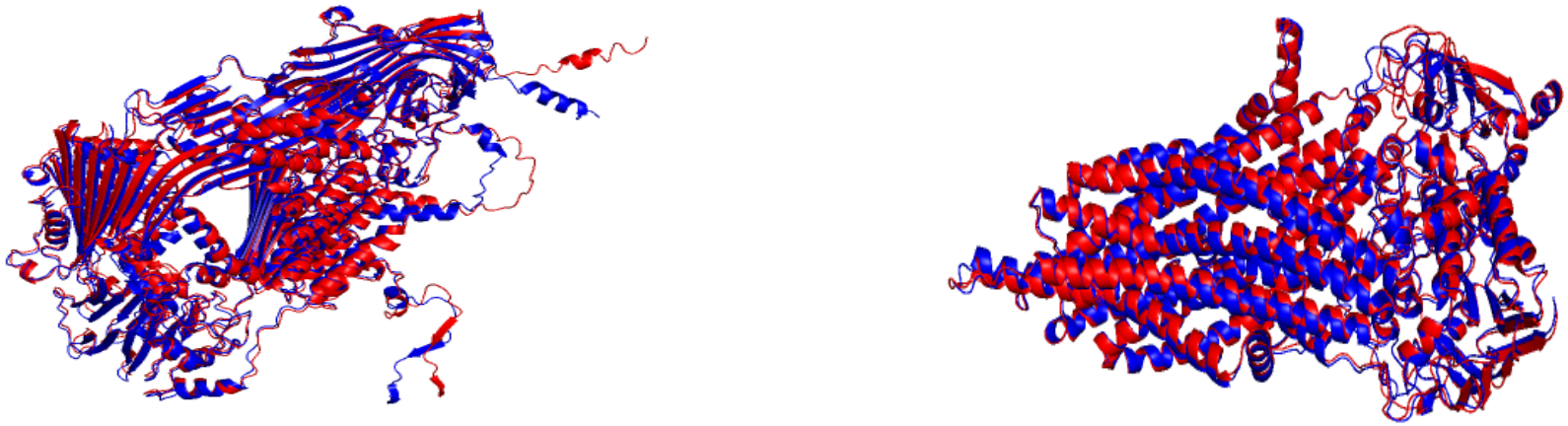


Fig. 2. An illustration of ChemExin, an intron-exon boundary detection scheme with physico-chemical approach supplemented with machine learning. The physico-chemical parameters obtained from MD simulations are used to train a 3D-Convolutional Neural Network (3D-CNN) model involving deep learning. This model outperforms the other state of the art tools across the three organisms studied. The parameter scores are: Sensitivity = 93.1, Specificity = 91.9 for *H.Sapiens*; Sensitivity = 80.4, Specificity = 79.4 for *M.Musculus*; and Sensitivity = 91.9 and Specificity = 92.0 for *C.elegans* [20].

3 Protein Structure Prediction

We have developed an all atom energy based methodology named *Bhageerath* for tertiary structure prediction of small proteins and subsequently extended it to larger proteins by combining homology methods (*BhageerathH*) [21-23]. More than 100 research groups have also made significant contributions to this area as chronicled in the biennial CASP experiments [24,25]. Some of the popular servers are Rosetta, Quark, I-Tasser, Multicom, Intfold, Distill, HHGG, Phyre, Floudas, Raptor, Pcons and so on. Full list is available at https://predictioncenter.org/. The overall accuracy, not long ago, however was around 30% in terms of less than 3 Å RMSD (root mean square deviation) of predictions in relation to crystal structures. This was the time when ANNs, although not fully ready for protein tertiary structure prediction (PSP), were touching near 80% accuracy in secondary structure prediction [26]. In came AI based Alphafold and its next version couple of years ago developed by DeepMind which took the PSP field by storm with near 100% accuracy for soluble proteins [27,28]. Interestingly, most successful PSP softwares participating in CASP15 experiment, assimilated Alphafold codes into their algorithms. *BhageerathH* however, fielded forcefield and homology based models and encouragingly enough, 49 out of 94 structures predicted by *BhageerathH* were within 3 Å of Alphafold structures (Fig. 3). Noting that the average size of the targets released in CASP15 was 478, the 52% accuracy by PCMs was good progress. The accuracies reported are underestimates since these are CASP targets with higher difficulty level in modelability [29]. Continuing with *Bhageerath*, we conceived of creating a computational PDB [30] and did so for Pvax in 2018 comprising over 2205 structures of soluble proteins and 3664 structures of soluble proteins of PvP01 in 2020 [31, 32]. Soon enough in 2022-23, Alphafold created tertiary structures of the entire uniport protein sequence database available from https://alphafold.ebi.ac.uk/. During our journey along the physico-chemical pathway, we were excited to discover that just four physico-chemical properties of amino acids not only explained the existence of the magic number 20 for naturally occurring amino acids, but also set very high water marks for homology modelling and function prediction [33, 34]. We also discovered that higher order Ramachandran maps carried sufficient information to build tertiary structures [35]. Further analyses of the configurational space of proteins suggested, when higher order spatial correlations of amino acid residue neighbours were included, that protein folding was a convergent problem as opposed to the then conventional wisdom [36]. Alphafold has now proved it. We were equally excited to discover that millions of proteins chronicled till date share a similar stoichiometry with very small standard deviations in their amino acid compositions which we called as the margin of life [37,38]. Equally interesting was the observation that there was a universality in the spatial distribution of the C-alpha atoms for any pair of amino acids defying the notion that preferential interactions among amino acid side chains dictated protein folding. This suggested to us a new view that inter-side chain interactions in proteins were offset by desolvation effects and folding thus was dictated by shape considerations of side chains along the sequence and exclusion by solvent water. Much remains to be done with PCMs in the PSP area to further understand the mechanism of folding at a molecular level. In the area of PSP, it is needless to say that AI inspired Alphafold and its variants, as of now, are leading the way.



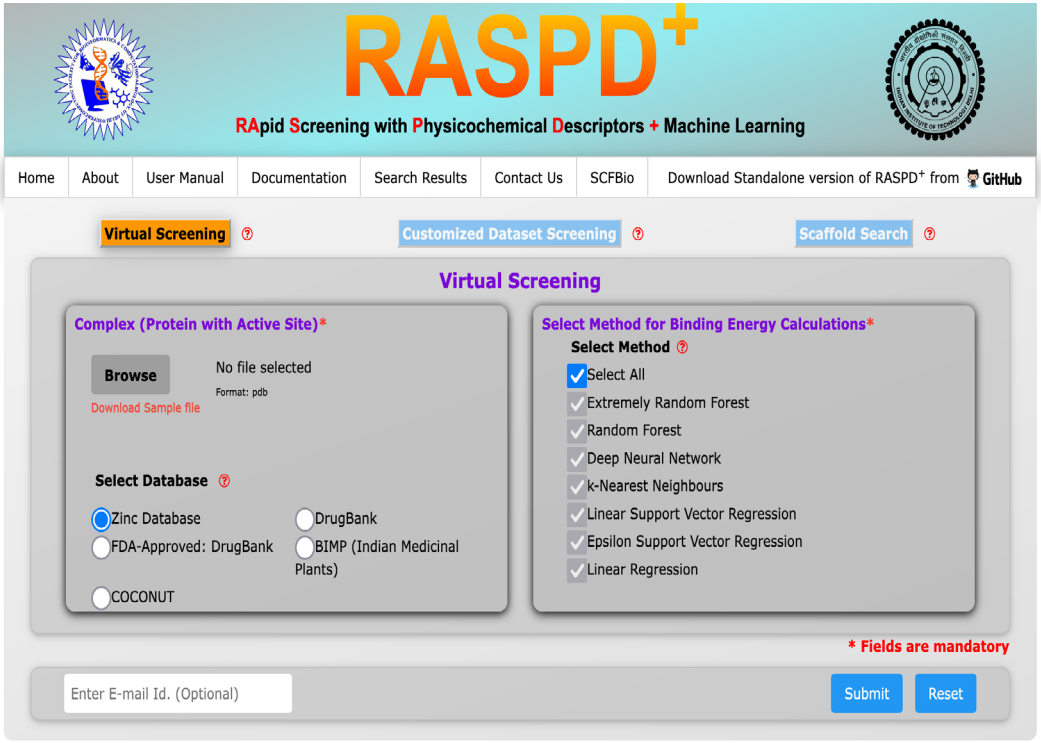
(A) (B)

Fig. 3. A superposition of the tertiary structures predicted by BhageerathH-Pro (blue) and Alphafold (red) for CASP15 targets: (A) Target T1191 with 1770 residues (RMSD: 0.97 Å) and (B) Target T1193 with 1297 residues (RMSD: 0.65 Å).

4 Drug Design

As a part of our efforts to develop a freely accessible, experimentally validated, complete drug discovery web-suite *Sanjeevini* [39-42] and particularly to facilitate a rapid screening of candidate molecules in the binding pocket of a target protein in search of good inhibitors, we developed an all atom binding energy based method for scoring named BAPPL [43] as an alternative to the conventional molecular free energy simulations [44,45]. Given the small size of experimental data sets then, and fewer groups in the field, BAPPL performed reasonably well. A decade later, data size increased, groups have increased and this has led us to revisit BAPPL, augment with ML methods (random forest in particular) and create BAPPL+ which seemed to outperform several other similar utilities [46]. We however, continue to recommend the performance of molecular dynamics (MD) simulations on the protein-ligand docked complexes followed by BAPPL+ analyses of the binding energies as ensemble averages over the MD trajectories for more reliable results [47]. The atomic level modelling of protein-ligand interactions is feasible if the number of candidates is less than 100 or 1000 given today's compute power. To scan large libraries such as a million compound library, against a target protein, in a matter of minutes, without actually docking, some novel strategies are required. We thus built a method (RASPD) based on a QSAR type equation in the space of physico-chemical properties lining the binding pocket of the target protein and the candidate molecules [48]. More recently, i.e. almost a decade later, we have introduced RASPD+ which incorporates several ML algorithms. eRF together with PCMs is seen to perform the best [49]. Fig. 4 illustrates the changing times for virtual screening. It is almost as if, integration of PCMs and MLs is the way to go for improved accuracies and especially

for accelerating new molecule discovery utilizing large libraries of small molecules. It may be noted that *Sanjeevini* protocols have already delivered experimentally validated low micromolar compounds against HAV [50], HBV [51,52]and CHIKV [53] infections, breast cancer [54], Alzheimer's [55] and malaria [56] and fungal infections [57].



Holderbach S, Adam L, Jayaram B, Wade RC, Mukherjee G. **RASPD+: Fast Protein-Ligand Binding Free Energy Prediction Using Simplified Physicochemical Features.** Front Mol Biosci. 2020 Dec 17;7:601065. PMID: 33392260; PMCID: PMC7773945. https://doi.org/10.3389/fmolb.2020.601065

Fig. 4. A screen shot of the front-end of the RASPD+ web-server developed at SCFBio for a rapid virtual screening of small molecule libraries against a protein target (a million compounds in minutes!). The methodology searches for complementarity in the physico-chemical properties of the active site residues of protein targets and small molecules. User has the option to choose the scoring function trained with several models such as extremely random forest, deep neural network, linear support vector regression etc. [48,49]. RASPD+ is freely accessible at http://www.scfbio-iitd.res.in/raspd+/.

5 Conclusions

Physico-chemical methods and molecular simulations have a long history in protein structure area and in computer aided drug discovery. They are relatively more recent in genome annotation. AI/ML methods have been the mainstay for genome annotation for more than two decades. However, these can significantly benefit with structural energetic feature inclusion as mentioned in section 2 above. Despite the emergence of Alphafold, work must go on with physico-chemical methods to arrive at unique native structures from sequences as there is no substitute to PCMs for new knowledge generation as adverted to briefly in section 3 above. There are many areas which have been impacted by AI/ML in CADD [58-63]. One area is toxicity. The chemistry and biology of toxicity remain a black box hindering the growth of PCMs although a few ideas and hypotheses prevail such as off target binding [64] and metabolism of xenobiotics [65] etc.. This is where AI/ML can play a tremendous role [66-68] in curtailing the failure rates of promising candidates bringing down the cost and time involved in drug discovery.

Acknowledgements

The author acknowledges the CoE support received from the Department of Biotechnology, Govt of India to SCFBio, IIT Delhi, and support from Ministry of Information Technology, Govt. of India and CDAC under NSM project to SCFBio, IIT Delhi. The author sincerely places on record his gratitude to all his students (http://www.scfbio-iitd.res.in/training/training.htm) who have contributed to the various stages of development of the *Dhanvantari* pathway over the years summarized in this article and also to his mentors and collaborators, Prof. David Beveridge, Prof. Barry Honig and Prof. Aditya Mittal, in particular, for making pursuit of science even more exciting. The author is also grateful to the Guest editors, Dr. A.K. Saxena and Dr. S. Nandi for giving us the opportunity to share our research findings with the readership of the GTHTM-2024 conference proceedings sponsored by Springer.

References

1. Bhat, R., Kaushik, R., Singh, A., DasGupta, D., Jayaraj, A., Soni, A., Shandilya, A., Shekhar, V., Shekhar, S., Jayaram, B.: " A comprehensive automated computer-aided discovery pipeline from genomes to hit molecules", *Chemical Engineering Science*, 222, 115711 (2020). https://doi.org/10.1016/j.ces.2020.115711
2. Soni, A., Menaria, K., Ray, P., Jayaram, B.: "Genomes to Hits in Silico - A Country Path Today, A Highway Tomorrow: A Case Study of Chikungunya", *Curr. Pharma. Des.*, 19, 4687-4700 (2013). http://www.eurekaselect.com/111973/article
3. Shenoy, S., Jayaram, B., Latha, N., Narang, P., et al.: "From gene to drug: A proof of concept for a plausible computational pathway", *Proceedings, IEEE Computer Society,* 1147-1152 (2006), 6th international conference on intelligent systems design and applications, Jinan, China, Oct. 16-18, 2006. https://dl.acm.org/doi/abs/10.1109/ISDA.2006.154
4. Salzberg, S., Delcher, A., Kasif, S., White, O.: "Microbial gene identification using interpolated Markov models", *Nucleic Acids Research*, 26(2), 544-548 (1998).
5. Burge, C., Karlin, S.: "Prediction of complete gene structures in human genomic DNA", *J. Mol. Biol.,* 268, 78-94 (1997).
6. Bhattacharya A, Bhattacharya S, Joshi A, Ramachandran S and Ramaswamy R.: "Identification of Parasitic Genes by Computational Methods", *Parasitol. Today,* 16, 127–130 (2000).
7. Gremme, G., Brendel, V., Sparks, M. E., Kurtz, S.: "Engineering a software tool for gene structure prediction in higher organisms", *Information and Software Technology*, 47(15), 965-978 (2005).
8. Uberbacher, E. C., Hyatt, D., Shah, M.: "GrailEXP and Genome Analysis Pipeline for Genome Annotation", *Current protocols in Human Genetics*, Wiley (2004). https://doi.org/10.1002/0471142905.hg0605s39
9. Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B.: "AUGUSTUS: a web server for gene finding in eukaryotes", *Nucleic Acids Research*, 32, W309–W312 (2004).
10. Wang, R., Bao, L., Liu, S., Liu, Z.: "Genome Annotation: Determination of the Coding Potential of the Genome", *Bioinformatics in Aquaculture, Principles and Methods*, Chapter 4, Book Editor(s):Zhanjiang (John) Liu, (2017), Wiley online library https://doi.org/10.1002/9781118782392.ch4.
11. Korf, I.: "Gene finding in novel genomes", *BMC Bioinformatics* 5, 59 (2004).
12. Dutta, S., Singhal, P., Agrawal, P., Tomer, R., Kritee, Khurana, E., Jayaram, B.: "A Physico-chemical model for analyzing DNA sequences", *J. Chem. Inf. Model., 46,* 78-85 (2006). http://pubs.acs.org/doi/abs/10.1021/ci050119x
13. Singhal, P., Jayaram, B., Dixit, S. B., Beveridge, D. L.: "Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations", *Biophys J.***,** *94,* 4173-4183 (2008). http://www.cell.com/biophysj/retrieve/pii/S0006349508700754
14. Khandelwal, G., Gupta, J., Jayaram, B.: "DNA energetics based analyses suggest additional genes in prokaryotes", *J. Bio Sc., 37,* 433-444 (2012). http://link.springer.com/article/10.1007%2Fs12038-012-9221-7
15. Jayaram, B.: "Beyond the wobble: the rule of conjugates", *J. Mol. Evol., 45,* 704-705 (1997).
16. Singh, A., Mishra, A., Khosravi, A., Khandelwal, G., Jayaram, B.: "Physico-chemical fingerprinting of RNA Genes", *Nucleic Acids Research*, 45 (7), e47 (2016). doi:10.1093/nar/gkw1236
17. Mishra, A., Siwach, P., Misra, P., Jayaram, B., Bansal, M., Olson, W. K., Thayer, K. M., Beveridge, D. L.: "Towards a universal structural and energetic model for prokaryotic promoters", *Biophysical Journal*, 115(7): 1180-1189 (2018). DOI:10.1016/j.bpj.2018.08.002
18. Mishra, A., Siwach, P., Misra, P., Dhiman, S., Pandey, A. K., Srivastava, P., Jayaram, B.: "Intron exon boundary junctions in human genome have in-built unique structural and energetic signals", *Nucleic Acids Research*, 49(5), 2674-2683 (2021). https://doi.org/10.1093/nar/gkab098

19. Sharma, D., Sharma, K., Mishra, A., Siwach, P., Mittal, A., Jayaram, B.: "Molecular Dynamics simulation based trinucleotide and tetranucleotide level structural and energy characterization of functional units of genomic DNA", *Physical Chemistry Chemical Physics (PCCP)*, 25, 7323-7337 (2023). DOI: https://doi.org/10.1039/D2CP04820E
20. Jayaram, B., Sharma, D., Aslam, D., Sharma, K., Mittal, A.: "Exon-Intron Boundary Detection Made Easy by Physicochemical Properties of DNA", manuscript submitted, (2024) https://www.researchsquare.com/article/rs-4359229/v1.
21. Narang, P., Bhushan, K., Bose, S., Jayaram, B.: "A computational pathway for bracketing native-like structures for small alpha helical globular proteins", *Phys. Chem. Chem. Phys., 7,* 2364-2375 (2005). http://pubs.rsc.org/en/content/articlelanding/2005/CP/b502226f
22. Jayaram, B., K. Bhushan, K., Shenoy, S., Narang, P., 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., 34,* 6195-6204 (2006). http://nar.oxfordjournals.org/content/34/21/6195.long
23. Jayaram, B., Dhingra, P., Mishra, A., Kaushik, R., Mukherjee, M., Singh, A., Shekhar, S.: "Bhageerath-H: A homology ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins", BMC *Bioinformatics*, 15(Suppl 16):S7 (2014). http://www.biomedcentral.com/1471-2105/15/S16/S7
24. Moult, J., Fidelis, K., Kryshtafovych, A., Tramontano, A.: "Critical assessment of methods of protein structure prediction (CASP)—round IX", *Proteins, Structure, Function and Bioinformatics*, 79, 1-5 (2011). https://doi.org/10.1002/prot.23200
25. Simpkin, A. J., Mesdaghi, S., Rodríguez, F. S., Elliott, L., Murphy, D. L., Kryshtafovych, A., Keegan, R. M., Rigden, D. J.: "Tertiary structure assessment at CASP15", *Proteins*, 91(12):1616-1635 (2023). https://doi.org/10.1002/prot.26593
26. McGuffin, L. J., Bryson, K., Jones, D. T.: "The PSIPRED protein structure prediction server", *Bioinformatics*, 16 (4), 404–405 (2000). https://doi.org/10.1093/bioinformatics/16.4.404
27. Jumper, J., Evans, R., Pritzel, A. et al.: "Highly accurate protein structure prediction with AlphaFold", *Nature,* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
28. Yang, Z., Zeng, X., Zhao, Y. et al.: "AlphaFold2 and its applications in the fields of biology and medicine", *Sig Transduct Target Ther*, 8, 115 (2023). https://doi.org/10.1038/s41392-023-01381-z
29. Kaushik, R., Jayaram, B.: "Structural difficulty index: A reliable measure for modelability of protein tertiary structures", *Protein Engineering Design and Selection*, 29(9), 391-97 (2016). doi: 10.1093/protein/gzw025.
30. Jayaram, B., Dhingra, P.: "Towards creating complete proteomic structural databases of whole organisms", *Current Bioinformatics*, *7*, 424-435 (2012). https://doi.org/10.2174/157489312803900992
31. Singh, A., Kaushik, R., Kuntal, H., Jayaram, B.: "PvaxDB: a comprehensive structural repository of Plasmodium vivax proteome", *Database*, 2018, bay021, (2018). doi:10.1093/database/bay021.
32. Singh, A., Kaushik, R., Chaurasia, D. K., Singh, M., Jayaram, B.: "PvP01-DB: computational structural and functional characterization of soluble proteome of PvP01 strain of Plasmodium vivax", *Database*, 2020, baaa036 (2020). https://doi.org/10.1093/database/baaa036
33. Kaushik, R., Singh, A., Jayaram, B.: "Where informatics lags chemistry leads", *Biochemistry*, 55(5): 503-505 (2018). DOI: 10.1021/acs.biochem.7b01073
34. Amita, P., Toran, R., Avinash, E., Jayaram, B.: "Mask Blast with a new chemical logic of amino acids for improved protein function prediction", *Proteins: Structure, Function and Bioinformatics* , 89 (8), 922-924 (2021). https://doi.org/10.1002/prot.26069.

35. DasGupta, D., Kaushik, R., and Jayaram, B.: "From Ramachandran Maps to Tertiary Structures of Proteins", *J. Phys. Chem. B* , 119 (34),11136 – 11145 (2015). DOI: 10.1021/acs.jpcb.5b02999
36. DasGupta, D., Kaushik, R., Jayaram, B.: "Protein folding is a convergent problem!", *Biochemical and Biophysical Research Communications*, 480 (4), 741-44 (2016). doi: http://dx.doi.org/10.1016/j.bbrc.2016.10.119.
37. Mittal, A., Jayaram, B., Shenoy, S. R., and Bawa, T. S.: "A Stoichiometry driven universal spatial organization of backbones of folded proteins: Are there Chargaff's rules for protein folding?", *J. Biomol. Struc. Dyn.*, *28*, 133-142 (2010). http://www.tandfonline.com/doi/abs/10.1080/07391102.2010.10507349
38. Mittal, A., Jayaram, B.: "Backbones of folded proteins reveal novel invariant amino-acid neighborhoods", *J. Biomol. Struc. Dyn.*, *28,* 443-454 (2011). http://www.tandfonline.com/doi/abs/10.1080/073911011010524954
39. Jayaram, B., Latha, N., Jain, T., Sharma, P., Gandhimathi, A., and Pandey, V. S.: "Sanjeevini: A comprehensive Active-Site directed lead design software", *Ind. J. Chem., 45A,* 1834-1837 (2006).http://www.niscair.res.in/sciencecommunication/researchjournals/rejour/ijca/ijca2k6/ijca_jan06.asp#p21
40. Jayaram, B., Singh, T., Mukherjee, G., Mathur, A., Shekhar, S., Shekhar, V.: "Sanjeevini: a freely accessible web-server for target directed lead molecule discovery", *BMC Bioinformatics*, *13,* S7 (2012). http://www.biomedcentral.com/1471-2105/13/S17/S7
41. Singh, A., Shekhar, S., Jayaram,B.: "CADD: Some success stories from Sanjeevini and the way forward", in Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design, Singh, S. K. (ed.), *Springer Nature*, Singapore , (2021). https://doi.org/10.1007/978-981-15-8936-2_1
42. Jayaram, B., Chaurasia, D. K., Anjum, R., Kesarwani, A., Shekhar, S., Mittal, A.: "*Sanjeevini* 3.0: A Comprehensive Automated Web-server for Computer Aided Drug Design", manuscript submitted (2024).
43. Jain, T., Jayaram, B.: "An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes", *FEBS Letters, 579,* 6659-6666 (2005). https://www.sciencedirect.com/science/article/pii/S0014579305012913
44. Kalra, P., Reddy, T. V., and Jayaram, B.: "Free energy component analysis for drug design: A case study of HIV-1 protease-inhibitor binding", *J. Med. Chem., 44,* 4325-4338 (2001).
45. Latha, N., Jain, T., Sharma, P., Jayaram, B.: "A free energy based computational pathway from chemical templates to lead compounds: a case study of COX-2 inhibitors", *J. Biomol. Struct. Dyn., 21,* 791-804 (2004).
46. Soni, A., Bhat, R., Jayaram, B.: "Improving the binding affinity estimations of protein-ligand complexes using machine-learning facilitated force field method", *J Comput Aided Mol Des*, 34, 817-830 (2020). https://doi.org/10.1007/s10822-020-00305-1
47. Bhat, R., Jayaraj, A., Soni, A., Jayaram, B.: "An overview of protein-ligand docking and scoring algorithms", in Protein Interactions: Computational Methods, Analysis and Applications, Gromiha, M. (Ed.), pg: 371-389 (2019), *World Scientific Publishing, Singapore*, https://doi.org/10.1142/11596
48. Mukherjee, G., Jayaram, B.: "A Rapid Identification of Hit Molecules for Target Proteins via Physico-Chemical Descriptors", *Phys. Chem. Chem. Phys.*, 15, 9107-16 (2013). http://pubs.rsc.org/en/Content/ArticleLanding/2013/CP/C3CP44697B
49. Holderbach, S., Adam, L., Jayaram, B., Wade, R. C., Mukherjee, G.: "RASPD+: Fast protein-ligand binding free energy prediction using simplified physicochemical features", *Front. Mol. Biosci.*, 7-2020 (2020). https://doi: 10.3389/fmolb.2020.601065
50. Banerjee K, Bhat, R., Rao, V. U. B., Nain, A., Rallapalli, K. L., Gangopadhyay, S., Singh, R. P., Banerjee, M., Jayaram, B.: "Towards development of generic inhibitors against the 3C proteases of picornaviruses", *The FEBS Journal*, 286(4):765-787 (2019).
51. Kiruthika, S., Bhat, R., Dash, R., Rathore, A. S., Perumal, V., Jayaram, B.: "A Novel Piperazine derivative that targets Hepatitis B Surface Antigen effectively inhibits Tenofovir Resistant Hepatitis B Virus", *Scientific Reports*, 11, 11723 (2021). https://doi.org/10.1038/s41598-021-91196-1.

52. Kiruthika, S.; Bhat, R., Jayaram, B., Perumal, V.: "A small molecule targeting Hepatitis B surface antigen inhibits clinically relevant drug-resistant hepatitis B virus", *Journal of Antimicrobial Chemotherapy*, 77(8), 2120–2124 (2022).
53. Tripathi, P.K., Soni, A., Yadav, S.P.S., Kumar A., Gaurav, N., Siva R.B., Sharma, P., Sunil., S., Ashish, Jayaram, B., Patel A. K.: "Evaluation of novobiocin and telmisartan for anti-CHIKV activity", *Virology*, 548, 250-260 (2020). doi.org/10.1016/j.virol.2020.05.010
54. Bhatnagar, S., Soni, A., Kaushik, S., Rikhi, M., Kumar, T.R.S., Jayaram, B.: "Non-steroidal estrogen receptor iso form selective biphenyls", *Chemical Biology & Drug Design,* 91(2), 620-630 (2018).
55. Kumar, J., Gill, A., Shaikh, M., Singh, A., Shandilya, A., Jameel, E., Sharma, N., Mrinal, N., Hoda, N., Jayaram, B.: "Pyrimidine-Triazolopyrimidine and Pyrimidine-Pyridine Hybrids as Potential Acetylcholinesterase Inhibitors for Alzheimer's Disease", *Chemistry Select: Medicinal Chemistry & Drug Discovery*, 3(2):736-747 (2018).
56. Shandilya, A., Hoda, N., Khan, S., Jameel, E., Kumar, J., Jayaram, B.: "De novo lead optimization of triazine derivatives identifies potent antimalarial", *J. Mol. Graphics & Modelling,* (71), 96-103 (2017).
57. Nishikawa, J.L. et al.: "Inhibiting fungal multidrug resistance by disrupting an activator-mediator interaction", *Nature*, 530, 485-489 (2016).
58. Mak, K.K., Wong, Y. H., Pichika, M.R.: "Artificial Intelligence in Drug Discovery and Development", In: Hock, F.J., Pugsley, M.K. (eds) Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays. Springer, Cham. (2023) https://doi.org/10.1007/978-3-030-73317-9_92-1
59. Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, A., Garcia-Fandino, R.: "The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies", *Pharmaceuticals*, *16*(6), 891 (2023); https://doi.org/10.3390/ph16060891
60. Walters, W. P., Barzilay, R.: "Critical assessment of AI in drug discovery", *Expert Opinion on Drug Discovery*, *16*(9), 937–947 (2021). https://doi.org/10.1080/17460441.2021.1915982
61. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., Tekade, R.K.: "Artificial intelligence in drug discovery and development", *Drug Discov Today*., 26(1), 80-93 (2021). doi: 10.1016/j.drudis.2020.10.010.
62. Dara, S, Dhamercherla, S., Jadav, S.S., Babu, C.M., Ahsan, M.J..: "Machine Learning in Drug Discovery: A Review", *Artif Intell Rev.*, 55(3):1947-1999 (2022). doi: 10.1007/s10462-021-10058-4.
63. Abbas, M., Rassam, A., Abunora, R., Karamshahi, F., Abouseada, M.: "The Role of AI in Drug Discovery". *Chembiochem*. May 12:e202300816. (2024) doi: 10.1002/cbic.202300816.
64. Shaikh, S., Jain. T., Sandhu, G., Latha, N., Jayaram., B.: "*A physico-chemical pathway from targets to leads"*, *Current Pharmaceutical Design*, 13, 3454-3470 (2007).
65. Mukherjee, G., Gupta, P. L., Jayaram, B.: "Predicting the binding modes and sites of metabolism of xenobiotics", *Molecular BioSystems*, 11, 1914 – 1924 (2015). DOI:10.1039/C5MB00118H
66. Sharma, A.K., Srivastava, G. N., Roy, A., Sharma, V. K.: "ToxiM: A Toxicity Prediction Tool for Small Molecules Developed Using Machine Learning and Chemoinformatics Approaches", *Frontiers Pharmacology*, 8 (2017). https://doi.org/10.3389/fphar.2017.00880.
67. Tran, T. T. V., Wibowo, A. S., Tayara, H., Chong, K. T. : "Artificial Intelligence in Drug Toxicity Prediction: Recent Advances, Challenges, and Future Perspectives", *J. Chem. Inf. Model.*, 63 (9), 2628−2643 (2023), DOI: 10.1021/acs.jcim.3c00200
68. Tonoyan, L., Siraki, A. G.: "Machine learning in toxicological sciences: opportunities for assessing drug toxicity", *Front. Drug Discov*., 4 (2024), doi.org/10.3389/fddsv.2024.1336025