Behind the Cover

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

Outline

Revolutionising Biology with Artificial Intelligence - 1

Suryadip Sarkar (IBAB Bengaluru)
Tathagata Chatterjee (RKMVERI Kolkata)

Suryadip and Tathagata offer an engaging exploration of how AI is transforming biological sciences. From the early days of rule-based AI to cutting-edge deep learning applications like AlphaFold, this article traces the evolution of AI's role in biology. It highlights AI's power in solving complex problems such as protein folding, drug discovery, and genomics. With groundbreaking examples like DeepVariant and DrugGPT, readers will gain insight into AI's monumental impact on research and healthcare. Dive into the future of biology where data meets innovation!

Revolutionising Biology with Artificial Intelligence – 1

Suryadip Sarkar (IBAB Bengaluru), Tathagata Chatterjee (RKMVERI Kolkata)

Artificial Intelligence (AI), the talk of the town right now, has revolutionised various fields and the life sciences is no exception. Before we go further into how exactly AI is changing the course of life as we speak, let us get up to speed with a few concepts.

Marvin Minsky defined artificial intelligence as "**the science of making machines do things that would require intelligence if done by men**". Ideally one would apply AI for problems where it is impossible to define consolidated rules. A very simple example could be the spam filters in our emails. It is virtually impossible for programmers to identify a set of rules to identify all types of spam. This is because these messages constantly keep evolving with newer wordings and patterns. Therefore AI models are trained using examples of both spam and legitimate emails to help identify subtle patterns and adapt to the new spam techniques over time. This is exactly why biology is such a beautiful candidate for AI applications, owing to its inherent complexity.

Machine Learning (ML) is a subset of AI that focuses on the development of algorithms that allow computers to learn from and make predictions based on data. Deep Learning (DL) is a further subset of ML that utilises neural networks (inspired from, but in no way similar to biological neurons) with many layers (hence "deep") to analyse various factors of data (especially non-linearity). These two branches of AI are the most prevalently used in biology to date.

Early Days: From Symbolic AI to ML and DL

It all dates back to the 1950s, when Alan Turing proposed the concept of machines being able to simulate any form of human reasoning through algorithmic approaches [1]. As a result, during the early years, AI was largely rule-based (aka Symbolic AI) which was based on logical representations of the world. This led to the birth of expert systems, which used knowledge bases of rules to solve specific problems. One such expert system was called MYCIN (1978) (Fig 1) [2]. It was developed by Edward Shortliffe as part of his doctoral dissertation, under the guidance of Bruce G. Buchanan, Stanley N. Cohen, and others at Stanford University. MYCIN was used to identify the bacteria causing severe bacterial infections like meningitis and bacteremia and to subsequently recommend appropriate antibiotics and their dosages according to the patient's body weights.

Fig 1: This is a really long caption which should demonstrate how lines break in captions.

The advent of the internet and a substantial increase in computational power led to the growth of enormous volumes of data. This led to the next big break in the early 2000s in the form of Machine Learning algorithms. These algorithms were able to learn from data and perform tasks such as predicting outcomes, classifying objects or clustering them based on similarity, as opposed to following predefined rules. Further, the introduction of DL opened even more avenues and a whole new range of applications in tasks such as image processing, signal processing, natural language processing etc. Nowadays, AI has extended its branches into any and all fields of human endeavour with an especially serial impact on healthcare, medicine, and biological research.

AI and the era of Structural Biology

Applications of AI for solving complex problems in the life sciences can be dated back to the late 1990s. This was the era of structural biology. Proteins are the molecular machines that carry out all biological processes in a cell, and their structure dictates their function [3] Therefore, scientists were devoted to finding solutions to identify and manipulate said protein structures.

Proteins are unbranched polymers constructed from 22 standard amino acids. They have four levels of structural organisation (primary, secondary, tertiary and quaternary). Primary structure refers to the amino acid sequence that is specified by the genetic information contained within the DNA. As the polypeptide folds, it forms certain localised arrangements of adjacent amino acids that constitute of the secondary structures (mainly in the form of α-helices and **β**-sheets). Tertiary structure refers to the three-dimensional shape of a protein formed by the overall folding of the polypeptide chain onto itself. It results from interactions among the various R groups (side chains) of the amino acids, including hydrogen bonds, ionic bonds, van der Waals forces, and disulfide bridges. A protein quaternary structure

arises when two or more polypeptide chains (subunits) come together to form a larger functional protein complex. The arrangement and interaction of these subunits are crucial for the protein's function. Together, these levels of structure are essential for a protein's biological activity and functionality (Fig 2) [4].

During this time, X-Ray Crystallography and Nuclear Magnetic Resonance Spectroscopy (NMR-Spec) were heavily being used to determine the complicated 3D structures of proteins. Due to the complexity of the protocols associated with these techniques coupled with low success rate, scientists started to look for more feasible solutions. In 1990, the labs of Ross King (Chalmers University of Technology, Sweden) and Michael Sternberg (Imperial College London) came up with an ML program called PROMIS (PROtein Machine Induction System) that could generalise rules to characterise the relationship between primary and secondary structures of globular proteins. These rules could further be used to predict unknown secondary structures from a known primary structure of a protein [5]. Quantitative Structure Activity Relationship Modeling (QSAR) is an extensively used approach in drug design and discovery that devises mathematical models connecting the biological activity of these drugs to their complex chemical features [6]. In 1994, the pair of King and Sternberg modelled the QSAR of a series of drugs using a technique called Inductive Logic Programming (ILP) [7]. ILP is a machine learning technique that uses a combination of logic programming and data driven decision making [8]. With the application of ILP, they were able to elucidate complex and interpretable patterns within the data that might have originally been overlooked while investigating using traditional statistical approaches.

Further, in the year 2000, the lab of R. Casadio used a class of artificial neural networks known as a Feed Forward Network (FFN), to predict protein folding and structure from only their corresponding amino acid sequence [9]. A feed forward network is basically composed of several layers of interconnected computational units called neurons (inspired from biological neurons) (Fig 3). Each neuron in a layer has a weight factor (think of it as a scalar value that denotes the neuron's contribution to the output) associated with it and receives numerical inputs from other neurons in the previous layer, processes them by applying a weighted sum and an activation function, and passes the result to the neurons in the next layer. This allows the network to model complex, non-linear patterns [10]. Therefore, hypothetically speaking, the non-linear relationship between a protein's primary amino acid sequence and its folded 3D structure can be modelled efficiently using this FFN. By capturing these subtle relationships, FFNs provided insights into how specific sequence motifs correspond to structural motifs underscoring the massive applications of AI in the field of structural biology.

Genomics and the Rise of AI The Human Genome Project (HGP) was an international research initiative with the goal of sequencing the entire human genome. Launched in 1990, the HGP was completed in 2003 with the submission of the 1st draftt of the human genome. Its goals included identifying all human genes, determining their sequences, and exploring the functions of these genes. The project has had profound implications for medicine, genetics, and our understanding of human biology. Following that, in the early 2000s, vast amounts of genomic data were generated which warranted the need for advanced analytical methods. Michael Schatz and his colleagues applied ML techniques to genome assembly, enhancing the efficiency of sequence alignment [11]. In 2004, David Heckerman demonstrated the use of Bayesian Networks, a type of probabilistic graphical model, to predict gene functions from microarray data [12]. Bayesian networks as the name suggests were based on the concept of Bayesian statistics that provides a way to update the probability of a hypothesis as new evidence becomes available, forming the foundation of Bayesian inference [13]. Heckerman used microarray data that contains information about the activity level (how active or inactive the gene is) of thousands of genes across specific conditions. In his approach, variables such as the genes were represented as nodes and the probabilistic dependencies between them as edges, making them particularly effective for capturing complex, uncertain relationships (Fig 4). This enabled Heckerman to identify the complex crosstalks between genes and their functional relationships to each other based on expression patterns.

Through these networks, it became possible to model how changes in the expression of one gene influence others, uncovering hidden dependencies and regulatory mechanisms [12]. In 2016, Angermueller's lab utilised Convolutional Neural Networks (CNN), which were traditionally used to process image data, on DNA sequences to predict gene expression, with accuracies comparable to traditional methods [14]. CNNs are comprised of layers of neurons that utilise convolutional filters to capture local patterns and hierarchical features. A convolutional filter is a matrix of weights that is slid along the entire sequence effectively learning hidden patterns and features from them, much like detecting edges or textures within an image which is nothing but a sequence of pixel values. Moreover, since the same filter is being slid along the entire sequence, it can capture both local and global dependencies among these sequences potentially identifying important motifs such as transcription factor binding sites [15].

Most recently in 2018, Google developed a DL based tool called the DeepVariant that employs a deep CNN architecture to identify variants in DNA sequences, outperforming traditional state of the art tools and pipelines such as GATK and Augustus [17] (Fig 6).

AI in Biomedical Imaging, Diagnostics and Drug Discovery CNNs being at the forefront of computer vision at the time also inspired biologists to actually use them for biomedical image analysis and predictive modelling based on biomedical image data. In 2017, Andre Esteva and his colleagues developed a CNN model that could

Fig 2: This is a really long caption which should demonstrate how lines break in captions.

diagnose and differentiate skin cancer from other standard skin lesions with accuracy comparable to dermatologists by analysing dermoscopic images [18]. U-Net, a CNN architecture developed by Olaf Ronneberger, is widely used nowadays for segmenting biological images, especially in the field of histopathology which is the study of tissue changes and diseases at the microscopic level, often involving the examination of biopsies to diagnose conditions (see Fig. 5). It has shown remarkable performance in analysing histological slides accurately diagnosing cancerous tissues, thereby improving patient outcomes [19].

ML has contributed significantly towards the drug discovery space as well, including applications in drug target identification, biomarker discovery, QSAR modelling and predicting efficacy of drug candidates, thereby accelerating the drug development process. Biomarker discovery—the identification of measurable indicators such as genes, proteins, or metabolites linked to specific diseases—has benefited from ML's ability to analyze complex datasets and uncover subtle patterns, aiding in personalized medicine and early diagnosis [20]. Yuesen Li and his colleagues in 2023 developed DrugGPT focusing on chemical space exploration of protein ligand complexes, which refers to navigating the enormously complex universe of chemical compounds to identify potential drug molecules targeting specific proteins[21].

A Leap Forward with AlphaFold In the realm of structural biology, we have come a long way as well. In the year 2021, John Jumper and his team at Google Deepmind rocked the world of biology with their groundbreaking model called AlphaFold. AlphaFold redefined the age-old problem of predicting protein 3D structures from just their amino acid sequence information. It uses a pseudo-bayesian genetic algorithm based deep neural network to model the physical and geometric properties of proteins from just their amino acid sequence and gives highly accurate 3D structures of proteins solving the bottleneck of time, complexity and low success rate associated with methods like X-ray crystallography, NMR-spec and Cryo Electron microscopy (Cryo-EM) [22]. In 2024, Jumper was awarded the Nobel Prize in Chemistry for AlphaFold emphasizing the importance and relevance of AI in biology even further.

Biology Inspiring AI models Interestingly enough, the relationship between AI and biology has been quite symbiotic. It is not only biology that has gained from AI models, biology has also inspired AI models on more than one occasion. The most prevalent of them all is actually that of the neural networks (the building blocks of DL). In 1943, Warren McCulloch and Walter Pitts built a mathematical model of the functioning of a single neuron [23]. Taking inspiration from human perception, Frank Rosenbahlt modelled the first neural network in 1957 that was able to recognize some handwritten symbols effectively and could model basic logical operations such as AND and OR gates. This was known as the Perceptron and it would lay the foundation for further development of DL and neural network architectures [24].

Our immune system only allows lymphocytes, that recognize certain antigens, to be cloned and proliferated with such identical antigenic receptors. This phenomenon known as clonal selection presents itself as a learning problem driven by context (antigen) and an appropriate response. This led to the invention of the Artificial Immune System, a class of computationally intelligent, rule-based machine learning systems, inspired by the intricacies of the immune system of vertebrates [25]. Akin to their living counterparts these models excel in recognizing something presented to it as an "antigen" and take the most appropriate response making them ideal for applications such as antiviruses and anti-spams.

Towards the Future The synergy between AI and biology has set the stage for groundbreaking discoveries and innovations. From protein structure prediction with AlphaFold to the use of neural networks in genomics and drug discovery, biology has been transformed into a data-driven, predictive and multi-disciplinary branch of science. Whether it's predicting disease, designing drugs, or deciphering the mysteries of life at the molecular level, AI is now an indispensable tool in every biologist's toolkit. So, if you're curious, now's the time to dive into this captivating intersection of science and technology—where the future of biology is being driven by AI.

Prof. Sengupta is a strong proponent of interdisciplinary and collaborative research. Her group at IISER Kolkata (mCED) probes molecular machinations of diverse biological systems with physics and data based methods.

Discussions with mCED research group members, and critical feedback from Dr. Kumar Vanka (CSIR-NCL, Pune) is acknowledged. Abhirup Mukherjee and Pousukhi Bagchi are credited with Figs 1-4, and with Fig 5, respectively.

References

1. Akman, V., & Blackburn, P. (2000). Alan Turing and artificial intelligence. Journal of Logic, Language, and Information, 391-395.
2. Shortliffe E. H. (1977). Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases. Proceedings of the Annual Symposium on Computer Application in Medical Care, 66–69.
3. Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Analyzing Protein Structure and Function
4. The Protein Folding Problem: A Structural Perspective, S. J. Baker et al. Nature Reviews Molecular Cell Biology 2006
5. King, R. D., & Sternberg, M. J. (1990). Machine learning approach for the prediction of protein secondary structure. Journal of molecular biology, 216(2), 441–457.
6. Tandon, H., Chakraborty, T., & Suhag, V. (2019). A concise review on the significance of QSAR in drug design. Chemical and Biomolecular Engineering, 4(4), 45-51.
7. Sternberg, M. J., King, R. D., Lewis, R. A., & Muggleton, S. (1994). Application of machine learning to structural molecular biology. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 344(1310), 365-371.
8. Cropper, A., & Dumančić, S. (2022). Inductive logic programming at 30: a new introduction. Journal of Artificial Intelligence Research, 74, 765-850.
9. Casadio, R., Compiani, M., Fariselli, P., Jacoboni, I., & Martelli, P. L. (2000). Neural networks predict protein folding and structure: artificial intelligence faces biomolecular complexity. SAR and QSAR in environmental research, 11(2), 149–182.
10. Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. Chemometrics and intelligent laboratory systems, 39(1), 43-62.
11. Initial Sequencing and Analysis of the Human Genome Authors: International Human Genome Sequencing Consortium :Nature 2001
12. Schatz, M. C., Delcher, A. L., & Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. Genome research, 20(9), 1165–1173.
13. Heckerman, D. (1998). A Tutorial on Learning with Bayesian Networks. In: Jordan, M.I. (eds) Learning in Graphical Models. NATO ASI Series, vol 89. Springer, Dordrecht.
14. Stephenson, Todd Andrew. An introduction to Bayesian network theory and usage. (2000).
15. Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. Molecular systems biology, 12(7), 878.
16. MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.
17. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019
18. Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., ... & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nature biotechnology, 36(10), 983-987.
19. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. nature, 542(7639), 115-118.
20. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention
21. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ... & Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature reviews Drug discovery, 18(6), 463-477.
22. Li, Y., Gao, C., Song, X., Wang, X., Xu, Y., & Han, S. (2023). DrugGPT: A GPT-based strategy for designing potential ligands targeting specific proteins. bioRxiv, 2023-06.
23. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. nature, 596(7873), 583-589
24. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5, 115-133.
25. Rosenblatt, F. (1957). The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory.
26. Nicosia, G., Cutello, V., Bentley, P. J., & Timmis, J. (2004, September). Artificial immune systems. In Third International Conference, ICARIS (Vol. 3239).
27. Eisenhaber F, Persson B, Argos P. Protein structure prediction: recognition of primary, secondary, and tertiary structural features from amino acid sequence. Crit Rev Biochem Mol Biol. 1995;30(1):1-94. doi: 10.3109/10409239509085139. PMID: 7587278

Revolutionising Biology with Artificial Intelligence - 1

Suryadip

Suryadip and Tathagata offer an engaging exploration of how AI is transforming biological sciences. From the early days of rule-based AI to cutting-edge deep learning applications like AlphaFold, this article traces the evolution of AI's role in biology. It highlights AI's power in solving complex problems such as protein folding, drug discovery, and genomics. With groundbreaking examples like DeepVariant and DrugGPT, readers will gain insight into AI's monumental impact on research and healthcare. Dive into the future of biology where data meets innovation!

Revolutionising Biology with Artificial Intelligence - 1

Suryadip

Artificial Intelligence (AI), the talk of the town right now, has revolutionised various fields and the life sciences is no exception. Before we go further into how exactly AI is changing the course of life as we speak, let us get up to speed with a few concepts.

Marvin Minsky defined artificial intelligence as "**the science of making machines do things that would require intelligence if done by men**". Ideally one would apply AI for problems where it is impossible to define consolidated rules. A very simple example could be the spam filters in our emails. It is virtually impossible for programmers to identify a set of rules to identify all types of spam. This is because these messages constantly keep evolving with newer wordings and patterns. Therefore AI models are trained using examples of both spam and legitimate emails to help identify subtle patterns and adapt to the new spam techniques over time. This is exactly why biology is such a beautiful candidate for AI applications, owing to its inherent complexity.

Machine Learning (ML) is a subset of AI that focuses on the development of algorithms that allow computers to learn from and make predictions based on data. Deep Learning (DL) is a further subset of ML that utilises neural networks (inspired from, but in no way similar to biological neurons) with many layers (hence "deep") to analyse various factors of data (especially non-linearity). These two branches of AI are the most prevalently used in biology to date.

Early Days: From Symbolic AI to ML and DL

It all dates back to the 1950s, when Alan Turing proposed the concept of machines being able to simulate any form of human reasoning through algorithmic approaches [1]. As a result, during the early years, AI was largely rule-based (aka Symbolic AI) which was based on logical representations of the world. This led to the birth of expert systems, which used knowledge bases of rules to solve specific problems. One such expert system was called MYCIN (1978) (Fig 1) [2]. It was developed by Edward Shortliffe as part of his doctoral dissertation, under the guidance of Bruce G. Buchanan, Stanley N. Cohen, and others at Stanford University. MYCIN was used to identify the bacteria causing severe bacterial infections like meningitis and bacteremia and to subsequently recommend appropriate antibiotics and their dosages according to the patient's body weights.

Fig 2: This is a caption. This is a caption. This is a caption.

The advent of the internet and a substantial increase in computational power led to the growth of enormous volumes of data. This led to the next big break in the early 2000s in the form of Machine Learning algorithms. These algorithms were able to learn from data and perform tasks such as predicting outcomes, classifying objects or clustering them based on similarity, as opposed to following predefined rules. Further, the introduction of DL opened even more avenues and a whole new range of applications in tasks such as image processing, signal processing, natural language processing etc. Nowadays, AI has extended its branches into any and all fields of human endeavour with an especially serial impact on healthcare, medicine, and biological research.

AI and the era of Structural Biology

Applications of AI for solving complex problems in the life sciences can be dated back to the late 1990s. This was the era of structural biology. Proteins are the molecular machines that carry out all biological processes in a cell, and their structure dictates their function [3] Therefore, scientists were devoted to finding solutions to identify and manipulate said protein structures.

Proteins are unbranched polymers constructed from 22 standard amino acids. They have four levels of structural organisation (primary, secondary, tertiary and quaternary). Primary structure refers to the amino acid sequence that is specified by the genetic information contained within the DNA. As the polypeptide folds, it forms certain localised arrangements of adjacent amino acids that constitute of the secondary structures (mainly in the form of α-helices and **β**-sheets). Tertiary structure refers to the three-dimensional shape of a protein formed by the overall folding of the polypeptide chain onto itself. It results from interactions among the various R groups (side chains) of the amino acids, including hydrogen bonds, ionic bonds, van der Waals forces, and disulfide bridges. A protein quaternary structure arises when two or more polypeptide chains (subunits) come

together to form a larger functional protein complex. The arrangement and interaction of these subunits are crucial for the protein's function. Together, these levels of structure are essential for a protein's biological activity and functionality (Fig 2) [4].

During this time, X-Ray Crystallography and Nuclear Magnetic Resonance Spectroscopy (NMR-Spec) were heavily being used to determine the complicated 3D structures of proteins. Due to the complexity of the protocols associated with these techniques coupled with low success rate, scientists started to look for more feasible solutions. In 1990, the labs of Ross King (Chalmers University of Technology, Sweden) and Michael Sternberg (Imperial College London) came up with an ML program called PROMIS (PROtein Machine Induction System) that could generalise rules to characterise the relationship between primary and secondary structures of globular proteins. These rules could further be used to predict unknown secondary structures from a known primary structure of a protein [5]. Quantitative Structure Activity Relationship Modeling (QSAR) is an extensively used approach in drug design and discovery that devises mathematical models connecting the biological activity of these drugs to their complex chemical features [6]. In 1994, the pair of King and Sternberg modelled the QSAR of a series of drugs using a technique called Inductive Logic Programming (ILP) [7]. ILP is a machine learning technique that uses a combination of logic programming and data driven decision making [8]. With the application of ILP, they were able to elucidate complex and interpretable patterns within the data that might have originally been overlooked while investigating using traditional statistical approaches.

Further, in the year 2000, the lab of R. Casadio used a class of artificial neural networks known as a Feed Forward Network (FFN), to predict protein folding and structure from only their corresponding amino acid sequence [9]. A feed forward network is basically composed of several layers of interconnected computational units called neurons (inspired from biological neurons) (Fig 3). Each neuron in a layer has a weight factor (think of it as a scalar value that denotes the neuron's contribution to the output) associated with it and receives numerical inputs from other neurons in the previous layer, processes them by applying a weighted sum and an activation function, and passes the result to the neurons in the next layer. This allows the network to model complex, non-linear patterns [10]. Therefore, hypothetically speaking, the non-linear relationship between a protein's primary amino acid sequence and its folded 3D structure can be modelled efficiently using this FFN. By capturing these subtle relationships, FFNs provided insights into how specific sequence motifs correspond to structural motifs underscoring the massive applications of AI in the field of structural biology.

Genomics and the Rise of AI The Human Genome Project (HGP) was an international research initiative with the goal of sequencing the entire human genome. Launched in 1990, the HGP was completed in 2003 with the submission of the 1st draftt of the human genome. Its goals included identifying all human genes, determining their sequences, and exploring the functions of these genes. The project has had profound implications for medicine, genetics, and our understanding of human biology. Following that, in the early 2000s, vast amounts of genomic data were generated which warranted the need for advanced analytical methods. Michael Schatz and his colleagues applied ML techniques to genome assembly, enhancing the efficiency of sequence alignment [11]. In 2004, David Heckerman demonstrated the use of Bayesian Networks, a type of probabilistic graphical model, to predict gene functions from microarray data [12]. Bayesian networks as the name suggests were based on the concept of Bayesian statistics that provides a way to update the probability of a hypothesis as new evidence becomes available, forming the foundation of Bayesian inference [13]. Heckerman used microarray data that contains information about the activity level (how active or inactive the gene is) of thousands of genes across specific conditions. In his approach, variables such as the genes were represented as nodes and the probabilistic dependencies between them as edges, making them particularly effective for capturing complex, uncertain relationships (Fig 4). This enabled Heckerman to identify the complex crosstalks between genes and their functional relationships to each other based on expression patterns.

Through these networks, it became possible to model how changes in the expression of one gene influence others, uncovering hidden dependencies and regulatory mechanisms [12]. In 2016, Angermueller's lab utilised Convolutional Neural Networks (CNN), which were traditionally used to process image data, on DNA sequences to predict gene expression, with accuracies comparable to traditional methods [14]. CNNs are comprised of layers of neurons that utilise convolutional filters to capture local patterns and hierarchical features. A convolutional filter is a matrix of weights that is slid along the entire sequence effectively learning hidden patterns and features from them, much like detecting edges or textures within an image which is nothing but a sequence of pixel values. Moreover, since the same filter is being slid along the entire sequence, it can capture both local and global dependencies among these sequences potentially identifying important motifs such as transcription factor binding sites [15].

Most recently in 2018, Google developed a DL based tool called the DeepVariant that employs a deep CNN architecture to identify variants in DNA sequences, outperforming traditional state of the art tools and pipelines such as GATK and Augustus [17] (Fig 6).

AI in Biomedical Imaging, Diagnostics and Drug Discovery CNNs being at the forefront of computer vision at the time also inspired biologists to actually use them for biomedical image analysis and predictive modelling based on biomedical image data. In 2017, Andre Esteva and his colleagues developed a CNN model that could diagnose and differentiate skin cancer from other standard

Fig 4: This is a caption. This is a caption. This is a caption.

skin lesions with accuracy comparable to dermatologists by analysing dermoscopic images [18]. U-Net, a CNN architecture developed by Olaf Ronneberger, is widely used nowadays for segmenting biological images, especially in the field of histopathology which is the study of tissue changes and diseases at the microscopic level, often involving the examination of biopsies to diagnose conditions (see Fig. 5). It has shown remarkable performance in analysing histological slides accurately diagnosing cancerous tissues, thereby improving patient outcomes [19].

ML has contributed significantly towards the drug discovery space as well, including applications in drug target identification, biomarker discovery, QSAR modelling and predicting efficacy of drug candidates, thereby accelerating the drug development process. Biomarker discovery—the identification of measurable indicators such as genes, proteins, or metabolites linked to specific diseases—has benefited from ML's ability to analyze complex datasets and uncover subtle patterns, aiding in personalized medicine and early diagnosis [20]. Yuesen Li and his colleagues in 2023 developed DrugGPT focusing on chemical space exploration of protein ligand complexes, which refers to navigating the enormously complex universe of chemical compounds to identify potential drug molecules targeting specific proteins[21].

A Leap Forward with AlphaFold In the realm of structural biology, we have come a long way as well. In the year 2021, John Jumper and his team at Google Deepmind rocked the world of biology with their groundbreaking model called AlphaFold. AlphaFold redefined the age-old problem of predicting protein 3D structures from just their amino acid sequence information. It uses a pseudo-bayesian genetic algorithm based deep neural network to model the physical and geometric properties of proteins from just their amino acid sequence and gives highly accurate 3D structures of proteins solving the bottleneck of time, complexity and low success rate associated with methods like X-ray crystallography, NMR-spec and Cryo Electron microscopy (Cryo-EM) [22]. In 2024, Jumper was awarded the Nobel Prize in Chemistry for AlphaFold emphasizing the importance and relevance of AI in biology even further.

Biology Inspiring AI models Interestingly enough, the relationship between AI and biology has been quite symbiotic. It is not only biology that has gained from AI models, biology has also inspired AI models on more than one occasion. The most prevalent of them all is actually that of the neural networks (the building blocks of DL). In 1943, Warren McCulloch and Walter Pitts built a mathematical model of the functioning of a single neuron [23]. Taking inspiration from human perception, Frank Rosenbahlt modelled the first neural network in 1957 that was able to recognize some handwritten symbols effectively and could model basic logical operations such as AND and OR gates. This was known as the Perceptron and it would lay the foundation for further development of DL and neural network architectures [24].

Our immune system only allows lymphocytes, that recognize certain antigens, to be cloned and proliferated with such identical antigenic receptors. This phenomenon known as clonal selection presents itself as a learning problem driven by context (antigen) and an appropriate response. This led to the invention of the Artificial Immune System, a class of computationally intelligent, rule-based machine learning systems, inspired by the intricacies of the immune system of vertebrates [25]. Akin to their living counterparts these models excel in recognizing something presented to it as an "antigen" and take the most appropriate response making them ideal for applications such as antiviruses and anti-spams.

Towards the Future The synergy between AI and biology has set the stage for groundbreaking discoveries and innovations. From protein structure prediction with AlphaFold to the use of neural networks in genomics and drug discovery, biology has been transformed into a data-driven, predictive and multi-disciplinary branch of science. Whether it's predicting disease, designing drugs, or deciphering the mysteries of life at the molecular level, AI is now an indispensable tool in every biologist's toolkit. So, if you're curious, now's the time to dive into this captivating intersection of science and technology—where the future of biology is being driven by AI.

abstract: Suryadip and Tathagata offer an engaging exploration of how AI is transforming biological sciences. From the early days of rule-based AI to cutting-edge deep learning applications like AlphaFold, this article traces the evolution of AI's role in biology. It highlights AI's power in solving complex problems such as protein folding, drug discovery, and genomics. With groundbreaking examples like DeepVariant and DrugGPT, readers will gain insight into AI's monumental impact on research and healthcare. Dive into the future of biology where data meets innovation!

Interview with Artificial Intelligence - 4

Suryadip Sarkar (IBAB Bengaluru)
Tathagata Chatterjee (RKMVERI Kolkata)

Suryadip and Tathagata offer an engaging exploration of how AI is transforming biological sciences. From the early days of rule-based AI to cutting-edge deep learning applications like AlphaFold, this article traces the evolution of AI's role in biology. It highlights AI's power in solving complex problems such as protein folding, drug discovery, and genomics. With groundbreaking examples like DeepVariant and DrugGPT, readers will gain insight into AI's monumental impact on research and healthcare. Dive into the future of biology where data meets innovation!

Interview with Artificial Intelligence – 4

Suryadip Sarkar (IBAB Bengaluru), Tathagata Chatterjee (RKMVERI Kolkata)

SS: Hello. Didi. So, you have had a remarkable journey, I believe, in IISER, Kolkata.

SR: Yes.

AD: So, you are 15 MS.

SR: No, I am 14 MS.

AD: 14 MS! So, you graduated here in 2019. That means when you just graduated, the 19MS batch of BSMS first entered IISER Kolkata.

SR: Yes.

SS: I believe the IISER Kolkata campus was not here back then as it stands now. Right?

SR: When I started my journey at IISER, Nivedita Hostel wasn't there yet. However, the LHC was! In fact, during our first year, construction for the LHC had just begun.

So, as you all know, we did our classes at Lecture Hall 4, the great LH4. The upstairs canteen is now located there.

SS: Mess, as we call it, today.

AD: Can you walk us through the progression of the facilities during your time at university?

SR: Sure! Most of our classes were held at the LHC in our second year. By the time we reached our fourth or fifth year, the RC was ready, but the TRC was still under construction. So, for a while, the RC and LHC were quite similar regarding available facilities, but the TRC wasn't yet in the picture.

AD: So, just to clarify, the RC was up and running, but the TRC wasn't available yet?
SR: Yes, exactly. The RC was available, but the TRC wasn't. It can be tricky sometimes to remember the exact timeline, especially after all these years, but that's how it unfolded.

SS: When you first came to IISER, did you think you would go towards earth sciences or petrology?

SR: No, no, not at all. Initially, I thought I would choose chemistry. Before coming to IISER, I had attended another college for about a month. The curriculum had already started there, and I had enrolled as a chemistry student. So, when I shifted to IISER after being selected, I thought I would also continue studying chemistry here.

But as you know, in our first year, we have to study all the subjects. All the subjects. That is true for you guys as well, I believe. So, after that time, I was introduced to Earth Science and I liked the subject. But, still, I was confused.

So, in the second year, I took Chemistry and Earth Science. And then, in the third year onwards, I chose to.

AD: So, did you have ECB as your premajor?

SR: No, not ECB. PCE. Physics, Chemistry, and Earth Science.

SS: The same goes as well as for my roommate. But he is now in Physics.

AD: So, what fascinated you about petrology? And, Petrology is associated with petroleum…….

SR: No, no, petro means rocks. It is associated with rocks, and the formation of rocks. Not anyway related with Indian Oil and Bharat Petroleum.

AD: No, not that. What I meant is, that is the norm we think about and come across.

SR: So, It is actually difficult to say and pinpoint any particular reason. I think the way teachers taught us and the books, those things were, I mean, I really enjoyed reading the books or the classes were very like, what should I say, very good. Interesting.

SS: So, in those days, did you have SDG?

SR: No.

SS: You did not have him?

SR: No. So, we studied petrology under Tapobrata Sir.

SS: Tapobrata Sarkar.

SR: Yes, and I think Somnath Dasgupta, sir, came when I was doing my master's thesis. Before that, he was the Vice-Chancellor at Assam University in Silchar.

SS: So, after leaving IISER, where have you been? Where did you pursue your PhD?

SR: I pursued my PhD at the Australia National University in Canberra, Australia.

And there I finished my PhD this year very recently.

SS: So, normally, I mean, from where my partner Aritra and

I stand, we see, the main target is either the USA or Europe. We both are from the Department of Physical Sciences. I mean, I am taking the UK in Europe only. So, mainly these two places. Australia National University does come to mind, but it comes mainly for that internship purpose. But to pursue PhD, I do not think a good chunk thinks of Australia or other places. So, why did you?

SR: Yes, actually, because I was more inclined to the experimental side of petrology.

And, if you know, the experimental petrology labs are not very common. So, in the USA, I think only a few places have experimental petrology labs. And in Australia, actually, ANU has a very big experimental petrology lab.

That's why, when selecting universities, ANU was one of my top choices, primarily because of the excellent experimental setup there.

SS: So, like Europe, you applied to a professor and got it or like the USA, you applied to an institute?

SR: No. So, like in Europe, I had to email some professors, and they conducted interviews with me. After that, I had to go through the application process and complete all the necessary steps before finally getting selected and enrolment.

SS: So, in your view, how are the Australian academic system and the Indian academic system somewhat similar or somewhat different?

SR: In the Australian academic system, we do not have to do any coursework.

So, directly we would like to enter the research as an RA, or Research Assistant.

AD: So, that is what postdocs do here.

SR: No, no. So, you may have heard that in the USA, when you enter, there is a TAship, and you would be working as a teaching assistant and then later on the student will get the RA, which is basically the research assistant. It is the same here as well.

SS: So, that means that at that time, you are getting the scholarship from the research and not from your teaching.

SR: So, in Australia, directly you will enter the RA, and you will not like to do any teaching courses. So, you did not have to do a TA. It is kind of optional.

As for my case, I did like a little bit of TAship duty, but that was an optional exercise. So, I just did it for my experience. Thus, the question that you asked earlier was how the academic systems in Australia as compared to India.

So, I would say that in Australia, the work-life balance is better.

AD: Better in the sense?

SR: Better in the sense, I mean, maybe for example in Australia, after 5 o'clock or after 4.30, nobody will, or rather should I put this in this way, technically should not send you any official email or communicate. And, in my whole PhD, I never get any kind of official email from either my supervisor or my professor, co-supervisor, or anyone after 5. So, these kinds of things were there.

SS: In India, we all get important mail at midnight. Yes. And we submit our assignments also at 12.

SR: I am not sure about the assignment deadline there because I never submitted that. So, basically, here I did more coursework, and there I did research. So, I would not compare it like that. Actually, it will be difficult for me to compare. Yes. But this is a kind of thing, I think in Europe also this balance is very much maintained, but in the USA, this is not and maybe not in India as much as well.

AD: Yes.

SS: So, did you give any admission interviews in universities other than ANU?

SR: In Australia? No. In Australia, I just gave an interview at ANU.

AD: In other places?

Srijita Das: In Europe, I gave, I think, an interview at a university in France, but I forgot the name.

SS: Is there any fundamental difference between the interview process in Europe and Australia?

SR: No. The interview process is very similar. So, in Europe, the principal supervisor and the co-supervisor asked me questions, and that was one round only, and in Australia also the same.

SS: What about the funding sources in Australia? So, the funding source is like a university can fund you, or if a professor has your fund, then that is a wa,y or there is another way like the government you can get funded.

AD: Australian government?

SR: Yes, the Australian government.

SS: And is that a handsome amount or just when you want to sustain yourself?

SR: No, no. Actually, the amount that you will get is sufficient to pay your rent and then like to have food and everything, and after that also you can save some money. So, it is not a very big proble,m and also in Australia, the scholarship is tax-free.
So you do not need to pay taxes. So, that is a plus point. That is a boon compared to other places.

SS: If you do not take it personally, I have heard from some teachers, petrology has no such future because petrology is mainly used in places like where the oil industry is very much associated. If you say region it is the Middle East or Russia. If you say company it is Oman oil, Indian Oil ,and similar.. So, in the upcoming 100 years, it is going to be a dead subject. You, as a new researcher, how do you see this comment?

SR: Well, I think actually I do not fully agree with this one because I mean you know about the renewable energy sources like the critical minerals. So, I mean those are, so it is very important to like to extract those in a cheap and economic way, and right now actually China is the, I mean one of the main suppliers for the overall world. So, basically in order to understand how we can extract those critical minerals and to understand their basic science, we need to understand and we need to go through the petrology.

So, in that way, actually, it is very useful, and I can say, in places like Australia, we do have lots of these rare earth mineral ores, and like many of the petrology labs are now trying to understand how they can extract those ores more cheaply. So, not only China will supply the world with those minerals, but other countries, rich in natural resources, can also take part in the game. For that, knowledge, fundamental research, and the development of technology are the utmost requirements of the time. So, in that way, I do not think that petrology is dying or something.

SS: So, now coming to little away from the academics, we have had COVID-19 and perhaps many of us not have seen the video we have come across a story, an incident where a doctor, student, or scientist is crying, accusing the world of saying. "Now Why don't you ask all the celebrities to fund you or to save you? For society, doing science and fundamental research is a thankless job?"

I mean, have you heard or seen?

SR: No. I have seen that during the COVID. It is not in India, and it is somewhere else.

SS: I mean, where we were all looking towards doctors and scientists for the doctors for the immediate cure to save the lives and scientists for mainly biologists, scientists. Yes, for the medication. For the medication, but then this came up.

So, this kind of situation perhaps does not fall directly on earth scientists, but as a member of that society, it is, again, I mean scientific society, budding researcher. How do you see this? Do you feel that we, as budding researchers, are doing something thankless?

SR: In science, especially when working in basic research, it can be very challenging to communicate your work to those outside the scientific community. People may not immediately see the value of your research and might even think of it as unimportant or insignificant.
However, in science, progress is often made incrementally, step by step. So, if you're doing basic science research—whether in Chemistry, Physics, or Earth Science—it can sometimes feel like a thankless job, especially during your PhD. However, as more people contribute to your work over time, the significance of your research becomes clearer. Eventually, others will recognize that it wasn't a thankless job but rather part of a gradual, step-by-step process that advances knowledge.
I hope that answers your question.

SS: That is ok. As a part of that question, have you ever felt that it would have been better to pursue engineering than this? I mean, at your age when your friends who have opted for engineering might have found a lucrative job with a high-paying salary.

SR: Honestly, during my PhD, I questioned myself several times—pretty much every day, if not more often. But now, as I'm nearing the end, having submitted my thesis and everything, I feel differently. It seems that what I was doing was the right thing after all.
Regarding the question of freedom or independence, I don't necessarily see it as having the "upper hand." Still, in academia, there's a certain autonomy that allows for deeper exploration and understanding. For instance, in engineering, many people pursue a PhD and take on similar challenges, but there's a difference in the kind of work they do compared to those in academia. Engineers working a 9-to-5 job might not have the same opportunity to contribute to the deeper theoretical understanding that comes from academic research.
Another point is that, in academia, there is typically more freedom in terms of time—like vacation flexibility—compared to a corporate job. So, while there are challenges, freedom and independence are valuable aspects of the academic path.
SS: So, in your view or not exactly your view in reality, how does the funding in India and Australia differ?
SR: To be honest, I am not very sure about the Indian funding.

So, for me, it was basically the panel member or the supervisor. So they have already funded the project and where I joined.

So, that's how I was a part of that funded project throughout my PhD. I didn't have to write the proposal and like to bring the funding. So, that's why maybe I was not aware of this funding situation or other things.

SS: See, the fact is that is why I am asking this for

the Australian case. Yes your scholarship was tax-free. Scholarship is tax-free in India also, but keeping it as a scholarship, there are other things you have to very much important your funding for your project and all.

And for the last few years, especially after this year, I mean the import duty and GST and whatever added together gave more than 150 percent in India. Are you aware of that?

SR: No.

SS: Ok. So, this is the present scenario. In this way, if the price goes up I mean that is that is the scenario in India I mean after the budget and all. So, how do you see I mean is there any kind of similar scenario in Australia? What is your view on this?

SR: I am not very sure about this topic because I was unaware of the budget or the present situation.

AD: Ok, never mind.
Have you ever felt that, despite the changing times, the challenges faced by the scientific community have mainly remained the same—especially when it comes to attacks from various parts of society?
For example, after COVID-19, former U.S. President Donald Trump stated that the United States would no longer be a part of the WHO, claiming that the organization was not acting under his wishes. Similarly, there have been instances where the U.S. has been allowed to violate environmental laws with minimal repercussions, often just paying a small amount of money, as if doing the world a favor. On the other hand, lower- and middle-income countries, especially in the Global South—such as those in Africa and South Asia—don't have the same privileges and often face harsh consequences for similar actions.
In South Asia, countries like India, Pakistan, Bangladesh, and Nepal still experience practices that undermine science and rationality. For example, there are cases of witch-hunting, where people, often women, are accused of being witches and subjected to violence. Malpractices in medicine also persist, with people resorting to quacks and superstitions. Even the practice of alchemy still exists in some places.
Furthermore, institutions like IITs and IISERs often suffer from inadequate funding. And unfortunately, there are still sections of the world that believe in discriminatory practices, such as apartheid, and argue that certain groups can't become better scientists or contribute to research as effectively as others.
Given all this, do you feel that the attacks on science and the scientific community persist, even today?

SR:, To be honest, I never felt, never felt that.
SS: So, after this PhD, what's your plan?

SR: So, I will join as a postdoc at Rice University in Texas. So, and that's a period of? For two years. After that? After that, basically, basically, I want to explore the industry as well because I have never been totally in academia like throughout.

So, I wanted to explore the industry, but I also wanted to open both of my options in academia and industry.

AD: Yes. Now, you are moving to the US. We hope that you will have plans to come back to India.

SR: Yes.

AD: So, now, lame question. Assume you are the president and the prime minister of the country. So, what changes would you, I mean from sitting today, in today's position, would you like to see so that the differences that you can see between the USA and India can be moved, rubbed out, rubbed off?

SR: So, maybe first, I would like to put more money into industries.

SS: What kind of industries?

SR: I would like research, so I may want to put more funding in the critical minerals or in the renewable energy or green technology in that sort of thing in India.

So, if someone wants to come back, we only have maybe IIT, IISERS, IACS, and a few institutes. Some of the DST and department of science.

So, maybe I will try to increase the number of central government institutes.

SS: So, now a follow-up question comes in: if more industry has to come up, are we not falling prey to some kind of thing in the hands of the businessmen? In the present case, if not in India, there are also other countries where, compared to the lower middle class, industries tend to govern and direct the ways science should work.

SR: I mean applied science, engineering science if you know engineering science fully, some application-based sciences, not fundamental discoveries or stuff like that. Also, they may be inclined towards some, something where. But I mean, when I am saying industry, it is not always like that. Research industries and national labs are included as well.

SS: So, national labs. Again, coming back to academia, right?

SR: So, the national lab is a link between academia and industry. So, it is not like full academia, or it is not like the entire industry.

So, it is kind of a link or bridge between these two. And like there are, I mean, there are other industries in the USA, which is more like research industries. Of course, there is a part of a business team.

But I think it is not business. I am not sure if I answered it satisfactorily.

SS: No. That is fine. I just wanted your opinion in this regard.

AD: So, what, if we finally come towards the end of the thing, will your advice to your fellow juniors in IISERs be?

SS: How many students used to be in earth science departments in your days?

SR: I think I had 32 in my batch.

SS: Really?

SR: Yes.

SS: 14 MS and 32? 14 MS. Major?

SR: Major, yes. And for you?

In earth science?

I have 20 MS, and only 10 are in that department. A similar thing goes for 19 MS and 18 MS. 22 MS. 22MS has again gone back to 32 or 37.

We were 14 MS.; I believe 15 MS was the highest total, around 42.

AD: Have you ever thought why earth science shows these fluctuations where physics and biology, departments show all-time high demands? Chemistry is not high demand, but it is a fair good moderate every year.

SR: I think it is because we learn physics, chemistry, biology, and mathematics from childhood. So, people usually like when they come here, they usually already love either physics or mathematics or chemistry like that, and they usually go there.

AD: In Australia, it was not a hostel life, right? You had to stay in apartments?

SR: So, in Australia, we stayed in studio apartments.

So it was a big room with your kitchen and bathroom, everything. And then, like at the side, you have another room with a kitchen and bathroom. So there was no sharing.

AD: So, is it one bedroom, one BHK type?

SR: Not BHK because there was no partitioning between the kitchen and bedroom. So it is kind of, so studio apartment is kind of arranged. For example, if this is like one room, then this side is the kitchen, and this side is the bathroom. The bathroom is the door, if you were asking that.

And one side has the bed.

SS: And what was the cost of staying there?
SR: So since I stayed on campus, that's why I had to pay... On campus means like us? Yes. So that's why I had to pay around $1200. I mean $1200 for a month.

SS: And how much did you earn? I mean, if you don't mind.

SR: Yes, I get around 4,000 per month.

AD: 4000 per month. So you could save 1000?

SR: Yes, the saving was fine.

SS: Okay, so with this we conclude. Thank you.

InScight Digest

Summarising The Frontiers In Research

Earth has its own electric field weaker than a pencil battery

Collinson, G.A., Glocer, A., Pfaff, R. et al., Nature 632, 1021–1025 (2024)

Contributed by **Chitradeep Saha (CESSI, IISER Kolkata)**

Our planet Earth also has a global electric field – as fundamental as its gravity and magnetic fields. Researchers have successfully measured this electric field for the first time. The quest started over half a century ago when a steady stream of outgoing plasma particles was detected near the Earth's poles. Theories have been proposed and refined over time to understand this peculiar phenomenon. Peculiar. Because the temperature of the outflowing plasma is too cold to evaporate due to solar radiation. Therefore, existing knowledge of global energy fields failed to explain it. Alternatively, the existence of an independent, electric field was hypothesized. However, more mature technology was required to make precise measurements and test the hypothesis. The research team flew a suborbital rocket through the arctic skies that touched the exosphere while sampling various ionospheric properties. The photoelectron spectrometer onboard detected a minute change in electric potential of 0.55 volts – less than that of a standard AA battery – across an altitude range of ~500 km, confirming the existence of such a global electric field. Due to an asymmetric gravitational pull on the lighter electrons and heavier ion cores, charge separation occurs in the atmosphere. The Coulombic force partially counteracts this charge separation; the associated electric field is ambipolar, as it works in both directions. The net effect of this global ambipolar field is to puff up the atmosphere, lifting some ions high enough to escape through the polar caps and giving rise to polar winds.

Mapping Strain in Laser-Written Diamond Waveguides Using Optically Detected Magnetic Resonance

Phys. Rev. Applied 22, 024055 (2024)

Contributed by **M. Sahnawaz Alam (RWTH Aachen University, Germany)**

Color centers in diamond, such as the nitrogen-vacancy (NV) center—which consists of a nitrogen atom adjacent to a vacancy in the carbon lattice—are renowned for their exceptional quantum properties, making them ideal candidates for applications in nanoscale sensing and quantum information processing at room temperature. However, NV centers often suffer from low contrast in experiments, which hampers their performance and limits practical applications. Integrating NV centers with laser-written optical waveguides enhances the coupling of light to these quantum systems, facilitating more efficient manipulation and readout of their states. This integration addresses the low-contrast issue by improving the interaction between NV centers and light. However, the process of laser writing photonic structures inherently introduces strain into the diamond lattice. This strain can alter the electronic and spin properties of the NV centers, affecting their performance and the fidelity of quantum operations. Until now, the full impact of this strain on defect centers has not been thoroughly understood. In our combined experimental and theoretical study, we demonstrate that optically detected magnetic resonance (ODMR) spectroscopy—a technique commonly used to probe the spin states of NV centers—provides sufficient information to fully characterize the spatial distribution of strain within laser-written diamond waveguides. Remarkably, this characterization is possible even without the application of an external constant magnetic field. Our findings present an accessible and non-invasive tool for mapping strain in diamond-based photonic devices. By utilizing ODMR spectroscopy, researchers can gain detailed insights into strain distributions, enabling the optimization of device fabrication processes and the improvement of quantum device performance. This advancement is a significant step forward in the development of diamond-based quantum technologies, potentially impacting a wide range of applications from high-precision sensing to quantum communication and computation.

GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run

R. Abbott et al., Phys. Rev. X 13, 041039 (2023)

Contributed by **Swarnendu Saha (CESSI, IISER Kolkata)**

This paper, “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, is an excerpt from the third Gravitational Wave Transient Catalog. The report was published in December 2023 in the journal Physical Review X that reports on observation of gravitational waves by the LIGO, Virgo, and KAGRA for the second half of the period of the third observing run. The interval covered is from November 1, 2019, through March 27, 2020. Our very own Professor Rajesh Kumble Nayak has been a part of this team from IISER Kolkata.

The paper also lists 35 new gravitational-wave events detected during that period, all the result of a "compact binary coalescence" in which pairs of black holes or neutron stars orbit ever tighter until they eventually merge. This brings the number of gravitational-wave detections across three observation runs to 90. Most of the events in the catalog arise from the merger of black holes, which can be pretty large, but this observing run also marks the first definite identification of neutron star–black hole (NSBH) mergers. Curiously, however, no BNS mergers were confirmed during this period.

The researchers selected these signals with advanced algorithms and data calibration, estimating probabilities for each event to be of astrophysical rather than noise artifacts. As there is always some inevitable noise, the probability for these signals to be due to a non- astrophysical origin is estimated to be around 10-15%. All the data are made available in the public domain for use by the community through the Gravitational Wave Open Science Center. The expanding dataset from GWTC-3 gives unparalleled insight into the properties and behaviors of black holes and neutron stars, offering unique insights into the population in the universe and providing informative input into the theory governing their formation, structure, and evolution.

Strange Metal and Quantum Spin Liquid in Heavy-Fermion Material: An Array of Exotic Phases

Hengdi Zhao et al., Phys. Rev. Lett. 132, 226503 (2024)

Contributed by **Abhirup Mukherjee (IISER Kolkata)**

Strange metals and spin liquids constitute deviations from the "standard model" of condensed matter physics. For most of the twentieth century, metals were believed to be smoothly connected to non-interacting electrons at low-energies (the so-called Landau's Fermi liquid theory), and insulators and superconductors were believed to arise from spontaneous symmetry breaking (the ground state does not have all symmetries of the Hamiltonian). Violations of these ideas were observed in the 1980s, a notable example being the discovery of high-temperature superconductivity in copper oxide materials. The strange metal phase of the copper oxides, while being a metal, displayed a linear-in- temperature resistivity, in contrast to the quadratic-in-temperature resistivty of Landau Fermi liquids. Due to the large transition temperature of the material and the proximity to an electronic-correlation driven Mott insulator, Phillip Anderson hyppthesised that the ground state of the Mott insulator was close to a "spin liquid", where the system does not settle into any particular configuration (in contrast to the symmetry-broken insulators) but keeps shifting (in the sense of a quantum superposition) between various configurations.

 While these exotic phases typically emerge in different materials, the authors of the present work have experimentally realised these exotic phases in crystals of the material Ba4 Nb1-xRu3+xO12, where x is the hole- doping concentration. By tuning the doping concentration, the material undergoes transition from a heavy strange metal phase to a heavy Fermi liquid phase to finally a spin liquid phase. The "heaviness" arises from the fact that these are heavy-fermion materials in which the inter-electron interactions increase the "inertia" of the quasiparticles. Other results suggest that the excitations in all three phases are described by spinons - spin-1/2 charge-neutral objects. These spinons are fractionalised excitations (to see why, recall that flipping a spin from -1/2 to 1/2 creates a spin-1 excitation). At the heart of these exotic phenomenon in this material is the underlying triangular lattice that leads to geometric frustration (it is not simple to obtain an energy-minimising configuration of spins on this lattice) and the emergence of novel elementary excitations (the spinons). Such a material provides a wonderful platform to realise and study these highly-correlated phases of matter.