The Protein Science Revolution: Implications for human health and mortality assumptions

By Craig Armstrong

Summary

What is the protein science revolution? Researchers now have access to new AI tools that allow them to understand and manipulate protein structure and function much more efficiently.

Why is it important? Proteins are significant drug targets and also can be used as components of medical technologies.

How does this impact mortality assumptions? In isolation, this is a breakthrough in basic science, and any impacts will be seen over the medium-to-long term. Assigning future improvements specifically to advances in protein science is questionable, but this represents a new source from which future improvements might arise, at a time when historical sources appear to be faltering in some countries.

Introduction

Proteins are fundamental to biological functions, and so it is unsurprising that many diseases are rooted in their misbehavior, and that many medicines and therapeutics target proteins at the biochemical level. With advancements in AI, notably those seen in language models and image-generation tools, the ability to understand, modify, and design proteins is rapidly improving. The scientific journal Nature highlighted improvements in the ability to design and manipulate proteins as one of its seven breakthroughs to watch in 2024,[1] but to our knowledge, the topic remains relatively unexplored in actuarial circles. This article considers the potential implications of these breakthroughs for human health and mortality.

Background

What is a protein?

Proteins are large molecules that play diverse and vital roles in our bodies. They are made of chains of building blocks called amino acids, of which there are 20 distinct varieties.[2] Because these chains can contain hundreds of amino acids, the number of possible distinct proteins is almost countless. To illustrate this, consider a chain of just eight amino acids (such a short protein would be more accurately described as a peptide) and calculate the number of possible unique combinations of amino acids. The result: 20^8 combinations, or more than 25 billion – three times the population of the earth.

What makes each amino acid unique is the chemical group attached to it, known as the sidechain. The nature of these chemical groups, and their order, gives rise to the myriad functionalities observed in proteins.

Protein folding

Proteins are not merely long, amorphous chains of amino acids; they adopt specific three-dimensional shapes. The transition from a linear string of amino acids to a structured form is called protein folding. The shape a protein takes is largely determined by its amino acid sequence. Even though a protein could theoretically fold in countless ways, it is striking that proteins with the same sequence consistently fold into a specific shape. Misfolding, though, can and does happen, and as we will see, such errors can impact our health.

The shape, or structure, of a protein is central to its function in the body:

- **Chemical reactions** – Some proteins (enzymes) have shapes that allow particular chemical reactions to occur, while preventing others. They achieve this by positioning specific amino acids close together.
- **Transport** – Proteins like hemoglobin carry molecules, in this case oxygen, to parts of the body where they are needed.
- **Communication** – Some proteins sit across cell barriers and help in passing signals or substances between the inside and outside of cells.
- **Defense** – Antibodies are proteins that recognize and neutralize foreign invaders like bacteria and viruses. Their structure allows them to bind to, and often neutralize, the intruding pathogen.

Proteins fulfill many more functions, such as support, movement, regulation, and storage. The myriad roles performed by proteins are intricately linked to their structure and amino acid sequence.

DNA, genes, and proteins

Proteins are encoded by DNA. Segments of the DNA called genes encode the sequence of amino acids that comprise a protein, and complex cellular mechanisms called transcription and translation convert the gene to its corresponding protein. This link between DNA and proteins has important implications for diseases; variations in genes potentially can give rise to defective proteins, which in turn can lead to diseases. Some of these variations are inherited, while others might arise from external factors like smoking or UV light exposure.

Proteins and human health

Proteins are essential for our health, but when they do not function as intended, medical issues can develop. A notable example is cystic fibrosis, caused by a mutation in the CFTR protein. This mutation disrupts the protein's role in ion regulation, resulting in thick mucus in the lungs and complications in other organs. Mutations to proteins such as p53 are associated with heightened cancer risk, and misfolding of proteins can lead to protein plaques such as those found in the brains of Alzheimer's patients.

Many therapeutic strategies target proteins to modulate health outcomes, even if the protein itself is not the underlying issue. For instance, statins work by reducing the amount of cholesterol produced by an enzyme in the liver. The enzyme is not itself problematic, but its inhibition is beneficial for some patients because it lowers cholesterol. Furthermore, proteins themselves can be therapeutics. Insulin is the classic example of this, and recent advances include immunotherapies that deploy antibodies to target cancer cells.

The COVID-19 pandemic brought proteins' relevance to human health squarely into the public eye. Vaccines targeting the SARS-CoV-2 spike protein have been pivotal in disease control. The emergence of mRNA vaccines, which get translated into proteins in human cells to elicit an immune response, marks a significant advancement.

Finally, proteins can function as components of diagnostic tools, such as the lateral flow tests used to detect infections during the pandemic. In this example, labeled antibodies bind to virus components present in a sample.

Another set of immobilized antibodies binds to the virus components farther along the strip, causing the labeled antibodies to concentrate in this region and form a visible line on the test.

The protein science revolution

Background

Protein research is well established. Elucidating protein sequences is straightforward, now that gene-sequencing technology is mature, and experimental techniques such as x-ray crystallography and nuclear magnetic resonance spectroscopy have allowed researchers to solve the three-dimensional structures of proteins. Structural characterization, however, takes time. Even with the advent of robotics and increasingly sophisticated x-ray beams,[3] it can take many months to establish a new protein structure. Naturally, researchers are interested in building tools and models that can predict protein structures, and a competition called CASP (critical assessment of structure prediction) runs every two years to assess the quality of the predictions.

AlphaFold

The first tremors of the protein revolution came at CASP 13 in 2018,[4] when Google DeepMind's AlphaFold entry substantially improved on previous attempts to predict protein structure, even for those proteins for which no known similar structures exist. In CASP 14, AlphaFold 2 made further substantial improvements, to the extent that some considered the prediction on par with experimental methods. AlphaFold utilizes deep-learning techniques to make structure predictions and incorporates information from known protein structures, as well as from protein sequences similar to the target sequence. AlphaFold's source code was released as open source, allowing the protein design community to leverage the tool in their own work and build upon existing methods.

From structure to function

Determining the structure of proteins is important, but manipulating them requires further research. We knew the structures of tens of thousands of proteins before AlphaFold, and so does knowing the structure of thousands more merely amount to scientific stamp-collecting? The key observation is that the same tools that help identify the structure of a protein could provide

insights into its functions and how to modulate those functions, and even allow for designing new and useful proteins unseen in nature.

Large language models

Researchers began testing whether the same tools used to predict protein structures could be repurposed to design them.[5] They found that structure-prediction tools could indeed be reverse-engineered to generate amino acid sequences. Others leveraged the idea that a sequence of amino acids is a bit like words in a sentence and used large language models akin to ChatGPT to make the link between amino acid sequence and the function of a protein.

Diffusion models

The most-recent breakthroughs in protein science come from the same domain as image-generating models like DALL-E. These can turn noisy versions of an image into its original form and then build on this by turning random noise into images representative of a text prompt. In a remarkable repurposing of this strategy, researchers have developed techniques that can generate novel protein structures by starting from random noise.[6] The same techniques can be used to generate proteins of a user-specified shape, much in the same way DALL-E can generate pictures based on user input.

Implications for scientific and medical research

Protein science has progressed with breathtaking speed since AlphaFold burst onto the scene. Scientists now have access to tools that were barely imaginable just a few years ago, and they are demonstrating that the outputs from these methods accurately reflect reality. It is reasonable to assume that the coming months and years will be punctuated by further breakthroughs, given the novelty of the technology and broad interest from the scientific community.

Proteins as therapeutic targets

Emerging evidence shows that this new breed of protein-design techniques can create proteins that bind to other proteins.[7] This is a key strategy for modulating protein function in biology, and tailoring this to specific needs may be a route to more precise and potent therapeutics. It is also possible to use predicted protein structures to elucidate potential drug designs.[8,9]

Defense against pandemics

Tentative reports tout the use of AI to understand the proteins associated with emerging viruses and to predict ways in which proteins in viruses such as SARS-CoV-2 could change to facilitate immune escape. These kinds of tools could become valuable weapons in preventing or mitigating future pandemics by accelerating efforts to combat viruses with effective drugs and vaccines.[10]

Elucidating new biological functions

The biological role of proteins is typically elucidated by techniques such as altering them and observing the impacts on the host organism or cells. For example, scientists can then drill down into why proteins play their roles – what they bind to and where they are found in the cell. Recent breakthroughs could move some of this work – such as finding proteins that bind to one another – out of the wet lab and into computers. A greater understanding of the roles of different proteins provides new insights into health and diseases, and new targets for therapeutic interventions.

Correcting protein defects

Understanding the implications of protein mutations or the ways in which they misfold can open avenues for new treatments. Knowing how cancer-causing mutations are impacting proteins can help with targeting rogue proteins and may form part of a personalized treatment strategy.

Better diagnostic tools

Natural proteins are incredibly effective at selectively sticking to things they are supposed to stick to. The human cell contains myriad things a given protein should not interact with, and relatively few that it should. Replicating this kind of specificity could, in principle, produce tests for any small molecules or biomarkers we want to detect. The interest around multi-cancer early detection tests for identifying abnormalities demonstrates the potential power of diagnostic tests; proteins can bind to a wider range of molecules than just DNA, and this could be the basis of future multi-disease tests that are powerful and non-invasive.

Reasons for caution

The protein science revolution is currently one of basic research, and we should be cautious about extrapolating breakthroughs into medical applications.

A good analogy for the protein science revolution of today may be the completion of the Human Genome Project in 2003. This was a monumental breakthrough in scientific research, which undoubtedly led to further breakthroughs that made a positive impact on human health. However, most mortality improvements since 2003 in advanced countries such as the U.K. have come from cardiovascular health, often largely attributed to statins rather than a better understanding of the genome. Another material driver of improvements has been reductions in lung cancer, a result of public health measures rather than a pharmacological breakthrough.

It is unclear whether a reductionist approach to understanding and manipulating biology is an efficient way to drive medical breakthroughs.[11] Researchers and pharmaceutical companies have long used their knowledge of protein structures to try to design new drugs that will bind to important sites and modulate function. Despite some notable successes, only about 8% of drugs that enter clinical trials end up registered as new drugs.[12] Of course, any increase in this efficiency is welcomed, but it is not clear how many newly registered drugs go on to materially increase the health span or lifespan of the population. In addition, promising results at the protein level might not be practical for clinical trials for many reasons; for example, therapeutics need to be synthesized, transported, and stored, and some proteins might not be amenable to this. Some promising therapeutics identified by computers will fail for these kinds of practical reasons, but we should acknowledge there have been AI-driven breakthroughs in this area too.[13]

Accessibility of treatments for the general public following any material breakthroughs must also be considered. Immunotherapies and personalized medicine are expensive, and some therapeutic strategies identified by new AI techniques may fall into the same category.

These concerns are not enough to dismiss the potential benefits promised by the protein science revolution; rather, they are a reminder to temper expectations for what it might deliver.

Implications for actuaries and biometric assumption-setting

What should actuaries make of the protein science revolution? Should they be rushing to revise mortality trend estimates, considering recent breakthroughs?

Headwinds and tailwinds

The major strength of the protein science revolution is also its major weakness: It is currently a breakthrough in fundamental science. This means the scope for impact is very large, given that basic science underpins a lot of medical science research. However, the "revolution" is also a long way removed from having a significant impact on human health, and we probably should not anticipate any material impacts on short-term mortality or morbidity improvement rates.

Meeting or beating expectations?

As with any medical breakthrough, we must consider whether it represents progress that was anticipated in our existing basis, or whether it is above and beyond our previous best estimate. A continuous stream of medical breakthroughs might be needed to meet a particularly bullish view of future mortality improvements, whereas a more pessimistic outlook could be exceeded with only modest progress.

Most mortality-improvement bases will assume positive improvements in future years, and even if it is impractical to assign these improvements to specific drivers with any precision, we must recognize the improvements have to come from somewhere. In the U.K. (and many other countries), mortality improvements have slowed, and it is unclear how mortality improvements will evolve as we exit the pandemic.[14,15] This is a stark reminder that rapid mortality improvements are not a fundamental law of nature. Either the drivers of previous improvements have run out of steam, or new drivers of negative improvements have emerged, or both.

Hope or hype?

The protein science revolution offers a lot of promise, but it would be very bold indeed to attribute assumed future improvements to this alone. Rather, it is part of a landscape of potential sources of future improvements – some of which may materialize, and some of which may not. Newspaper articles,

press releases, editorials, and opinion pieces from past decades feature a wealth of touted medical breakthroughs. Some of these will have had tangible impacts at the population level – statins, for example. Others will have made a positive difference in the lives of many individuals but not enough to alter population mortality trajectories (in isolation). Others will have fallen by the wayside, never to be mentioned again.

Conclusion

The seemingly random nature of what succeeds and what does not brings us neatly back to the main strength of the breakthrough: It could boost the efficiency of many threads of medical research and hence improve the odds of winning the "numbers game," in which many potential breakthroughs are required to improve population mortality. The odds of accumulating many small positive drivers of improvements, or finding the next blockbuster drug, improve with a greater number of candidates, or if the quality of individual candidates is slightly but tangibly better than those that came before.

It is unlikely that actuaries will be scrambling to revise their mortality assumptions based on basic scientific research. Rather, the potential for new sources of medical breakthroughs will be a welcomed source of optimism that potential drivers of future mortality improvements are still out there, even if recent improvements have been underwhelming.

AUTHOR

Craig Armstrong

Director, Actuarial Analytics, Risk and Behavioral Science

References

1. https://www.nature.com/articles/d41586-024-00173-x
2. There are more types of amino acids but just 20 that are commonly found in human proteins.
3. And the emerging role of cryo-electron microscopy.
4. https://doi.org/10.1002/prot.25823
5. https://www.nature.com/articles/s41587-022-01618-2
6. https://www.nature.com/articles/s41586-023-06415-8
7. https://www.nature.com/articles/s41586-023-06953-1
8. https://pubs.rsc.org/en/content/articlelanding/2023/sc/d2sc05709c
9. https://www.nature.com/articles/s41596-023-00885-w
10. https://www.nature.com/articles/d41586-023-03201-4
11. A reductionist approach is one that looks to study the component parts of a system in isolation, whereas biology is a product of many

complex systems of individual components. Studying component parts, such as individual proteins, risks overlooking important insights that a holistic (or top-down) approach might yield.

12. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10173933
13. https://pubs.acs.org/doi/full/10.1021/acscatal.3c02743
14. https://www.rgare.com/knowledge-center/article/beyond-the-pandemic-a-driver-based-view-of-mortality-rates-and-their-implications-for-actuaries
15. https://www.rgare.com/knowledge-center/article/u.k.-mortality-projections--practical-implications-of-cmi-proposals