helium) through nuclear fusion. The metals are ejected into the surroundings through stellar winds (outflows of gas from the star) and supernova explosions, which can occur at the end of a star's life. The next generation of stars forms from interstellar material that is mixed with this ejected matter. The metal content of stars can therefore tell us indirectly about their stellar predecessors[8].

The gravitational pull of the GCDs' dark-matter halos caused gas to collapse into stars earlier than in globular clusters, but the objects were not massive enough to initiate further inflows of gas. This meant that after the first burst of star formation, there was not enough fuel for a subsequent generation of stars in the GCDs (Fig. 1) – unlike in regular dwarf galaxies that, in some cases, host continuous star formation up to the present day.

The GCDs had lower abundances of metals and formed, on average, earlier than did the globular clusters in the simulation. The authors predict that if GCDs are found, they will also be older than the oldest dwarf galaxies and might contain evidence of the earliest population of stars, which would be composed almost entirely of hydrogen and helium[9]. The existence of such 'metal-free' stars in the early Universe is widely accepted, but they have never been observed.

Although this study uses world-class computational models, the results still leave open questions. Current cosmological models cannot reproduce some observed properties of globular clusters, such as the stellar abundances of light chemical elements, including sodium and aluminium[10]. Furthermore, the simulations did not include globular clusters that form in galaxies of higher mass than dwarf galaxies[11], which would be computationally expensive to include in the model. Approaches such as graphic- processing-unit acceleration or machine-learning techniques could be explored to speed up the calculations.

However, the authors suggest that it might be possible to put their results to the test by looking in our own 'cosmic backyard'. They point out that some of the Local Group dwarf galaxies, as well as the Milky Way's C-19 – a very metal-deficient chain or 'stream' of stars – might be GCDs. If this is the case, these objects present an avenue for astronomers to constrain the properties of dark matter and study the nature of the first stars, giving scientists a better understanding of the fundamental processes that have shaped our Universe.

Natalia Lahén is at the Max Planck Institute for Astrophysics, Garching bei München 85748, Germany.
e-mail: nlahen@mpa-garching.mpg.de

1. VandenBerg, D. A. *et al. Astrophys. J.* **775**, 134 (2013).
2. Forbes, D. A., Read, J. I., Gieles, M. & Collins, M. L. M. *Mon. R. Astron. Soc.* **481**, 5592–5605 (2018).
3. Taylor, E. D. *et al. Nature* **645**, 327–331 (2025).
4. Agertz, O. *et al. Mon. Not. R. Astron. Soc.* **491**, 1656–1672 (2020).
5. Sameie, O. *et al. Mon. Not. R. Astron. Soc.* **522**, 1800–1813 (2023).
6. Gutcke, T. A. *Astrophys. J.* **971**, 103 (2024).
7. Calura, F. *et al. Astron. Astrophys.* **698**, A207 (2025).
8. Heger, A. & Woosley, S. E. *Astrophys. J.* **567**, 532 (2002).
9. Frebel, A. & Norris, J. E. *Annu. Rev. Astron. Astrophys.* **53**, 631–688 (2015).
10. Carretta, E., Bragaglia, A., Gratton, R. & Lucatello, S. *Astron. Astrophys.* **505**, 139–163 (2009).
11. Zinn, R. *Astrophys. J.* **293**, 424 (1985).

The author declares no competing interests.

Immunology

Common feature found for allergy-triggering proteins

Bart N. Lambrecht

Why do only certain molecules drive allergies? A feature now identified among some allergy-causing proteins is their ability to form pores in cells. **See p.475**

Everyone knows someone with an allergy, and the chances are that you're allergic to something yourself. Allergies arise as a result of the immune system overreacting to allergens: otherwise harmless substances such as dust mites, pollen or certain foods[1]. Various substances can act as allergens, and one of the enduring puzzles in allergy research is why only some of the many proteins we encounter reliably trigger allergic responses[2]. On page 475, Shi *et al.*[3] address part of this mystery by identifying a property shared by some allergens – the ability to form large pores in epithelial cells that line airways. These pores enable calcium to enter and trigger the release of molecules that alert the immune system to danger and can promote allergy[4].

"This supports the possibility that pore formation is a widespread mechanism that triggers allergies."

Allergic reactions often occur rapidly, with redness, swelling and intense itching shortly after exposure. These symptoms are driven mainly by release of the molecule histamine from immune cells called mast cells (Fig. 1). These are activated when allergens react with a type of antibody called immunoglobulin E, which binds to a receptor on mast cells, causing the release of histamine[1].

Although most allergic reactions are mild, severe responses can be life-threatening, as can occur with allergies to substances such as peanuts. Chronic exposure to allergens can lead to inflammatory diseases such as asthma or eczema, which impair quality of life. In these chronic conditions, histamine-blocking medications are often insufficient for successful treatment, because other immune cells, particularly eosinophils, contribute to sustained inflammation[1].

Allergic responses are associated with a pathway known as type 2 inflammation[5]. This is characterized by the presence of immunoglobulin E, eosinophils and immune-signalling molecules (cytokines) such as IL-4, IL-5 and IL-13. Type 2 inflammation is thought to have evolved to counter venoms and worm (helminth) infections, but when this protective system is misdirected, allergies arise[6]. To understand the root causes of these conditions, researchers are seeking to define the molecular triggers that drive type 2 inflammation.

Despite decades of study, it remains unclear what makes a protein an allergen. Allergens and venoms encompass many unrelated protein families, and no unifying structural motif has been found previously[2]. Some allergens function as enzymes called proteases, which degrade proteins essential for maintaining the external barrier formed by epithelial cells in the skin or lungs. Such damage aids allergen entry and, crucially, stimulates epithelial cells to release cytokines called alarmins[7]. These cytokines drive immune signalling that promotes the production of immunoglobulin E and type 2 inflammation.

One well-characterized alarmin is IL-33, which is normally stored inside cells and released only when tissue is under stress[8]. IL-33 lacks the usual protein segment that enables molecules to be transported from a cell by vesicle-mediated secretion, and its release mechanism is a long-standing enigma. Extensive cell death could explain IL-33 leakage after helminth infections, but most allergens do not

From the archive

A persuasive first-hand encounter with mosquitoes, and an effort to standardize the printing of mathematical formulae.

100 years ago

A new phase in the control of the mosquito problem was reached last week when members of the Section of Zoology of the British Association attended the opening of the British Mosquito Control Institute which has been founded and equipped by Mr. J. F. Marshall ... at Hayling Island ... As a result of the preventive work undertaken, which has consisted largely of draining the inter-tidal areas, Hayling Island, which formerly was infested with 'saltmarsh' mosquitoes, has now been practically cleared ... Sir Ronald Ross, in opening the Institute, described his own experiences at Hayling Island some three years ago. He had, he admitted, been sceptical as to the accusations made against *Ochlerotatus detritus*, the salt-water mosquito. Mr. Marshall, however, had convinced him by taking him to a part of his garden and inviting him to observe the insects flying about in broad daylight. In the course of his observations, Mr. Marshall had proved his point by directing his attention to the fact that three of the insects were simultaneously extracting blood from the back of his neck.

From *Nature* 12 September 1925

150 years ago

[T]he Committee on Mathematical Printing was appointed to report on mathematical notation and printing, with the view of leading mathematicians to prefer in optional cases such forms as are more easily put into type, and of promoting uniformity of notation ... The report ... also attached diagrams showing the mechanical operation of setting up mathematical expressions in type, so that when there were two forms equally satisfactory from the mathematical point of view, writers might choose the one that would give the printer less trouble; as everything that tended to cheapen mathematical printing tended to the spread of the science. With regard to notation ... the matter should be discussed by a larger committee.

From *Nature* 9 September 1875

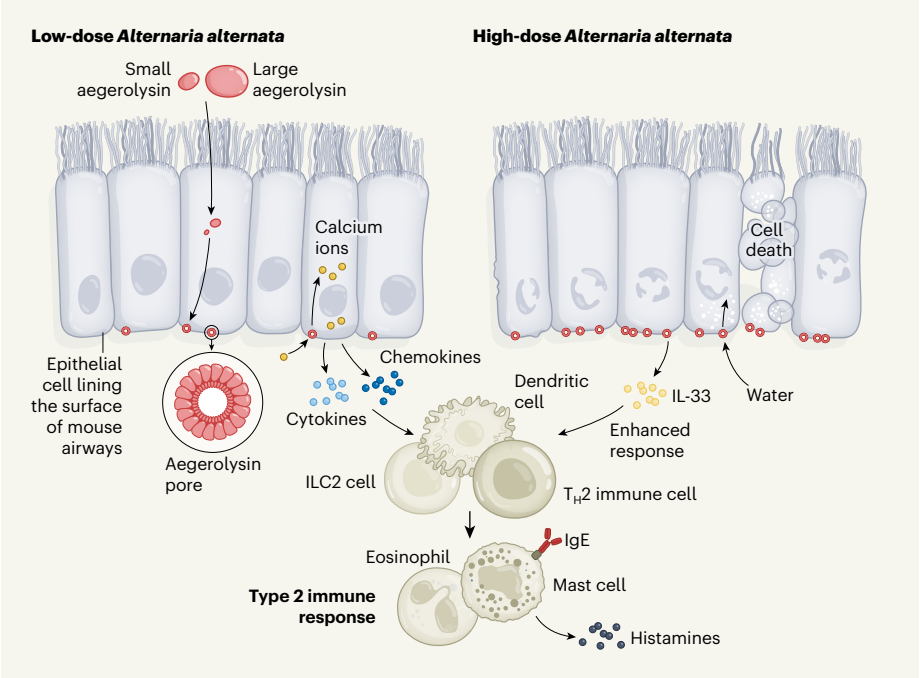

Figure 1 | Pore-forming proteins cause allergies. It was unclear why only certain types of protein cause allergies. Shi *et al.*[3] examined how the fungus *Alternaria alternata* triggers allergies in mouse airways. They found that these allergies are induced by two proteins (one small and one large) called aegerolysins. These form large transmembrane pores comprising many aergerolysin proteins. Where these pores form in the cell membrane is unknown. If the animals received a low dose of fungal extracts, calcium ions entered the pores, and the epithelial cells lining the airways released immune-signalling molecules called cytokines and chemokines. These attracted immune cells called dendritic cells (DCs), type 2 innate lymphoid (ILC2) cells and T helper 2 (T_H2) cells. These cells drove an allergic response called type 2 inflammation, mediated by eosinophils and mast cells: immune cells that release the molecule histamine when a receptor binds antibody-related immunoglobin E (IgE) proteins. If the animals received a high dose of fungal extracts, the immune-signalling molecule IL-33 was released, which resulted in a more intense allergic response. Water entered the pores, resulting in epithelial cell death.

cause such overt damage. This raises the question of how IL-33 is released during allergies.

Shi *et al.* tackled this question using allergen-containing extracts from a fungus called *Alternaria alternata*. The authors tested which components trigger IL-33 release and calcium influx in epithelial cells, focusing on the two key signals required for type 2 responses *in vivo*. They found that two proteins were necessary. These are small and large forms of a fungal protein called aegerolysin, which is part of a protein family known for forming pores in cell membranes.

Using biochemical experiments and cryo-electron microscopy, the authors showed that these aegerolysins assemble into pores in the cell membrane, each consisting of 36 small subunits and 18 large ones. At high concentrations of protein, the pores induced inflammation, cell death mediated by water influx, and leakage of IL-33 and other alarmins. At lower concentrations, the pores triggered calcium influx without killing cells, leading to inflammatory signalling (Fig. 1). When mice inhaled these proteins, they developed features that are a hallmark of allergic asthma, including immunoglobulin E production and inflammation mediated by eosinophils. Crucially, genetic disruption of either pore-forming protein in *A. alternata* prevented the fungal extracts from inducing type 2 immunity in mice.

The authors extended their findings by showing that unrelated pore-forming toxins, from the fungus *Aspergillus niger* or the sea anemone *Actinia equina*, also drive IL-33 release and allergic responses. Cleavage of IL-33 by proteases can make it more active[7]. However, the authors did not address whether IL-33 is further activated after release and, if so, how. It is unclear whether pore formation could initiate such activation, perhaps by release of proteases.

These findings suggest that membrane perforation might be a common characteristic that explains why structurally diverse proteins act as allergens. Pore-forming proteins such as gasdermins, which are normally present in human cells, have been implicated in IL-33 release, type 2 inflammation and asthma[9,10]. This supports the possibility that pore formation is a widespread mechanism that triggers allergies[9,10]. Understanding the structural and molecular determinants of pore formation, and its immunological consequences, might offer a path towards developing therapeutics

that could intervene at the earliest steps of allergic reactions.

Bart N. Lambrecht is at the VIB Center for Inflammation Research, Ghent University, 9000 Ghent, Belgium.
e-mail: bart.lambrecht@ugent.be

1. Gao, Y.-D. *et al. Allergy* **80**, 1848–1877 (2025).
2. Kopp, E. B., Agaronyan, K., Licona-Limón, I., Nish, S. A. & Medzhitov, R. *Immunity* **56**, 687–694 (2023).
3. Shi, K. *et al. Nature* **645**, 475–483 (2025).
4. Van Dyken, S. J. *et al. Nature Immunol.* **17**, 1381–1387 (2016).
5. Lambrecht, B. N., Ahmed, E. & Hammad, H. *Nature Immunol.* https://doi.org/10.1038/s41590-025-02212-9 (2025).
6. Marichal, T. *et al. Immunity* **39**, 963–975 (2013).
7. Cayrol, C. *et al. Nature Immunol.* **19**, 375–385 (2018).
8. Lambrecht, B. N., Hammad, H. & Fahy, J. V. *Immunity* **50**, 975–991 (2019).
9. Yang, L. *et al. Immunity* **57**, 1056–1070 (2024).
10. Zhao, M. *et al. Immunity* **55**, 623–638 (2022).

The author declares no competing interests.
This article was published online on 30 July 2025.

Protein engineering

AI expands the CRISPR-protein universe

Pascal Notin

A generative artificial-intelligence tool has designed a synthetic CRISPR system that successfully edits human DNA and sharply reduces off-target effects. **See p.518**

Despite emerging clinical successes, current genome editors suffer from off-target effects and can trigger unwanted responses from the immune system, limiting their broader therapeutic applications. On page 518, Ruffolo *et al.*[1] present OpenCRISPR-1, the first AI-generated CRISPR–Cas protein to edit human DNA successfully. The work demonstrates how machine learning can be used to engineer functional biological systems that extend beyond those found in nature.

CRISPR is a prime example of serendipity in modern biology. What began with an observation[2] of unusual repeating DNA sequences in the bacterium *Escherichia coli* in 1987 led to the discovery of a sophisticated adaptive immune system. Bacteria capture viral DNA fragments as 'spacers' in repetitive genomic sequences called CRISPR arrays. This creates a genetic memory that directs CRISPR-associated (Cas) proteins that cleave DNA to destroy matching viruses on reinfection.

In 2012, Emmanuelle Charpentier and Jennifer Doudna demonstrated that Cas9 from the bacterium *Streptococcus pyogenes* (SpCas9) could be repurposed into a programmable gene-editing tool[3], with guide RNAs that direct Cas9 to precise genomic locations (Fig. 1a). Subsequent demonstrations of functionality in human cells[4] opened the door to therapeutic applications, culminating in regulatory approvals for therapies, such as Casgevy for treating sickle-cell disease[5] (see go.nature.com/45ngyfk).

However, for all its power, SpCas9 remains a wild tool rather than a tame one. It evolved for bacterial defence, not for the precision required in human therapeutic contexts. SpCas9 can tolerate mismatches between its guide RNA and DNA sequences, leading to off-target cuts that pose serious safety concerns[6]. As a protein from a common pathogen, SpCas9 triggers pre-existing immunity in most individuals, potentially neutralizing the effects of therapies[7]. Its large size makes it difficult to package into convenient delivery systems, such as adeno-associated virus vectors. Furthermore, SpCas9 can make cuts only adjacent to short DNA motifs called PAM sequences, which limits the regions it can target to a subset of genomic sites[8].

> **"The authors' approach could yield editors that are less visible to the immune system."**

These shortcomings are most acute for *in vivo* editing, in which cells are modified directly inside the body, unlike *ex vivo* approaches in which cells are removed, edited and returned to the individual. The success of *ex vivo* therapies such as Casgevy has intensified the search for better genome editors for *in vivo* applications, which offer greater scalability but demand near-perfect specificity and a minimal immunogenicity (the ability to trigger immune responses).

Efforts to overcome these limitations have used several engineering strategies. Rational design uses knowledge about the structure of Cas9 to make specific mutations, yielding high-fidelity variants, such as SpCas9-HF1, that trade on-target activity for reduced off-target cleavage[9]. Directed evolution mimics natural selection through the screening of vast mutant libraries[10], but it remains labour-intensive and explores fairly restricted regions of sequence space around the starting protein. Bioprospecting mines microbial biodiversity for naturally superior enzymes, but any discovered enzyme probably still carries the same fundamental problems as SpCas9 in terms of off-target effects and immunogenicity. The dream of a bespoke editor – efficient, specific, deliverable and non-immunogenic – has so far remained elusive under conventional approaches.

Ruffolo and colleagues pursue a fourth paradigm: *de novo* design using generative artificial intelligence (AI; Fig. 1b). Their approach uses protein language models, which are trained on large protein-sequence databases[11]. Analogous to large language models such as GPT, which learn patterns of human language[12], protein language models process hundreds of millions of examples to learn the implicit 'grammar' of protein evolution – the complex statistical relationships between amino acids that characterize functional natural proteins.

The authors recognized that any AI model's performance is fundamentally limited by the quality and scale of its training data. Instead of relying on existing databases, they undertook a considerable data-mining effort to construct the CRISPR–Cas Atlas, a resource built by sifting through 26.2 trillion bases of assembled microbial genomes. This effort yielded more than 1.2 million CRISPR 'operons' – functional units that include Cas protein sequences, CRISPR arrays, CRISPR-associated RNAs and PAMs. It also resulted in a fourfold expansion in the number of Cas9 sequences compared with those recorded in the protein database UniProt.

Ruffolo *et al.* next implemented a hierarchical training strategy. A protein language model called ProGen2 – which was pretrained on hundreds of millions of protein sequences from the databases UniRef and BFD – was fine-tuned on the CRISPR–Cas Atlas to learn the particular sequence constraints that underlie the function of Cas9 proteins. The resulting model was then used to generate a repertoire of synthetic Cas-like proteins that were almost five times more diverse than known variants and included thousands of candidates unlike any variants found in nature.

Are all of these proteins functional? Probably not. Do these types of model-generated repertoire contain new sequences with desirable qualities for target applications? To demonstrate that this is the case, the researchers further fine-tuned their model by training it exclusively on around