AIMS Biophysics, 13(2): 225−230.
DOI: 10.3934/biophy.2026014
Received: 18 June 2026
Revised: 22 June 2026
Accepted: 23 June 2026
Published: 24 June 2026

http://www.aimspress.com/journal/biophysics

Editorial

Microproteins: How DNA's dark matter is reshaping the drugs of the future

Giovanni Colonna[1,*] and Domenico Aprile[2,*]

[1] Medical Informatics Unit, AOU Università degli Studi della Campania, L. Vanvitelli, Naples, Italy
[2] Università degli Studi Link Campus University, Dipartimento di Scienze della Vita, della Salute e delle Professioni Sanitarie, Rome, Italy

* **Correspondence:** Email: giovanni.colonna@unicampania.it, d.aprile@unilink.it.

Preface

Approximately 1000 to 4000 genes that encode small proteins (less than 100 amino acids) are still officially uncataloged in the human proteome. These sequences, often called smORFs (small open reading frames), have eluded identification for decades, which has relegated them to the simple category of "noncoding RNA" [1]. These sequences have eluded identification for three main reasons:
1. Arbitrary criteria: Historically, bioinformaticians have set a minimum threshold of 100 amino acids to exclude genetic "noise".
2. Technical difficulties: Small proteins degrade rapidly and are difficult to detect using standard mass spectrometry.
3. Misclassification: Many are hidden within regions previously classified as "noncoding RNA" (ncRNA) or in segments overlapping with other genes.

Today, we know that this "genetic noise" hides precision regulators critical to molecular biology, development, and metabolism [2]. Integrating polymer physics (physical-chemical properties) and deep learning (AlphaFold) is one of the few ways to scale up the study of the 1000−4000 missing microproteins. Without these models, we would have to study them one by one with crystallography, which would take centuries.

1. Key roles of physicochemical and biophysical properties

Many microproteins are intrinsically disordered (IDP); they do not possess a defined three-dimensional structure until they encounter their molecular partner [3]. Traditional X-ray crystallography would require centuries to map the entire catalog.

In this scenario, analyzing the physicochemical properties of the amino acid sequence is among the best predictive tools for guiding research. Key biophysical parameters include:

• Hydrophobicity: This allows us to predict the microprotein's propensity to integrate into lipids in cell membranes.

• Net electrical charge: This helps identify the affinity and localization in specific cellular compartments, such as the nucleus or cytosol.

• Amino acid residue content: This reveals the flexibility, chemical stability, and folding capacity of the amino acid chain.

• Alpha-helical propensity: Identifies structural motifs typical of ion channel modulators, guiding the understanding of their function.

2. Artificial intelligence as a biological accelerator

Integrating biophysical properties and deep learning has eliminated the bottleneck that hindered the transition from genomics to functional biology [4]. Artificial intelligence operates on multiple levels:

• Complex prediction (AlphaFold-Multimer): AI algorithms do not analyze the microprotein in isolation but process its chemical and physical properties to simulate how it changes conformation when it fits into the molecular "pockets" of larger partner proteins.

• Computational interactomics: AI rapidly screens microproteins against the entire known proteome, identifying potential biological complexes before they even enter the lab [5].

• De novo computational design: Advanced algorithms enable the design of synthetic microproteins that perfectly mimic natural biophysical mechanisms, accelerating pharmaceutical research by years.

3. From in silico models to laboratory validation

The modern scientific workflow for mapping the "dark side" of the genome follows a precise synergy, as follows:

Bioinformatics filter and AI ➔ physico-chemical property analysis ➔ structural prediction (AlphaFold) ➔ experimental validation (BioID/Ribo-seq).

In predictions, AI and physico-chemical analysis suggest where the microprotein is located and with whom it interacts. Predicted models are verified in vivo and validated by using proximity labeling techniques (such as BioID), which mark nearby proteins in the cell, and Ribo-seq, which shows the effective translation of the RNA by following the rhythm of the triplet periodicity [6].

The following diagram illustrates the process of molecular translation, the fundamental biological step traced by Ribo-seq to show that a "junk" sequence (smORF) is indeed read by ribosomes to produce a stable microprotein:

Schematic of the integrated workflow from screening to drug

[HUMAN GENOME] ➔ 98% DNA previously considered "junk"

▼

[BIOINFORMATIC FILTER] ➔ Detection of smORFs (<100 amino acids)

▼

[TRANSLATIONAL VALIDATION] ➔ Ribo-seq (mapping of ribosomes in action)

▼

[BIOPHYSICAL CHARACTERIZATION] ➔ Analysis of hydrophobicity, charge, and IDP fluctuations

▼

[ARTIFICIAL INTELLIGENCE] ➔ AlphaFold-Multimer (interactomic prediction)

▼

[BIOMEDICAL TARGETING] ➔ Design of micro-vaccines and therapeutic peptides

4. Emblematic examples of success

• DWORF: Initially identified as a noncoding RNA, conservation analysis and in silico interaction studies revealed its ability to bind to the calcium pump SERCA, increasing cardiac contractility.
• PIGBOS: Composed of only 54 amino acids, PIGBOS has been shown, according to chemical and physical models, to have a strong transmembrane domain, allowing us to identify its role in the mitochondrial membrane in regulating cellular stress.
• MRLN (Myoregulin): Bioinformatic analysis revealed a single alpha-helical structure (typical of ionic modulators), confirming its key role in muscle efficiency and energy expenditure.
• MRI-2: A microprotein that regulates DNA repair. Simulation of its biophysical interlocking with the Ku70/Ku80 complex has accelerated the understanding of how tumors use it to resist chemotherapy.

5. Drug design of the future: a therapeutic revolution

The pharmaceutical industry is investing heavily in microproteins and their synthetic analogs because they overcome the structural limitations of traditional drugs, combining the specificity of monoclonal antibodies with the versatility of small-molecule chemicals [7].

5.1. Overcoming the "undruggable" effect (inaccessible targets)

Many human proteins implicated in serious diseases have flat surfaces or complex structures that traditional small-molecule chemistry cannot bind tightly to.

• Structural flexibility: Microproteins, due to their intrinsically disordered nature (IDP) or specific alpha-helical propensities, possess unique flexibility.

• Interlocking mechanism: Acts as a "biomimetic molecular key", adapting its shape to fit perfectly into structural pockets previously considered unreachable [8].

5.2. Optimizing pharmacokinetic and biophysical properties

The design of microprotein-based drugs focuses on the targeted manipulation of their chemical and physical parameters:

• Tissue penetration and the blood–brain barrier: Compared to large antibodies, the small size of microproteins allows for exceptional diffusion in dense tissues (such as solid tumors). By modifying their net electrical charge and surface hydrophobicity, researchers are designing micropeptides that can cross the blood–brain barrier to treat neurodegenerative diseases.

• Low toxicity and biocompatibility: Being composed of natural amino acids, the human body can metabolize and eliminate them. This dramatically reduces the risk of liver or kidney toxicity and minimizes side effects due to nonspecific off-target binding.

• Stability engineering: Natural microproteins degrade rapidly within the cell. Through computational design, scientists introduce artificial chemical modifications (such as peptide cyclization or the insertion of non-canonical amino acids) to "shield" them from proteolytic enzymes, increasing their biochemical stability and residence time in the organism [9].

5.3. Cancer immunotherapy and microprotein-derived vaccines

Cancer immunotherapy and "micro-vaccines" are among the most advanced frontiers in the development of personalized therapeutic vaccines against cancer. They use hidden neoantigens: some tumors survive by expressing specific microproteins that standard tests cannot detect. The mechanism of action comprises identifying these tumor smORFs via Ribo-seq [6] and, when possible, synthesizing peptide "micro-vaccines" in the laboratory. Once administered, these fragments train the patient's immune system to recognize and selectively destroy only malignant cells displaying that specific microprotein on their surface, sparing healthy tissue.

6. Conclusions

In conclusion, global interest in microproteins has exploded over the last 5–10 years, moving from academic curiosity to a true biomedical frontier. The international scientific community is investing heavily in microproteins due to the following main reasons:

• New therapeutic targets: Many of these proteins regulate vital processes such as DNA repair, muscle contraction, and inflammation. Because they are small, they are ideal models for designing new drugs or can act as drugs (therapeutic peptides).

• The "dark matter" of the genome: Identifying thousands of new genes means rewriting biology

textbooks. It is the missing piece to truly complete the mapping of the human genome, which began in 2000.

• Oncology: It has been discovered that some tumors survive due to specific microproteins that standard tests cannot detect. Targeting them could open new avenues for immunotherapy.

• Evolution: The study of smORFs allows us to understand how genes arise "from scratch" (de novo gene birth), a process previously thought to be extremely rare.

At the institutional level, projects like OpenProt and international consortia [10] are working to standardize the cataloging of these small peptides, while major pharmaceutical companies are creating divisions dedicated to screening them. They play a role in specific diseases; microproteins are not simple byproducts, but "precision" regulators, currently used in oncology and neurodegenerative and cardiovascular diseases. They are the potential drugs of the future, representing a revolutionary class of drugs thanks to their high specificity. Unlike small chemical molecules (which often target similar sites, leading to side effects), they mimic the body's natural mechanisms with extreme precision. Some startups are developing "micro-vaccines" that train the immune system to recognize specific microproteins found only on the surface of cancer cells.

Use of generative-AI tools declaration

During preparation of this manuscript, the authors used generative-AI tools only for language editing and improvement of readability. The authors reviewed and edited all AI-assisted output and take full responsibility for the content of the manuscript.

Conflict of interest

Giovanni Colonna and Domenico Aprile are the guest editor of special issue "The Dark Side of the Human Genome: Microproteins" for AIMS Biophysics and were not involved in the editorial review or the decision to publish this article. The authors declare no conflict of interest.

Author contributions

Conceptualization, Giovanni Colonna and Domenico Aprile; writing - original draft, Giovanni Colonna and Domenico Aprile; writing - review and editing, Giovanni Colonna and Domenico Aprile. Both authors have read and agreed to the final version of the manuscript.

References

1. Deutsch E W, Lindskog C, Budnik B, et al. (2026) The 2025 report on the human proteome from the HUPO human proteome project. *J Proteome Res* 25: 539–555. https://doi.org/10.1021/acs.jproteome.5c00759
2. Ko S H, Cho B L, Shin D (2025) Microproteins in metabolic biology: emerging functions and potential roles as nutrient-linked biomarkers. *Int J Mol Sci* 26: 11883. https://doi.org/10.3390/ijms262411883
3. Mohsen JJ, Martel AA, Slavoff SA (2023) Microproteins—Discovery, structure, and function. *Proteomics* 23: 2100211. https://doi.org/10.1002/pmic.202100211

4. AlQuraishi M, Sorger PK (2021) Differentiable biology: using deep learning for biophysics-based and data-driven modeling of molecular mechanisms. *Nat Methods* 18: 1169–1180. https://doi.org/10.1038/s41592-021-01283-4
5. Miller B, de Souza EV, Pai VJ, et al. (2025) ShortStop: a machine learning framework for microprotein discovery. *BMC Methods* 2: 16. https://doi.org/10.1186/s44330-025-00037-4
6. Khoroshkin M, Buyan A, Dodel M, et al. (2024) Systematic identification of post-transcriptional regulatory modules. *Nat Commun* 15: 7872. https://doi.org/10.1038/s41467-024-52215-7
7. Porozova NO, Petukhova NV, Gureev MA, et al. (2025) Beyond monoclonal antibodies: constraints and the case for alternative PD-1/PD-L1-targeting formats. *Front Immunol* 16: 1729468. https://doi.org/10.3389/fimmu.2025.1729468
8. Du E, Tang X, Zhang W, et al. (2025) Emerging mechanically interlocked cages. *Nat Rev Chem* 9: 506−522. https://doi.org/10.1038/s41570-025-00721-7
9. Baumgart M, Röpke M, Mühlbauer ME, et al. (2021). Design of buried charged networks in artificial proteins. *Nat Commun* 12: 1895. https://doi.org/10.1038/s41467-021-21909-7
10. Perez-Riverol, Y (2022) Proteomic repository data submission, dissemination, and reuse: key messages. *Expert Rev Proteomic* 19: 297−310. https://doi.org/10.1080/14789450.2022.2160324

AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)