REVIEW ARTICLE

NUTRITION IN MEDICINE

Chemical Complexity of Food and Implications for Therapeutics

Giulia Menichetti, Ph.D.,[1-3] Albert-László Barabási, Ph.D.,[2,4] and Joseph Loscalzo, M.D., Ph.D.[1,5]

Author affiliations are listed at the end of the article. Dr. Loscalzo can be contacted at jloscalzo@bwh.harvard.edu or at Brigham and Women's Hospital, NRB, 77 Avenue Louis Pasteur, Room 630H, Boston, MA 02115.

N Engl J Med 2025;392:1836-45.
DOI: 10.1056/NEJMra2413243

Poor nutrition is a leading cause of illness in the United States, contributes to more than half a million deaths each year[1,2] and affects half of all American adults who have one or more preventable noncommunicable diseases, such as cardiovascular disease, hypertension, type 2 diabetes mellitus, cancer, and poor bone health.[3] Beyond increasing the risk of disease, poor nutrition has broad societal effects, driving up health care costs and decreasing productivity, with expenditures for obesity alone reaching $173 billion annually.[2] By contrast, adopting a healthy diet and lifestyle can significantly counteract even a strong genetic predisposition to coronary heart disease and reduce the relative risk by nearly 50%.[4]

Growing evidence further highlights the importance of dietary quality in disease prevention, particularly amid a global surge in early-onset cancers — an emerging research priority identified by the National Cancer Institute.[5] Indeed, since the 1990s, the incidence of cancer among adults under the age of 50 years has risen worldwide, despite unchanged hereditary cancer rates, a finding that underscores the influence of environmental and lifestyle factors.[6] Early-life exposures have shifted markedly over the past several decades and reflect trends toward overweight, obesity, and Western-style diets, even among children and adolescents.[6]

These epidemiologic patterns highlight the limitations of focusing on genetics alone. Although the Human Genome Project greatly expanded our understanding of hereditary contributions to disease, genes appear to account for approximately 10% of disease risk, with the remainder determined largely by environmental and dietary factors.[7] To build on the wealth of discoveries from the Human Genome Project, we must recognize that food is not just a source of calories and vitamins but also carries a vast array of chemicals with health implications that extend beyond those currently known. Indeed, food chemicals can regulate the activity of human and microbial proteins, a fact that brings nutrition studies closer to the resolution typical of "omics" research. Filling this fundamental knowledge gap will help us redefine how we measure dietary quality and enhance the precision of dietary interventions. As the Food Is Medicine initiative progresses,[8] promising better health through nutritious food, it becomes crucial to map food molecules and accurately identify the mechanisms that contribute to our well-being.

Here, we review the exceptional chemical complexity of our diet and discuss its implications for health. Our analysis is anchored in the Nutrition Dark Matter (NDM) library, a curated and harmonized database assembled by our group, which catalogues thousands of distinct food chemicals, along with their unique structural identifiers, physicochemical properties, and food annotations. Drawing from

KEY POINTS

CHEMICAL COMPLEXITY OF FOOD

- More than 139,000 molecules in food, termed "nutrition dark matter" (NDM), hold immense, largely untapped therapeutic potential.
- Food molecules affect nearly half the human proteome and play a role as broad modulators of biologic processes.
- Approximately 2000 food molecules are already used as drugs, which highlights the pharmaceutical relevance of food chemicals.
- Advanced artificial intelligence (AI) tools and network medicine frameworks could help researchers decode the health relevance of NDM and enable predictions of molecular targets and biologic mechanisms.
- The lack of systematic food–chemical mapping limits progress; strategic funding and scalable AI-based approaches are urgently needed.
- Mapping NDM could revolutionize dietary science and accelerate drug discovery, which would complement the achievements of the Human Genome Project.

diverse biologic repositories such as PubChem,[9] DrugBank,[10] and BRENDA (Braunschweig Enzyme Database),[11,12] the NDM library provides an updated perspective on the vast biochemical landscape of foods and the emerging opportunities to explore novel therapeutics. Evidence suggests that more than 139,000 chemicals in food together modulate a large number of human proteins. Approximately 2000 of these food molecules are currently used as drugs. Thus, there is an enormous pool of chemicals with subcellular roles that remain unknown, and this pool may serve as the starting point for future therapies. Finally, we also explore the current roadblocks to addressing this massive gap in our understanding of diet and potential avenues for overcoming them (Fig. 1).

MOLECULAR TAXONOMY OF FOOD

Nutrients, defined as the chemical substances essential for sustaining the basic functions of the body, are typically categorized into six major classes: carbohydrates, lipids, proteins, vitamins, minerals, and water.[13] National food databases, such as that of the U.S. Department of Agriculture (USDA), focus primarily on these chemical classes, providing detailed information on up to 150 essential micronutrients and macronutrients, mostly linked to energy intake and metabolism. Although this core nutritional panel has long shaped our understanding of food, it remains limited in capturing the full chemical richness within the staple ingredients of human diets. Indeed, the raw ingredients of food, representing organisms that were once alive, have exceptional chemical diversity. This is particularly true for plants that, unable to use locomotion to escape predators, have developed extensive secondary metabolic pathways that generate distinctive textures, colors, and smells that fend off some predators while attracting others. These metabolic pathways produce thousands of polyphenols, terpenoids, and alkaloids that cannot be generated by the human metabolism yet are essential for our well-being.

Despite increased investment in research since 2003 to characterize small-molecule content,[14-17] the full range of compounds in food remains largely unknown. In 2019, this realization led us to coin the term "nutrition dark matter", which highlights the vast number of food molecules largely untracked by food composition databases and epidemiologic research.[14,18] Since then, we have intensified efforts to document the presence of small molecules in food systematically. These low-molecular-weight organic compounds, typically ranging in mass from 50 to 1500 daltons, have druglike characteristics and often serve as substrates or products in biologic processes.[19]

We have now catalogued 139,443 chemicals in the NDM library, each identified by a valid international chemical identifier (InChIKey), distributed across more than 3000 common foods and 17,000 species in the National Center for Biotechnology Information taxonomy.[20] Creating this resource required aggregating and disambiguating data from various scientific fields, including annotations drawn from a vast array of specialized literature,[21] untargeted mass-spectrometry experiments,[22] and diverse compo-

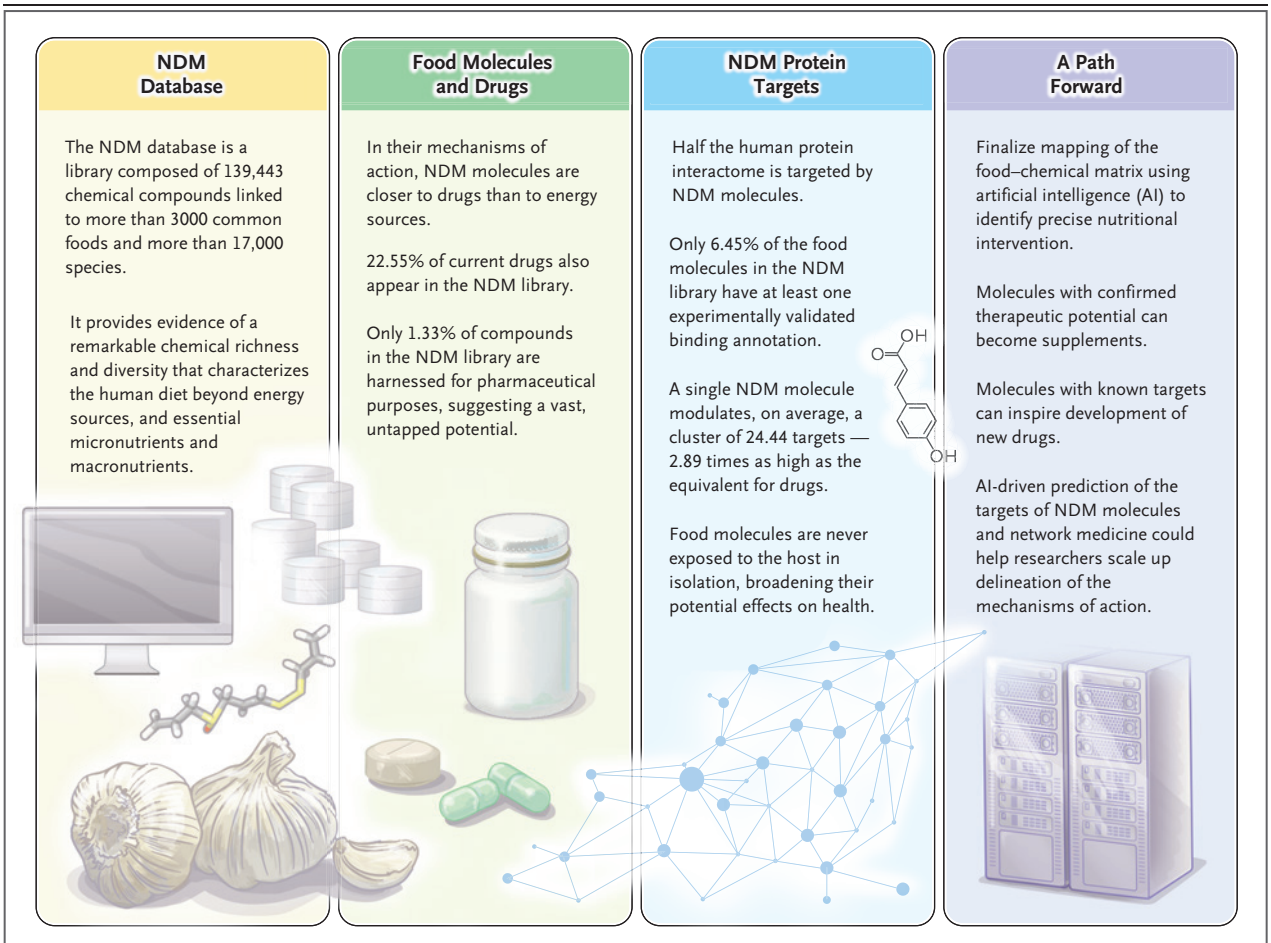

Figure 1. Overall Approach to Food Molecule Research, from Nutrition Dark Matter (NDM) to Food as Medicine.

As the Food Is Medicine initiative progresses, promising better health through nutritious food, it becomes essential to go beyond standard nutritional components and map the full spectrum of food molecules in order to identify the mechanisms that support well-being. By characterizing the bioactivity of food compounds as compared with that of drugs and exploring the broader therapeutic potential of NDM, we can deepen our understanding of the effect of diet on health and contribute to the progress of systems pharmacology.

sition databases.[23-25] Genomic and pathway predictions contributed to this comprehensive effort.[26] The resulting database provides chemical identifiers linking annotations to external libraries ranging from the *Dictionary of Food Compounds*[24] to FooDB,[25] PhytoHub,[27] and the USDA databases,[28] to name but a few.

An interactive graphic is available at NEJM.org

Chemical language models such as MoLFormer,[29] which represent each food molecule as a vector embedded in a high-dimensional space (768 dimensions for MoLFormer), allow us to visualize the structural and functional similarities of food molecules in the NDM library. As shown in Image 1A in the interactive graphic, the NDM library contains a wide variety of lipidlike structures that serve as intracellular messengers, including terpene glycosides (3662), diacylglycerols (2076), steroids (2942), and sphingolipids (258). The secondary metabolism of plants is also well represented, particularly by the chemical superclass[30] of phenylpropanoids and polyketides (12,827), which includes known compound families such as flavonoids (7624), isoflavonoids (1106), cinnamic acids (916), coumarins (876), tannins (405), and stilbenes (340). Additional superclasses are linked to the secondary metabolism of plants, such as alkaloids (910), lignans and neolignans (669), and organosulfur compounds (547), underscoring the ex-

ceptional diversity of the chemicals we consume daily.

Consider garlic, a staple in many cuisines and used medicinally for thousands of years, dating back to ancient Egypt.[31] The USDA SR Legacy database lists 69 nutrients for raw garlic, but it does not track key organosulfur compounds such as allicin[32] (responsible for the distinct aroma of freshly crushed garlic) and ajoene,[33] both of which contribute to the cardioprotective and antimicrobial properties of garlic.[34-37] In addition, the database does not include p-coumaric acid,[38] a hydroxycinnamic acid that belongs to the broad class of polyphenols and is known for its protective effects against carcinogenesis and inflammation.[39,40] As we have recently shown,[14] these three compounds are just a small fraction of the 6802 small molecules documented in raw garlic (Image 1B in the interactive graphic), many of which are secondary metabolites that serve as the chemical defense of the plant against extreme weather and predators.

FROM FOOD MOLECULES TO DRUGS

The 150 nutritional components tracked by the USDA are essential to our health; they serve as energy sources, provide essential amino acids for growth and repair, and regulate enzyme activity by acting as cofactors or cosubstrates. Given this backdrop, one might reasonably surmise that the vast spectrum of molecules catalogued in the NDM library is not part of the essential nutrient focus of the USDA for the simple reason that these molecules are not central components of human metabolism and are therefore considered to be inert. In reality, however, many of these molecules have well-documented implications for health. For example, rather than engaging directly with central metabolic pathways, many polyphenols, which are absorbed from the diet to differing extents, bind to specific human proteins, modulate the composition of gut microbiota, and regulate a vast array of subcellular processes. In other words, with respect to mechanisms of action, the NDM molecules may be closer to drugs than to the energy sources conventionally studied by nutritional biochemists.

Many approved drugs are NDM chemicals. The origins of aspirin, of course, can be traced back to the salicylic acid found in willow bark and certain fruits and vegetables.[41] Similarly, lovastatin is linked to compounds found in red yeast rice, a dietary staple in some cultures,[42] and quinine, a traditional remedy for malaria, is derived from the bark of the cinchona tree and is still used in tonic water today.[43] These are not isolated examples. Of the 8219 small-molecule drugs reported in DrugBank with known targets,[44] 22.55% also appear in the NDM library, implying that they are molecules found in natural ingredients that we consume.

In Image 2A in the interactive graphic, molecular embeddings (or the conversion of molecular structures into vectorial representations, as in MoLFormer) show how molecules in the current family of drugs cluster alongside naturally occurring compounds documented in the NDM library. Take, for example, rosmarinic acid, a polyphenol recognized for its antithrombotic effects exerted by binding to signaling proteins involved in platelet activation.[45] This compound, currently overlooked by the USDA, aligns closely with clopidogrel, a synthetic antiplatelet drug, in the embedding space; as shown in Image 2A in the interactive graphic, drug compounds tend to cluster in a distinct region of the NDM space, particularly near amino acids, peptides, and benzoic acid derivatives, which predominantly have hydrophilic properties. Virtually no area of this drug-related space is untouched by food molecules, which reflects the fact that many current drugs have been inspired by natural compounds. Yet currently, only 1.33% of compounds in the NDM library are harnessed for pharmaceutical purposes, which suggests a vast, untapped potential within the broader chemical landscape of food.

The tendency of drug molecules to cluster in specific zones of the NDM library can be partly attributed to the methods of drug design. One such method, known as "me-too chemistry,"[46] involves iterative modifications of existing compounds that are often tailored for precise protein targets, which limits the diversity of chemical structures used in drug design. As an example, the 115 adrenergic antagonists reported in DrugBank are concentrated in a small, localized region of the NDM library, which indicates not only their roots in food but also their narrow chemical spectrum. Indeed, some of the earliest adrenergic antagonists were derived from natural sources, particularly plant alkaloids, such as reserpine, an alkaloid extracted from the plant

Rauwolfia serpentina,[47] and ergot alkaloids, derived from the fungus *Claviceps purpurea*[48] (which, of course, can serve as adrenergic antagonists and adrenergic agonists). However, most modern adrenergic antagonists (96 of 115) are synthetic, since advances in organic chemistry and pharmacology have enabled the design of new molecules that specifically block alpha- or beta-adrenergic receptors and their subtypes. Yet these new molecules remain chemically similar to the natural molecules that inspired them and, hence, are clustered in the same neighborhood of the NDM library (Image 2B in the interactive graphic).

In contrast to drugs, food molecules — shaped by millions of years of evolution — have greater chemical diversity, allowing them to interact with a wider array of protein targets. These food compounds evolved in dynamic ecosystems, where they varied in concentration and competed for different biologic targets. Consequently, food molecules often provide chemical scaffolds that are optimized for protein binding[49] but also have a broad range of structures and features. An example is shown in Image 2C in the interactive graphic, where we highlight 12,889 polyphenols,[21] molecules that have multiple health benefits, such as antioxidant or pro-oxidant activity, which they achieve by binding to proteins, modulating signal-transduction pathways through cross-kingdom signaling,[49,50] and interacting with gut microbiota. As a result, they are not limited to a narrow neighborhood of the NDM library but instead extend across a wide area of the chemical space.

Taken together, the examples in Image 2 in the interactive graphic illustrate the broad chemical diversity inherent in food molecules as compared with the much narrower diversity of approved drugs. This difference highlights the untapped therapeutic potential of a comprehensive compilation of small molecules in food. As we discuss below, this resource would not only enrich our understanding of the health implications of diet but also greatly enhance the research landscape of systems pharmacology.

PROTEINS AND NDM

The wide chemical diversity of food molecules and their nonmetabolic, regulatory role raise the question: what part of the proteome do they modulate? This issue is depicted in Image 3 in the interactive graphic, where we use node2vec to visualize the human interactome[51,52] — a subcellular map cataloguing 354,659 physical interactions among 18,659 human proteins. Given the large number of interactions, it is impossible to show the full network legibly. Instead, the layout of the figure reflects the underlying network, with interacting proteins placed closer to each other than proteins that are not interacting. Most important, the figure shows all protein targets that have known, experimentally validated binding annotations with small molecules in the NDM library,[44] indicating that approximately half the interactome (8997 proteins) is targeted by food molecules. Given that only 6.45% of the food molecules in the NDM library have at least one experimentally validated binding annotation, the true fraction of the proteins modulated by food molecules is likely to be much higher. There is a fundamental difference, however, between the roles of drugs and food molecules: at any given time, only a few proteins are modulated substantially by drugs taken in response to a disease. In contrast, food molecules influence the activity of much of the proteome throughout a person's life span, with individual variability driven by dietary choices (amount of food ingested, timing of ingestion, matrix composition, other coingested foods, and so forth).

Although drugs and dietary compounds share characteristics as small molecules, protein targets for drugs are highly specific, since they are either selected or designed to limit associations and mitigate potential side effects. Natural compounds in food have greater structural diversity and promiscuity, which allow them to bind to a broad range of targets. To address these challenges, we turn to network medicine,[52,53] which defines diseases as localized changes in the subcellular interaction networks. We identify the relative effect of a drug or a food molecule on the interactome by measuring the size of the largest subgraph of interacting proteins targeted by the same chemical.[54,55] Here, we focus on molecules simultaneously annotated by both the NDM library and DrugBank (1853 molecules), offering a more unbiased assessment by reporting the bioactivity of food molecules that have received as much scientific attention as that paid to drug molecules. We find that, on average, the subgraphs generated by the targets of food molecules connect 24.44 targets. This is 2.89 times as high as the number of interacting proteins targeted by mole-

cules found in DrugBank that are not rooted in food, which is only 8.47 (Mann–Whitney P value, 3.35×10^{-41}; effect size, 0.14). The larger interactome neighborhood modulated by single food molecules reflects their broader molecular effect, which cannot be easily interpreted through the narrow lens typically applied in drug design. Food molecules act as promiscuous modulators of cellular molecular networks, simultaneously influencing multiple processes and pathways.

In addition, food molecules, unlike drugs, encounter the interactome not in isolation but in conjunction with other food molecules that are often interdependent. These interdependencies are shown in Image 4A in the interactive graphic, which shows the Spearman correlation matrix of chemical concentrations measured for 108 raw fruits and vegetables with USDA-catalogued concentrations in multiple foods. We observe distinct clusters of highly correlated compounds that often reflect shared pathways in plant metabolism. For example, vitamin K_1 and lutein, which have a difference of at least one order of magnitude in their average concentrations,[12] show a strong correlation because of their shared biosynthetic pathway and common precursor, geranylgeranyl pyrophosphate. In contrast, delphinidin, an anthocyanidin, belongs to a separate cluster, which indicates the absence of a shared biosynthetic pathway. These findings suggest that many food molecules produced together by the metabolism of an organism share structural and functional features and often target the same areas of the human interactome. Anthocyanidins and anthocyanins, for example, differ only with respect to the attachment of sugar molecules. Their presence in food shows a strong Spearman correlation, ranging from 0.4839 to 0.7283, and they also modulate a common neighborhood of the interactome through the action of 10 shared protein targets (Image 4 in the interactive graphic).

THE LANGUAGE OF FOOD

The remarkable chemical complexity of our diet is illustrated by the 139,443 molecules documented in food, a number that will continue to grow with further research. Although each of the vast majority of these molecules was detected in one or, at most, a few food ingredients, they are likely to be present in multiple foods. Thus, to truly understand diet and its potential health implications, we need to finalize the currently incomplete food–chemical matrix, offering accurate information on which of these biochemicals are found in specific food ingredients, along with their concentrations.

Although we are only at the beginning of this journey, the Periodic Table of Food Initiative (PTFI; https://foodperiodictable.org/)[56] — led by the American Heart Association[57] — has taken an important first step by striving to standardize "foodomics" and comprehensively characterize the chemical composition of foods, using modern "omics" techniques (see the Supplementary Appendix, available with the full text of this article at NEJM.org). Yet as of August 15, 2024, data were available for only 328 of the 1650 expected foods. Furthermore, whereas the PTFI reference dataset quantifies 165 broad chemical classes or individual compounds — substantially overlapping with the USDA nutrient panel — the PTFI discovery dataset, which involves the use of untargeted metabolomics, comprises 18,000 compounds. However, only 462 of these compounds are structurally characterized. The remainder are defined only by molecular formulas, which limits our ability to conduct precise mechanistic studies and characterize the bioactivity of these molecules. Integrating the PTFI with the NDM library yielded a mere 41 additional compounds with fully resolved InChIKeys, which indicates that the database largely captures previously documented substances. This contrast makes it clear that establishing a reproducible, standardized foodomics pipeline differs markedly from exploring NDM to its full extent. Given the time and scalability constraints of the current experimental platforms, the most effective immediate step may be to consolidate and rationalize existing data. Such efforts can guide downstream tasks, including prioritizing experimental validation and more rapidly filling critical knowledge gaps.

We can follow an alternative path and achieve greater scalability by leveraging rapid advances in artificial intelligence (AI). As we have argued elsewhere,[14] integration of the information collected in massive mass spectrometry repositories on different food ingredients with the phylogenetic relationships between species offers an avenue for mapping the full food–chemical matrix and for obtaining an accurate description of which of these molecules are in which ingredients. Achieving this goal remains unlikely in the

near term, but if the project is completed, its effect on our understanding of diet and biology (or pathobiology) will be comparable to the effect of the Human Genome Project on medicine.

We do not need the full ingredient–chemical matrix to begin unlocking the potential of NDM. Indeed, the 139,443 molecules already documented in the NDM library are generally safe when consumed in regular serving sizes in concentrations naturally present in food, and many of them have known human protein targets capable of modulating specific subcellular processes. Collectively, this natural chemical space is roughly 10 times as large as the current pool of approved and experimental drugs. If clinical research confirms the therapeutic value of any of these compounds, their established safety profile at standard dietary levels could enable their use as supplements and potentially circumvent the need for lengthy, costly safety trials required for approval by the Food and Drug Administration. Furthermore, knowledge of the targets of NDM molecules with validated mechanisms of action can help identify possible new targets for therapeutic interventions, speeding up target identification, which is a bottleneck in drug discovery. NDM molecules with known targets and therapeutic potential can also be further modified for stronger binding to lead to new classes of drugs.

What are the current challenges to activating the exceptional body of knowledge that the NDM library offers? First, we need to know the protein targets of each food molecule. Second, we need to predict their potential therapeutic effect — that is, what cellular processes they modulate and what their mechanism of action is. Recent advances in AI and network medicine offer the tools to complete this journey.

A key step toward understanding the potential therapeutic effect of food molecules is understanding where they bind. Since experimentally validated binding annotations are available for only 6.45% of the NDM molecules, we need to predict the binding interactions of food-derived small molecules with both human and microbial proteins[58,59] reliably; the latter would allow us to understand how food molecules modulate the microbiome beyond the data from standard studies of the relative abundance of microbial species.

To address these issues, it becomes imperative to invest in AI models capable of predicting binding interactions with accuracy and efficiency at large scale. Yet state-of-the-art AI models in drug discovery have been shown to be prone to shortcut learning,[60,61] in which the model fails to grasp ligand–protein binding principles fully and struggles to generalize to novel molecules. Thus, to improve the generalization capabilities of these models for food molecules, future efforts should emphasize training strategies that encourage a deeper understanding of ligand–protein binding — such as incorporating more diverse training datasets, refining model architectures, designing more challenging cross-validation, and rigorously validating predictions with experimental assays.

This development should occur in parallel with large-scale mapping of the molecular composition of food to provide a meaningful biologic context for each molecule identified, to reveal interdependencies, and to accelerate our understanding of their health effects. AlphaFold 3[62] is one of the most promising advances in this domain, but its scalability and precision in predicting interactions for more than 139,000 small molecules in food and more than 18,000 proteins in the human interactome remain untested. Ultimately, AI tools will need to be combined with experimental assays to test predictions and achieve the accuracy and clarity provided by the combination of yeast two-hybrid and mass spectrometry purification assays, which have helped map the protein interaction network.[63]

To uncover the mechanisms of action for small molecules, we can turn to network medicine, which has contributed to the development of falsifiable predictions that can be validated experimentally. Again consider, for example, rosmarinic acid (Image 2A in the interactive graphic), which targets FYN along with PDE4D, CD36, and APP, vascular disease proteins associated with platelet function.[14] The connected component that these targets form within the human interactome suggests that the food molecule could affect platelet thrombus formation in vascular disease. In vitro experiments have confirmed this prediction and shown that rosmarinic acid inhibits collagen-mediated platelet aggregation and α-granule secretion by inhibiting protein tyrosine phosphorylation through its interaction with FYN.[19]

Taken together, advances in AI-driven prediction of the targets of NDM molecules, combined with the tools of network medicine, allow us to activate the exceptional knowledge contained in the NDM library. Ultimately, this strategy may allow us to identify food molecules that could have a direct therapeutic effect on specific diseases and could represent the starting point for the development of new drugs and therapies.

CONCLUSIONS

The sequencing of the human genome has revolutionized biology by offering a platform to explore genomic variations, as well as to understand how these variations lead to disease. This information is essential but not sufficient to understand human disease, as illustrated by the fact that genetic effects account for only 10% of disease occurrences; the bulk of the remaining occurrences can be linked to environmental and dietary factors.[7] To harness fully the potential for the role of diet in improving health and longevity, we must characterize the vast molecular repertoire of foods. The approach we describe — systematically identifying, cataloguing, and analyzing tens of thousands of food molecules — represents a critical step forward. By unveiling the molecular mechanisms through which these compounds interact with human biologic pathways, this strategy not only refines our definition of dietary quality but also opens new avenues for targeted therapies and precision nutrition. We are only at the beginning of this journey. Although we have thus far documented more than 139,000 molecules in food, for the vast majority of them, we do not yet know if they are absorbed after ingestion, how they are metabolized (by the gut microbiome, the host, or both), to which proteins they bind, and what cellular processes they affect. Unlocking this knowledge could revolutionize the way we think about the role of food in health.

We currently lack an institutional or funding blueprint to achieve this goal and map the molecular composition of food. For example, the first goal of the National Institutes of Health (NIH) 2020–2030 Strategic Plan for NIH Nutrition Research[64] is titled "Spur Discovery and Innovation through Foundational Research — What do we eat and how does it affect us?" However, the plan mainly emphasizes identifying unknown metabolites arising from the microbiome and host metabolism, without addressing the need to understand the chemical compounds in food. In fiscal year 2019, the NIH spent an estimated $1.9 billion on nutrition research and funded approximately 4600 projects across at least 25 of the 27 NIH institutes, centers, and offices.[65] Although this investment is laudable, without systematic efforts to map NDM, the effort will have a questionable effect, since it relies on very limited information on food composition. Given the critical role of nutrition in disease prevention and management, the lack of concentrated investment in foundational projects that map the chemical makeup of food continues to limit our ability to turn nutrition science into an accurate, data-driven, predictive discipline.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

AUTHOR INFORMATION

[1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston; [2]Network Science Institute and Department of Physics, Northeastern University, Boston; [3]Harvard Data Science Initiative, Harvard University, Boston; [4]Department of Network and Data Science, Central European University, Budapest, Hungary; [5]Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston.

REFERENCES

1. Department of Agriculture. Food and Nutrition Security (https://www.nal.usda.gov/human-nutrition-and-food-safety/nutrition-security).
2. Centers for Disease Control and Prevention. Fast facts: health and economic costs of chronic conditions. July 12, 2024 (https://www.cdc.gov/chronic-disease/data-research/facts-stats/index.html).
3. Office of Disease Prevention and Health Promotion. Food is medicine: a project to unify and advance collective action (https://health.gov/our-work/nutrition-physical-activity/food-medicine#4).
4. Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N Engl J Med 2016;375:2349-58.
5. National Cancer Institute. Past CSSI programs. June 17, 2024 (https://www.cancer.gov/about-nci/organization/cssi/research/past-programs#provocative-questions-initiative).
6. Ugai T, Sasamoto N, Lee H-Y, et al. Is early-onset cancer an emerging global epidemic? Current evidence and future implications. Nat Rev Clin Oncol 2022;19:656-73.
7. National Institute for Occupational Safety and Health. Exposome and exposomics.

Atlanta: Centers for Disease Control and Prevention, August 19, 2022 (https://archive.cdc.gov/www_cdc_gov/niosh/topics/exposome/).
8. Volpp KG, Berkowitz SA, Sharma SV, et al. Food is medicine: a presidential advisory from the American Heart Association. Circulation 2023;148:1417-39.
9. National Library of Medicine. PubChem. 2004 (https://pubchem.ncbi.nlm.nih.gov/).
10. Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008;36:D901-D906.
11. Placzek S, Schomburg I, Chang A, et al. BRENDA in 2017: new perspectives and new tools in BRENDA. Nucleic Acids Res 2017;45(D1):D380-D388.
12. Menichetti G, Barabási A-L. Nutrient concentrations in food display universal behaviour. Nat Food 2022;3:375-82.
13. Morris AL, Mohiuddin SS. Biochemistry, nutrients. In: StatPearls. Treasure Island, FL: StatPearls Publishing, 2023 (https://www.ncbi.nlm.nih.gov/books/NBK554545/).
14. Menichetti G, Barabási A-L, Loscalzo J. Decoding the foodome: molecular networks connecting diet and health. Annu Rev Nutr 2024;44:257-88.
15. Sebastian R, Wilkinson Enns C, Goldman JD, Steinfeldt L, Martin CL, Moshfegh AJ. Flavonoid values for USDA survey foods and beverages 2007–2010. Beltsville, MD: Department of Agriculture, Agricultural Research Service, 2016 (https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/fndds/FlavonoidDB_documentation_0710.pdf).
16. Bhagwat S, Haytowitz DB, Wasswa-Kintu SI, Holden JM. USDA develops a database for flavonoids to assess dietary intakes. Procedia Food Sci 2013;2:81-6 (http://linkinghub.elsevier.com/retrieve/pii/S2211601X1300014X).
17. Bhagwat S, Haytowitz DB, Holden JM. USDA database for the flavonoid content of selected foods: release 3. Beltsville, MD: Department of Agriculture, Agricultural Research Service, 2011 (https://www.ars.usda.gov/arsuserfiles/80400525/data/flav/flav_r03.pdf).
18. Barabási AL, Menichetti G, Loscalzo J. The unmapped chemical complexity of our diet. Nat Food 2020;1:33-7 (https://doi.org/10.1038/s43016-019-0005-1).
19. European Bioinformatics Institute. Metabolomics: small molecules (https://www.ebi.ac.uk/training/online/courses/metabolomics-introduction/what-is/small-molecules/).
20. National Center for Biotechnology Information. Taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy).
21. Hooton F, Menichetti G, Barabási A-L. Exploring food contents in scientific literature with FoodMine. Sci Rep 2020;10:16191.
22. European Bioinformatics Institute. MetaboLights (https://www.ebi.ac.uk/metabolights/).
23. Rothwell JA, Perez-Jimenez J, Neveu V, et al. Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. Database (Oxford) 2013;2013:bat070.
24. Yannai S. Dictionary of food compounds. 2nd ed. New York: CRC Press, 2012.
25. WishartLab. FooDB (http://foodb.ca/).
26. Ofaim S, Menichetti G, Sebek M, Barabási AL. Genomics-based annotations help unveil the molecular composition of edible plants. August 9, 2023 (https://www.biorxiv.org/content/10.1101/2022.01.24.477528v2). preprint.
27. PhytoHub (https://phytohub.eu/).
28. Department of Agriculture. USDA food composition databases. 2018 (https://ndb.nal.usda.gov/ndb/).
29. Ross J, Belgodere B, Chenthamarakshan V, Padhi I, Mroueh Y, Das P. Large-scale chemical language representations capture molecular structure and properties. Nat Mach Intell 2022;4:1256-64 (https://www.nature.com/articles/s42256-022-00580-7).
30. Djoumbou Feunang Y, Eisner R, Knox C, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminform 2016;8:61.
31. Block E. Garlic and other alliums: the lore and the science. London: Royal Society of Chemistry, 2010.
32. Department of Agriculture. Dr. Duke's phytochemical and ethnobotanical databases: allicin (https://phytochem.nal.usda.gov/chemical-allicin).
33. Department of Agriculture. Dr. Duke's phytochemical and ethnobotanical databases: ajoene (https://phytochem.nal.usda.gov/chemical-ajoene).
34. Gao Y, Wang B, Qin G, et al. Therapeutic potentials of allicin in cardiovascular disease: advances and future directions. Chin Med 2024;19:93.
35. Sánchez-Gloria JL, Arellano-Buendía AS, Juárez-Rojas JG, et al. Cellular mechanisms underlying the cardioprotective role of allicin on cardiovascular diseases. Int J Mol Sci 2022;23:9082.
36. Banerjee SK, Maulik SK. Effect of garlic on cardiovascular disorders: a review. Nutr J 2002;1:4.
37. Rendu F, Daveloose D, Debouzy JC, et al. Ajoene, the antiplatelet compound derived from garlic, specifically inhibits platelet release reaction by affecting the plasma membrane internal microviscosity. Biochem Pharmacol 1989;38:1321-8.
38. Department of Agriculture. Dr. Duke's phytochemical and ethnobotanical databases: p-coumaric-acid (https://phytochem.nal.usda.gov/chemical-p-coumaric-acid).
39. Pragasam SJ, Venkatesan V, Rasool M. Immunomodulatory and anti-inflammatory effect of p-coumaric acid, a common dietary polyphenol on experimental inflammation in rats. Inflammation 2013;36:169-76.
40. Scalbert A, Johnson IT, Saltmarsh M. Polyphenols: antioxidants and beyond. Am J Clin Nutr 2005;81:Suppl 1:215S-217S.
41. Mahdi JG, Mahdi AJ, Mahdi AJ, Bowen ID. The historical analysis of aspirin discovery, its relation to the willow tree and antiproliferative and anticancer potential. Cell Prolif 2006;39:147-55.
42. Klimek M, Wang S, Ogunkanmi A. Safety and efficacy of red yeast rice (Monascus purpureus) as an alternative therapy for hyperlipidemia. P T 2009;34:313-27.
43. Gachelin G, Garner P, Ferroni E, Tröhler U, Chalmers I. Evaluating Cinchona bark and quinine for treating and preventing malaria. J R Soc Med 2017;110:31-40.
44. Piras A, Chenghao S, Sebek M, Ispirova G, Menichetti G. CPIExtract: a software package to collect and harmonize small molecule and protein interactions. July 5, 2024 (https://www.biorxiv.org/content/10.1101/2024.07.03.601957v1). preprint.
45. do Valle IF, Roweth HG, Malloy MW, et al. Network medicine framework shows that proximity of polyphenol targets and disease proteins predicts therapeutic effects of polyphenols. Nat Food 2021;2:143-55.
46. Laszlo P. Me-too chemistry. In: A life and career in chemistry. Cham, Switzerland: Springer, 2021;159-69.
47. Lobay D. Rauwolfia in the treatment of hypertension. Integr Med (Encinitas) 2015;14:40-6.
48. Majrashi M, Ramesh S, Deruiter J, et al. Multipotent and poly-therapeutic fungal alkaloids of Claviceps purpurea. In: Medicinal plants and fungi: recent advances in research and development. Singapore: Springer, 2017;229-52 (https://link.springer.com/chapter/10.1007/978-981-10-5978-0_8).
49. Nasirian F, Menichetti G. Molecular interaction networks and cardiovascular disease risk: the role of food bioactive small molecules. Arterioscler Thromb Vasc Biol 2023;43:813-23.
50. Lacroix S, Klicic Badoux J, Scott-Boyer M-P, et al. A computationally driven analysis of the polyphenol-protein interactome. Sci Rep 2018;8:2232.
51. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12:56-68.
52. Menche J, Sharma A, Kitsak M, et al. Disease networks: uncovering disease-disease relationships through the incomplete interactome. Science 2015;347:1257601.

53. Loscalzo J, Barabási A-L, Silverman EK. Network medicine: complex systems in human disease and therapeutics. Cambridge, MA: Harvard University Press, 2017.
54. Aldana A, Sebek M, Ispirova G, et al. NetMedPy: a Python package for large-scale network medicine screening. September 8, 2024 (https://www.biorxiv.org/content/10.1101/2024.09.05.611537v1). preprint
55. Barabasi A-L. Network science. Cambridge, United Kingdom: Cambridge University Press, 2016 (http://networksciencebook.com/).
56. Jarvis A, Gallo-Franco J, Portilla J, et al. Periodic Table of Food Initiative for generating biomolecular knowledge of edible biodiversity. Nat Food 2024;5:189-93.
57. American Heart Association. Precision medicine platform (https://precision.heart.org/ptfi-data-resource).
58. Balint D, Brito IL. Human-gut bacterial protein-protein interactions: understudied but impactful to human health. Trends Microbiol 2024;32:325-32.
59. Blasco T, Pérez-Burillo S, Balzerani F, et al. An extended reconstruction of human gut microbiota metabolism of dietary compounds. Nat Commun 2021;12:4728.
60. Menichetti G. An AI pipeline to investigate the binding properties of poorly annotated molecules. Nat Rev Phys 2022;4:359 (https://www.nature.com/articles/s42254-022-00471-1).
61. Chatterjee A, Walters R, Shafi Z, et al. Improving the generalizability of protein-ligand binding predictions with AI-Bind. Nat Commun 2023;14:1989.
62. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024;630:493-500.
63. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. Cell 2011;144:986-98.
64. Nutrition Research Task Force. 2020-2030 strategic plan for NIH nutrition research: a report of the NIH Nutrition Research Task Force. Bethesda, MD: National Institutes of Health (https://dpcpsi.nih.gov/onr/strategic-plan).
65. Fleischhacker SE, Woteki CE, Coates PM, et al. Strengthening national nutrition research: rationale and options for a new coordinated federal research effort and authority. Am J Clin Nutr 2020;112:721-69.