Learning from nature by leveraging integrative biomateriomics modeling toward adaptive and functional materials

Sofia E. Arevalo and Markus J. Buehler*

Biological systems generate a wealth of materials, and their design principles inspire and inform scientists from a broad range of fields. Nature often adapts hierarchical multilevel material architectures to achieve a set of properties for specific functions, providing templates for difficult tasks of understanding the intricate interplay between structure–property–function relationships. While these materials tend to be complex and feature intricate functional interactions across scales, molecular-based multiscale modeling, machine learning, and artificial intelligence combined with experimental approaches to synthesize and characterize materials have emerged as powerful tools for analysis, prediction, and design. This article examines materiomic graph-based modeling frameworks for assisting researchers to pursue materials-focused studies in a biological context, and provides an overview of methods that can be applied to bottom-up manufacturing, including a historical perspective of bioinspired materials research. Through the advent of novel modeling architectures and diverse systems from nature, there is potential to develop materials with improved properties.
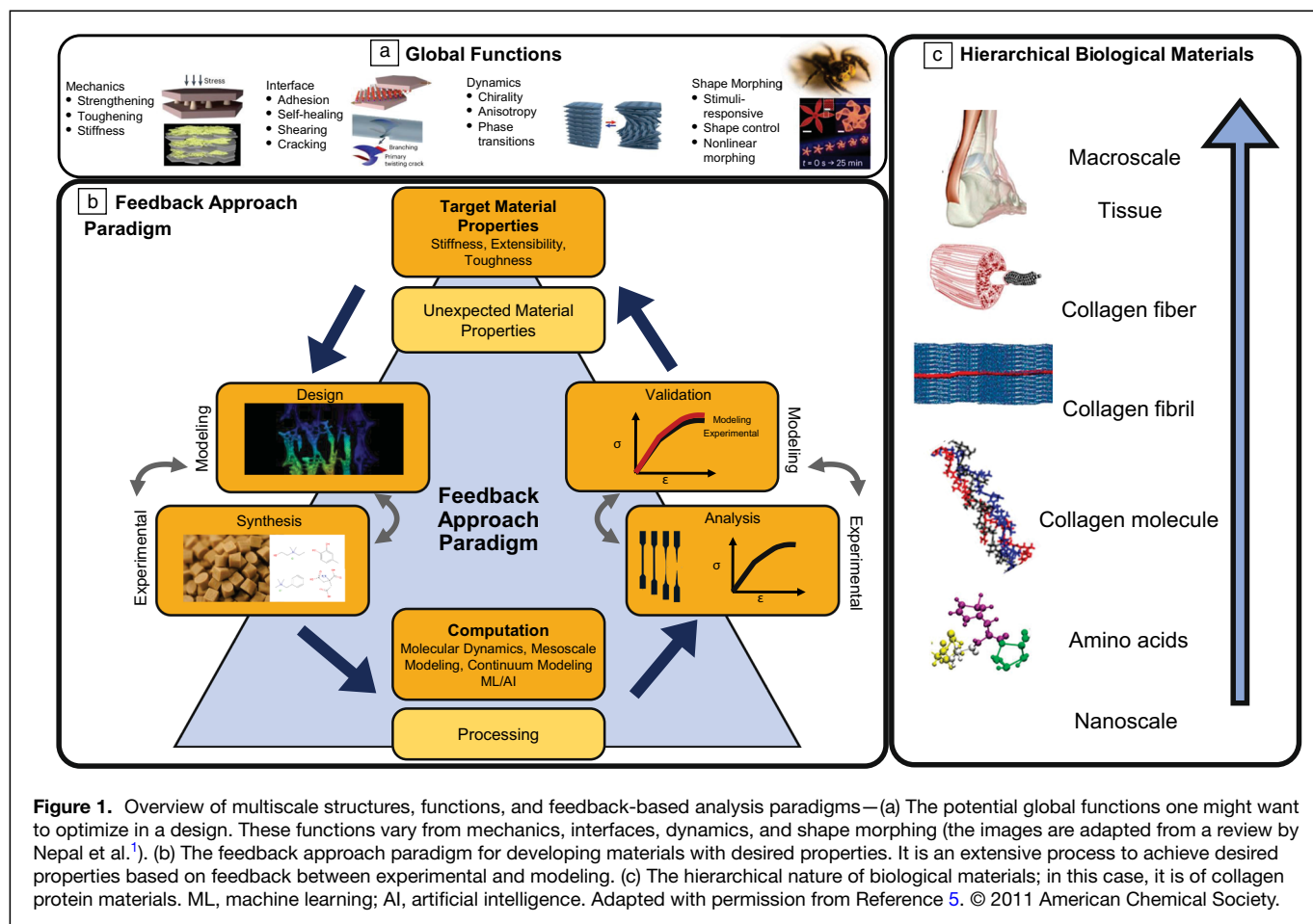
Introduction

Materials science historically focuses on a wide array of materials, from solids to liquids, and in various contexts, from fundamental scientific aspects to practical solutions to advance civilization. While biomaterials have always been of keen interest, their complexity results in challenges that have largely precluded their rigorous analysis just decades ago. That has changed, and now we can analyze complex biomaterials and design new ones, create synergistic relationships between natural and synthetic materials, and even engineer *de novo* living materials from first principles[1–4] (**Figure 1**).

Humans have effectively utilized a variety of materials such as wood, clay, and later, human-created materials such as steel and ceramics, and highly integrated architectures such as computer chips. However, nature continues to serve as a source of inspiration for material development, especially when it comes to understanding the synthesis of living materials and integrated systems. This is because biological systems often have outstanding sets of qualities that exceed those of synthetic materials, which often include not only high-performance targets but notably, features such as tunability, changeability combined with favorable energy balances, and synthesis cost, to name a few examples. The properties of biological materials are owed to the complex distinct hierarchical structures at the nano-, micro-, and mesolevel.[1,6–10] As a result, biological materials inspire materials engineers to develop materials with optimized functions, properties, and structures for specific applications.[11,12] Notable behaviors exhibited by biological materials and highly desirable for mimicking are their self-organization, multifunctionality, and self-healing capabilities,[13,14] often derived from a set of relatively simple building blocks arranged in highly complex patterns. Developing bioinspired materials is motivated by optimizing the behavior and properties for a specified application, such as improved adhesion, optical properties, toughness, and strength.[12,15–17] Traditional approaches to tailoring synthetic materials properties and functions involve many design-test iterations until desired properties are achieved. This often results in a time-consuming, labor-intensive, and expensive process to discover *de novo* materials with desired properties, and sometimes includes human bias toward existing engineering solutions. However, this process is not efficient at exploring the vast

Sofia E. Arevalo, Laboratory for Atomistic and Molecular Mechanics (LAMM), Massachusetts Institute of Technology, Cambridge, USA; sarevalo@mit.edu
Markus J. Buehler, Laboratory for Atomistic and Molecular Mechanics (LAMM), Massachusetts Institute of Technology, Cambridge, USA; mbuehler@mit.edu
*Corresponding author
doi:10.1557/s43577-023-00610-8

Figure 1. Overview of multiscale structures, functions, and feedback-based analysis paradigms—(a) The potential global functions one might want to optimize in a design. These functions vary from mechanics, interfaces, dynamics, and shape morphing (the images are adapted from a review by Nepal et al.[1]). (b) The feedback approach paradigm for developing materials with desired properties. It is an extensive process to achieve desired properties based on feedback between experimental and modeling. (c) The hierarchical nature of biological materials; in this case, it is of collagen protein materials. ML, machine learning; AI, artificial intelligence. Adapted with permission from Reference 5. © 2011 American Chemical Society.

search space as it involves trial and error and often lacks a directed approach without human biases. There is also a renewed interest in biomaterials because of their biocompatibility and innate sustainability of their life cycle.[4,18,19]

A high-throughput approach to addressing these challenges would be advantageous to expedite the discovery of novel materials with targeted, optimal properties (Figure 1a).[20] Such approaches emerged in recent years, as highlighted in Figure 1b, enabling a feedback-based paradigm, and facilitating a more efficient process for discovering novel materials.[20,21] The feedback approach paradigm relies on the synergy between modeling and experiments to iteratively traverse through materials until the target materials are achieved.[20] The approach relies on bioinformatics tools to accelerate searching and discovering new materials, and to offer a comprehensive, integrative, and systematic analysis paradigm to seek deep insights into hierarchical relationships in living materials and their translational potential to engineering (Figure 1c). And with the advent of machine learning (ML) and artificial intelligence (AI), navigation processes can become more efficient. AI, a term coined in the 1950s, refers to the concept that machines have the ability to mimic human capabilities. ML is a subfield of AI that uses algorithms to learn from and recognize patterns from data. As such, ML/AI tools provide a

complement to existing modeling and also hold the promise to integrate experimental and simulation/theory data to further the predictive power of models.

The types of models that may be employed with ML/AI are statistical experimental design, pattern recognition, operations research, reinforcement, and active learning.[22–24] This feedback approach paradigm is implemented in the design of *de novo* protein sequences, the 3D structure of spider webs, learning of the classical mechanical materials design problems, and bridging the gap between the microstructure–design–physical performance of materials.[22,25–27] Understanding the structure–property relationships of biological materials is essential for developing flexible armor inspired by pangolin dermal armor, conch shells, and fish scales, a few notable examples.[28,29] Other studies focus on the hierarchical structure of abalone shells, lobster exoskeletons, antler bones, and silica sponges for understanding crack deflection mechanisms at interfaces,[30] an essential behavior for mitigating catastrophic failures in engineering materials. Furthermore, performing mechanical analysis on plant stems, porcupine quills, toucan beaks, and feather rachis, may provide insightful knowledge on design principles and material composition found in biological structures that are resistant to torsion, bending, and buckling.[30] Owed to the immensity of biological materials and

their complex structure, mimicking nature to achieve desired properties or structure is an immense task that may require ML/AI to explore the vast material composition and design space. Additionally, ML/AI may extend to serve as a diagnostic tool to learn about failure modalities of complex architected materials such as bone,[31] among many other example applications.

ML and AI research is a rapidly evolving field transcending many research areas from materials design to failure analysis.[21,32–37] Current research shows the use of ML to answer pivotal research questions and in conjunction with other numerical or mathematical modeling can complement to uncover critical mechanistic relationships and insights.[38,39] Further, it may provide an approach for high-throughput materials discovery enabling inexpensive and efficient methods for designing novel materials, fueled by advances in large language models that incorporate attention mechanisms. While nature often excels at adapting properties for a specific function (or sets of functions) over long time scales, we can investigate those nature-occurring mechanisms to design materials that are on par or even better than nature, and using synthetic raw materials and building blocks. Other motivations include finding ways to integrate design across evolutionary lineages, from deep time, into modern engineering and ways to combine disparate properties (e.g., enzymatic activity, strength, toughness, resilience, antimicrobial, etc.).

The aims of this article are to provide examples, relevance, and properties of interest in biological materials, and to survey the state-of-the-art modeling techniques (e.g., molecular dynamics [MD], coarse-graining, ML/AI) for the design of novel materials. Additionally, we provide a general landscape of the type of research that implements ML and multiscale modeling to assess and develop bioinspired materials while presenting a case study on metal-coordinated protein materials. Although ML/AI is an essential tool in materials research, there are various challenges for implementation in bioinspired materials design that will be discussed in more detail.

Surveying numerical and mathematical modeling in research

Modeling, experimentation, and theory development of biological materials pose a complex challenge owing to the biological materials' intricate multiscale structure and associated multifunctionality. However, understanding the mechanisms found in biological materials through modeling and experimentation may offer insights into developing synthetic materials with desired properties and behavior.[20] As such, predictive multiscale modeling and simulation are employed to characterize existing biological materials and subsequently the development of novel systems. Selecting the appropriate multiscale model is dependent on the length and time scale of the specific research query and the goal of the investigation. **Table I** shows a set of mathematical modeling methods for the molecular, meso-, and macroscale along with the application for each method.

Insights gained at the molecular scale to advance bioinspired research

Mathematical and numerical modeling such as molecular dynamics can be used to investigate the behavior and properties at the molecular scale. As summarized in Table I:

- Steered molecular dynamics (SMD) is implemented for measuring mechanical properties, such as elasticity, strength, and unfolding pathways of systems.[40]
- Replica exchange molecular dynamics (REMD) can overcome high-energy barriers more easily and thus, sample the conformational space of proteins.[41]
- Umbrella sampling is implemented for assessing systems undergoing systematic change as it can sample states often not captured by normal molecular dynamics due to its intrinsic time-scale limits.[42]
- Well-tempered meta-dynamics uses Gaussian meta-dynamics to explore the free-energy landscape of a system.[43]

Table I. Summary of the types of chemistry/physics modeling from the molecular scale to the macroscale.			
Scale	Molecular Scale	Mesoscale	Macroscale
Methods	Molecular dynamics (MD): Energy minimization and equilibration Steered molecular dynamics (SMD) Replica exchange molecular dynamics (REMD) Umbrella sampling Well-tempered meta-dynamics Reactive force fields	Quasi-continuum method: Kinetic Monte Carlo method Coarse-grained molecular dynamics Dissipative particle dynamics (DPD)	Continuum methods: Solid and computational fluid dynamics (CFD) (finite-difference method, finite element method, finite volume method, boundary element method) Multi-body simulation (MBS)
Application	Conformation Chemical reactions mechanical properties	Size effects Self-assembly Radius of gyration Hydrodynamic radius Fracture properties	Nonlinear structure analysis Global thermal and electrical analysis Multibody problems

- Reactive molecular dynamics allows the simulation of chemical reactions and reaction pathways.[44]

Molecular dynamics (MD) can serve as a powerful method to investigate conformation, chemical reactions, and mechanical property measurement at the molecular scale, especially to gain mechanistic insights.[45] In MD simulations, the behavior of atoms is observed to understand the molecular phenomena that govern macroscopic behavior using Newton's equation. As a result, MD can calculate molecular structure, interactions at interfaces, mechanical properties, and thermodynamic properties over time.[10,46] Through MD simulations, a molecular landscape can be obtained about the relationship between adhesive strength and adhesive density and investigates the bond formation and breaking of different atoms.[45]

From the perspective of materials design, MD simulations may provide crucial insight into the mechanisms and interactions of atoms and how they yield specific properties.[45] In particular, MD modeling is applied to understand biological materials, such as crystalline composites commonly found in mollusks, which are interesting to scientists because of their varying structures at different scales and hierarchic architecture.[47] Further, implementing MD simulations allow for an understanding of the effect of misorientation in architecture that yields increased fracture toughness and serve as a source of inspiration for the synthesis and design of bioinspired materials resistant to fracture.[10,47]

When larger atomistic molecular systems need modeling to capture the behavior of millions of atoms, modeling techniques such as coarse-graining and particle simulations are used. Coarse-graining modeling minimizes the number of degrees of freedom over large length and time scales, for example, by grouping atoms into a single bead to capture the interactions of the macroscopic behavior. By reducing the number of degrees of freedom, sampling is improved and decreases computational time. As a result, researchers implemented coarse-grained molecular analysis to assess the mechanical properties of the proteins designed via end-to-end deep learning methods,[39] whereas other researchers used coarse-graining to predict the equilibrium properties of bioelastomers.[48]

As with much bioinspired research, there is a need to connect the microscopic to the macroscopic behavior of the material system. Researchers utilize molecular simulation to understand the protein–surface interactions and connect them to nanoparticle interactions.[49] Further, they posit that understanding the relationship between proteins and nanoparticles in various conditions may provide opportunities to develop hierarchical materials inspired by nature.[49] On a similar note, performing MD simulations provide an understanding of the hierarchical structure effects on the mechanical behavior of diatoms algae, essential knowledge that can be transferred to

the design of materials.[50] A key takeaway from studying diatoms at the molecular scale using MD is understanding that introducing hierarchical structures with inherently strong but brittle materials may result in a tough, strong, and ductile system.[50] Through years of evolution, nature is adept at refining the hierarchical structures of a system to obtain outstanding properties for its application. Employing MD simulations to study a system in nature allows for directed and well-informed mimicking of synthetic materials and structures. It can also be a powerful source for synthetic data sets that can be upscaled, and generalized, using ML.

Insights gained at the macroscale to advance bioinspired research

Simulating at the macroscale level involves using continuum models. For example, finite element modeling (FEM) provides a structural analysis by solving differential equations. Analysis at the macroscale provides an opportunity to optimize the macroscale properties and tailor the design. An interesting application of FEM is the possibility of complementing it with ML, by providing synthetic ground-truth data sets, to determine the properties of 2D composites and design new functional composites.[22] Further, an ML model can be validated against MD, coarse-grained, or FEM descriptions to assess the accuracy of designing and optimizing the composites, utilizing what is referred to as synthetic data sets.[22] Continuum models simulate larger systems with a substantial reduction in computational cost, in comparison to smaller length-scale models. While researchers utilize FEM, others first use atomistic MD and those results are then input into continuum models to have an understanding of the macroscopic mechanical behavior of two-phase silk fibers.[51] Combining simulation at two different length scales aids in providing a complete narrative of how the various features in the complex hierarchical structure affect the overall macroscopic behavior. This narrative provides essential information for future designs

Coupling physics-based modeling with ML

Biological materials are multiscale in nature, whose complex behavior spans across multiple spatiotemporal scales. It is a challenge to connect the interplay between the temporal and spatial relationship of systems, and often analytical-style theories are not (yet) known, or difficult to elucidate (e.g., in protein folding, protein property predictions, and other problems in biomaterials). Such relationships may sometimes be more efficiently discovered and exploited by ML algorithms that thrive in finding patterns and relationships between disparate data sources. Thus, to move the field toward one that better mimics the nature of biological materials, there is a need to combine ML tools with current mathematical modeling and experiments to develop high-throughput approaches for future material discovery[52] (**Figure 2**).

ML models to support bioinspired material analysis and design

There are several categories of ML that can be broadly classified into supervised learning, unsupervised learning, reinforcement learning, and semi-supervised learning, whereas their integration is emerging as powerful paradigms into multimodal multiparadigm models as done in the development of multimodal language models[25,53–55] (Figure 2a–b). The type of ML classification will depend on the type of data set and the research query. As in the case of supervised learning, knowledge is known ahead of time about a set of data, referred to as a training data set, which is used to predict the characteristics of another set of data. This entails that the success of the model is dependent on the selection of the training data set for the desired outcome prediction. Common approaches that are in the category of supervised learning include the
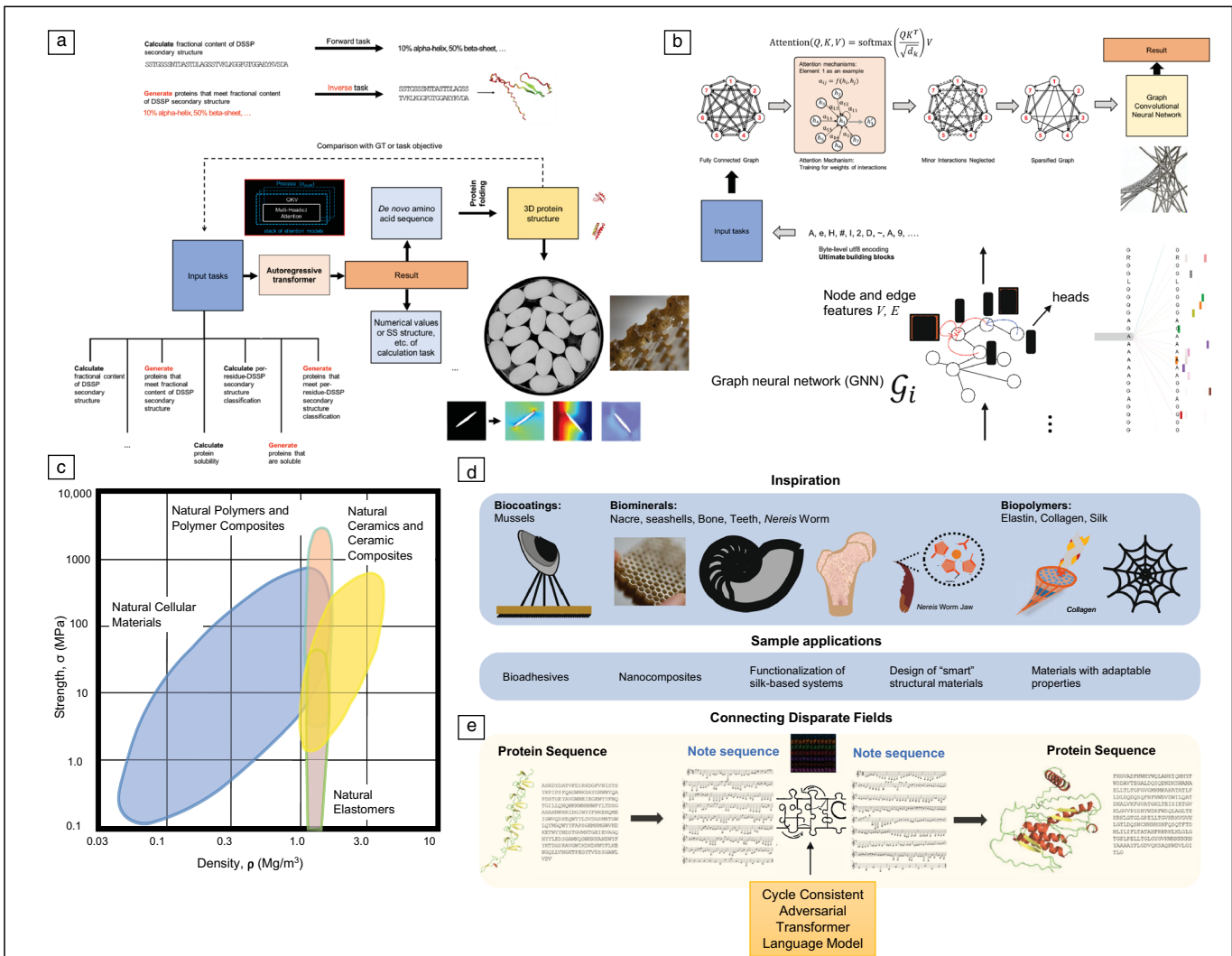

Figure 2. Overview of modeling strategies that incorporate a mix of physics and data-driven approaches. Deep learning and generative artificial intelligence can be effective means to solve forward, inverse, and general biomaterials design problems, formulated within generalized, integrated, and multitasking models.[70] (a) Shows how multitask forward and inverse generative pretrained transformer (GPT) models can be used to solve a variety of modeling challenges.[70] (b) More detailed visualizations of the graph-forming capacities of such multiheaded attention models, where dynamic graphs are discovered, and then used, in the modeling process, uncovering the materiomic makeup of materials. These tools can be applied to describe biological materials, formalizing them as a source of inspiration for developing synthetic materials with improved and functional properties. (c) An Ashby plot, showing the strength of materials with respect to density for a range of natural materials from cellular to natural ceramics and composites. It provides a landscape with a wide range of properties that can be achieved by mimicking nature. (d) By observing biological materials such as mussels, nacre, seashells, and silk, we can learn to design synthetic materials with improved adhesive strength, smart structural materials, and materials with adaptable properties. (e) Connecting disparate fields is another area of interest as it widens the scope of possibilities for designing novel materials. As an example, protein sequence and the connection to music are studied as a fundamental first step for connecting principles between art and science, offering a gateway for connecting other disparate fields. The connector between the two areas is machine learning, as this example shows the use of deep learning methods using cycle-consistent transformer neural networks.[80] SS, secondary structure.

nearest-neighbor algorithm, decision trees and random forests, logistics regression and linear discriminant analysis, support vector machines, and artificial neural networks. Neural networks have historically been effective in solving computer vision tasks, such as image or voxel data analysis. Long–short-term memory (LSTM) is a type of neural network that was originally developed to learn logical trajectories and relationships within data sets. In one example, researchers implemented a LSTM model with variational autoencoders to design optimal materials.[56] Another wide form of a neural network is based on convolutional operators (e.g., convolutional neural networks [CNNs]), which are often used to classify and predict mechanical properties—by modeling a learned coarse-graining strategy—to yield information on tissue deformation and fracture, as an example.[31] When formulated as a U-net architecture (see inset in **Figure 3**), they can also be effectively used to transform high-dimensional data into other high-dimensional data via bottlenecks, such as applied in stress–strain field predictions.[57] Such models can be further developed to incorporate neural operators, such as Fourier neural operators (FNOs), where these models belong to a class of methods that output functions by learning the mapping from any functional parametric dependence to the solution.[58,59]
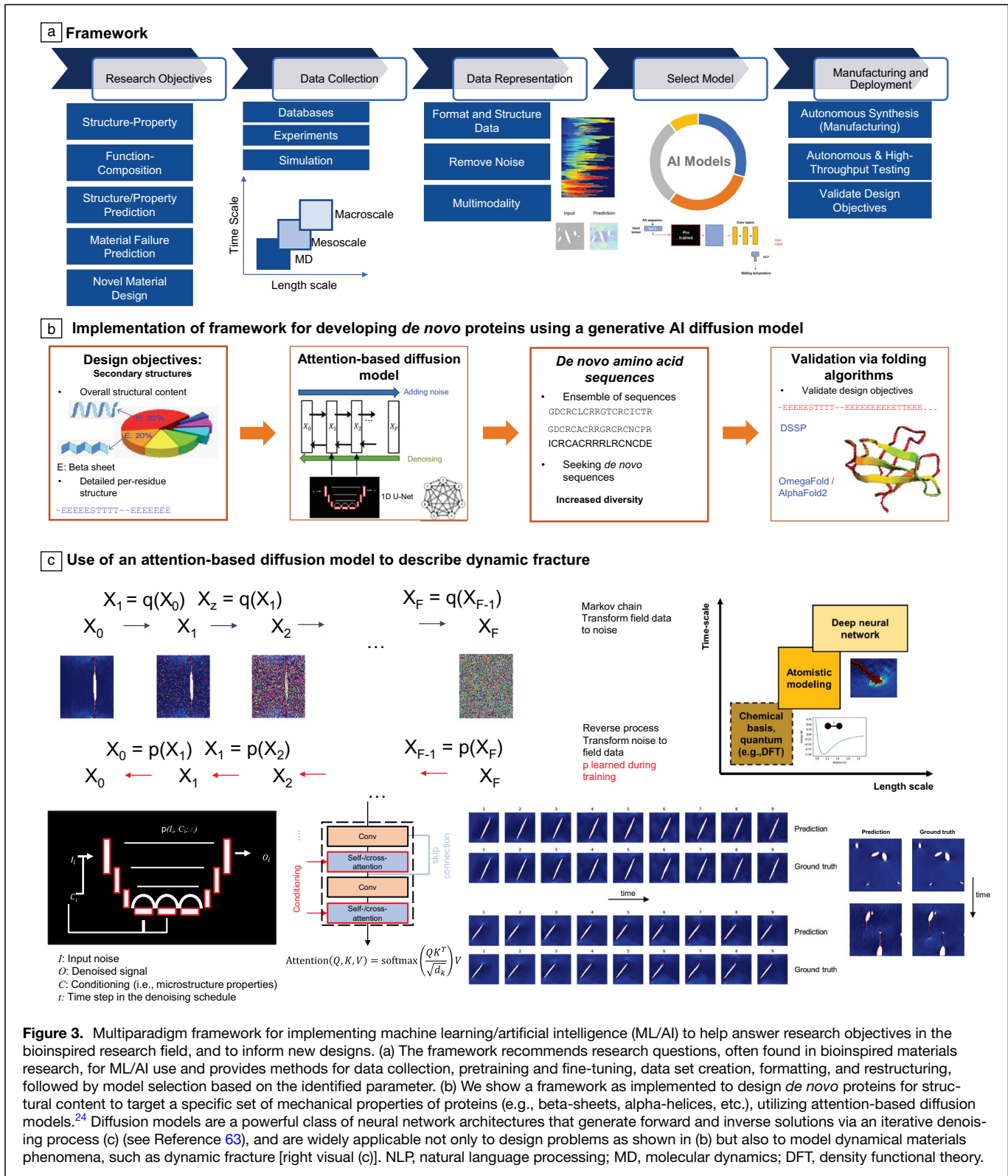
Attention-based language in materials design

Emerging tools in this domain further include the use of attention-based language models such as the transformer architecture that allows pretraining with unlabeled data or poorly structured data and then fine-tuning, which involves providing a small data set that is specific and of high quality relevant to solve a variety of specific tasks[60] (Figures 2a, 3a). As a result, transformer models are generalizable and can be applied to a myriad of predictive tasks. Often, transformer models are compared to the way humans solve scientific problems; this is done so by focusing attention on a specific aspect of information and observing how it relates to the output.[60] The transformer architecture involves a neural olog description that describes emerging materiomic relationships between building blocks.[61] The structure of an olog, or ontology log, stems from a branch of mathematics known as category theory. Ologs are similar to database schemas and due to their effective graph representation often serve as data repositories, and have many advantages. For example, ologs of different materials or systems can be compared by using functors and can be extended with new information, and used in various materials tasks ranging from analysis to design (see Reference 62). Neural ologs aim to generate categories using graph-generating algorithms such as transformer models. As such, by using neural ologs, the model discovers the relationships based on the data and uses them to make predictions or to develop analogies between dissimilar or seemingly unrelated representations. Embedding features provide information based on the organization of the building blocks that make up a material, including also a description of the processes used.

Through the embeddings, important information is provided, this allows for the transformer model to learn high-dimension embeddings that are then used to solve tasks. In materials research, transformer models are used to predict materials properties. And implementing semi-supervised learning is useful for developing new formulations with desired properties. As an example, using graph neural network (GNN) models, graph-structured data, in conjunction with transformer models can be applied for rapid mechanical property prediction of a material's strength and toughness, and *de novo* design of 3D web structures.[1,61,63,64] By the same token, using attention-based transformer networks can be applied to solve problems in protein folding research,[65,66] and molecular property and field predictions,[67,68] and interaction with human language, logic, and mathematics.

The transformer model architecture is emerging as an important tool for materials research, and when combined with other architectures can aid to predict materials properties and the design of novel materials. Furthermore, because transformer models belong to natural language processing (NLP), it allows designers to obtain biological analogies and supports biomimetic design as multiple data modalities can easily be mixed (numbers, instructions, processing steps, sequences, images, and so on). Specific to bioinspired materials research, NLP has the potential to bridge biological information and biomimetic design, as the biological information can yield meaning to inform the design. Notable examples of NLP are various types of transformer models such as encoder-only approaches (e.g., BERT) and decoder-only strategies (e.g., GPT-3/4, LLaMA, Falcon, etc.).[60] Researchers also discuss the opportunity to use natural language models through attention neural networks to transform information into features and into properties without requiring labeled data or large amounts of data, relying on a variety of pretraining strategies that endow these models with general knowledge about materials that can then be fine-tuned to solve specific challenges.[60]

For instance, flexible language models have recently been proposed to solve various both forward and inverse problems for modeling proteins.[69] The model is based on attention neural networks that integrate transformer and graph convolutional architectures to take advantage of the graph-forming capacity of attention models. This results in a generative pretrained model, often abbreviated as "GPT" and also found as the basis for ChatGPT, for instance, but it is scientifically formulated and can be a powerful platform to solve multimodal tasks. The goal of the model reported in Reference 69 is to predict secondary structure content, protein solubility, and sequencing tasks, and the same model can be used to design proteins with desired properties as well.[70] Because this model is prompt-based, it is able to process, discover, and synergize within diverse information from sequence data to numbers.[69] With the use of interpretable methods, we can mine attention maps and glean insights into the unitary mechanics of how predictions are derived (Figure 2a–b).

Figure 3. Multiparadigm framework for implementing machine learning/artificial intelligence (ML/AI) to help answer research objectives in the bioinspired research field, and to inform new designs. (a) The framework recommends research questions, often found in bioinspired materials research, for ML/AI use and provides methods for data collection, pretraining and fine-tuning, data set creation, formatting, and restructuring, followed by model selection based on the identified parameter. (b) We show a framework as implemented to design *de novo* proteins for structural content to target a specific set of mechanical properties of proteins (e.g., beta-sheets, alpha-helices, etc.), utilizing attention-based diffusion models.[24] Diffusion models are a powerful class of neural network architectures that generate forward and inverse solutions via an iterative denoising process (c) (see Reference 63), and are widely applicable not only to design problems as shown in (b) but also to model dynamical materials phenomena, such as dynamic fracture [right visual (c)]. NLP, natural language processing; MD, molecular dynamics; DFT, density functional theory.

Other applications of deep learning models are used to predict mechanical properties and properties of deep eutectic solvents,[64] in which the objective was to predict the sequence structure of proteins given desired properties, referred to as the inverse model, and the forward model, for which the properties are predicted based on a sequence structure. A diffusion model and autoregressive transformer were implemented and observed the model's ability to yield accurate predictions

when the model was trained on disparate data.[64] This result is promising for the future of modeling, especially when an application domain or field contains relatively small data sets.

Unsupervised learning in materials design

In unsupervised learning, there is often no labeled "ground truth," but rather the goal is to extract and summarize the content of the data to capture internal relationships and organization; for instance, to discover innate physical relationships in a reduced-dimensional space. Within unsupervised learning, there is cluster analysis and dimensionality reduction. In cluster analysis, the data are placed into groups based on similarity, resulting in each object being assigned to a bin. The number of bins is unknown and the objects are high-dimensional vectors. The goal is to minimize the Euclidean distance of each object to the centroid of its assigned cluster. As for dimensionality reduction, the objective is to find a simplified representation of the complex multidimensional data objects in a lower-dimensional space.

Examples of unsupervised methods include k-means clustering, autoencoders, diffusion models, and generative adversarial networks. Generative adversarial networks are a type of deep neural network that uses data statistics of the training set to generate new data. In the generative adversarial network, there is a generator and a discriminator that is trained against each other using game theory. The generator is tasked with producing candidates, whereas the discriminator evaluates the candidates. Generative models are used in materials research to generate collagen sequences with specific properties such as melting temperatures[23] and to model and synthesize artificial and bioinspired 3D web structures.[71] These methods are instrumental to developing novel structures, materials for manufacturing, and biomedical applications and may serve as an approach to the materials by design paradigm.

Although generative adversarial networks are initially unsupervised learning, labels can be added as constraints, which leads to conditional generative adversarial networks (cGANs). To contextualize this method, a U-net serves as a generator that translates high-dimensional microstructure input data into high-dimensional output field data, via the use of a series of convolutional operators that effectively act as coarse-graining mechanisms. The generative model uses paired images as the constraint. The field images produced by the generator have random noise, and the discriminator evaluates the field images by comparing them with real field images. In the field of materials research, cGAN is implemented to predict physical fields directly from a material's microstructure and predicts field data and materials properties.[57] The implementation of cGAN aids to bridge the gap between the microstructure–design–physical performance of a material. Further, it is a powerful technique to improve the efficiency and evaluation of physical properties for materials with hierarchical structures. A widely used example of unsupervised learning is the set of autoencoders, which is a three-building-block neural network primarily used to extract features from an input, for example, extracting features from an image. The three building blocks are (1) the input layer, (2) the hidden layer and bottleneck, and (3) the output layer. The network thereby implements encoder and decoder functionalities. The encoder function maps the given input to a hidden, latent, and typically lower-dimensional representation and the decoder function maps the latent representation to the output. The decoder is then tasked with reconstructing the input representation. In applications of such a model, autoencoders may be used to explore the discovery of a coarse-grained representation of phenomena in latent design space, which can be used to better understand key physical mechanisms, and also used for optimizing a material.[56] Autoencoders can be useful for design and discovery as they can learn the problem of a material and can be used to reconstruct complex data from the low-dimensional encoding of only a few variables. These models can be combined with language models, for instance via the use of discrete autoencoder methods (e.g., vector-quantized autoencoders) that express latent space representations as a set of symbols that can be understood as an abstract presentation.[69,72]

Reinforcement learning in materials design

Reinforcement learning involves an intelligent agent interacting with the environment to learn and decide the subsequent actions to maximize the cumulative reward. For example, researchers Yu et al. proposed a new method that uses reinforcement learning to design bioinspired composite materials.[73] A breakthrough of their research is the ability to achieve high-resolution designs in their composite, reducing stress concentrations at the crack tips, and thereby enhancing mechanical behavior.[73] The motivation to design materials using fewer computational resources with desired properties that mimic biological materials has led to an influx of novel machine learning methods. Most notably, reinforcement learning is often implemented as a way to optimize properties or synthetic reaction routes for multistep reactions, as an example in the semiconductor nanoparticles.[74–76] The incentive for implementing reinforcement learning is to continue accelerating the materials discovery process with minimal intervention. This can also be combined with autonomous laboratories to establish self-driving experimentation setups that have the potential to significantly accelerate research and discovery.

Survey of bioinspired materials research: Computational multiscale modeling toward ML/AI

While conventional computational models focus on utilizing multiscale modeling techniques (e.g., MD, quasi-continuum methods, and continuum methods), newer research often focuses on complementing the modeling with ML/AI to design and analyze hierarchical composites and architected materials. Many computational models are computationally expensive. Thus, there is a need to find alternative ways, such as ML/AI, to understand, predict, and control the properties of materials, as a complement to conventional methods.

The use of molecular modeling in conjunction with machine learning has been widely employed in non-bioinspired materials. For example, deep learning has been applied to predict fracture patterns in crystalline solids,[77] and fracture mechanisms in graphene[34] and polycrystalline 2D materials.[78] Although deep learning is used to predict fracture in nonbiological materials, the concepts and algorithms developed may be applied to predict fracture behavior in biological composites and the design of high-performance materials.[77]

An interesting field of using machine learning is finding patterns and finding relationships across disparate knowledge domains. For example, an unsupervised deep learning method has been proposed that can relate design strategies across disparate modalities, such as seen in art and science, in particular, they are connecting musical data to develop novel protein sequences.[79,80] In doing so, it provides an avenue for connecting contrasting fields to develop novel materials, for example, finding utility in other engineering domains, evolutionary patterns, or other types of sequences and proteins.[80]

It is common for researchers to utilize simulated data to feed into an ML algorithm. Specifically, in earlier work, we used a deep neural network architecture that uses data from coarse-grained simulations as a training data set of hierarchical microstructures.[72] This allows for the analysis and design of multiscale architected materials and the ability to solve both forward and inverse research problems.[72] Similarly, ML was used as an alternative to coarse-graining (mathematical modeling) to design hierarchical materials with superior toughness and strength.[81] Researchers have also shown that while MD simulation may take a significant amount of time to complete, ML offers an alternative solution to answering research questions of fracture mechanics much faster.[61]

Based on prior research, ML is used to solve important research questions. For example, Maurizi and colleagues use the inverse-design approach to find an optimal design for the bucking resistance of lattice structures. The inverse-design approach involves finding optimal combinations of properties by manipulating the architecture of materials. The design space for finding those optimal properties is a large design landscape, thus, remaining a complex challenge for designers. Researchers used deep neural networks and genetic algorithms to efficiently and effectively find the optimal properties.[81–83] While some researchers aim to use ML tools to optimize structures, others use ML to observe nature's patterns (e.g., spider webs, protein patterns) to develop bioinspired designs.[84] Similarly, in other work inspiration from nature was used—here, specifically leaf microstructures, and employed via powerful generative adversarial network models to develop novel 2D and 3D architected materials.[85] They then optimized the properties of the generated architected materials using genetic algorithms.[85] Creating composites with tunable materials to achieve superior mechanical properties is a grand challenge when using a trial-and-error approach. However, ML techniques have paved the way for exploring the large design space

and accelerating the development of functional and tailorable materials.[22]

Bioinspiration, synthesis, and experimental validation

Engineers draw inspiration from biological materials to create superior designs for materials. Examples of interesting biological behavior and materials worth imitating are biocoatings, biominerals, and biopolymers. After thousands of years of evolution, nature has developed biological materials that offer exciting opportunities in engineering. As illustrated in Figure 2, there is a wide span of materials properties existing in nature (Figure 2c) and biological materials that exhibit desirable properties (Figure 2d) with potential applications as bioadhesives, nanocomposites, and in "smart" structural materials, to name a few. Further, with the continual emergence of powerful ML architectures, connecting disparate fields, such as art and science, to develop novel materials is now possible, as shown in Figure 2e.

Notably, the mussel's adhesive ability and its coatings that provide extensibility, particularly in an underwater environment and exposed to constant perturbation of waves, are of great interest to scientists. Researchers have investigated in great depth the mechanisms and architecture of mussels' byssus. Mussels use the byssus thread as a tethering device that withstands high strains from waves. The byssus thread is stiff while being able to experience high extensibility, a difficult combination of properties to imitate in synthetic materials. Typical mussels have about 50–100 threads that are attached near the foot base, allowing them to attach to nearby foreign objects and extend in multiple directions.[86] Although only a small handful of biological lessons translate to technological breakthroughs, mussels have provided insight into designing synthetic polymers displaying high stiffness, high hardness, high extensibility, and self-healing properties.[87] A key finding demonstrated the presence of metals in the mussel's cuticle, a coating of the byssus thread, results in increased hardness.[88] As with organisms that utilize metal complexation with dopa, they serve as noncovalent cross-linking agents.[88] From nature, polymer networks containing metal in addition to the complex structures allow for mimicking biological behaviors in synthetic materials.[87,89,90] The mechanical properties of advanced structural materials can be engineered using supramolecular cross-linking, a wide-ranging class that includes coordination with transition metals. By utilizing metal coordination bonding with transition metals, scientists are developing polymeric hydrogels that are capable of self-healing as their bonds can reform even after rupture.[87,89,90] In metal coordination, the metal ions coordinate with ligands to bond by donating two electrons from the ligand to the metal ion.[91] Compared to other bonding types, tuning properties is relatively easier with metal coordination bonding. It is possible to adjust these properties by altering the chemistry, and pH, and exchanging the metal ions. By adjusting the properties, metal coordination bonding

may span across a range of strengths and time scales.[91] In addition to controlling the dynamic mechanical properties, chelators can be used to "switch off" the coordination complex. Broomell and colleagues showed that chelating Zn from *Nereis*'s jaws results in a 65% decrease in modulus and hardness.[92] By understanding the role of transition metals as a hardening strategy in nature, synthetic materials can be engineered to improve stiffness, wear resistance, hardness, and toughness.[92] Although Zn was provided in this review, other transition metals such as copper, manganese, and cadmium can also contribute to the hardening of jaws in biological organisms.[92] Besides studying marine organisms and their interactions with metal ions, it is worth finding other biological organisms to engineer materials.

A concerted effort is to construct multiscale design principles to connect chemical dynamics to material mechanics via experimentation and computation. Predicting the mechanical properties of metal coordination bonding, a common design motif seen in nature, remains a challenge in the field. However, recent efforts show the use of meta-dynamics and macroscopic relaxation time for a metal-coordinated hydrogel.[93] Their findings show a quantitative empirical relationship between bond energy landscape and bulk network relaxation time.[93] This research is fundamental to the understanding of the microscopic origin of the macroscopic behavior that would enable improved load-bearing design materials composed of metal coordination in the future.[93] Although molecular modeling offers an opportunity to investigate the microscopic origin of the behavior of metal coordination, exploring the bond dynamics remains challenging because of the limitations of the metal ion force fields and long sampling methods to obtain relevant time scales. As engineers implement metal coordination to develop self-healing, it is imperative to develop structure–property relationships. Understanding the structure-property relationships for these materials are important as there are various parameters such as bond energy, bond dynamics, and coordination number that may influence the mechanical and self-healing properties.[89] Further, the authors argue that there is a need to characterize the exact contributions of geometric arrangements of metal coordination bonds on mechanical behavior.[94] As such, they explored the rupture force of histidine and nickel(II) and found that the bonds rupture in groups of two or three bonds.[94] Applications of polymeric-containing-coordination complex are for building deformable composites for use in crash protection and armored clothing.[95] Similarly, Li et al. used a 2,6 pyridinedicarboxamide ligand coordinated with Fe(III) to develop highly stretchable and autonomous self-healing materials.[90] Another fascinating biological material highly explored is collagen because of its contribution to the mechanical properties of various biomaterials (tendon, bone, skin, blood vessels, and many others), owing to its crucial structural proteins and hierarchical organization. To forecast the thermal stability of collagen, researchers employed transformer models.[96] They identified which model is most effective at predicting specific biophysical properties with a limited training data set.[96] The inverse of the problem, identifying collagen sequences with specific thermal temperatures, can be achieved using genetic algorithms and experimental validation, providing *de novo* design strategies for the discovery of specific proteins to meet target properties.[23] To predict the thermal stability of collagen, researchers utilize ML, namely, two distinct types of transformer models. The goal is to identify which model is best to predict specific biophysical properties with a limited training data set. This research was the first to use transformer models for small data sets to predict specific biophysical properties.[23]

Silk is another alluring biological material as it exhibits superior mechanical properties and biocompatibility.[97] Researchers have implemented artificial intelligence to predict mechanical behavior based on the amino acid sequence, with the goal of optimizing its mechanical properties for desired applications.[38] However, the challenge to duplicate silk into synthetic materials remains. Spider silks can have a variety of mechanical properties as they are tailored for specific needs. Some forms of silk may exhibit properties in the range of high-tech materials, up to 1.7 GPa in strength. Another example is the dragline silks, which are lightweight and when hydrated undergo super-contraction.[98] When the toughness is compared to steel and Kevlar by weight, dragline silks outperform.[97]

Framework: Implementation of ML to bioinspired materials research

ML models are particularly of interest to implement in materials science as a tool to predict and optimize properties. Researchers in materials science, particularly those focusing on bioinspired design, implement ML to predict and optimize properties based on both experimental and simulation data. Mechanical properties such as strength, toughness, failure mode, yield stress, and strain are often optimized using convolutional neural networks, recurrent neural networks, genetic algorithms, and Gaussian processes.[60,78,99] Reinforcement learning may also be used to automate the design of digital composites and the discovery of composite structures. Modern language models include recurrent neural networks, LSTM, attention-based transformer networks for protein folding, and molecular property prediction.

The general framework for implementing ML methods is by identifying the research question. As illustrated in Figure 3a, researchers in bioinspired design and biomaterials have used these methods to obtain structure–property relationships, and as a way to design novel materials, to name a few. However, an essential parameter to then consider is the type of data and how to amass and organize the data to train models. The model selection, whether it is supervised, unsupervised, reinforcement, or semi-supervised will depend on the research question. One approach is to apply the data set to existing ML algorithms such as diffusion models, generative models, etc. However, these models are not a one-size fit and may require a combination of existing programming languages, libraries,

and tools. These tools provide prebuilt functions and modules that help build, train, and implement ML models. These open-source libraries may aid in tasks such as simplifying the process of building and training the models and may be applied to analyze, model, cluster, and develop dimensionality reduction algorithms that may be used in computer vision, and natural language processing. By following a typical framework for implementing ML/AI, it can be applied to develop *de novo* sequences as shown in Figure 3b. A more in-depth step-by-step guideline for the implementation of ML in materials science is provided by Wang et al.[100] Although Wang et al. provide fundamental guidelines for best practices, the reader is encouraged to understand the more modern architectures, as ML is a fast-moving field that brings important methodological advances. As such, an instrumental and deeper dive into the theory behind AI with a specific emphasis on quantum, atomistic, and continuum systems for biological systems is provided by Zhang et al.[101]

The challenges with implementing machine learning are the collection of data using high-throughput methods, validation using multiscale models or experiments, and further research on the generalization potentials of such models. An area of intense interest is mining deep models for insights gained, which is especially interesting for transformer architectures where attention maps can yield important graph-based insights into structure–function relationships. Further, a prevalent challenge with bioinspired design is the translation of knowledge from the biological material to developing an engineering solution, owing to limitations in the manufacturing and large-scale implementation of these complex materials. Bioinspired design is a fascinating field with vast opportunities for mimicking structures and applying them to a wide array of engineering designs.[14] Although a significant number of reviews survey hierarchical structures of biological materials, there is much more to discover and learn from biological materials to apply to bioinspired design.[102–105] As a result, techniques such as ML/AI that enable the identification of structural patterns across an immense array of existing biological organisms, and the study of the fundamental mechanisms across multiscale are imperative for expeditious surveying and implementation of bioinspired design (Figure 3c).
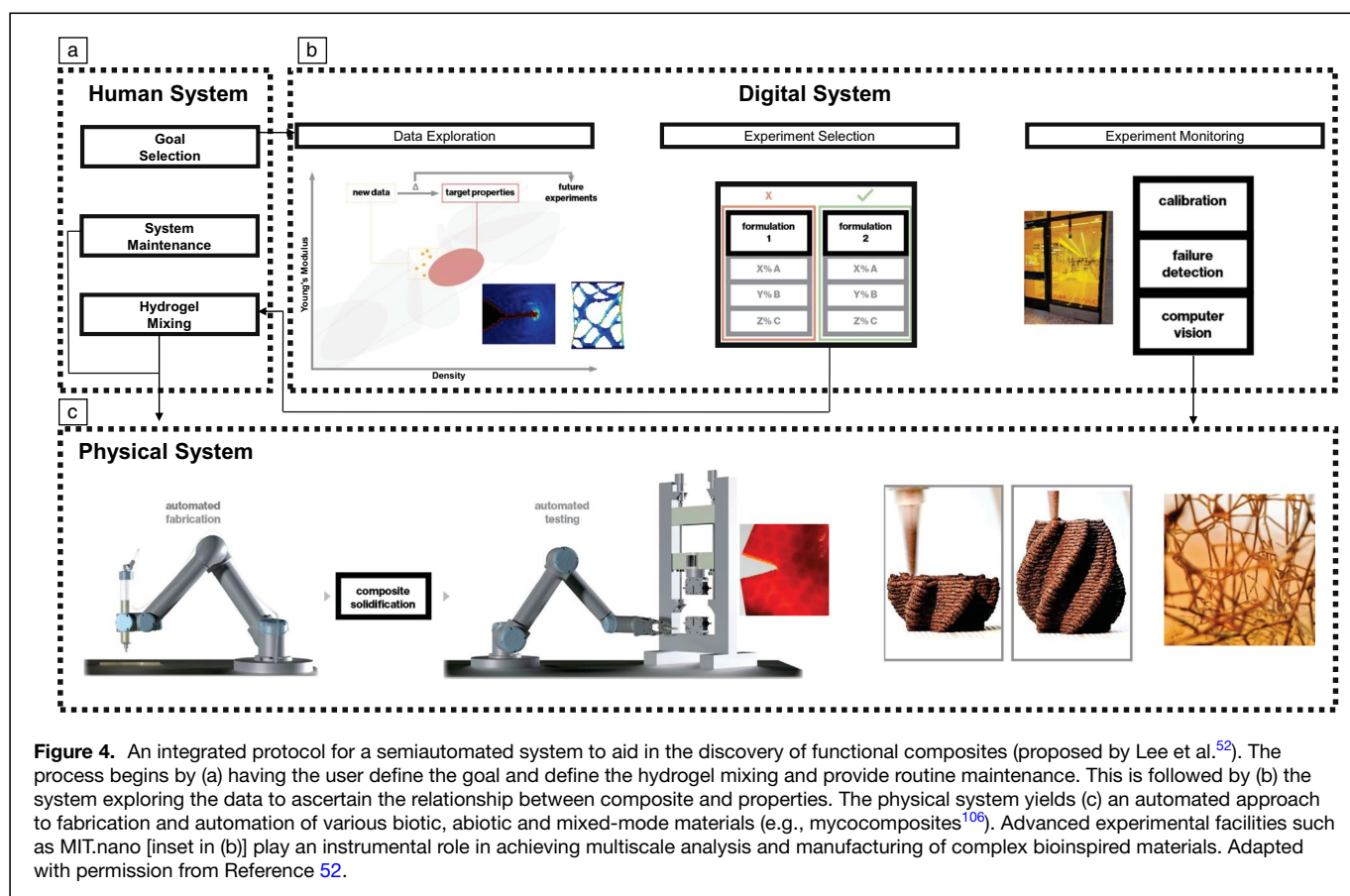
Conclusion and future perspectives

Multiscale modeling, complemented, supported, or enhanced by ML/AI methods, has been instrumental in advancing scientific knowledge in a variety of biomaterials fields ranging from tissue engineering to *de novo* materials design, as an avenue to design biomimetic constructs and other innovations in materials design, inspired by nature such as marine worms, silks, and biominerals. ML is often implemented in materials science to identify suitable material candidates for a desired application, gain insights into mechanisms in biological or bioinspired materials, and develop novel modeling techniques to address materials science questions. A particularly powerful combination is to use data-driven with

physics-based modeling and supported by experimental synthesis or data collection as it allows us to deal with the high degree of complexities.

Implementing ML remains a challenge because such frameworks often rely on large amounts of data for training to increase accuracy. Although limited data continue to be a challenge for the implementation of ML in many fields, the advent of novel architecture, such as the transformer architecture allows for using these models with limited, or distinct types of data. Transformer modeling architecture allows for the use of pretraining on a subset of low-quality data but is then followed by a smaller subset of high-quality data, in a process referred to as fine-tuning,[63] providing an effective way to build materiomic graph-based material models (e.g., Figure 2a–b). This further allows for solving multiple tasks such as the inverse and forward problem, within one model, and even capturing multiple modalities such as text, images, and numbers.[69] Most importantly, in terms of materials design, transformer models are adept at solving mechanics problems because of their flexible framework to capture complex relationships. They do so by using a language approach, which breaks down mechanics into elementary building blocks (Figure 2) and can be further augmented by physics-inspired diffusion modeling approaches (Figure 3). Transformer models are interesting because of their generalizability and ability to be applicable to other problems, besides those they have been trained for, including straightforward unsupervised pretraining and intrinsic multimodality (from text to images to symbolism, including math). The push for mining data autonomously and using machine learning to draw relationships between data for biological materials is because biological designs do not follow known mathematical laws.

Developing an automated method to obtain large amounts of data is especially important when publicly available data from databases are scarce. Thus, to mitigate the scarcity of data sets for training an ML algorithm, laboratories are proposing developing high-throughput methods to generate large volumes of data from experiments.[52,57,105] A proposed protocol is summarized in **Figure 4**. In this method, the user identifies the objective and the digital system selects the appropriate experimental parameters, and the physical system fabricates and tests the samples.[52] Alternatively, mathematical simulations at the desired length scale can supplement and produce large amounts of data for training. Aside from accessibility to data sets, methods to address and clean up noisy data sets, standardization, data sharing, and interpreting new ML models also play an integral role in slowing down innovation progress. Nevertheless, major advancements have been made in developing novel ML models that overcome small-data set issues. In sum, ML/AI have enabled the development of novel materials and answer key questions spanning from proteins to developing novel composites.[84] Further, deploying these novel computational frameworks to develop materials inspired by nature can lead to breakthroughs in various industries and contribute to a more sustainable future.

Figure 4. An integrated protocol for a semiautomated system to aid in the discovery of functional composites (proposed by Lee et al.[52]). The process begins by (a) having the user define the goal and define the hydrogel mixing and provide routine maintenance. This is followed by (b) the system exploring the data to ascertain the relationship between composite and properties. The physical system yields (c) an automated approach to fabrication and automation of various biotic, abiotic and mixed-mode materials (e.g., mycocomposites[106]). Advanced experimental facilities such as MIT.nano [inset in (b)] play an instrumental role in achieving multiscale analysis and manufacturing of complex bioinspired materials. Adapted with permission from Reference 52.

Funding

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access

References

1. D. Nepal, S. Kang, K.M. Adstedt, K. Kanhaiya, M.R. Bockstaller, L.C. Brinson, M.J. Buehler, P.V. Coveney, K. Dayal, J.A. El-Awady, L.C. Henderson, D.L. Kaplan, S. Keten, N.A. Kotov, G.C. Schatz, S. Vignolini, F. Vollrath, Y. Wang, B.I. Yakobson, V.V. Tsukruk, H. Heinz, *Nat. Mater.* **22**(1), 18 (2023)
2. H.D. Espinosa, J.E. Rim, F. Barthelat, M.J. Buehler, *Prog. Mater. Sci.* **54**, 1059 (2009)
3. M.J. Buehler, *J. Comput. Theor. Nanosci.* **7**, 1203 (2010)
4. S.C. Shen, E. Khare, N.A. Lee, M.K. Saad, D.L. Kaplan, M.J. Buehler, *Chem. Rev.* **123**, 2242 (2023)
5. A. Gautieri, S. Vesentini, A. Redaelli, M.J. Buehler, *Nano Lett.* **11**(2), 757 (2011)
6. P. Fratzl, *J. R. Soc. Interface* **4**, 637 (2007)
7. O. Tokareva, M. Jacobsen, M. Buehler, J. Wong, D.L. Kaplan, *Acta Biomater.* **10**, 1612 (2014)
8. K. Guo, M.J. Buehler, *Matter* **1**, 302 (2019)
9. M.A. Meyers, P.-Y. Chen, A.Y.-M. Lin, Y. Seki, *Prog. Mater. Sci.* **53**(1), 1 (2008)
10. E. Beniash, C.A. Stifler, C.-Y. Sun, G.S. Jung, Z. Qin, M.J. Buehler, P.U.P.A. Gilbert, *Nat. Commun.* **10**(1), 4383 (2019)
11. U.G.K. Wegst, H. Bai, E. Saiz, A.P. Tomsia, R.O. Ritchie, *Nat. Mater.* **14**, 23 (2015)
12. Z. Liu, M.A. Meyers, Z. Zhang, R.O. Ritchie, *Prog. Mater. Sci.* **88**, 467 (2017)
13. N. Holten-Andersen, M.J. Harrington, H. Birkedal, B.P. Lee, P.B. Messersmith, K.Y.C. Lee, J.H. Waite, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2651 (2011)
14. S.E. Naleway, M.M. Porter, J. McKittrick, M.A. Meyers, *Adv. Mater.* **27**, 5455 (2015)
15. R.O. Ritchie, *Nat. Mater.* **10**, 817 (2011)
16. S. Tadepalli, J.M. Slocik, M.K. Gupta, R.R. Naik, S. Singamaneni, *Chem. Rev.* **117**, 12705 (2017)
17. S. Zhou, K. Jin, T. Khan, Z. Martin-Moldes, D.L. Kaplan, M.J. Buehler, *Adv. Eng. Mater.* **25**(6), 2201134 (2023)
18. D. López Barreiro, Z. Martín-Moldes, J. Yeo, S. Shen, M.J. Hawker, F.J. Martin-Martinez, D.L. Kaplan, M.J. Buehler, *Adv. Mater.* **31**, 1 (2019)

19. Y. Pei, W. Yang, K. Tang, D.L. Kaplan, *Biotechnol. Adv.* **63**, 108099 (2023)
20. W. Huang, D. Ebrahimi, N. Dinjaski, A. Tarakanova, M.J. Buehler, J.Y. Wong, D.L. Kaplan, *Acc. Chem. Res.* **50**(4), 866 (2017)
21. A.J. Lew, K. Jin, M.J. Buehler, *NPJ Comput. Mater.* **9**(1), 80 (2023). https://doi.org/10.1038/s41524-023-01036-1
22. G.X. Gu, C.T. Chen, M.J. Buehler, *Extreme Mech. Lett.* **18**, 19 (2018)
23. E. Khare, C.-H. Yu, C. Gonzalez Obeso, M. Milazzo, D.L. Kaplan, M.J. Buehler, *Proc. Natl. Acad. Sci. U.S.A.* **119**(40), e2209524119 (2022). https://doi.org/10.1073/pnas.2209524119
24. B. Ni, D.L. Kaplan, M.J. Buehler, *Chem.* **9**(7), 1828 (2023). https://doi.org/10.1016/j.chempr.2023.03.020
25. K. Guo, Z. Yang, C.H. Yu, M.J. Buehler, *Mater. Horiz.* **8**, 1153 (2021)
26. C.-H. Yu, E. Khare, O. Prakash Narayan, R. Parker, D.L. Kaplan, M. Buehler, *J. Mech. Behav. Biomed. Mater.* **125**, 104921 (2022)
27. W. Lu, Z. Yang, M.J. Buehler, *J. Appl. Phys.* **132**(7), 074703 (2022). https://doi.org/10.1063/5.0097589
28. B. Wang, W. Yang, V.R. Sherman, M.A. Meyers, *Acta Biomater.* **41**, 60 (2016)
29. T. Van Le, A. Ghazlan, T. Ngo, T. Nguyen, A. Remennikov, *Compos. Struct.* **225**, 111172 (2019)
30. M.A. Meyers, J. McKittrick, P.-Y. Chen, *Science* **339**, 773 (2013)
31. S.C. Shen, M. Peña Fernández, G. Tozzi, M.J. Buehler, *J. Mech. Behav. Biomed. Mater.* **123**, 104761 (2021)
32. Y.-C. Hsu, M.J. Buehler, *APL Mach. Learn.* **1**(2), 026105 (2023)
33. M. Abolhasani, K.A. Brown, *MRS Bull.* **48**(2), 134 (2023)
34. A.J. Lew, C.-H. Yu, Y.-C. Hsu, M.J. Buehler, *NPJ 2D Mater. Appl.* **5**(1), 48 (2021). https://doi.org/10.1038/s41699-021-00228-x
35. K.G. Reyes, B. Maruyama, *MRS Bull.* **44**(7), 530 (2019)
36. C.P. Gomes, B. Selman, J.M. Gregoire, *MRS Bull.* **44**(7), 538 (2019)
37. P. Ball, *MRS Bull.* **44**(5), 335 (2019)
38. D. López Barreiro, J. Yeo, A. Tarakanova, F.J. Martin-Martinez, M.J. Buehler, *Macromol. Biosci.* **19**(3), 1970007 (2019). https://doi.org/10.1002/mabi.201970007
39. Z. Yang, Y.C. Hsu, M.J. Buehler, *J. Mech. Phys. Solids* **170**, 105098 (2023)
40. S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, K. Schulten, "Steered Molecular Dynamics," in *Computational Molecular Dynamics: Challenges, Methods, Ideas, Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling* (Springer, Berlin, 1999), chap. 1, p. 39
41. T.Q. Yu, J. Lu, C.F. Abrams, E. Vanden-Eijnden, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11744 (2016)
42. J. Kästner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**(6), 932 (2011)
43. G. Bussi, A. Laio, *Nat. Rev. Phys.* **2**, 200 (2020)
44. C. Li, G.A. Voth, *J. Phys. Chem. B* **125**, 10471 (2021)
45. S.A. Hollingsworth, R.O. Dror, *Neuron* **99**, 1129 (2018)
46. O. Büyüköztürk, M.J. Buehler, D. Lau, C. Tuakta, *Int. J. Solids Struct.* **48**, 2131 (2011)
47. A.J. Lew, C.A. Stifler, A. Tits, C.A. Schmidt, A. Scholl, A. Cantamessa, L. Müller, Y. Delaunois, P. Compère, D. Ruffoni, M.J. Buehler, P.U.P.A. Gilbert, *Adv. Mater.* **35**(28), 2300373 (2023). https://doi.org/10.1002/adma.202300373
48. M.C. Demirel, M. Cetinkaya, A. Pena-Francesch, H. Jung, *Macromol. Biosci.* **15**, 300 (2015)
49. X. Qi, Y. Zhao, K. Lachowski, J. Boese, Y. Cai, O. Dollar, B. Hellner, L. Pozzo, J. Pfaendtner, J. Chun, F. Baneyx, C.J. Mundy, *ACS Nano* **16**, 1919 (2022)
50. A.P. Garcia, D. Sen, M.J. Buehler, *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **42**, 3889 (2011)
51. M. Cetinkaya, S. Xiao, F. Gräter, *Soft Matter* **7**, 8142 (2011)
52. N.A. Lee, S.C. Shen, M.J. Buehler, *Matter* **5**(11), 3597 (2022)
53. Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, B. Yu, *Neurocomputing* **323**, 37 (2019)
54. Y. Hu, M.J. Buehler, *APL Mach. Learn.* **1**(1), 010901 (2023)
55. M.J. Buehler, Diatom-inspired architected materials using language-based deep learning: Perception, transformation and manufacturing (2023), Preprint, arXiv:2301.05875
56. A.J. Lew, M.J. Buehler, *Forces Mech.* **5**, 100054 (2021)
57. Z. Yang, C.-H. Yu, M.J. Buehler, *Sci. Adv.* **7**(15), eabd7416 (2021). https://doi.org/10.1126/sciadv.abd7416
58. M.S. Khorrami, J.R. Mianroodi, N.H. Siboni, P. Goyal, B. Svendsen, P. Benner, D. Raabe, *NPJ Comput. Mater.* **9**, 37 (2023). https://doi.org/10.1038/s41524-023-00991-z
59. Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations (2020), Preprint, arXiv:2010.08895
60. M.J. Buehler, *Acc. Chem. Res.* **55**, 3387 (2022)
61. M.J. Buehler, *Mater. Today* **57**, 9 (2022)
62. D.I. Spivak, R.E. Kent, *PLoS One* **7**, e24274 (2012)
63. M.J. Buehler, *J. Appl. Mech.* **89**(12), 121009 (2022)
64. R.K. Luu, M. Wysokowski, M.J. Buehler, *Appl. Phys. Lett.* **122**, 234103 (2023)
65. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **596**, 583 (2021)
66. M. AlQuraishi, *Cell Syst.* **8**, 292 (2019)
67. E.L. Buehler, M.J. Buehler, *Biomed. Eng. Adv.* **4**, 100038 (2022)
68. C.H. Yu, Z. Qin, F.J. Martin-Martinez, M.J. Buehler, *ACS Nano* **13**, 7471 (2019)
69. M.J. Buehler, MeLM, a generative pretrained language modeling framework that solves forward and inverse mechanics problems (2023), Preprint, https://doi.org/10.48550/arXiv.2306.17525
70. M.J. Buehler, Generative pretrained autoregressive transformer graph neural network applied to the analysis and discovery of novel proteins (2023), Preprint, arXiv:2305.04934
71. W. Lu, N.A. Lee, M.J. Buehler, *Proc. Natl. Acad. Sci. U.S.A.* **120**(31), e2305273120 (2023). https://doi.org/10.1073/pnas.2305273120
72. M.J. Buehler, *Model. Simul. Mater. Sci. Eng.* **31**, 054001 (2023)
73. C.-H. Yu, B.-Y. Tseng, Z. Yang, C.-C. Tung, E. Zhao, Z.-F. Ren, S.-S. Yu, P.-Y. Chen, C.-S. Chen, M.J. Buehler, *Adv. Theory Simul.* **5**(11), 2200459 (2022)
74. A.A. Volk, R.W. Epps, D.T. Yonemoto, B.S. Masters, F.N. Castellano, K.G. Reyes, M. Abolhasani, *Nat. Commun.* **14**(1), 1403 (2023)
75. I.D. Lutz, S. Wang, C. Norn, A. Courbet, A.J. Borst, Y.T. Zhao, A. Dosey, L. Cao, J. Xu, E.M. Leaf, C. Treichel, P. Litvicov, Z. Li, A.D. Goodson, P. Rivera-Sánchez, A.M. Bratovianu, M. Baek, N.P. King, H. Ruohola-Baker, D. Baker, *Science* **380**, 266 (2023)
76. F. Sui, R. Guo, Z. Zhang, G.X. Gu, L. Lin, *ACS Mater. Lett.* **3**, 1433 (2021)
77. Y.C. Hsu, C.H. Yu, M.J. Buehler, *Matter* **3**, 197 (2020)
78. A.J. Lew, M.J. Buehler, *Appl. Phys. Rev.* **8**, 041414 (2021). https://doi.org/10.1063/5.0057162
79. M. Milazzo, G.I. Anderson, M.J. Buehler, *Bioinspir. Biomim.* **17**, 015001 (2022)
80. M.J. Buehler, *Patterns* **4**, 100692 (2023)
81. G.X. Gu, C.T. Chen, D.J. Richmond, M.J. Buehler, *Mater. Horiz.* **5**, 939 (2018)
82. M. Maurizi, C. Gao, F. Berto, *NPJ Comput. Mater.* **8**, 247 (2022). https://doi.org/10.1038/s41524-022-00938-w
83. Y.H. Chiu, Y.H. Liao, J.Y. Juang, *Polymers* (Basel) **15**, 281 (2023)
84. M.J. Buehler, *Oxf. Open Mater. Sci.* **2**, 33 (2022)
85. S.C. Shen, M.J. Buehler, *Commun. Eng.* **1**(1), 37 (2022)
86. J.H. Waite, N.H. Andersen, S. Jewhurst, C. Sun, *J. Adhes.* **81**, 297 (2005)
87. H. Li, P. Yang, P. Pageni, C. Tang, *Macromol. Rapid Commun.* **38**, 1700109 (2017)
88. N. Holten-Andersen, T.E. Mates, M.S. Toprak, G.D. Stucky, F.W. Zok, J.H. Waite, *Langmuir* **25**, 3323 (2009)
89. D.P. Wang, J.C. Lai, H.Y. Lai, S.R. Mo, K.Y. Zeng, C.H. Li, J.L. Zuo, *Inorg. Chem.* **57**, 3232 (2018)
90. C.H. Li, C. Wang, C. Keplinger, J.L. Zuo, L. Jin, Y. Sun, P. Zheng, Y. Cao, F. Lissel, C. Linder, X.Z. You, Z. Bao, *Nat. Chem.* **8**, 618 (2016)
91. E. Khare, N. Holten-Andersen, M.J. Buehler, *Nat. Rev. Mater.* **6**, 421 (2021)
92. C.C. Broomell, M.A. Mattoni, F.W. Zok, J.H. Waite, *J. Exp. Biol.* **209**, 3219 (2006)
93. E. Khare, S.A. Cazzell, J. Song, N. Holten-Andersen, M.J. Buehler, *Proc. Natl. Acad. Sci. U.S.A.* **120**(4), e2213160120 (2023)
94. E. Khare, D.S. Grewal, M.J. Buehler, *Nanoscale* **15**, 8578 (2023)
95. J.-C. Lai, X.-Y. Jia, D.-P. Wang, Y.-B. Deng, P. Zheng, C.-H. Li, J.-L. Zuo, Z. Bao, *Nat. Commun.* **10**(1), 1164 (2019)
96. E. Khare, C. Gonzalez-Obeso, D.L. Kaplan, M.J. Buehler, *ACS Biomater. Sci. Eng.* **8**, 4301 (2022)
97. C. Vepari, D.L. Kaplan, *Prog. Polym. Sci.* **32**, 991 (2007)
98. J.M. Gosline, M.W. Denny, M.E. DeMont, *Nature* **309**, 551 (1984)
99. A. Krizhevsky, I. Sutskever, G.E. Hinton, *Commun. ACM* **60**, 84 (2017)
100. A.Y.T. Wang, R.J. Murdock, S.K. Kauwe, A.O. Oliynyk, A. Gurlo, J. Brgoch, K.A. Persson, T.D. Sparks, *Chem. Mater.* **32**, 4954 (2020)
101. X. Zhang, L. Wang, J. Helwig, Y. Luo, C. Fu, Y. Xie, M. Liu, Y. Lin, Z. Xu, K. Yan, K. Adams, M. Weiler, X. Li, T. Fu, Y. Wang, H. Yu, Y. Xie, X. Fu, A. Strasser, S. Xu, Y. Liu, Y. Du, A. Saxton, H. Ling, H. Lawrence, H. Stärk, S. Gui, C. Edwards, N. Gao, A. Ladera, T.

Wu, E.F. Hofgard, A.M. Tehrani, R. Wang, A. Daigavane, M. Bohde, J. Kurtin, Q. Huang, T. Phung, M. Xu, C.K. Joshi, S.V. Mathis, K. Azizzadenesheli, A. Fang, A. Aspuru-Guzik, E. Bekkers, M. Bronstein, M. Zitnik, A. Anandkumar, S. Ermon, P. Liò, R. Yu, S. Günnemann, J. Leskovec, H. Ji, J. Sun, R. Barzilay, T. Jaakkola, C.W. Coley, X. Qian, X. Qian, T. Smidt, S. Ji, Artificial intelligence for science in quantum, atomistic, and continuum systems (2023), Preprint, arXiv:2307.08423

102. L.J. Gibson, *J. R. Soc. Interface* **9**, 2749 (2012)
103. B. Bar-On, H.D. Wagner, *J. Struct. Biol.* **183**, 149 (2013)
104. Q. Chen, N.M. Pugno, *J. Mech. Behav. Biomed. Mater.* **19**, 3 (2013)
105. R.A. Patel, M.A. Webb, *ACS Appl. Bio Mater.* (2022). https://doi.org/10.1021/acsabm.2c00962
106. S.C. Shen, N.A. Lee, W.J. Lockett, A.D. Acuil, H.B. Gazdus, B.N. Spitzer, M.J. Buehler, Robust myco-composites as a platform for versatile hybrid-living structural materials (2023), Preprint, arXiv:2305.12151 □

Markus J. Buehler is the Jerry McAfee Professor of Engineering at the Massachusetts Institute of Technology. He pursues modeling, design, and manufacturing approaches for advanced materials that offer greater resilience and controllable properties from the nano- to the macroscale. He pioneered the field of materiomics and demonstrated broad impacts in the study of mechanical properties of complex materials. His numerous awards include the Feynman Prize in Nanotechnology, the American Society of Mechanical Engineers Daniel C. Drucker Medal, and the J.R. Rice Medal. He is a member of the National Academy of Engineering. Buehler can be reached by email at mbuehler@mit.edu.

Sofia E. Arevalo is a postdoctoral fellow in the School of Engineering at the Massachusetts Institute of Technology. Her research focuses on modeling protein-metal complexes and implementing machine learning for developing novel biomaterials. She received her PhD degree in mechanical engineering from the University of California, Berkeley, in 2022, where she studied nanomechanical characterization of orthopedic polymers. Arevalo can be reached by email at sarevalo@mit.edu.