Please cite this article in press as: Tamasi and Gormley, Biologic formulation in a self-driving biomaterials lab, Cell Reports Physical Science (2022), https://doi.org/10.1016/j.xcrp.2022.101041

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Perspective

Biologic formulation in a self-driving biomaterials lab

Matthew J. Tamasi[1] and Adam J. Gormley[1,*]

SUMMARY

Biologics such as monoclonal antibodies (mAbs) and RNA therapeutics have revolutionized standard of care but present stability challenges due to their fragile structure. This is particularly true considering the demanding manufacturing, storage, distribution, and administration requirements that far exceed the otherwise stable biological environment from which they are derived. Therefore, the pharmaceutical industry routinely implements a suite of experiments to optimize formulations using a standard set of excipients that are known to enhance stability. While this process has been productive, the complexity of biologic-excipient interactions prevents an efficient transition to precise and tailored formulations. Recent advances in laboratory automation, high-throughput analytics, and artificial intelligence/machine learning (AI/ML) now provide a unique opportunity to fully automate the design process and provide next-generation formulations with remarkable durability. Here, we put forth a plan to develop a biomaterials acceleration platform (BioMAP) (i.e., self-driving biomaterials lab) focused initially on biologic formulation.

INTRODUCTION

Therapeutic proteins and vaccines, known as biologics, have proven themselves as fundamental technologies to human health. This impact is impressive considering how fragile many of these biomolecules are to standard pharmaceutical storage and handling conditions. To be an effective therapy, biologics must retain their complex structures from production to administration. If biologics denature during any point in this process, their therapeutic efficacy is lost. Thus, to ensure their stability, monumental formulation efforts using additives such as small-molecule stabilizers, polymer excipients, and surfactants are routine. In one clear example, hydrolytically unstable mRNA was encapsulated in lipid nanoparticles (LNPs) to provide the world population with remarkably effective vaccines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).[1] Similarly, many monoclonal antibodies (mAbs) aggregate at clinically relevant concentrations and temperatures, causing a significant drop in efficacy.[2] For this reason, the pharmaceutical industry invests significant resources in formulating stabilization solutions for new biologics.

Current approaches for biologic formulation rely on either human-based rational design, high-throughput (HTP) screening experiments, or some combination of the two. When using rational design, a team of scientists rely on their collective expertise and experience to predict the best combination of additives for a given therapeutic and objective. This approach can be challenging, as it pulls from a standard set of excipients and is unable to anticipate higher-order interactions and synergies that govern stabilization of these complex macromolecules. This limitation is

[1]Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

*Correspondence: adam.gormley@rutgers.edu

https://doi.org/10.1016/j.xcrp.2022.101041

Please cite this article in press as: Tamasi and Gormley, Biologic formulation in a self-driving biomaterials lab, Cell Reports Physical Science (2022), https://doi.org/10.1016/j.xcrp.2022.101041

partially addressed in HTP screening studies that aim to survey excipient combinations until an acceptable performance threshold is reached. The challenge is that screening studies are inherently inefficient and may not map the design space with enough resolution to uncover highly specific non-linear synergies to maximize performance. Even when thoughtful consideration is given to screening efforts to maximize information gain from a small number of experiments, such as design of experiments, the non-linear behavior of formulations continues to act as a barrier to their development and optimization.

To combat these obstacles, the potential of artificial intelligence (AI) and robot scientists in biologics formulation and biomaterial engineering has long been recognized.[3,4] Early proof-of-principle robot scientists, such as Adam, used to autonomously study functional genomics[5] have provided a glimpse at how automated experimentation can be combined with data-driven decision-making to investigate complex problems in biology. With many of the challenges in biologic formulation stemming from the inability to predict how the choice of excipients and formulation process ultimately effect performance, these tools are increasingly being adopted to map such connections.[6] In recent examples, machine learning (ML) models have been utilized to predict complicated phenomena such as aggregation temperature, viscosity, and long-term stability of mAbs at high concentration.[7,8] Further, models have even been adopted to iteratively optimize biologic formulations toward multiple objectives, enabling optimizations in both protein thermal stability and interfacial stability to be acquired simultaneously.[9] While these are exciting developments, such examples are likely only the beginning of the potential advancements that AI and robotics will bring to improved formulation.

BIOMATERIALS ACCELERATION PLATFORM (BioMAP)

In this white paper,[10] we propose developing a BioMAP for precision biologic formulation (Figure 1). This concept builds on recent work from our lab where we used a data-driven Design-Build-Test-Learn workflow to formulate life science enzymes[11] and therapeutic proteins.[12] In these studies, we optimized tailored polymer additives using HTP experimentation coupled with ML to formulate multiple unique proteins toward improved thermal stability. Robotic tools were used to manufacture polymer excipient libraries,[13] while scientists performed enzyme assays, trained ML models, exported AI/ML predicted designs, and initiated the cyclic workflow. Inspired by other efforts in developing MAPs for accelerating scientific discovery,[14] we believe it is possible to move beyond this proof-of-concept work and realize a fully autonomous BioMAP. Further, we seek to expand our material library to include generally recognized as safe (GRAS) excipients, allowing formulations to be optimized for both performance and speed to market. Finally, we aim to expand the impact of our BioMAP by launching new pipelines for LNP formulation. This center of excellence that combines the best of current practice with the power of ML and automated chemistry will allow industry partners to benefit from a step change in their formulation stability with unprecedented agility.

In designing a BioMAP capable of biologic formulation, we first consider how to organize a closed-loop Design-Build-Test-Learn workflow so that an efficient and fully autonomous discovery campaign can be implemented. As formulation efforts are directed at a diverse set of target biologics, an ideal BioMAP must contain a flexible array of automated instrumentation that is advantageous for each biologic type. mAbs and protein biologics can be primarily handled utilizing a flexible liquid handling robot, which will serve as the core instrument for combining reagents

Please cite this article in press as: Tamasi and Gormley, Biologic formulation in a self-driving biomaterials lab, Cell Reports Physical Science (2022), https://doi.org/10.1016/j.xcrp.2022.101041

CellPress
OPEN ACCESS

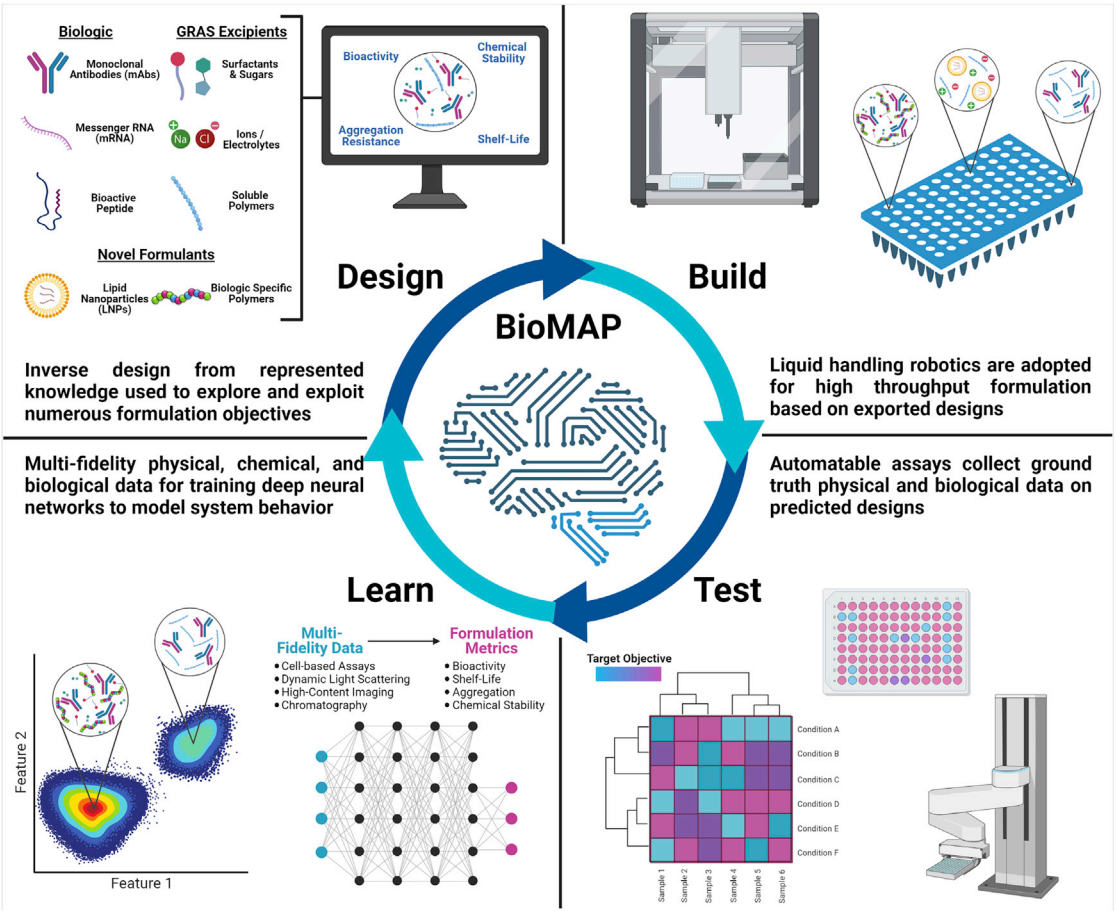

Figure 1. Overview of a closed-loop, self-driving biomaterials acceleration platform (BioMAP) for biologic formulation

Biologic formulation is performed entirely through autonomous workflows. Multisource data from physical and biological experiments are exploited by deep neural networks to map complex structure-function landscapes and inform downstream design campaigns. Figure created with BioRender.

and performing sterile cell culture. To increase the versatility of the core module, a multimode reagent dispenser, plate heater/shaker, plate sealer, and a vacuum filtration system will be available as supportive modules. Meanwhile, for LNP production, a parallelized microfluidic device[15] would be connected to a continuous-flow fluidics platform to form LNPs of varied compositions and lipid chemistry on demand. This automated instrumentation would enable direct formulation of both protein and RNA therapeutics (mRNA and small interfering RNA [siRNA]) in an HTP format for downstream functional testing. Once libraries of biologic formulations are made, analytical testing and characterization to understand formulation behavior can be readily performed in a highly reproducible automated workflow. An automated microplate incubator with humidity and CO_2 regulation can be included to test the stability of formulations over a range of storage and handling conditions. This same platform can also be used to support cell-based assays. On the analytical end, we will include a suite of standard instruments for continuous monitoring of new formulations and biological activity. This includes UV-visible (UV-vis) and dynamic light scattering (DLS) plate readers, size-exclusion chromatography (SEC), and a high content imager (HCI). Further analytical modules to expand the scope of the platform may include equipment for genomics, transcriptomics, proteomics, and flow cytometry. With this suite of automatable equipment, we would have the necessary hardware to implement a fully self-driving BioMAP.

Please cite this article in press as: Tamasi and Gormley, Biologic formulation in a self-driving biomaterials lab, Cell Reports Physical Science (2022), https://doi.org/10.1016/j.xcrp.2022.101041

To complement the Build-Test portion of the design campaign, data from experiments can be leveraged to inform Learn-Design efforts. To be machine interpretable, it is critical to first consider how features and data should be represented. As biologic formulations contain a diverse combination of small and macromolecules, feature engineering strategies that are amenable to all of these systems will be required. Initial approaches could include simple one-hot encoding strategies or the use of SMILES[16] paired with packages such as Mordred[17] to create chemical description vectors for formula components. These computationally efficient approaches have been shown to be effective for learning the behavior of macromolecular systems.[11,18] Further, as many experiments for assessing the physical, chemical, and biological properties of biologics have well-defined experimental outputs, collected data will be readily paired with these descriptors and immediately used for supervised ML. As deep neural networks have shown to be effective at learning in high-dimensional and non-linear data environments, we will couple Bayesian neural networks with Bayesian optimization strategies to learn from collected data and autonomously propose new experiments. This implementation will be critical as the potential permutations of formulation options are near infinite. Thus, Bayesian neural networks can be utilized for active learning approaches that intelligently explore and exploit formulation and biomaterial parameters in BioMAPs.[19] This approach enables an agile and efficient method for optimizing key formulation properties for biologics such as shelf-life, aggregation resistance, and chemical stability.

PLANS FOR PRACTICAL IMPLEMENTATION VIA HEILMEIER'S CATECHISM

Our long-term goal is to develop a comprehensive BioMAP capable of autonomously designing new biomaterials for drug delivery, immunoengineering, and regenerative medicine applications. To start, we will focus initially on the formulation of mAbs with GRAS additives and mRNA with LNPs. These biologics were selected due to strong clinical need and the opportunity for direct clinical translation.

Therapeutic mAbs are a critically important class of drugs. Precision formulation may provide new opportunities for concentrated mAb solutions with greater resistance to aggregation when shipped and stored at more cost- and energy-efficient conditions. Similarly, mRNA LNPs have proven their value during the COVID-19 pandemic; however, these critical vaccines were only possible due to outstanding formulations by Moderna and BioNTech that still required storage at extremely low temperatures, limiting their utility in rural and emerging markets.[20] As new variants or diseases emerge, we will need a more efficient and precise process for formulating new therapies and vaccines that enable rapid and widespread distribution.

Our proposed approach complements HTP automation with AI/ML for the data-driven design of new formulations. In this instance, all of the collected data will be used to continuously train deep Bayesian neural network models whose Bayesian optimizations guide new experimental campaigns within the closed-loop circuit. In this way, challenging formulation objectives with high dimensionality and non-linear behavior can be modeled and optimized. We believe that this approach will be successful, as proof-of-concept has already been implemented and validated via human-machine collaboration in our recent work.[11,12] Once properly resourced, we plan to make our facility open for collaboration so that a broad range of discovery campaigns can proceed in parallel, supporting existing demand from both

Please cite this article in press as: Tamasi and Gormley, Biologic formulation in a self-driving biomaterials lab, Cell Reports Physical Science (2022), https://doi.org/10.1016/j.xcrp.2022.101041

academia and industry. This would provide the entire biologics community with access to state-of-the-art AI/ML-directed formulations as an established center of excellence service, without needing to establish expertise or infrastructure in house.

Current barriers to fully deploy a BioMAP center around the overall complexity required to integrate experimental and analytical automation modules into seamless workflows. Similar to existing MAP examples, custom application programming interfaces (APIs) for connecting modules and controlling experimental processes will need to be developed. Further, strategic plans to incorporate automated validation and redundancies to minimize error will be paramount to inter-experimental reproducibility. Fortunately, the drug discovery community already has decades of experience building similar systems for HTP drug screening, potentially allowing for significant technology transfer toward BioMAPs. Feature engineering optimization is another area for growth and opportunity, as the goal for biologic formulation would be to build a flexible model that can transfer learn from each design campaign rather than generate new models for each new objective.

The financial barrier to full implementation and staffing of a biologic formulation BioMAP is modest, considering the opportunities for pharmaceutical industry savings. We estimate it will require $2.5 M in hardware and another $2 M in personnel over three years to establish the proposed BioMAP. To ensure consistent progress over this three-year period, the BioMAP will require a collaborative team of software engineers, biomedical engineers, pharmaceutical scientists, and data scientists. Assembling and resourcing this multidisciplinary and dynamic team represents the most significant go/no-go milestone for building and testing a BioMAP.

POTENTIAL FOR SOCIETAL IMPACT

In their 2021 Strategic Plan, the Materials Genome Initiative (MGI), a United States federal multiagency initiative, outlined an ambitious plan to unify the materials innovation infrastructure, harness the power of materials data, and educate, train, and connect the materials research and development (R&D) workforce.[21] Central to this plan is the strategic integration of laboratory automation and AI/ML into MAPs, which they predict may be one of our greatest opportunities to enter the next age of materials discovery.[22] While MAPs have been demonstrated for other critical materials science efforts,[10,14,23] no developed MAPs currently exist for the explicit development of bioformulations. As a field with enormous impact on health and well-being in society, we believe that advancing MAPs toward the invention of novel biologic formulations and biomaterials at large is a pressing need for the biomedical and materials sciences communities. If successful, BioMAPs could play a pivotal role in accelerating the rate at which new mAbs, RNA, and other biological therapies are brought to market. Further, as cold-chain logistics needed for transporting sensitive biologics are disproportionately lacking in underdeveloped regions of the world, novel stable formulations could improve global access to life-saving therapies.[24]

The outlined BioMAP builds on existing and synergistic projects in collaboration with the MGI through the National Science Foundation Designing Materials to Revolutionize and Engineer our Future (DMREF) program.[25] In this multidisciplinary and collaborative project, AI/ML methods are being developed for optimized protein formulation with a long-term goal of creating generalizable AI/ML models for protein formulation based on downloadable PDB files. This approach draws inspiration from DeepMind's AlphaFold[26] by aiming to publish models as readily accessible cyberinfrastructure for the broader research community. In line with the MGI's

mission, specific attention is focused on the featurization of polymers[18] and proteins so that past, present, and future data can be equally used to continuously evolve models as new knowledge is gained. To be machine interpretable, this project adopts a standardized approach to data handling and storage[27] so that data is findable, accessible, interoperable, and reusable (FAIR) through the National Materials Data Network (NMDN).[21] Ultimately, these AI/ML methods will serve as the core cyber platform for the proposed BioMAP, which will enshrine it as an active participant and contributor to the community of open science. The proposed BioMAP will also directly serve the scientific community as physical infrastructure. Companies and academics in need of biologic formulation may engage the BioMAP by providing material, an assay, and campaign objectives. Those interested in augmenting the discovery pipeline could gain access to the BioMAP as users to help install methods for new applications. In this way, the BioMAP can be continuously improved and updated to ensure it meets new demands by the scientific community.

VISIONS FOR A DISTRIBUTED BioMAP INFRASTRUCTURE

The field of biomaterials science and drug delivery is diverse and actively engaged in solving many scientific and medical challenges. This includes regenerative medicine, immunoengineering, medical diagnostics, and targeted cancer therapy. Successful validation of the proposed BioMAP for biologic formulation would motivate diversification into these new areas. Each field is likely to benefit from AI/ML, and it is possible that an expanded BioMAP could play a central role in these data-driven projects. As major elements of the platform are validated for diverse applications, we envision the installation of new BioMAP infrastructure around the world using this validated blueprint like the expansion of HTP drug screening facilities in the 1990s and 2000s. If expanded, BioMAP infrastructure in core facilities, national labs, and companies could become an integral part of biomedical research facilities worldwide as patients ultimately benefit from a realized dream of an automatable and AI-driven future.

ACKNOWLEDGMENTS

The authors acknowledge funding support from the National Science Foundation (DMREF-2118860).

AUTHOR CONTRIBUTIONS

M.J.T. and A.J.G. contributed equally to the conception, writing, and revision of this perspective.

DECLARATION OF INTERESTS

M.J.T. and A.J.G. have filed a PCT patent application and are co-founders of Plexymer, Inc.

REFERENCES

1. Schoenmaker, L., Witzigmann, D., Kulkarni, J.A., Verbeke, R., Kersten, G., Jiskoot, W., and Crommelin, D.J. (2021). mRNA-lipid nanoparticle COVID-19 vaccines: structure and stability. Int. J. Pharm. *601*, 120586. https://doi.org/10.1016/j.ijpharm.2021.120586.

2. Vázquez-Rey, M., and Lang, D.A. (2011). Aggregates in monoclonal antibody manufacturing processes. Biotechnol. Bioeng. *108*, 1494–1508. https://doi.org/10.1002/bit.23155.

3. Vasilevich, A., and de Boer, J. (2018). Robot-scientists will lead tomorrow's biomaterials discovery. Curr. Opin. Biomed. Eng. *6*, 74–80. https://doi.org/10.1016/j.cobme.2018.03.005.

4. Le, T.C., and Winkler, D.A. (2015). A bright future for evolutionary methods in drug design. ChemMedChem *10*, 1296–1300. https://doi.org/10.1021/cr200066h.

5. King, R.D., Rowland, J., Aubrey, W., Liakata, M., Markham, M., Soldatova, L.N., Whelan, K.E., Clare, A., Young, M., and Sparkes, A. (2009). The robot scientist Adam. Computer *42*, 46–54. https://doi.org/10.1109/MC.2009.270.

6. Bannigan, P., Aldeghi, M., Bao, Z., Häse, F., Aspuru-Guzik, A., and Allen, C. (2021). Machine

learning directed drug formulation development. Adv. Drug Deliv. Rev. *175*, 113806. https://doi.org/10.1016/j.addr.2021.05.016.

7. Gentiluomo, L., Roessner, D., and Frieß, W. (2020). Application of machine learning to predict monomer retention of therapeutic proteins after long term storage. Int. J. Pharm. *577*, 119039. https://doi.org/10.1016/j.ijpharm.2020.119039.

8. Lai, P.-K., Fernando, A., Cloutier, T.K., Gokarn, Y., Zhang, J., Schwenger, W., Chari, R., Calero-Rubio, C., and Trout, B.L. (2021). Machine learning applied to determine the molecular descriptors responsible for the viscosity behavior of concentrated therapeutic antibodies. Mol. Pharm. *18*, 1167–1175. https://doi.org/10.1021/acs.molpharmaceut.0c01073.

9. Narayanan, H., Dingfelder, F., Condado Morales, I., Patel, B., Heding, K.E., Bjelke, J.R., Egebjerg, T., Butté, A., Sokolov, M., and Lorenzen, N. (2021). Design of biopharmaceutical formulations accelerated by machine learning. Mol. Pharm. *18*, 3843–3853. https://doi.org/10.1021/acs.molpharmaceut.1c00469.

10. Seifrid, M., Hattrick-Simpers, J., Aspuru-Guzik, A., Kalil, T., and Cranford, S. (2022). Reaching critical MASS: crowdsourcing designs for the next generation of materials acceleration platforms. Matter. https://doi.org/10.1016/j.matt.2022.05.035.

11. Tamasi, M.J., Patel, R.A., Borca, C.H., Kosuri, S., Mugnier, H., Upadhya, R., et al. (2022). Machine learning on a robotic platform for the design of polymer-protein hybrids. Adv. Mater. *34*, 2201809. https://doi.org/10.1002/adma.202201809.

12. Kosuri, S., Borca, C.H., Mugnier, H., Tamasi, M., Patel, R.A., Perez, I., et al. (2022). Machine-assisted discovery of chondroitinase ABC complexes toward sustained neural regeneration. Adv. Healthc. Mater. *11*, 2102101. https://doi.org/10.1002/adhm.202102101.

13. Tamasi, M., Kosuri, S., DiStefano, J., Chapman, R., and Gormley, A.J. (2020). Automation of controlled/living radical polymerization. Adv. Intell. Syst. *2*, 1900126. https://doi.org/10.1002/aisy.201900126.

14. Stach, E., DeCost, B., Kusne, A.G., Hattrick-Simpers, J., Brown, K.A., Reyes, K.G., Schrier, J., Billinge, S., Buonassisi, T., and Foster, I. (2021). Autonomous experimentation systems for materials development: a community perspective. Matter *4*, 2702–2726. https://doi.org/10.1016/j.matt.2021.06.036.

15. Shepherd, S.J., Warzecha, C.C., Yadavali, S., El-Mayta, R., Alameh, M.-G., Wang, L., Weissman, D., Wilson, J.M., Issadore, D., and Mitchell, M.J. (2021). Scalable mRNA and siRNA lipid nanoparticle production using a parallelized microfluidic device. Nano Lett. *21*, 5671–5680. https://doi.org/10.1021/acs.nanolett.1c01353.

16. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. *28*, 31–36. https://doi.org/10.1021/ci00057a005.

17. Moriwaki, H., Tian, Y.-S., Kawashita, N., and Takagi, T. (2018). Mordred: a molecular descriptor calculator. J. Cheminf. *10*, 1–14. https://doi.org/10.1186/s13321-018-0258-y.

18. Patel, R.A., Borca, C.H., and Webb, M.A. (2022). Featurization strategies for polymer sequence or composition design by machine learning. Mol. Syst. Des. Eng. https://doi.org/10.1039/D1ME00160D.

19. Jospin, L.V., Laga, H., Boussaid, F., Buntine, W., and Bennamoun, M. (2022). Hands-on bayesian neural networks—a tutorial for deep learning users. IEEE Comput. Intell. Mag. *17*, 29–48. https://doi.org/10.1109/MCI.2022.3155327.

20. Mills, M.C., and Salisbury, D. (2021). The challenges of distributing COVID-19 vaccinations. eClinicalMedicine *31*. https://doi.org/10.1016/S0140-6736(20)32466.

21. Materials Genome Initative. https://www.mgi.gov/.

22. de Pablo, J.J., Jackson, N.E., Webb, M.A., Chen, L.-Q., Moore, J.E., Morgan, D., Jacobs, R., Pollock, T., Schlom, D.G., and Toberer, E.S. (2019). New frontiers for the materials genome initiative. npj Comput. Mater. *5*, 1–23. https://doi.org/10.1038/s41524-019-0173-4.

23. Wagner, J., Berger, C.G., Du, X., Stubhan, T., Hauch, J.A., and Brabec, C.J. (2021). The evolution of Materials Acceleration Platforms: toward the laboratory of the future with AMANDA. J. Mater. Sci. *56*, 16422–16446. https://doi.org/10.1007/s10853-021-06281-7.

24. Yu, Y.B., Briggs, K.T., Taraban, M.B., Brinson, R.G., and Marino, J.P. (2021). Grand challenges in pharmaceutical research series: ridding the cold chain for biologics. Pharmaceut. Res. *38*, 3–7. https://doi.org/10.1007/s11095-021-03008-w.

25. Design of Stabilized Protein-Polymer Hybrids by Combinatorial Experimentation, Molecular Modeling, and Machine Learning. https://dmref.org/projects/1571.

26. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

27. Webb, M., Patel, R., Gormley, A., Tamasi, M., Borca, C., and Kosuri, S. (2022). Data on Enzyme Activity Retention in Glucose Oxidase, Lipase, and Horseradish Peroxidase. https://doi.org/10.34770/h938-nn26.