REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
01-12-2022	Final Report	5-May-2018 - 30-Apr-2021

4. TITLE AND SUBTITLE
Final Report: Design of protein biomaterials through tailored shape and packing strategies of patchy particles

5a. CONTRACT NUMBER
W911NF-18-1-0167

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Michigan - Ann Arbor
3003 South State Street

Ann Arbor, MI 48109 -1274

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
72954-BB.7

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Sharon Glotzer
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 734-615-6296

RPPR Final Report
as of 01-Dec-2022

Agency Code: 21XD

Proposal Number: 72954BB **Agreement Number: W911NF-18-1-0167**
INVESTIGATOR(S):

Name: Sharon Glotzer
Email: sglotzerkjc@umich.edu
Phone Number: 7346156296
Principal: Y

Organization: **University of Michigan - Ann Arbor**
Address: 3003 South State Street, Ann Arbor, MI 481091274
Country: USA
DUNS Number: 073133571 EIN: 386006309
Report Date: 31-Jul-2021 Date Received: 01-Dec-2022
Final Report for Period Beginning 05-May-2018 and Ending 30-Apr-2021
Title: Design of protein biomaterials through tailored shape and packing strategies of patchy particles
Begin Performance Period: 05-May-2018 **End Performance Period:** 30-Apr-2021
Report Term: 0-Other
Submitted By: Sharon Glotzer Email: sglotzerkjc@umich.edu
Phone: (734) 615-6296
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 1 **STEM Participants:** 5

Major Goals: The ability to control the assembly of materials is paramount to the design and production of smart materials with novel properties. However, programmed assembly has proven difficult: nucleic acids can be programmed to assemble, but in general do not form structures extensive enough to be useful for many applications. Conversely, polymers that can form extensive structures, such as plastics, have interactions that are relatively non-directed and are not sufficiently information rich to program. Our research aims to ascertain whether proteins that can be produced in bulk can be generally programmed to self-assemble into higher order structures based on relatively non-directed electrostatic interactions, effectively encompassing the information rich nature of nucleic acids and the relatively non-directed and extensive assembly capabilities of plastics and other materials.

Under a previous grant from ARO, we developed a computational model of a supercharged green fluorescent protein (a protein with added charges on specific sites) capable of reproducing and predicting experimental findings by our collaborator at UT Austin. Our computational model was based on treating interacting proteins as “patchy particles.” Prior to that grant, the state-of-the-art in simulating proteins as patchy particles was based on models developed by others in which proteins are treated as spheres. Because we know how important shape is to the assembly of nanoparticles and colloids, we aimed to develop the first patchy protein model where protein shape is explicitly included. We successfully parameterized our patchy shape model to account for anisotropic molecular shape and attractive interactions corresponding to the crystallographic interfaces of these proteins. The model was successful in rationalizing the experimentally observed formation of protomers (small assemblies of proteins) and a joint paper was published in Nature Chemistry and was featured as the front cover.

Building on those prior results, the overarching goal of the current grant was to develop generalizable design principles for protein assembly and determine how to program simple proteins to self-assemble into higher order structures different from those they might ordinarily form in the absence of programming. Charge complementarity through supercharging is the main route to programmability we proposed to explore because excess charges can be placed in designated places along the protein to promote associations that might differ from those of the original (un-supercharged) protein. However, it is not the only route, as other methods of engineering non-native inter-protein associations are possible today through the judicious placement of interaction “patches” relying on other forces, such as van der Waals attraction, along the proteins.

Our prior work taught us that developing a generalizable approach to engineering proteins to self-assemble into arbitrary target structures required a deep-dive exploration of the various forces contributing to protein association, and especially understand the role of protein shape and shape complementarity. The grant identified the following major goals:

RPPR Final Report
as of 01-Dec-2022

1. Extending hierarchical assembly

2. Generalization to other proteins

3. Hierarchical assembly of protein complexes for materials production

Accomplishments: Please see attachment

Training Opportunities: Graduate students were trained in state-of-the-art methods in simulation-based engineering and science. Graduate students and postdocs were trained in proposal development and submission, and contributed to multiple successful OLCF/INCITE high-performance computing allocation proposals. Participants were also mentored in communicating research results to both expert and broad audiences and also gave oral presentations at professional society meetings.

Mary Silvio, a chemical engineering freshman in Michigan's SURE (Summer Undergraduate Research in Engineering) Program 2020, worked with graduate student Fengyi Gao to explore the contribution of shape to protein dimerization. Under Fengyi's mentorship, Mary learned to carry out Hard Particle Monte Carlo (HPMC) simulations with depletion interactions on high performance computing resources and analyze results. Mary gave a formal presentation of results at our research group meeting and presented a poster at the Michigan SURE symposium in October 2020.

RPPR Final Report
as of 01-Dec-2022

Results Dissemination: Publications:

1. Gao, F., Glaser, J. & Glotzer, S. C., "The role of complementary shape in protein dimerization," Soft Matter (2021). DOI: 10.1039/D1SM00468A.

* DOE Office of Science highlight, “Can Proteins Bind Based Only on Their Shapes?” https://www.energy.gov/science/listings/science-highlights; shared on https://www.eurekalert.org/doe/ and https://www.newswise.com/doescience/

* Featured by Oak Ridge National Laboratory in an OLCF highlight,” Shape-Based Model Sheds Light on Simplified Protein Binding,” https://www.olcf.ornl.gov/2021/08/09/shape-based-model-sheds-light-on-simplified-protein-binding/

2. Glaser, J., Schwendeman, P. S., Anderson, J. A. & Glotzer, S. C., "Unified memory in HOOMD-blue improves node-level strong scaling," Comput. Mater. Sci. 173, 109359 (2020). DOI: 10.1016/j.commatsci.2019.109359

3. Glaser, J., Zha, X., Anderson, J. A., Glotzer, S. C. & Travesset, A., "Pressure in rigid body molecular dynamics," Comput. Mater. Sci. 173, 109430 (2020). DOI:10.1016/j.commatsci.2019.109430

4. Gao, F., & Glotzer, S. C., "The role of entropy and protein shape in the hierarchical crystallization pathway of insulin," Proc. Nat. Acad. of Sci., Under Review.

Invited talks:

1. “The Complexity & Diversity of Entropic Colloidal Crystals,” Nanoscience Global Lecture Series on the 20th anniversary of Nano Letters, online, March 2021.

2. “From self assembly to colloidal robots,” NanoGe 2021 Spring Meeting Invited Speaker, Symposium title: Self-Organization at All Scales: from Nano and Micro Structures to Functional Devices, online March 2021.

3. “Engineering Colloidal Matter & The Entropic Bond,” John R. and Donna S. Hall Engineering Lecture, Vanderbilt University, Nashville, February 2020.

4. “Engineering Colloidal Matter & The Entropic Bond,” 2020 Patten Distinguished Seminar Speaker, Department of Chemical and Biological Engineering, University of Colorado Boulder, January 2020.

5. “Engineering Colloidal Matter Through Computation,” Chhabra-Landau Distinguished Lecture, Department of Physics and Astronomy, University of Georgia, Athens, GA, January 2020.

6. Jens Glaser, “Using computer simulation to study the assembly of protein crystals and complexes”, 31th annual workshop on Recent Developments in Computer Simulation Studies in Condensed Matter Physics, University of Georgia, Athens, 2019.

7. S. C. Glotzer, “Crystallization Pathways for Protein Crystals and Colloidal Assemblies,” Symposium Honoring Nicholas Peppas, ACS Spring Meeting, Orlando, April 2019.

8. S. C. Glotzer, “Sustainable Software for Computational Molecular Science,” PHYS Division Invited Talk, ACS Spring Meeting, Orlando, April 2019.

9. S. C. Glotzer, “Digital Alchemy, Machine Learning and Inverse Design for Self Assembly,”Aneesur Rahman Prize Talk, APS March Meeting, Boston, MA, 2019.

10. S. C. Glotzer, “Assembly Engineering of Colloid and Protein Crystals and Superstructures,” Dale Pearson Lecturer, Department of Chemical Engineering, UC Santa Barbara, February 2019.

Oral presentations at Professional Conferences:

1. Fengyi Gao, Jens Glaser and Sharon C. Glotzer, "The Role of Complementary Shape in Protein Dimerization," 2021 AIChE Annual Meeting, Boston, November 2021.

2. Fengyi Gao, Jens Glaser and Sharon C. Glotzer, "The Role of Complementary Shape in Protein Dimerization," American Physical Society March Meeting, Virtual, March 2021.

3. Jens Glaser and Sharon C. Glotzer, "Controlling Protein Crystallization through Wetting Near a Critical Point," AICHE Annual Meeting, Orlando, Nov 2019.

4. Jens Glaser, Peter Schwendeman, Joshua A. Anderson, Sharon C. Glotzer, "Using unified memory and NVLINK for node-level strong scaling of molecular dynamics simulations," AICHE Annual Meeting, Orlando, Nov 2019.

5. Jens Glaser and Sharon C. Glotzer, "Large-Scale Simulations of Protein Self-Assembly," American Physical Society March Meeting, 2019.

Poster presentations:

1. Jens Glaser, Fengyi Gao and Sharon C. Glotzer, “Shape Driven Phase Separation and Crystallization of Proteins,” Gordon Research Conference on Chemistry and Physics of Liquids, Holderness School, NH, 2019, August 2019.

2. Fengyi Gao, Anna J. Simon, Yi Zhou, Vyas Ramasubramani, Jens Glaser, Arti Pothukuchy, Jimmy Gollihar, Jillian C. Gerberich, Janelle C., Leggere, Barrett R. Morrow, Cheulhee Jung, David W. Taylor, Andrew D. Ellington, Sharon C. Glotzer, “Modelling Organized Assembly of Supercharged Green Fluorescent Proteins”, Life Science Symposium, Ann Arbor, MI, September, 2019

3. Mary Silvio, Fengyi Gao and Sharon C. Glotzer, "Using Shape Complementarity to Drive Protein Dimerization," Summer Undergraduate Research Experience Symposium, University of Michigan, Ann Arbor, October 2020.

Honors and Awards: Sharon C. Glotzer – Elected to National Academy of Engineering, 2019.

Sharon C. Glotzer – 2019 Fred Kavli Distinguished Lectureship in Materials Science, Materials Research Society.

Sharon C. Glotzer – 2019 Alexander M. Cruickshank Lecturer, Gordon Research Conference on Physics and Chemistry of Liquids, Holderness, NH, August 2019.

Dr. Jens Glaser accepted a position as Computational Scientist, Materials at Oak Ridge National Laboratory (Oak Ridge Leadership Computing Facility), April 2020.

Graduate student Vyas Ramasubramani awarded University of Michigan J. Robert Beyster Computational Innovation Graduate Fellowship 2019 – 2020.

Sharon C. Glotzer -- Named or Otherwise Distinguished Lectures during the reporting period

1. “The Complexity & Diversity of Entropic Colloidal Crystals,” Nanoscience Global Lecture Series on the 20th anniversary of Nano Letters, online, March 2021.

2. “From self assembly to colloidal robots,” NanoGe 2021 Spring Meeting Invited Speaker, Symposium title: Self-Organization at All Scales: from Nano and Micro Structures to Functional Devices, online March 2021.

3. “Engineering Colloidal Matter & The Entropic Bond,” John R. and Donna S. Hall Engineering Lecture, Vanderbilt University, Nashville, February 2020.

4. “Engineering Colloidal Matter & The Entropic Bond,” 2020 Patten Distinguished Seminar Speaker, Department of Chemical and Biological Engineering, University of Colorado Boulder, January 2020.

5. “Engineering Colloidal Matter Through Computation,” Chhabra-Landau Distinguished Lecture, Department of Physics and Astronomy, University of Georgia, Athens, GA, January 2020.

6. “Inverse Design for Self-Assembly: Patchy Particles, Machine Learning, and the Truth about Entropy,” NSF Distinguished Lecture in Mathematical and Physical Sciences, Arlington, VA, December 2019.

7. “Engineering Matter Across Scales,” 2019 Fred Kavli Distinguished Lectureship in Materials Science, Materials Research Society Fall Meeting, Boston MA, Dec 2019.

8. “In search of the photonic band gap,” MESD Plenary Lecture, AIChE Annual Meeting 2019, Orlando FL, November 2019.

9. “Data Science for Assembly Engineering,” Invited talk, AIChE Annual Meeting 2019, Orlando FL, November 2019.

10. “26th EPSCoR Conference Keynote Address: Engineering Matter on Demand: Solving Grand Challenges through Convergent Research,” 26th NSF EPSCoR National Conference on Science and Partnerships Across Disciplinary Boundaries, Columbia SC, October 2019.

11. “Engineering Entropy in Colloidal Matter,” 2019 Parr Lecturer in Chemical and Biomolecular Engineering at the University of Illinois, October 2019.

12. University of Massachusetts Amherst Endowed Chemical Engineering Alumni Lectures, Amherst, MA, September 2019.

13. Alexander M. Cruickshank Lecture, Physics and Chemistry of Liquids Gordon Research Conference, August 2019.

14. John C. and Florence W. Holtz Lecture, Department of Chemical and Biomolecular Engineering, Johns Hopkins University, April 2019.

15. Reilly Lectures, Department of Chemical Engineering, Notre Dame University, April 2019.

16. Aneesur Rahman Prize Talk, American Physical Society March Meeting, Boston, MA, 2019.

17. Kavli Symposium Lecture, American Physical Society March Meeting, Boston, MA, 2019.

18. Dale Pearson Lecturer, Department of Chemical Engineering, UC Santa Barbara, February 2019.

Protocol Activity Status:

Technology Transfer: We developed an advanced capability of HOOMD-blue, our particle dynamics simulation code, to take advantage of OLCF Summit's multi-GPU node architecture. Work on the optimizations began in 2018 and new code was incorporated into HOOMD-blue 2.5 (released Feb 2019). A further focused development effort took place with OLCF staff at the Summit Training Workshop in December 2018, leading to MPI-unified memory interoperability being enabled in HOOMD-blue 2.5.2 (released April 2019). On Summit, we use the capability whenever possible – it enables us to obtain up to 25% speed up over previous code and made possible the simulations in this project.

We optimized the depletion functionality in HOOMD-blue for GPU strong scaling and improved Monte Carlo acceptance rates. These changes were publicly released in HOOMD-blue v3.0 (released March 2022, beta release was March 2021).

PARTICIPANTS:

Participant Type: PD/PI
Participant: Sharon C Glotzer
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: Y

Participant Type: Staff Scientist (doctoral level)
Participant: Jens S Glaser
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Brandon Butler
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Fengyi Gao
Person Months Worked: 11.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Luis Y Rivera-Rivera

Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yuan Zhou
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Vyas Ramasubramani
Person Months Worked: 7.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Domagoj Fijan
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Research Experience for Undergraduates (REU) Participant
Participant: Mary Silvio
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

ARTICLES:

RPPR Final Report
as of 01-Dec-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Nature Chemistry
Publication Identifier Type: DOI Publication Identifier: 10.1038/s41557-018-0196-3
Volume: 11 Issue: 3 First Page #: 204
Date Submitted: 8/5/19 12:00AM Date Published: 1/1/19 5:00AM
Publication Location:
Article Title: Supercharging enables organized assembly of synthetic biomolecules
Authors: Anna J. Simon, Yi Zhou, Vyas Ramasubramani, Jens Glaser, Arti Pothukuchy, Jimmy Gollihar, Jillian C.
Keywords: supercharged protein assembly; protein oligomers
Abstract: Symmetrical protein oligomers are ubiquitous in biological systems and perform key structural and regulatory functions. However, there are few methods for constructing such oligomers. Here we have engineered completely synthetic, symmetrical oligomers by combining pairs of oppositely supercharged variants of a normally monomeric model protein through a strategy we term ‘supercharged protein assembly’ (SuPrA). We show that supercharged variants of green fluorescent protein can assemble into a variety of architectures including a well-defined symmetrical 16-mer structure that we solved using cryo-electron microscopy at 3.47 Å resolution. The 16-mer is composed of two stacked rings of octamers, in which the octamers contain supercharged proteins of alternating charges, and interactions within and between the rings are mediated by a variety of specific electrostatic contacts. The ready assembly of this structure suggests that combining oppositely supercharged pairs of protein variants may
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Computational Materials Science
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.commatsci.2019.109359
Volume: 173 Issue: First Page #: 109359
Date Submitted: 9/2/20 12:00AM Date Published: 2/1/20 5:00AM
Publication Location:
Article Title: Unified memory in HOOMD-blue improves node-level strong scaling
Authors: Jens Glaser, Peter S. Schwendeman, Joshua A. Anderson, Sharon C. Glotzer
Keywords: CUDA; GPUs; Molecular dynamics; NVLINK; Rigid bodies; Unified memory
Abstract: Current supercomputer designs rely on increasing the compute density inside a node to maximize the performance of applications that tightly integrate the processors within a shared memory space. HOOMD-blue 2.5 enables molecular dynamics simulations that take advantage of multiple GPUs inside the same node which are connected via NVLINK. We describe the native implementation of CUDA unified memory in HOOMD-blue for strong scaling on this hardware, and provide performance benchmarks.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Computational Materials Science
Publication Identifier Type: DOI Publication Identifier: 10.1016/j.commatsci.2019.109430
Volume: 173 Issue: First Page #: 109430
Date Submitted: 9/2/20 12:00AM Date Published: 2/1/20 5:00AM
Publication Location:
Article Title: Pressure in rigid body molecular dynamics
Authors: Jens Glaser, Xun Zha, Joshua A. Anderson, Sharon C. Glotzer, Alex Travesset
Keywords: Molecular dynamics; Nanoparticles; Pressure; Rigid bodies; SPC/E water model
Abstract: We present a detailed derivation of the expression for the pressure in MD simulations that contain rigid bodies, where two equivalent formulations have been developed. One of these formulations was used in HOOMD-blue v1.x, but implemented incorrectly. We point out the precise reason for this implementation issue, the difference with the current and correct implementation in HOOMD-blue v2.x, and lessons learned. We perform numerical validation tests using dumbbell models, a mixture of cubic and spherical particles, and the SPC/E water model.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 01-Dec-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Soft Matter
Publication Identifier Type: DOI Publication Identifier: 10.1039/D1SM00468A
Volume: Issue: First Page #:
Date Submitted: 7/28/21 12:00AM Date Published:
Publication Location:
Article Title: Role of complementary shape in protein dimerization
Authors: Fengyi Gao, Jens Glaser, Sharon C Glotzer
Keywords: self assembly, protein binding, shape complementarity
Abstract: Shape guides colloidal np to form complex assemblies, but its role in defining interfaces in biomolecular complexes is less clear. We isolated the role of shape in protein complexes by studying the reversible binding processes of 46 protein dimer pairs, and investigate when entropic effects from shape complementarity alone are sufficient to predict the native protein binding interface. We employ a generic, implicit depletion model to amplify the magnitude of the entropic forces arising from lock-and-key binding and isolate the effect of shape complementarity in protein dimerization. For 13% of the complexes studied, protein shape is sufficient to predict native complexes as equilibrium assemblies. We elucidate the results by analyzing the importance of competing binding configurations and how it affects the assembly. A machine learning classifier, with a precision of 89.14% and a recall of 77.11%, is able to identify the cases where shape alone predicts the native protein interface.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 4-Under Review
Journal: Proceedings of the National Academy of Sciences
Publication Identifier Type: Other Publication Identifier: not yet available
Volume: Issue: First Page #:
Date Submitted: 11/29/22 12:00AM Date Published: 12/31/22 3:00PM
Publication Location:
Article Title: The role of entropy and protein shape in the hierarchical crystallization pathway of insulin
Authors: Fengyi Gao, Sharon C. Glotzer
Keywords: insulin | self-assembly | entropy | crystallization | protein
Abstract: Proteins often follow complex assembly pathways en route to a crystal. The same is true for colloidal particles, which can form arbitrarily complex crystals along nonclassical pathways even without explicit attraction, due solely to entropy maximization. What role might entropy play in protein crystallization? Here we isolate the role of entropy in crystallizing the insulin protein by simulating a model devoid of energetic interactions. Insulin is stored internally as a crystal, which forms via a hierarchical assembly pathway. We show that entropy alone can reproduce the native pathway from monomer to dimer to hexamer to crystal with no information other than protein shape. Thus for insulin, shape complementarity alone can predict the native interfaces at every stage of its assembly.
Distribution Statement: 4-Distribution authorized to the Department of Defense and U.S. DoD contractors only
Acknowledged Federal Support: **Y**

DISSERTATIONS:

Publication Type: Thesis or Dissertation
Institution: University of Michigan
Date Received: 30-Nov-2022 Completion Date: 8/11/22 7:03PM
Title: Colloidal and Protein Self-Assembly with Shape-Based Models
Authors: Fengyi Gao
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 01-Dec-2022

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Sharon C Glotzer
Signature Date: 12/1/22 6:07PM

Accomplishments Under Goals

For Goals 1 & 2 (*Extending hierarchical assembly and Generalization to other proteins),* we investigated the roles of different types of protein interactions (e.g. van der Waals, electrostatics, hydrophobicity/philicity, depletion, etc.) and protein shape complementarity in driving protein assembly. The significance of shape complementarity has been reported since the earliest days of protein structure determination. However, shape complementarity had yet to be included in simulations except for idealized shape models that cannot be mapped to real proteins. We hypothesized that isolating and investigating the role of shape complementarity in protein assembly in the absence of other interactions would be an important first step in understanding protein assembly. In particular, to engineer proteins with complementary interaction sites so as to promote non-native associations needed to achieve particular higher order structures, we need to know what associations protein *shape* favors. We thus aimed to answer the questions: Is shape alone ever sufficient to assemble protein native complexes, and, if so, can we identify the proteins for which this is possible and use shape complementarity to predict the native interface?
We answered these questions in Gao, F., Glaser, J. & Glotzer, S. C., "The role of complementary shape in protein dimerization," *Soft Matter* 17, 7376–7383 (2021). DOI: 10.1039/D1SM00468A.

To understand how complementary shape contributes to protein assembly, we carried out systematic assembly simulations of 50 protein dimer pairs in the Dockground dataset (http://dockground.compbio.ku.edu/) on the Summit leadership class supercomputer at Oak Ridge National Laboratory. In our model the shape complementarity of the protein is isolated using depletion interactions to mediate the protein dimer interface through the maximization of overlap volume (Fig. 1). Depletion forces are strictly entropic and thus reflective solely of protein shape complementarity. For each protein dimer, we studied how depletant size and concentration affect the assembly behavior and compared the assembled complex with the experimentally determined native complex. We evaluated the performance of our method on each protein dimer interface in terms of the yield of native assemblies found during simulation and found that six dimer pairs reached 50% average yield at the optimal depletant parameters. Thus for 13% of the dimer complexes we studied, protein shape *alone* was sufficient to predict native complexes as equilibrium assemblies. We further corroborated the role of shape in dimerization by analyzing the potential of mean force distribution for competing binding complexes, to elucidate how different factors affect the performance of the model. We further analyzed the importance of competing binding configurations and how they affect assembly. A machine learning classifier, with a precision of 89.14% and a recall of 77.11%, was able to identify the cases where shape alone predicts the native protein interface. Despite the simplicity of our model incorporating nothing more than protein shape and excluded volume interactions to elucidate the role of entropy in dimerization, it is predictive compared to patchy protein models that use existing structural information about the target protein complex that is only known *a posteriori*.

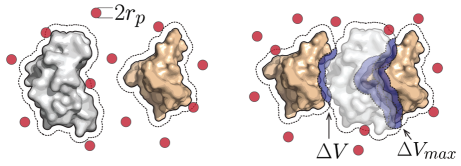

Figure 1: Our implemented model of depletion interactions, mediating shape complementarity at a protein dimer interface through the maximization of overlap volume ΔV, and hence entropy. (From the paper.)

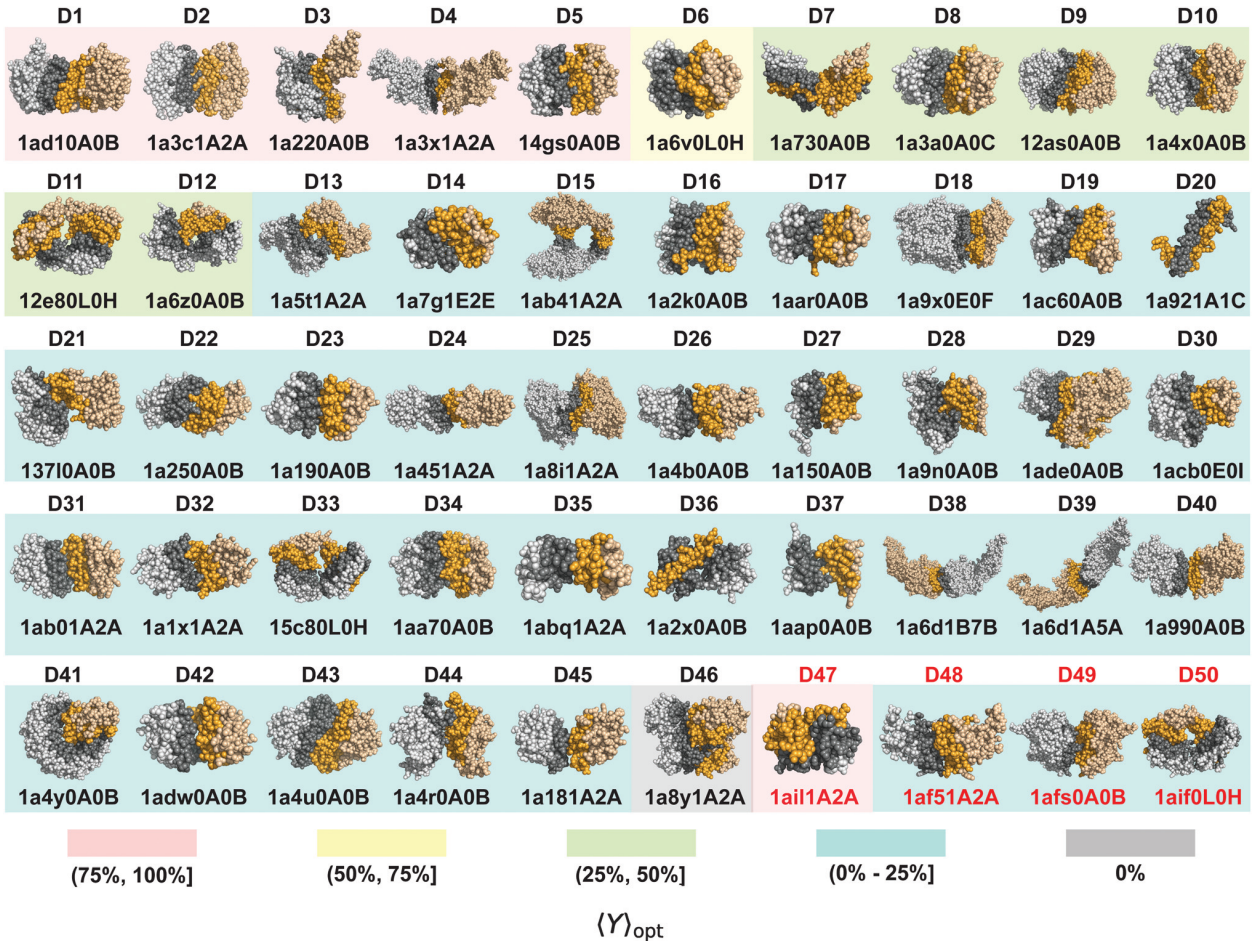

Figure 2: The 50 protein dimer complexes studied. The six dimer pairs for which shape alone could reliably predict the binding interface are indicated in pink and yellow. The last four pairs in the table (labeled in red) were excluded because inspection of the native configuration revealed that the path to assembly is topologically forbidden without reconfiguration due to their intertwined structure and a finite maximum transient yield. (From the paper.)

The implicit depletion algorithm used for the above protein dimer study was developed and published in 2015 (Glaser, J., Karas, A. S. & Glotzer, S. C., "A parallel algorithm for implicit depletant simulations," J. Chem. Phys. 143, 1–10 (2015). DOI: 10.1063/1.4935175) under MURI grant # W911NF-10-1-0518. To improve the algorithm for the protein study, we developed and implemented a novel acceptance criterion that obeys detailed balance while allowing for a tunable amount of inserted particles in the implicit depletion calculation, so as to improve sampling and reduce potential artifacts of kinetically trapped configurations. This accomplishment is described in the Soft Matter paper.

The above study also required a significant rewrite of HOOMD-blue's Hard Particle Monte Carlo (HPMC) implementation to treat the concave rigid shapes of proteins and to include implicit solvent interactions with optimized code. The Summit supercomputer, an IBM AC922 machine at Oak Ridge National Laboratory, which came online at the start of this grant, features a 'dense node' design that includes six GPUs in a single compute node. These GPUs are connected by the NVLINK 2 interconnect and support addressing of a shared memory space via the unified memory

feature of the CUDA programming model. Simulations of protein models tend to be compute and/or memory bandwidth bound, so it is important to enable fast access to the unified memory space between different GPUs. While HOOMD-blue was already able to exploit GPU architectures for speed and was already capable of performing a spatial sort of particle-based data along a Hilbert curve, Summit's then-new NVLINK technology allowed us to map the proteins onto multiple GPUs in a contiguous fashion thereby achieving strong-scaling on multiple GPUs of the same compute node. **This accomplishment was published in Glaser, J., Schwendeman, P. S., Anderson, J. A. & Glotzer, S. C., "Unified memory in HOOMD-blue improves node-level strong scaling,"** ***Comput. Mater. Sci.*** **173, 109359 (2020). DOI: 10.1016/j.commatsci.2019.109359** and **included in HOOMD-blue v2.5, which was disseminated open source.**

While undertaking these code improvements, we also revisited HOOMD-blue's pressure calculations for rigid bodies, which is important not only for proteins but also for nanoparticles or rigid molecules in solvent. We addressed methods of calculating the virial for rigid bodies so that both rigid and non-rigid particles could be accommodated within the same framework. We published a rigorous statistical mechanical derivation and describe its implementation in actual simulations, providing a reference for users of HOOMD-blue and others carrying out molecular simulations of proteins and other shapes and who require a deeper understanding of how to calculate the pressure. **This accomplishment was published in Glaser, J., Zha, X., Anderson, J. A., Glotzer, S. C. & Travesset, A., "Pressure in rigid body molecular dynamics,"** ***Comput. Mater. Sci.*** **173, 109430 (2020). DOI: 10.1016/j.commatsci.2019.109430, a special issue of the journal entitled, HOOMD-Blue: the first decade."**

To address **Major Goal 3** (*Hierarchical assembly of protein complexes for materials production*), we carried out an extensive and rigorous study of the insulin assembly pathway. We chose insulin (PDB ID: 4INS) as a model system for protein assembly due to its biological importance, its ability to form various native structural motifs, and the hierarchical nature of the motifs. As a critically important hormone, insulin regulates blood sugar levels and *in vivo* is found in four different forms. The monomeric form of insulin is biologically active, the dimeric form is responsible for endogenous delivery of insulin, and the body stores insulin both in hexameric and crystalline forms. Studies show that the form of insulin administered to patients elicits very different onset rates and response duration. Therefore, understanding and controlling insulin's assembly pathway is of fundamental importance for designing insulin formulations in diabetes treatment. We also hypothesized that the insulin assembly pathway would be an excellent exemplar for elucidating a complex hierarchical assembly processes of protein complexes that could be leveraged for materials production.

Previous computational studies have investigated the dissociation and stability of the insulin dimer and hexamer with atomistic force fields. However, due to the large computational requirements of simulations resolved with atomistic detail, predicting the complete insulin crystallization process from monomer all the way to crystal had not yet been possible. In particular, the relative importance of insulin shape in assembly compared with the various protein-protein interactions is unknown. For example, does shape complementarity favor the known native interfaces along the assembly pathway, and is it sufficient to predict the native interfaces? Our experience with colloidal crystallization, and our prior work under this grant, suggested that shape may play a determining role in the insulin assembly pathway.

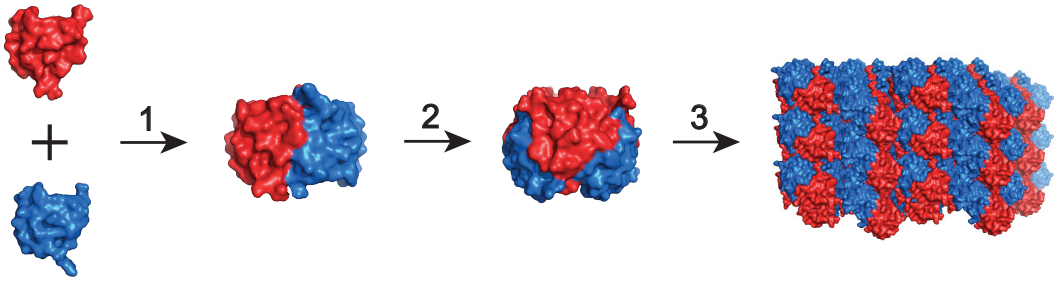

Figure 1: We investigated the hierarchical assembly pathway of the insulin crystal using a shape-based model. 1: Insulin monomers associate to form dimers; 2: Three dimers assemble into the hexamer with a spherical cap shape; 3: hexamers assemble into the insulin crystal with H3 space group.

We found that knowledge of insulin shape alone is indeed sufficient to predict the entire hierarchical assembly pathway, and that interactions beyond entropic ones are not needed anywhere along the pathway. We carried out hard particle Monte Carlo (HPMC) simulations where only the shape of insulin is explicitly modeled, and successfully reproduced all three steps of insulin crystallization: monomer to dimer, dimer to hexamer, and hexamer crystallization. In the first two steps, entropic depletion forces conspire with shape complementarity to drive dimer and, subsequently, hexamer assembly. In the final step, no depletant molecules were needed for hexamers to entropically assemble a crystal under crowded conditions. Among other things, the success of the entropic pathway suggests that the chemical interactions among insulin proteins are synergistic with entropic forces and tend to align proteins as entropy would. It also suggests an efficient way of modeling protein assembly in certain cases using the protein's shape, but foregoing expensive energetic interactions or the intrinsically floppy nature of proteins. We also demonstrated that Zn^{2+} ions further stabilize the insulin hexamer. **The results of this study will be published in Gao, F., & Glotzer, S. C., "The role of entropy and protein shape in the hierarchical crystallization pathway of insulin," Proc. Nat. Acad. of Sci., Under Review.**

The success of this rigorous study of the assembly pathway of the insulin protein demonstrates the validity and strength of the “patchy particle” model of protein association and assembly. We have design principles in mind to alter the intermediate insulin complexes through the addition of charges at strategic locations on the insulin protein. While the Covid-19 pandemic prevented us from exploring these ideas with our experimental collaborators, we hope to pursue them in the future. Additionally, a powerful code base now exists to simulate designer proteins of arbitrary shape and interaction patchiness assembling hierarchically through multiple stages of complex structures. While the pandemic thwarted our ability to obtain the experimental data necessary to develop and validate models of specific supercharged proteins and to carry out many of the original aims outlined in the proposal, we hope to continue this work in the future.

Additional accomplishments

As a complementary approach to developing shape-based models of proteins with interactions, in which the many interprotein forces are included explicitly, we explored the use of machine learning to aid the systematic development of a model. We use a deep learning model with point cloud representation to predict the protein binding interfaces. The machine learning model predicts the interface region on the interacting protein pairs in their dimer complexes and learns the

underlying physical complementarity driving molecular recognition. Building on these results, we are developing patchy protein models to reproduce the predicted binding interfaces and aid the understanding of protein association mechanisms. In 2021 we simulated and analyzed the binding distribution of 208 protein pairs with different depletant parameters, totaling to 7,488 statepoints and are comparing the simulation results to the machine learning predictions with shape-only inputs. OLCF Summit was used to benchmark, tune hyperparameters, and pre-process data for our protein interface prediction model. We pre-processed geometric data for the MaSIF dataset (5,332 protein pairs) and tested around 650 hyperparameter combinations per each iteration of our model. The best iteration of our geometry-only model has an AUC-ROC of 0.63, when trained on a highly curated subset of the dataset (225 protein pairs) and a small subset of points on the protein surface (4096 points). We aim to achieve higher AUC-ROC values by increasing the number of subsampled points from the protein surface, and training on larger, but less curated datasets, once the model performance is optimized. Unfortunately, we were unable to increase the number of subsampled points during training due to limitations in GPU memory, where the most GPU memory intensive task occurred during batch normalization with TensorFlow. We attempted to circumvent this issue by decreasing batch size, however, our model requires us to tune relevant hyperparameters to accommodate the change in batch size. We hope to continue this research with new funding.