



Perspective: Successes and Challenges in Developing a New Biomolecular Modeling and Design Summer High School Research Internship

OPEN ACCESS

Edited by:

Jorge Rodríguez-Becerra,
Universidad Metropolitana
de Ciencias de la Educación, Chile

Reviewed by:

Spencer A. Benson,
Education Innovations International,
LLC, United States
Szarecka Agnieszka,
Grand Valley State University,
United States
Anne M. Brown,
Virginia Tech, United States

*Correspondence:

Ashley Vater
awvater@ucdavis.edu

† These authors have contributed
equally to this work

‡ These authors share senior
authorship

Specialty section:

This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

Received: 31 March 2022

Accepted: 30 May 2022

Published: 01 July 2022

Citation:

Vater A, Caster R, Haddox H,
Olshefsky A, Said M, King NP and
Siegel JB (2022) Perspective:
Successes and Challenges
in Developing a New Biomolecular
Modeling and Design Summer High
School Research Internship.
Front. Educ. 7:908936.
doi: 10.3389/educ.2022.908936

Ashley Vater^{1*}, Ryan Caster^{1†}, Hugh Haddox^{2,3†}, Audrey Olshefsky^{2,4†}, Meerit Said^{2,3†}, Neil P. King^{2,3†‡} and Justin B. Siegel^{1,5,6†‡}

¹ Genome Center, University of California, Davis, Davis, CA, United States, ² Institute for Protein Design, University of Washington, Seattle, WA, United States, ³ Department of Biochemistry, University of Washington, Seattle, WA, United States, ⁴ Department of Bioengineering, University of Washington, Seattle, WA, United States, ⁵ Department of Biochemistry and Molecular Medicine, University of California, Davis, Davis, CA, United States, ⁶ Department of Chemistry, University of California, Davis, Davis, CA, United States

In response to the limited research experiences for young scholars during the COVID-19 pandemic and community interest, we developed the Pre-College Rosetta Internship Opportunity (PCR-IO). The mission of PCR-IO was to offer a program to increase equitable access to computational biomolecular research. The PCR-IO program engaged rising senior high school students in a protein therapeutic design project in which they produced novel structural models using the PyRosetta and Foldit software packages. The program comprised a year-long series of activities, with an immersive summer internship that involved students in research as the cornerstone. These activities aimed to support the overarching goal of the program by expanding participating students' social capital and technical skills, making them more likely to consider and succeed in STEM in their future endeavors. Here we describe the program's components and rollout and discuss successes and challenges in implementing a remote computational research-based educational high school program. We observed considerable student skill development and conclude that the program created real added value to student participants' education. We also uncovered issues associated with curriculum pace and found that the required mentorship effort exceeded our expectations. This perspective is intended to offer insight, share recommendations, and create dialog to increase propagation of research-based computational internships, and to shed light on how much novice students can accomplish with mentorship, structured curricula, and access to the research community.

Keywords: high school, biomolecular, protein science, computational, research-based education, Rosetta, Foldit

INTRODUCTION

In 2021, research opportunities for young scholars were scarce due to the COVID-19 pandemic and the resulting personnel capacity restrictions many labs faced. However, members of the Rosetta Commons, a consortium of biomolecular structure researchers, were well positioned to translate their expertise and mentor remote projects for novice students given the computational nature of their work (Koehler Lemman et al., 2020). The Rosetta Commons actively supports projects that improve equitable access to computational biology opportunities and recognizes that students from historically disadvantaged backgrounds are the least likely to consider and most likely to leave STEM. It was in this spirit that the Pre-College Rosetta Internship Opportunity (PCR-IO) was created. The overarching goal of the program was to welcome high school students into the protein science community who would otherwise not have been exposed to it and to engage them in a project that would advance their education and preparation for college. To reach this goal, PCR-IO aimed to: (1) lower barriers to access research experiences; (2) increase student participants' social capital and technical skills—both key components of career advancement that research experiences have the potential to build; and (3) extend the benefits of being a member of the protein science research community.

The pedagogy community recognizes that research experiences confer benefits to undergraduate students (Kuh, 2008; Brewer and Smith, 2011; Gentile et al., 2017), and particularly to those of historically marginalized backgrounds (Jones et al., 2010; Bangera and Brownell, 2014; Thompson et al., 2016). However, literature describing high school research recommends targeting the experiences to gifted and talented students (Kettler and Puryear, 2021), which may unintentionally exclude disadvantaged students. Further, challenges mentioned in the literature describing integrating computational biology into high school science curricula primarily reported issues of student motivation and limitations of teacher professional development (Gallagher et al., 2011; Freeman et al., 2014).

One appealing mechanism to introduce research experiences to students is through online, citizen science computer games. This is because of the games' clear scientific relevance, well-developed tools, and the nature of play. The citizen science computer game Foldit engages the power of human spatial intuition to solve protein structure prediction challenges, and Foldit players have outperformed purely computational methods in some situations (Cooper et al., 2010). As a result, Foldit became a central element of the PCR-IO program's research project. Similar citizen science games such as Phylo and EteRNA have enjoyed comparable scientific successes in multiple sequence alignment and RNA structure prediction (Kawrykow et al., 2012; Lee et al., 2014). Most of the work investigating online citizen science games focuses on increasing player participation and advancing scientific discovery (Shirk et al., 2012; Curtis, 2015; Follett and Strezov, 2015; Koepnick et al., 2019), but there are emerging reports of these tools' efficacy in the classroom (Horowitz et al., 2014; Dsilva et al., 2019; Miller et al., 2020). Outside of online gamified simulations, teaching tools for protein

modeling include those that are physical in nature and focused on conceptual demonstration (Bain et al., 2006) rather than advancing the field of protein science and small molecule-protein interactions (Helgren and Hagen, 2017). The intersection of game-play, novel research in protein design, and pedagogy remains relatively unexplored.

We found very limited published work describing research-based computational biology projects targeted to high school students. Although high school research-based computational biology projects as a pedagogy remains relatively unexplored, we hypothesize this educational experience has potential for considerable impact because it aligns with the theoretical framework of expanding students' social capital (Bourdieu and Passeron, 1977; Bourdieu, 1985; Thompson et al., 2016), in that increasing both *what* and *who* the students know compounds as a positive feedback loop. To summarize the relevance of this theory, readers may find it helpful to consider the example of students with limited social capital: students who are outsiders to the research community and as such are "not positioned to attain information vital for success" (National Academies of Sciences, Engineering, and Medicine, 2020). Social capital in the context of STEM mentorship can be broken down into four components: setting expectations, channeling information, internalizing norms, and integrating knowledge (National Academies of Sciences, Engineering, and Medicine, 2020). The PCR-IO program aimed to address these components to drive the initial formation of these skills, attributes, and knowledge to boost students' propensity for success in their future scientific endeavors.

Main

The PCR-IO program ran for a full calendar year in 2021 with four distinct phases: (1) Student recruitment and outreach, (2) a 3-part package application workshop series, (3) the cornerstone 6-week immersive summer internship, and (4) fall college application advising.

Prior-Year Planning

The program was planned and proposed in the last quarter of 2020 by a core team of two principal investigators, five graduate student and staff research mentors, and site program advisors, including lab safety and administrative personnel. Recruitment commenced in early 2021, through established high school program communication channels at UC Davis and the University of Washington: Young Scholars Program and Making Connections, respectively. Funding was secured from the Rosetta Commons to provide student participants with a modest stipend to encourage applications from students who might otherwise need to work a summer job.

Winter Engagement

From recruitment and outreach efforts, in February 58 students engaged in the Package Application Workshop series that coached students in navigating daunting package applications typical of immersive summer internships and college applications. The series included the following sessions: (1) Goal setting and aligning experiences, attributes, and

metrics; (2) crafting resumes/CVs and requesting letters of recommendation; (3) writing best practices for essays and personal statements.

Applicant Selection Process

Junior (3rd year) high school students were invited to apply to the intensive summer program. The application requested the students' resume, transcripts, a letter of recommendation, and short answer responses to a series of questions related to evaluation criteria. The program received 23 applications. In March, each application was reviewed by at least two different selection committee members based on a 5-point scoring rubric where scores from three categories (Impact, Preparation, and Success) were averaged, and contributed equal weight to a final score (Table 1).

Intensive Internship Design

During spring of 2021, the 6-week program cornerstone—the immersive internship—was planned in earnest. Students in our regions have a summer break that is only 8 weeks long. We expected students to be participating in the program full-time (40 h per week) while in session. We also expected that students would need time during the summer for other family commitments. This, combined with our mentors' summer availability, was the rationale for the 6-week program length. We considered priming the students with activities leading up to the internship, but we prioritized the philosophy of compensation for time-spent-learning as an equity measure, such that working students would not be disadvantaged from those who could afford the time to engage in extra homework outside of the funded summer program.

During the spring, mentors met monthly to brainstorm and review the overarching research project and the program's structure and lessons. The research project was selected because it was timely and scientifically relevant, integrated in-house expertise, and was technically tractable for high school students working on conventional laptops. The students were tasked to computationally redesign protein minibinders to more strongly adhere to SARS-CoV-2 variants of concern than the original designs created to target the wild-type strain of the virus spike. This work built off recent promising results (Cao et al., 2020) demonstrating that small, approximately 50-residue three-helix proteins create a “sticky” interface with the potential to tightly attach to the receptor binding motif of the SARS-CoV-2 spike. These miniprotein binders inhibit viral entry and thus block infection and present a potential therapeutic for COVID-19 (Cao et al., 2020; Case et al., 2021). However, the published minibinders were designed to attach to spike protein of the first described SARS-CoV-2 strain, whereas variants of concern, such as Beta, Gamma, Delta, and Omicron have point mutations in the receptor binding domain of the spike protein (Han et al., 2021) that might reduce efficacy of the binders. We hypothesized that novice students could use accessible, yet state-of-the-art computational structural biology tools to investigate binder-target interactions and propose new redesigns that accommodate mutations observed in the spike proteins of variants of concern.

To accomplish the research goal, students learned the fundamental concepts of protein folding, how to use protein structure visualization software, Python coding basics, and how to use relevant biomolecular modeling and design software packages and protocols: Foldit (Kleffner et al., 2017) and PyRosetta's FastRelax and FastDesign. Students also learned primary literature reading and searching skills as well as scientific community networking skills and results communication skills. Lessons were organized into modules designed to support the research goal of the program by sharing requisite knowledge and skills that would enable the students to investigate the biomolecular research goal. A week-long series of lessons comprised the modules (Table 2), which were each led by one or two research mentors.

Summer Internship: Program Cornerstone

For 6 weeks in June and July, the five selected applicants interned with the team of research mentor faculty, graduate students, and staff from the University of Washington's Institute for Protein Design and University of California Davis. All meetings were held remotely over Zoom. Communication with the students occurred during the Zoom session or over email. We did not use a messaging app such as Slack or Discord because these communications channels were not recommended by university policies regarding working with minors. Mentorship was distributed amongst the team: mentors rotated teaching ongoing mini class sessions in the morning; afternoon session modules were led by a single or pair of mentors for the entirety of the week. The mini class curriculum summaries follow:

¹Sci-Comm: This class spanned and supported students' progress in the internship and culminated with the final presentation in the last module “Communicating Results.” Other mini class topics included: intro to science communication, science and Twitter, making informative figures, and slide deck creation and presentation.

²Sci-Careers: Students met with professional researchers at UC Davis and University of Washington. Discussion topics include scientific projects, personal stories explaining how they arrived at protein science as a career, and recommendations for next steps beyond high school.

³Book club: With leadership and guidance from the mentor team, students read and discussed a mentor-curated collection of protein science articles that comprehensively built students' ability to interpret the data collected in their research project.

⁴Lab Livestream: Students were given a virtual tour of the Siegel Lab and Institute for Protein Design spaces to observe wet-lab work from a candid, unscripted, real-time weekly virtual visit.

⁵Friday Updates: Students prepared and presented a slide summarizing their outcomes and experiences from the week as a debrief that provided formative feedback on successes and challenges for the mentor team. This was a team meeting to discuss cumulative progress and setbacks with the group. However, students were receiving ongoing,

TABLE 1 | Applicant evaluation rubric summary of criteria under attribute categories: Impact, preparation, and success.

Impact attributes	Preparation attributes	Success attributes
(1) Student background: working while in school, demographics historically excluded from stem (2) Has never done research (3) LOR indicates that the experience will be particularly impactful	(1) Ready for rigorous and independent work: high GPA, has upward trending GPA if low; well-developed CV with intensive experiences; (2) Demonstrates effective communication skills: responds directly to the writing prompts, able to tell a story, word choice is successful (3) Good team player: likes to work with others, had strong teamwork experiences, expresses interest in becoming a better team player	(1) Enjoys computational work and/or explicitly remarks on enjoying problem solving (2) Clear motivation: is interested in science, is really excited about the program in particular (3) Has potential to grow confidence and appears to be teachable: expresses some doubt in abilities, open about limited experiences, wants to learn (4) Growth mindset: has grit, has recovered and grown from challenges, wants to try new things, open to trying new tasks

Examples and criteria details are described in the lower row.

TABLE 2 | Week-by-week progression of PCR-IO modules with learning activities and lessons designed to support the research outcome.

Week	Module	Learning objectives	Learning activities
1	Protein science 101 with Foldit	Develop protein science conceptual knowledge by using Foldit to explore biomolecular structure through modeling and design.	Protein Folding Basics (lecture)* Foldit: how to install, navigate, play, and model (live-demo, guide and activity)* Repacking and minimizing (guide and activity)* Foldit as a research tool (lecture)*
2	Python coding is easy in Google Colab	Read and code in Python to complete simple programming tasks necessary and sufficient to use PyRosetta Fastdesign package in downstream module.	Getting started in Google Colab (lecture and guide)* Python basics "follow along" (Google Colab interactive workbook)* Plotting in Python (Lecture and Google Colab interactive workbook)* Heatmaps in Python (Lecture and Google Colab interactive workbook)*
3	Proteins are beautiful in PyMOL	Create compelling clear and visually appealing images of protein structures in PyMOL	Intro to PyMOL: features and navigation, aligning structure files, and making informative figures (lecture)* Protein Data Bank orientation with structural models of SARS-CoV-2 variants of concern spike proteins (lecture, guide, activity)*
4	Expand your power with PyRosetta	Apply foundational biophysical computational modeling and design concepts by running FastDesign protocol in PyRosetta by appropriately modifying the notebook code	PyRosetta 101: basics of poses, score function, movers and navigation (lecture)* FastDesign notebook for investigating minibinders (Google Colab interactive workbook)* AlphaFold comparative model (Google Colab interactive workbook)*
5	Minibinder redesign project: bringing it all together	Compare and contrast results from Foldit, PyRosetta, and explore combining designs suggested from Foldit and PyRosetta to generate new structures computationally predicted to improve protein-protein binding. Analyze, troubleshoot and synthesize explorative work to iteratively improve research product: minibinder redesigns	Cao et al., 2020 minibinders designed for wild-type SARS-CoV-2 (primary literature reading activity) Foldit standalone (guide and activity)* Foldit-to-PyRosetta: Fast Design workflow (guide and activity)* Results summary workbook (Google Sheets repository for intra-team data sharing)
6	Communicating results	Produce a slide deck and present results: capturing and highlighting major findings and points of interest in the results, the process of obtaining them, and solicit feedback on next steps	Slide deck assignment (lecture and template slides, assignment worktime) Presentation assignment (lecture, guided practice and final presentation)

*Materials for learning activities can be found (Pre-College Rosetta Internship Opportunity: Workflow and Resource Hub, 2022).

formative support in the morning and afternoon research and skill-building sessions.

Daily sessions consisted of recurring mini-classes in the mornings followed by mentored, unstructured research time for skill development practice. Module learning activities, like lectures, tutorials, and workshops occurred in the afternoon. The core group of five students was joined in

the afternoon meeting sessions by a larger cohort of local students and students in a collaborating campus program to form a team of 12.

Mentorship Structure

Distributed mentorship, also described as Group Mentorship (Nowell, 2022), was central to the program. Students were not assigned a sole mentor, but engaged with the team: two

graduate students and a postdoctoral fellow from the University of Washington and two staff scientists from UC Davis. Mentors took turns leading sessions, which involved teaching content and coaching students on skill building, as well as serving as an advocate, support system, and advisor. The program benefited from the “it takes a village to raise a child” approach because it allowed mentors to leverage their expertise and students were able to learn from five perspectives as opposed to a single mentor perspective. We were aware that such a complicated, immersive, and challenging project could not be mentored asynchronously. As such, at least one mentor was present during all the sessions—morning and afternoon—every weekday, for the entire 6-week duration of the program. In most of the sessions, particularly the technical sessions in the afternoons, two mentors ran the session: one as a lead and one as a support. When individual student(s) needed extra help, the supporting mentor(s) would join them in a Zoom breakout room to help them learn at their pace. This team mentoring approach amounted to over 250 student-mentor contact hours; over a full working week of time for each mentor involved. Although the mentors were able to help students with technical difficulties during the program sessions, the mentors sometimes requested additional technical support from Rosetta developers. Troubleshooting the Rosetta software required expert knowledge and technical support provided by a campus IT group would not be conducive to resolving issues.

Mentored Research Progress

Students research projects were guided by the mentor team holistically. At the end of the program, students worked in small teams on a mini-binder theme (e.g., binder starting scaffold as researched by Cao et al., 2020) and each team was given the same set of questions: to investigate how their binder could be redesigned by hand in Foldit, with the purely computational PyRosetta protocol FastDesign, and using a combination of computer-generated and human intuition methods. Mentors worked closely with a single team during this final phase of the internship. While we recognize the power of the partner-pair relationship, we landed on distributed mentorship structure because it aligned with logistical needs and maximized mentorship expertise. As the 6 weeks rolled out, this became an integral tenant of the program.

The internship capstone assignment challenged the students to present their research results to their mentors and to the Siegel Lab graduate researchers. Students received feedback on their minibinder redesigns and their research process, thus experiencing a ubiquitous cultural institution of science: the research lab group meeting. Collectively, students presented tractable design concepts, some of which did indeed show predicted improved binding between the miniproteins and the viral variant spike.

Fall College Application Coaching

Through the fall in a series of monthly meetings, the mentor team continued to meet and advise the group summer interns

through their college application process by (a) honing students’ school selection process, (b) reviewing application materials, and (c) providing a supportive community. All PCR-IO students successfully submitted college applications.

In the process of building this program, the mentor team created a suite of resources for high school and novice college students that introduces computational protein modeling and design through active learning and the lens of an authentic research project. The materials are readily adaptable such that the research goal could seamlessly pivot to other timely projects or other labs’ investigations (Pre-College Rosetta Internship Opportunity: Workflow and Resource Hub, 2022).

Discussion

Throughout the summer 2021 internship, the mentees appeared to grow in their excitement and curiosity about protein design. This author team is optimistic that the mentees will remain active members of the science research community as they continue their studies in college. When the program was conceptualized, these were the two overarching student-learning goals of the program. This program did not conduct an institutional review board approved study of student outcomes due to the focus on first-iteration implementation. As such, any student related outcomes that we might report on are speculative. However, in aligning our mentorship activities with the social capital theoretical framework, it is plausible that benefit was conferred to the student participants in this program. We can report that the students were independently able to successfully (1) produce novel redesigned miniproteins, (2) propose a well-supported hypothesis for the efficacy of those designs using the modeling data they collected, (3) demonstrate primary literature reading and comprehension, and (4) create compelling protein structure visualizations. These skill-based outcomes were supported by their engagement with the research project and with the curriculum, ranging from the winter application workshop to attending the mini classes and module lessons. The students demonstrated their success in learning these skills in their final presentation, which we propose is an informal yet appropriate assessment measure in this initial implementation of the program.

We argue that the skills students learned were not developed in isolation from research cultural learnings, which we expect they also gained and which have shown to contribute to expanded social capital in other studies (Thompson et al., 2016). Through this structured mentorship program, expectations for students were established through clearly communicated milestones and by integrating common research culture practices like participating in group meetings. Resource and opportunity information was channeled to students via written and verbal communications within the research groups. Students were immersed in cultural norms of the research community and were able to adapt their behavior and language as demonstrated in their presentations. It was clear the students integrated their new knowledge because they successfully produced a viable computationally designed protein structure as a research output of the program.

While the teaching and learning materials that comprise the modules and this workflow were built around the SARS-CoV-2 minibinder redesign project, they can readily be adapted to explore other protein systems and mutational effects that might optimize or disrupt the system. These are available online as early version resources (Pre-College Rosetta Internship Opportunity: Workflow and Resource Hub, 2022), which this author team expects to hone and develop in the years to come through iterative teaching cycles in other educational contexts.

However, in addition to the successes of the program, we uncovered noteworthy challenges at every phase of the program that we describe here to inform future efforts in this space. We struggled to recruit students who would not be competitive college applicants or college-ready without the internship. In other words, our student participants were all college-bound and already deeply interested in science. It is unclear and perhaps unlikely that the program had the intended impact central to its mission in reaching students most at-risk of failing to access science. To improve our selection process, we feel that more data is needed, in terms of observing students in the program, such that we can identify elements of the application that best inform both (a) the impact the program would have on a student and (b) a student's propensity for success. We hypothesize that this will be possible to elucidate, but with only five students we were unable to observe such trends. The applications were extensive, including short essays, transcripts, a demographic questionnaire, and letters of recommendation; our hope would be to refine the application to make it less arduous to complete while retaining the most valuable information. The reason that we only accepted five students was because of limited funding for student stipends; most of the students applying appeared to be highly qualified.

While revising our applicant recruitment and selection processes may address this issue, we also became concerned that the accelerated content delivery and challenging research project may not be amenable to students less prepared than this group. We indeed found the curriculum pace and the Zoom format created barriers for students to engage and learn. We believe this would be exacerbated in situations where students have more limited resources. Some modules such as primary literature comprehension were unanimously reported by the students to be particularly challenging. We recommend that the data visualization with plots and heat maps, while excellent additions to a longer program, could be omitted from the curriculum and that students would still be able to complete the project. Streamlining the Foldit-to-PyRosetta workflows would also simplify the student's experience and make the content more accessible. Students in subsequent iterations will certainly benefit from better developed instructional materials; we were able to flesh out and resolve a great deal of ambiguities in the content. Finally, new programs could be piloted well prepared students; as teaching materials develop to become better organized overall, with tasks and instructions more explicit and content explained more clearly, those materials will be able to support students with more limited preparation.

Overall, and perhaps more relevantly, in many ways entering into research is analogous to entering a foreign culture, with a different language, norms, and values. Immersing students in the culture of research was a goal of the program, but we recognize that is a big ask and a huge challenge. It appeared that skills and experiences like prior coding work and foundational biology knowledge seemed to be the greatest mediators of students' integration into the project. We were pleased that the students were remarkably able to digest the interdisciplinary content seemingly regardless of prior knowledge. One consideration that might improve outcomes for less prepared students would be to incorporate training on best practices of inclusive teaching as professional development for the mentor team, who have a collective interest in teaching and education. We hypothesize this training would "grease the wheels" making it easier for students to learn and easier for mentors to teach, and which data suggests would most benefit students from marginalized backgrounds with limited preparation. However, adding extra demands to the mentor's responsibilities may be challenging. We learned that immense coordination and mentorship effort was needed to run such an immersive program. Without a designated staff person and a team of funded graduate students and postdoctoral fellow mentors, this research program would not be sustainable. From these learnings, we propose concrete recommendations and changes that we would make in future iterations of the program.

Create Your Own Outreach Channels to Target the Student Populations You Intend to Reach

Each existing educational program has a unique set of goals and priorities, and their recruitment pipelines, though established, may not facilitate your program mission. Spend the extra time and effort to create communication channels that will ensure you reach your target audience.

Pay Student Interns Competitively

The modest summer stipend we were able to provide amounted to less than what they would make in a full-time, minimum wage position in Washington and California. We may have limited our applicant pool by failing to offer competitive compensation.

Less Is More; Streamline Curriculum

To make the curriculum more accessible, particularly for students with limited preparation, the curriculum for a 6-week program should include only what is essential to complete the research project.

Budget for Staff and Mentorship Support

As energizing as the mentorship and new program development is for faculty, staff, and graduate student researchers, the effort needed to implement a program is considerable. An intensive program such as PCR-IO should fund a coordinator full-time during the program internship and part-time for much of the year. Graduate student mentors should receive summer funding for participation, receive credit in fulfillment of departmental teaching requirements, or the mentorship responsibilities should be more distributed across a larger group.

Expand to More Students

Because of the structured nature of the program and mentorship, increasing the size of the mentee cohort should be very tractable. Hosting closer to twenty students, with designated staff support and more mentors, would be quite feasible and would increase the impact of the program.

In conclusion, we reiterate that the student participants likely enjoyed the benefits of participation and will continue to realize those benefits as they remain in science and interface with the scientific community. We hope that by publishing this perspective article other labs can propagate and sustain the opportunities for pre-college students across the US and beyond.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AV wrote the first draft of the manuscript. RC, HH, AO, and MS wrote sections of the manuscript. All authors contributed conception, design and implementation of the educational

program described in this perspective article and contributed to manuscript revision, read and approved the submitted version.

FUNDING

The Rosetta Commons sponsored this program. University of California, Davis Library provided support for open access publications fees.

ACKNOWLEDGMENTS

In planning the program, we graciously received advice from Claudia Ludwig (Institute for Systems Biology), Senait Habte (Making Connections), Caroline Shelton and Laura Harrington (UW Office of the Youth Protection Coordinator), and Megan McKenzie Bettis (UC Davis Young Scholars Program). We acknowledge Kandise A. VanWormer and Celine Abell for their significant program operations support. We thank the Rosetta Commons for program sponsorship. Special thanks to Longxing Cao, Lance Stewart, Ian Haydon, Peishan Huang, Pamela Denish, and Adam Broerman for meeting with the students as guest scientist speakers. We thank all members of the King and Siegel Labs for welcoming and engaging the students in group meetings. Finally, we are grateful to our PCR-IO students for helping shape this new program with their highly engaged participation.

REFERENCES

- Bain, G. A., Yi, J., Beikmohamadi, M., Herman, T. M., and Patrick, M. A. (2006). Using physical models of biomolecular structures to teach concepts of biochemical structure and structure depiction in the introductory chemistry laboratory. *J. Chem. Educ.* 83:1322.
- Bangera, G., and Brownell, S. E. (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *CBE Life Sci. Educ.* 13, 602–606. doi: 10.1187/cbe.14-06-0099
- Bourdieu, P., and Passeron, J.-C. (1977). *Reproduction in Education, Culture and Society*. Beverly Hills, CA: Sage.
- Bourdieu, P. (1985). “The forms of capital,” in *Handbook of Theory and Research for the Sociology of Education*, ed. J. Richardson Greenwood, NY: Richardson Publ.
- Brewer, C. A., and Smith, D. (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*. Washington, DC: American Association for the Advancement of Science.
- Cao, L., Goresnik, I., Coventry, B., Case, J. B., Miller, L., Kozodoy, L., et al. (2020). De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* 370, 426–431.
- Case, J. B., Chen, R. E., Cao, L., Ying, B., Winkler, E. S., Johnson, M., et al. (2021). Ultrapotent miniproteins targeting the SARS-CoV-2 receptor-binding domain protect against infection and disease. *Cell Host Microbe* 29, 1151–1161.
- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., et al. (2010). Predicting protein structures with a multiplayer online game. *Nature* 466, 756–760. doi: 10.1038/nature09304
- Curtis, V. (2015). Motivation to participate in an online citizen science game: a study of Foldit. *Sci. Commun.* 37, 723–746.
- Dsilva, L., Mittal, S., Koepnick, B., Flatten, J., Cooper, S., and Horowitz, S. (2019). Creating custom Foldit puzzles for teaching biochemistry. *Biochem. Mol. Biol. Educ.* 47, 133–139. doi: 10.1002/bmb.21208
- Follett, R., and Strezov, V. (2015). An analysis of citizen science based research: usage and publication patterns. *PLoS One* 10:e0143687. doi: 10.1371/journal.pone.0143687
- Freeman, J., Magerko, B., McKlin, T., Reilly, M., Permar, J., Summers, C., et al. (2014). “Engaging underrepresented groups in high school introductory computing through computational remixing with earsketch.” in *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, Atlanta, 85–90.
- Gallagher, S. R., Coon, W., Donley, K., Scott, A., and Goldberg, D. S. (2011). A first attempt to bring computational biology into advanced high school biology classrooms. *PLoS Comput. Biol.* 7:e1002244. doi: 10.1371/journal.pcbi.1002244
- Gentile, J., Brenner, K., and Stephens, A. (2017). *Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities*. Washington, DC: The National Academies Press. doi: 10.17226/24622
- Han, P., Su, C., Zhang, Y., Bai, C., Zheng, A., Qiao, C., et al. (2021). Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. *Nat. Commun.* 12:6103. doi: 10.1038/s41467-021-26401-w
- Helgren, T. R., and Hagen, T. J. (2017). Demonstration of autodock as an educational tool for drug discovery. *J. Chem. Educ.* 94, 345–349. doi: 10.1021/acs.jchemed.6b00555
- Horowitz, S., Koldewey, P., and Bardwell, J. C. (2014). Undergraduates improve upon published crystal structure in class assignment. *Biochem. Mol. Biol. Educ.* 42, 398–404. doi: 10.1002/bmb.20811
- Jones, M. T., Barlow, A. E. L., and Villarejo, M. (2010). Importance of undergraduate research for minority persistence and achievement in biology. *J. Higher Educ.* 81, 82–115.
- Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., et al. (2012). Phylo: a citizen science approach for improving multiple sequence alignment. *PLoS One* 7:e31362. doi: 10.1371/journal.pone.0031362
- Kettler, T., and Puryear, J. S. (2021). “Research experiences for high school students: a curriculum for expertise and authentic practice,” in *Modern Curriculum for Gifted and Advanced Academic Students*, ed. T. Kettler (London: Routledge), 189–203.
- Kleffner, R., Flatten, J., Leaver-Fay, A., Baker, D., Siegel, J. B., Khatib, F., et al. (2017). Foldit Standalone: a video game-derived protein structure

- manipulation interface using Rosetta. *Bioinformatics* 33, 2765–2767. doi: 10.1093/bioinformatics/btx283
- Koehler Leman, J., Weitzner, B. D., Renfrew, P. D., Lewis, S. M., Moretti, R., Watkins, A. M., et al. (2020). Better together: elements of successful scientific software development in a distributed collaborative community. *PLoS Comput. Biol.* 16:e1007507. doi: 10.1371/journal.pcbi.1007507
- Koepnick, B., Flatten, J., Husain, T., Ford, A., Silva, D.-A., Bick, M. J., et al. (2019). De novo protein design by citizen scientists. *Nature* 570, 390–394. doi: 10.1038/s41586-019-1274-4
- Kuh, G. D. (2008). Excerpt from high-impact educational practices: what they are, who has access to them, and why they matter. *Assoc. Am. Coll. Univ.* 143, 28–29.
- Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., et al. (2014). RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2122–2127. doi: 10.1073/pnas.1313039111
- Miller, J. A., Khatib, F., Hammond, H., Cooper, S., and Horowitz, S. (2020). Introducing foldit education mode. *Nat. Struct. Mol. Biol.* 27, 769–770.
- National Academies of Sciences, Engineering, and Medicine (2020). *The Science of Effective Mentorship in STEMM*. Washington, DC: National Academies Press (US).
- Nowell, L. (2022). Beyond tradition: innovative mentorship models for higher education. *Pap. Postsecond. Learn. Teach.* 5, 1–8.
- Pre-College Rosetta Internship Opportunity: Workflow and Resource Hub. (2022). *github.io*. Available online at: <https://ashley-vater.github.io/PCR-IO/> (accessed June 7, 2022).
- Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., et al. (2012). Public participation in scientific research: a framework for deliberate design. *Ecol. Soc.* 17:29.
- Thompson, J. J., Conaway, E., and Dolan, E. L. (2016). Undergraduate students' development of social, cultural, and human capital in a networked research experience. *Cult. Stud. Sci. Educ.* 11, 959–990.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Vater, Caster, Haddox, Olshefsky, Said, King and Siegel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.