

Undergraduate Chemistry and Biology Students' Use of Causal Mechanistic Reasoning to Explain and Predict Preferential Protein–Ligand Binding Activity

Clare G.-C. Franovic,* Keenan Noyes, Jon R. Stoltzfus, Christina V. Schwarz, Tammy M. Long, and Melanie M. Cooper



Cite This: *J. Chem. Educ.* 2023, 100, 1716–1727



Read Online

ACCESS |



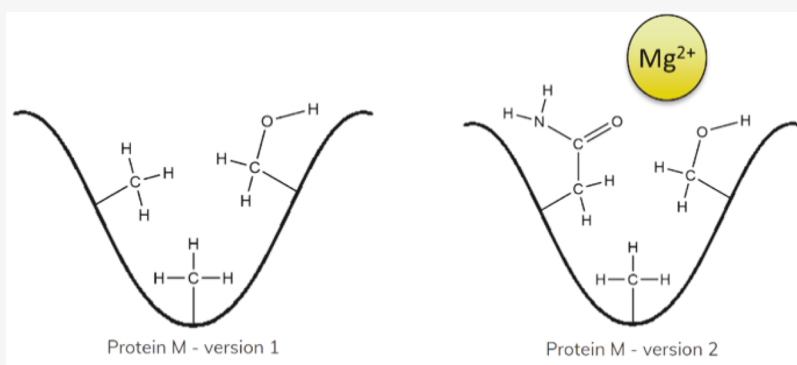
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: There is emerging recognition that mechanistic reasoning about scientific phenomena as well as developing an interdisciplinary understanding of science are valuable for student success. Our team of chemistry, biology, and teacher education researchers strives to support students in their interdisciplinary learning and engage them in causal mechanistic reasoning (CMR), a thinking strategy that involves identifying underlying entities and their properties and linking these ideas to explain or predict a target phenomenon. We aim to help students build connections between their undergraduate chemistry and biology courses through engagement with carefully developed activities, which can be administered in a range of courses. In this study, we explored how students in 7 different chemistry, biology, or biochemistry courses engaged in CMR about preferential protein–ligand binding, a biological phenomenon that requires an understanding of core chemistry ideas like atomic properties and electrostatics. Explanations varied from noncausal mechanistic to fully causal mechanistic across all courses; however, students co-enrolled in molecular biology and second semester general chemistry did not differ in their engagement in CMR when situated in each course. Student responses provided evidence of activation of different conceptual resources, including charge, polarity, and space, to explain this phenomenon. Lastly, correct predictions about the preferred binding site were highly correlated with fully engaged CMR, showing compelling evidence for the power of this thinking strategy. This task effectively elicits CMR in both chemistry and biology courses, thus making it a potential tool for helping students build connections across disciplinary silos they experience in undergraduate science curricula.

KEYWORDS: Chemical Education Research, First-Year Undergraduate/General, Second-Year Undergraduate, Interdisciplinary/Multidisciplinary, Transfer, Testing/Assessment, Molecular Properties/Structure, Noncovalent Interactions

■ BACKGROUND AND SIGNIFICANCE

Connecting Across Disciplines

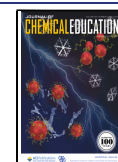
Having both an interdisciplinary understanding of science and a mechanistic understanding of scientific phenomena are valued ways of thinking in science education.^{1–5} In an increasingly complex world with problems and phenomena not bounded by disciplinary silos, it is important to think deeply about how and why they occur to make informed decisions and to recognize when and how to use and apply prior knowledge.

The importance of interdisciplinary learning at the undergraduate level has increased in science education discourse. These efforts include large projects (for example, the BRAID project⁶), as well as smaller, single-activity research articles,

Received: August 4, 2022

Revised: April 9, 2023

Published: April 24, 2023



such as those proposed by Underwood et al.^{7–10} While interdisciplinarity may focus on the integration of science with social and economic studies (for example, addressing climate change or sustainability issues), students should also be able to integrate ideas between their science courses, such as chemistry and biology, particularly prehealth students who need to leverage both disciplines in their future courses and careers. Students in undergraduate STEM majors frequently enroll in both chemistry and biology courses to meet their degree requirements. In fact, general chemistry often serves as a pre- or corequisite for introductory biology courses, implying interdisciplinary connections; however, these connections are not often obvious to students,^{11–13} making this an intriguing area of study.

The National Academies, in their 2012 consensus report, further emphasized the importance of integrating ideas across science courses, and to make this actionable, introduced three-dimensional learning (3DL). 3DL aims to engage science students in three dimensions during their learning experience: disciplinary core ideas, scientific practices, and cross-cutting concepts.³ Causal mechanistic reasoning (CMR) about phenomena can be used as a tool for instructors who are aiming to implement 3DL in coursework, because this reasoning strategy incorporates all three dimensions: a disciplinary core idea relevant to the phenomenon under consideration, the scientific practice of *constructing explanations*, and the crosscutting concept of *cause and effect*. While both chemistry and biology education research have implemented these ideas in their respective fields, there is limited work probing student engagement in CMR using chemical principles in biological contexts.¹⁴

Causal Mechanistic Reasoning

Mechanistic reasoning has been defined as a type of causal reasoning that involves explaining the sequential stages of the underlying causal events leading to a phenomenon, or how and why one or more factors behave to give rise to a phenomenon.^{15,16} Krist et al. describe this as an epistemic resource, or thinking strategy, consisting of three steps: (1) considering the scalar level below the phenomenon of interest, (2) identifying and unpacking the properties and behaviors of entities at that lower scalar level, and (3) connecting how those interactions and behaviors give rise to the phenomenon of interest.¹⁶ This understanding of CMR guided our work in both the activity development process (i.e., the type of thinking we aimed to elicit) and the assessment process (i.e., our interpretation of student explanations and predictions).

When considering a phenomenon, we call upon conceptual resources, or pieces of knowledge/information, such as the idea that opposite charges attract each other, in order to identify relevant concepts for specific situations.¹⁷ In addition to conceptual resources, we can leverage epistemic resources, such as CMR, which we use as a productive way of thinking about, explaining, or predicting phenomena in science and in our everyday lives.¹⁶ Our world is made up of mechanisms (in science as well as social and economic studies among others) and models that explain mechanisms, supporting this type of reasoning as a crucial part of education. Therefore, as educators, we must provide opportunities for students to use prior knowledge and promote their ability to construct causal mechanistic explanations, something that is critical in science and beyond.

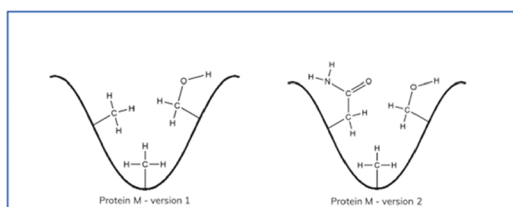
Several research findings from studies in general and organic chemistry courses emphasize the importance and utility of CMR.^{18–24} CMR can help focus attention to the important parts of the phenomena so that learners do not try to memorize every detail. This can help people learn “better” rather than “more” and understand phenomena more deeply. Such approaches can support all students and, in fact, a recent report provides evidence that mechanistic reasoning tasks are more equitable than tasks assessing rote knowledge or skills.²⁵ We posit that CMR is also useful for spanning the disciplinary boundary between chemistry and biology, because the two disciplines may emphasize the same phenomenon at different scalar levels, and the information at each scalar level is important for a deep understanding. In this work, we focus on the (macro)molecular level of proteins and define atoms or amino acids as the scalar level below that of proteins, since it is the behavior and interactions of relevant atoms and side chains which govern protein activity, such as the binding of ligands. Thus, it is the “chemical” ideas (electrostatic forces and interactions occurring at the atomic/electronic level) dictating how and why ligands bind to proteins at certain binding sites.

Structure, Properties, and Function

Structure, property, and function (SPF) relationships, in a chemical context, refer primarily to the structures of atoms or molecules, which govern their chemical properties, such as melting or boiling point. In biology, these ideas can be extended to macromolecules, such as proteins, which carry out specific functions as a result of their chemical properties, which as noted, are dictated by their structures. While the connections regarding these ideas between chemistry and biology may be accepted and understood by experts, these connections may not always be so explicit for students.^{11–13} For example, Kohn et al. interviewed students who were co-enrolled in second semester general chemistry and introductory molecular biology to investigate how they thought about the “big ideas” in each course. When discussing chemistry, these students identified the relationship between the molecular structure of a substance and its properties as central to understanding the discipline; however, these same students referred to “structure-function” as the big idea in biology. Thus, they saw these ideas as being discussed differently in the two courses. Further, these same students talked about how they actively separate ideas related to “energy” in biology versus chemistry.¹¹ In this study, one student said, “I know for biology what [the instructor] wants us to say and then for chemistry what we have to say”, suggesting that students may be constructing different responses for the same phenomena based on the course in which they are situated. While the students noted a disconnect between SPF relationships in these two disciplines, unlike with energy, they eventually recognized that these SPF ideas were quite similar and, even more encouraging, a number of students spontaneously voiced the idea that both chemistry and biology courses should attempt to explicitly help students make connections from structure to properties to function. That is, each course had part of the connection, but neither had the whole.

The fact that students suggested the idea that chemistry and biology courses should emphasize and clarify the connections between structure, properties, and function and that it is critical to do so for a sustained understanding and use of the content²⁶ makes this an important area of investigation. Research exploring students’ understandings of ideas about

The drawings below represent binding sites in two different versions of protein M showing only the atoms in relevant amino acid side chains. Consider a positively charged magnesium ion (Mg^{2+}). Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



Explain what causes the magnesium ion to bind to the protein making specific references to your drawing.

Explain why the protein you chose has the better magnesium binding site and how the structural differences in the site cause this difference in binding.

Figure 1. Full PL task. Students are first prompted to draw Mg^{2+} in the version that would better bind the ion. Then, they are prompted to explain what causes Mg^{2+} to bind and why the version they chose is the better binding site.²⁸

SPF relationships in the contexts of specific phenomena relevant to chemistry and biology is limited;^{10,14} however, in a recent publication, Yoho et al. found that faculty in chemistry, biology, and biochemistry all emphasized the importance of these relationships in each discipline but also the importance and challenges (via instructor opinions) of bridging the disciplines using these ideas.²⁷

PURPOSE OF THIS PAPER

Given (1) our interest in students building connections via CMR between chemistry and biology, (2) the importance of context in activating resources and how this might impact students' explanations, and (3) a gap in the literature studying student responses across these disciplines, we investigated student engagement in CMR using a task centered around protein–ligand binding, a biological structure–function phenomenon that can be explained using chemical principles associated with structure and properties. This task, however, serves as just one part of a three-part activity. Parts two and three also aim to elicit CMR, but the phenomena focus on how different proteins result in different functions (part 2) and how protein variation emerges in populations (part 3). Because of our large team of discipline-based education researchers, with varying interests involving student engagement in CMR across chemistry and biology undergraduate courses, we gathered responses from a range of chemistry, biology, and biochemistry courses. To our knowledge, a study of this magnitude and depth has not previously been done, fueling our interest to conduct an exploratory and descriptive investigation that can inform future work to support both CMR and interdisciplinary learning. In this manuscript, we focus solely on student responses to the protein–ligand binding task, which has deepened our understanding of the ideas students use to reason causal mechanistically about this phenomenon as well as the degree to which engagement in CMR influences overall predictions.

Research Questions

By collecting and analyzing responses from a wide array of courses, we could explore how students at different points in an undergraduate degree engaged in CMR about this phenomenon. In this analysis, we aimed to answer the following research questions:

1. How do students in chemistry and biology courses engage in causal mechanistic reasoning in the context of protein–ligand binding?
2. What conceptual resources do students enrolled in chemistry and biology courses use when explaining this phenomenon?
3. How does engagement in CMR relate to students' overall predictions?

METHODS

Development of the Task

Noyes et al. developed the protein–ligand task (hereon referred to as “PL task”) and the subsequent coding approach that we used to analyze the responses in this paper.²⁸ The task aimed to elicit both CMR (an epistemic resource or way of thinking) and the appropriate knowledge pieces (conceptual resources) that would be required to construct an accurate prediction and explanation, while also providing a balanced amount of information so that the task would sufficiently activate these resources without giving away the answer. We hoped to activate conceptual resources related to electrostatics in this context, so that the students could then use and advance those ideas while reasoning about this relevant, biological phenomenon. To scaffold the writing of the task, we used a modified version of evidence-centered design,^{29,30} a process outlined by Mislevy et al. in which the design of the task is based on statements that the designers deem acceptable evidence of learning.²⁹ In our case, we defined acceptable evidence as a response that appropriately leverages electrostatics and explicitly identifies lower-level entities, as well as how these entities and properties link to preferential binding. Based on the design process discussed by Noyes et al.,²⁸ this task seems to activate appropriate resources (conceptual and epistemic) for explaining this phenomenon and prompt students to include these ideas in their written explanations.

The PL Task

The PL task asks students to choose one of two protein binding sites to which a positive magnesium ion would most likely bind (Figure 1). The students are prompted to draw the Mg^{2+} in the site they chose, showing why it is binding there, then explain what causes the ion to bind to the protein, and finally, explain why the version they chose has the better

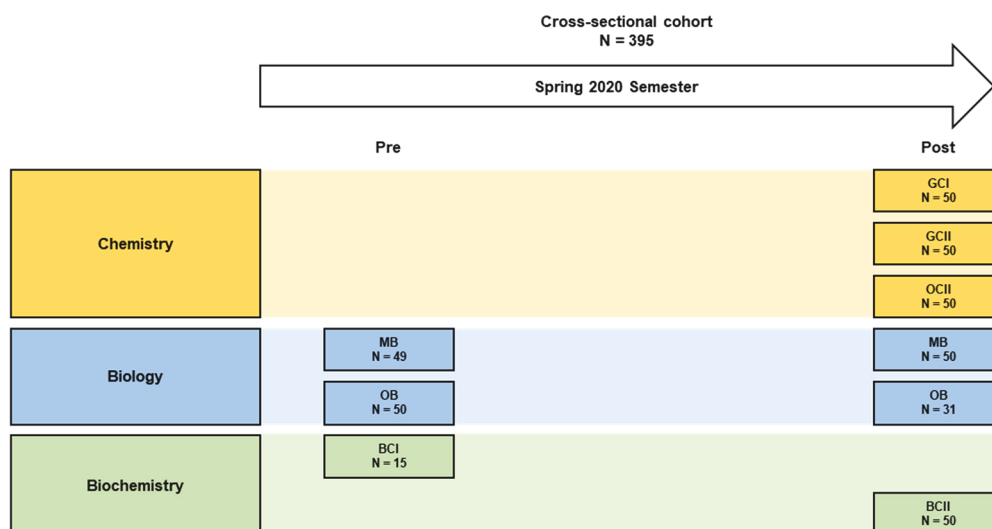


Figure 2. All students in the cross-sectional cohort took the PL task for the first time in the specified course and time point displayed.

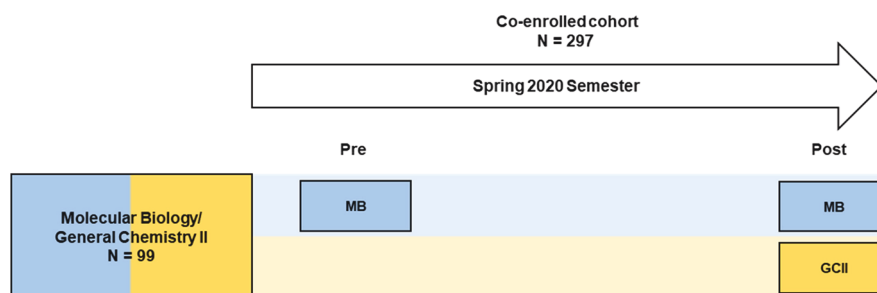


Figure 3. All students in the co-enrolled cohort completed the task in either MB or GCII at the three time points displayed. Ninety-nine students and $N = 297$ responses.

binding site and how structural differences cause the difference in binding.

Based on structure–property predictions and the information available in the diagrams, the correct protein version is version 2, because that binding site has two polar amino acids (as opposed to one in version 1) which have partial negative charges on the oxygen atoms. The two partially negative oxygen atoms attract the positive magnesium ion more strongly than the one partially negative oxygen in version 1. We recognize that this represents a simplified model of a complex system, one in which other factors like 3-dimensional protein folding, access to the binding site, potential solvation of Mg^{2+} , etc. may impact binding activity; however, here we only provided the students with structural images of two binding sites, so we did not expect them to consider anything outside of this given context.

Participants and Data Collection

During spring semester 2020, we administered the full activity containing the PL task to students in seven different courses at a large, public, research-intensive midwestern university. The courses in which we administered the task included general chemistry I (GCI), general chemistry II (GCII), molecular biology (MB), organismal biology (OB), organic chemistry II (OCII), biochemistry I (BCI), and biochemistry II (BCII). We chose to administer this activity to these courses so that we could create “co-enrolled”, “longitudinal”, and “cross-sectional” cohorts of students in order to address the broader interests of our research group. In this study, we are primarily interested in

the “co-enrolled” and “cross-sectional” cohorts of students; however, we explain the selection process of all three cohorts in [Supporting Information S1](#) as these processes influence one another. We provide additional information about the courses and data collection process in [Figures 2 and 3](#) and [Table 1](#).

Students completed the task on an online assessment platform called beSocratic, which allows students to provide both drawn and written responses.³¹ Students were provided a

Table 1. Number of Responses Collected and Analyzed in Each Course and Timepoint

course name	time in semester	number of responses collected	responses coded for cross-sectional cohort	responses coded for co-enrolled cohort
General Chemistry I	Post	1016	50	0
General Chemistry II	Post	837	50	99
Molecular Biology	Pre	584	49 ^a	99 ^a
Molecular Biology	Post	636	50	99
Organismal Biology	Pre	350	50	0
Organismal Biology	Post	368	31	0
Organic Chemistry II	Post	225	50	0
Biochemistry I	Pre	20	15	0
Biochemistry II	Post	56	50	0
Total		4092	395	297

^aThe 49 MB pre responses in the cross-sectional cohort were also included in the co-enrolled cohort.

Table 2. Expanded Analytic Rubric^a

bin	options	criteria
Attraction of oppositely charged species	Polarity, charge, both, no	Provides evidence of the attraction between oppositely charged species or polar species and ions
Negative or polar nature of atoms and amino acids in the binding sites	Polarity, charge, both, no	Identifies an atom or amino acid as polar or negative
Magnitude of charge/polarity causing the preferential binding of Mg ²⁺	Polarity, charge, both, no	Identifies the site they chose as more or less negative/polar than the other
Correct binding site (version 2)	Yes, no	Clearly (through drawing and/or text) indicates that Version 2 is the better Mg ²⁺ binding site.
Space or some physical attribute causing Mg ²⁺ to bind	Yes, no	Discusses the space or accessibility or some physical aspect of the binding site to explain Mg ²⁺ binding

^aThe original analytic rubric bins are bolded.

small amount of course credit for completing the task but could choose to exclude their responses from our study. All students in this study agreed and consented to our use of their responses according to our IRB protocol, and after removing incomplete or uncodable responses, there was a total of 4,092 student explanations. We deidentified the responses by assigning each response a random (using a random number generator) identification value. Of the 4,092 total responses, we decided to begin coding with the two smaller, more feasibly sized, cohorts that represent the focus of this study: the “cross-sectional cohort” and the “co-enrolled cohort”.

The Cross-Sectional Cohort

In the cross-sectional cohort, we targeted responses from students in all courses. After removing students who did not provide a complete or codable response, we aimed to limit other confounding factors by selecting responses from students who had only seen the task once up until that point (i.e., some students were enrolled in more than one of the listed courses or completed both a pre and post response). For example, the MB and OB post responses were constructed by students who did not complete a pre response. Finally, we randomly selected a maximum of 50 responses from each course/time point (Figure 2), resulting in a total of 395 responses for the cross-sectional cohort.

The Co-enrolled Cohort

To answer our broader questions about how students in chemistry and biology courses use CMR and conceptual/epistemic resources to explain this phenomenon, we also needed to see if the course in which the task is asked affects the students' responses. We included the co-enrolled cohort (students who were co-enrolled in GCII and MB) in this study, because (1) resources are contextually activated and (2) there is some evidence that students actively separate their chemistry and biology knowledge.¹¹ This cohort consists of 99 students, and while not the direct intent of selecting this cohort, because we administered the task at the beginning of the semester in MB, we could observe how the students' responses changed over the course of the semester. Each student in this cohort completed the task three times: once at the beginning of the semester (for their MB course) and twice at the end of the semester (for both their MB and GCII courses) (Figure 3). Specifically, we chose the same 49 students who were coded for the “MB pre” point in the cross-sectional cohort as well as an additional 50 students who fit this description. Thus, this set consisted of 297 total coded responses from these 99 students. We note that, while we administered the PL task in GCII prior to MB at the end of the semester, there was some slight overlap in the timing of the

tasks and three of these students completed their post response in MB prior to GCII. Table 1 outlines the course names, the total number of responses collected, the total number of responses coded, and the time point of administration of the task.

We conducted independent samples *t* tests to evaluate the difference in grade point average (GPA) and SAT scores between the students who constructed responses that were coded and the remaining students (not coded) in each respective group. Because of the small sample size coded for each course, we did not compare other factors (e.g., age, race/ethnicity, course history, major, etc.). Each of the independent sample *t* test results were not significant (*p* > 0.05). Thus, according to GPA and SAT scores, the responses we coded were constructed by students that are representative of each course population; the *p* values for each group are shown in Supporting Information S2.

COVID-19

During this semester, all courses at the university switched to an online format in response to the COVID-19 pandemic. Thus, pre responses (in BCI, MB, and OB) were constructed prior to the pandemic, while post responses (in GCI, GCII, OCII, MB, OB, and BCII) were constructed during online learning and amid the pandemic. This unprecedented event had (and continues to have) widespread effects on students that should not be ignored when considering the results; however, it is important to continue research on teaching and assessing students so that we may understand how to provide the best support possible.

Coding Rubric

Noyes et al. described the coding approach that we used to characterize students' engagement in CMR from their constructed explanations to the PL task.²⁸ The engagement in CMR (non-CM, partially CM, or fully CM) depends on the presence or absence of three key ideas in the explanation: (1) *the attraction of oppositely charged species*, (2) *the negative or polar nature of atoms and amino acids in the binding sites*, and (3) *the larger negative charge in one of the binding sites, causing the preferential binding of Mg²⁺*. If a response provided evidence of an understanding of all three key ideas, then we coded it as fully CM. If the response provided evidence of one or two of the ideas, but not all three, then we assigned the code partially CM. Lastly, if there was no evidence for any of these three key ideas, the response was deemed non-CM. For a more comprehensive description and student examples of the coding approach, we direct readers to our previous publication detailing its development and implementation.²⁸

Initially, the first and second authors used the CMR coding approach to code the deidentified responses from the cross-sectional cohort. However, while coding this larger set (as compared to the smaller sets used for development of the coding approach),²⁸ we noticed an additional resource being leveraged by some students: *space or some physical attribute causing Mg^{2+} to bind*, instead of, or in tandem with, electrostatics. Because our goal is to characterize both student engagement in CMR (RQ1, which uses the holistic codes) as well as the different resources that students leveraged when constructing their explanations for this phenomenon (RQ2), we expanded the analytic rubric to include a coding bin for space and specified the codes for the original analytic rubric to include charge, polarity, both, or neither (to identify resource(s) rather than just “yes” or “no”) (Table 2). For the holistic scheme, whether the students used charge or polarity to explain the binding of Mg^{2+} did not impact the code, since either resource is sufficient for explaining this phenomenon; however, we did identify which of the two (or both) was used in the responses. Finally, we added one last bin to see whether the students chose the correct binding site (version 2), allowing us to address RQ3. The expanded analytic rubric was used by the first and second authors to code the co-enrolled cohort as well.

Analysis of Responses and Inter-rater Reliability

During coding, the coders could not see the course or any other identifying student information. The first and second authors individually coded a set of 40 responses from the cross-sectional cohort, reaching percent agreement between 93% and 100% for each analytic bin. Then, after coding the remaining responses in the cross-sectional cohort individually, these authors discussed and came to consensus on any disagreements, which could then be used for the final analysis. We used Cohen's kappa values³² to calculate inter-rater reliability (IRR), as the holistic codes are ordinal and mutually exclusive. Our calculated Cohen's kappa was 0.814, suggesting “almost perfect” agreement³³ between the two coders. After coding the cross-sectional cohort, the two authors coded the co-enrolled cohort in the same manner (Cohen's kappa = 0.785). The agreement between the iterations of coding for each holistic code was 98% for non-CM, 98% for partially CM, and 100% for fully CM.

We calculated quantitative statistics associated with Cohen's kappa, Pearson's χ^2 , independent t tests, ANOVA, and sign tests with IBM SPSS Statistics Version 27.³⁴ For these tests, we used a significance threshold of 0.05 and, when appropriate, conducted a Bonferroni adjustment for increased risk of type I error (i.e., false positives). For significant Pearson's χ^2 tests, we calculated Cramér's V, a modified version of ϕ for contingency tables with more than 2 rows or columns,³⁵ to characterize the effect size.

RESULTS

In the following sections, we answer our research questions by leveraging both the holistic and analytic coding rubrics. We share the results for RQ1 using the holistic codes, which characterize student engagement in CMR. For RQ2, however, we use the analytic rubric to characterize student use of conceptual resources in their explanations. Lastly, RQ3 focuses on how student engagement in CMR (holistic codes) related to their binding site predictions.

RQ1: How Do Students in Chemistry and Biology Courses Engage in CMR in the Context of Protein–Ligand Binding?

To address our first research question, we coded responses from nine different courses/time points (Table 1) characterizing engagement in CMR. We present raw results (i.e., breakdown of non-CM, partially CM, and fully CM responses) for each of these time points and courses in Supporting Information S3, which shows an encouraging, stark increase in fully CM responses from GCI to BCII. In this section, we highlight comparisons between specific groups to illustrate how chemistry and biology students engage in CMR. Specifically, we focus on (1) the responses from the chemistry students in the cross-sectional cohort and (2) the responses from the co-enrolled cohort as representative of the biology students.

Chemistry Students' Engagement in CMR. Student Sample Population. To address this question, we analyzed data from students in the cross-sectional cohort who completed the PL task in one of three chemistry courses: GCI, GCII, and OCII. In order to determine whether students from each course were roughly equivalent, we used a one-way analysis of variance (ANOVA) to evaluate the difference in mean SAT scores between the students analyzed from each course. The results showed a significant difference ($p = 0.012$), with the GCI students having a lower mean SAT when compared to the GCII and OCII students (GCI mean = 1165, GCII mean = 1236, OCII mean = 1239; for more information, see Supporting Information S4). This is not surprising, given that a wider range of students enroll in GCI. However, we also did a one-way ANOVA to evaluate the difference in mean cumulative GPA, which showed no difference between the students in these three courses ($p = 0.888$). Supporting Information S4 provides descriptive statistics for each of these performance measures.

Students in GCII and OCII Construct More Fully CM Responses than Students in GCI. Chemistry course and engagement in CMR were found to be significantly related, with GCI students constructing the least amount of fully CM responses (Pearson's $\chi^2 = 26.450$, $p < 0.001$, Cramér's V = 0.297). The percentage of students constructing fully CM responses in GCI, GCII, and OCII were 2%, 24%, and 30%, respectively (Figure 4). We did follow-up pairwise comparisons to evaluate the difference among these proportions across courses (Table 3). While the GCI students constructed significantly different explanations from both the GCII and OCII students, the GCII and OCII students did not differ in

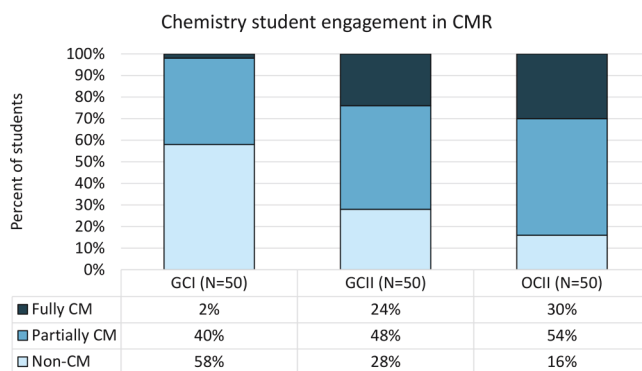


Figure 4. Chemistry student engagement in CMR (all cross-sectional cohort students who completed the PL task at the “post” time point).

Table 3. Pearson's χ^2 Test Results Exploring the Relationship between Chemistry Courses and Engagement in CMR

comparison	χ^2 value	p value ^a	Cramér's V ^b
GCI and GCII	14.904	$p < 0.001$	0.386
GCI and OCII	25.211	$p < 0.001$	0.502
GCII and OCII	2.146	$p = 0.342$	0.146

^aBonferroni-adjusted $\alpha = 0.017$ ^bSuggestions for interpreting Cramér's V:³⁶ small, 0.1; medium, 0.3; large, 0.5

their engagement in CMR ($p = 0.342$). Only one student in GCI constructed a fully CM response, while 24% and 30% of the GCII and OCII students did so. We posit that this is due to resource refinement; that is, students in GCI do not have complete access to the resources with which they need to construct a causal mechanistic explanation, while the GCII and OCII students have had both time and opportunity to (1) reason mechanistically about phenomena (i.e., develop this epistemic resource) and (2) refine their conceptual resources, so that they can be called on and used in appropriate contexts.

Biology Students' Engagement in CMR. Student Sample Population. To understand how students from biology engaged in CMR, we focused on the responses from the co-enrolled cohort. We chose to focus on this cohort, because students in the cross-sectional cohort had not seen the PL task prior; therefore, the post OB and MB students may not be representative of the rest of the group, since they had missed the pre administration of the PL task for some reason (e.g., transferred into the course late, poor engagement early in the semester, etc.).

MB Student Explanations Improved from Pre to Post. We coded responses from ninety-nine MB students who provided explanations both at the beginning and end of their MB course; however, we removed three of these students from this analysis, because they had not completed their GCII post response yet. Thus, to maintain consistency in the order in which each student completed their explanations, we only analyzed the responses constructed by these 96 students. We used the sign test for significance (a repeated-measures test) to compare the differences in engagement in CMR at these two time points (MB pre and post) for the related sample, showing 47 positive changes, 15 negative changes, and 34 ties ($p < 0.001$) (Figure 5). That is, nearly half (49%) of the students improved their response type from non-CM to partially or fully CM or from partially CM to fully CM.

These findings indicate that student engagement in CMR improved throughout the duration of their MB course; however, like the chemistry students, these students were enrolled in other courses during this time which may have also contributed to their improvement. Additionally, having seen the prompt twice already could have impacted student explanations by allowing more time to wrestle with these ideas.

Comparing Student Engagement in CMR for Chemistry and Biology. According to the resources perspective, context matters when constructing explanations or thinking about certain ideas. Additionally, Kohn et al. showed that students think about SPF relationships in different ways when prompted about their biology course versus chemistry course, and some actively separate ideas about energy for these two courses, even when they are enrolled in them at the same time.^{11,12} For these reasons, we expected that students would construct different responses "for their chemistry course"

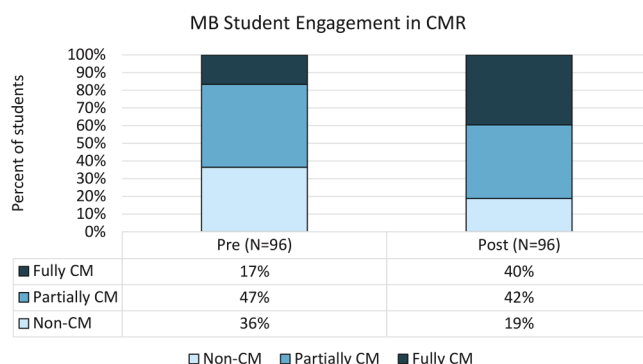


Figure 5. MB ($N = 96$) pre and post engagement in CMR. Note that it is the same cohort of students represented in the pre and post categories.

versus "for their biology course", even though both responses were completed on beSocratic. To test this, we examined the post responses from those students in the co-enrolled cohort who completed the PL task for both MB and GCII. All the students included in this analysis ($N = 96$ students, $N = 192$ responses) completed the GCII post task first, followed by completion of the MB post task within the next week.

We used a sign test to measure the difference between students' post engagement in CMR in their chemistry course and their biology course. This sign test indicated no significant difference between these groups, with 13 negative changes (e.g., partially CM response in chemistry changing to non-CM in biology) and 17 positive changes ($z = -0.548$, $p = 0.584$). After examining the responses from students who changed their engagement in CMR, we did not discern any trends in the slight differences observed. Thus, even though some students had marginally altered responses, from a statistical standpoint, the responses were the same regardless of the course in which they were constructed. Figure 6 displays this nearly identical engagement in CMR between courses.

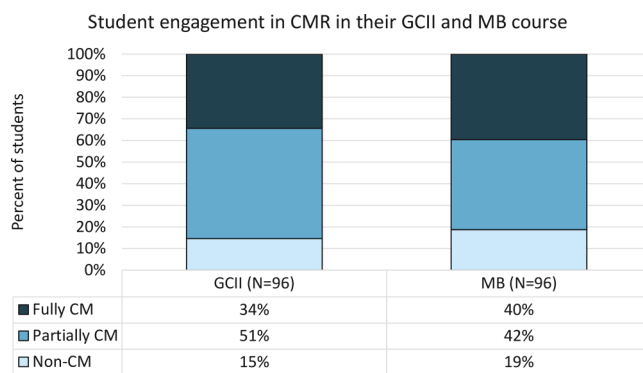


Figure 6. Frequency of response type constructed by students at the end (post) of their GCII course and end of their MB course. Recall that each student completed their response in GCII before completing it in MB.

We find this result encouraging, because it points to the interdisciplinary nature of the task, as it is equally effective at eliciting appropriate resources and CMR from students, regardless of the "course context". Further, the MB and GCII courses at this institution emphasize CMR in frequent formative and summative assessments, likely contributing to

this result. Had one of these two courses put less emphasis on CMR, we may have seen different results.

Ideally these students, regardless of context, leveraged ideas from both courses and, therefore, are building connections to refine their knowledge frameworks when explaining this phenomenon; however, the only evidence here is the nearly identical engagement in CMR in both courses. An additional study involving think-aloud interviews would provide more support for this by asking students how and why they call upon certain ideas.

RQ2: What Conceptual Resources Do Students Enrolled in Chemistry and Biology Use When Explaining This Phenomenon?

The similar engagement in CMR seen between the co-enrolled group of students suggests similar activation of electrostatic resources, such as opposite charges attracting each other. While most students in both cohorts leveraged these ideas, we also noted a handful of students discussing a related idea, polarity, and some students discussing an unrelated idea, space, which we could quantify using our analytic rubric. For example, student 1554 from OB leveraged charge (not polarity) in their response by writing, “*The partial negative charges on the carbonyl group and on the OH group will attract the positive charge on the Mg.*” In contrast, student 1764 from MB leveraged polarity instead of charge by writing, “*I chose version 2 for the better binding site because it has more polar side chains, thus making it easier for the magnesium to interact with the side chains within that protein.*” Another student (1267 from OCII) leveraged both space and charge by writing, “*The partial negative charge on the oxygen draws the positive magnesium ion. Protein 1 has less steric hindrance. Protein 1 has less steric hindrance and thus has more space for the magnesium ion to bind.*” We included this research question to investigate what resources the cross-sectional cohort students used to explain this phenomenon and whether there were differences in the resources used between disciplines.

Polarity versus Charge. We predicted that students in biology courses would more often invoke polarity when compared to their chemistry student peers, because polarity is the descriptor most used in biology to discuss the uneven distribution of electron density in a molecule, whereas in chemistry the underlying cause of the uneven distribution is discussed in much greater depth. With our expanded analytic rubric, we were able to identify whether students used charge, polarity, or both in their explanations. For this analysis, we removed the GCI ($N = 50$) students, since we saw from RQ1 that they did not have the content background needed to construct a fully CM explanation. Of the remaining 345 responses analyzed from the cross-sectional cohort, 52 of them invoked polarity in some way, with 48 of them coming from biology (MB or OB) or biochemistry (BCI or BCII) courses. Only one GCII student and three OCII students leveraged polarity. Because the number of responses per course from the cross-sectional cohort was not consistent, we grouped each discipline and calculated the percent of partially or fully CM responses that leveraged polarity (Table 4, non-CM responses removed). We conducted a Pearson's χ^2 test which showed an overall significant difference ($\chi^2 = 14.820$, $p < 0.001$, Cramér's $V = 0.236$) in resource activation between chemistry, biology, and biochemistry. Over a quarter of biology students used polarity in their responses; 23% of biochemistry students used polarity, and only 5% of chemistry students used polarity

Table 4. Number and Percent of Students Leveraging Polarity in Their Explanations

discipline	total partially CM and fully CM responses	number of responses using polarity	percent of responses using polarity (%)
Chemistry (GCII and OCII, all post, $N = 100$)	78	4	5
Biology (MB and OB, pre and post, $N = 180$)	136	36	26
Biochemistry (BCI pre and BCII post, $N = 65$)	53	12	23

(Table 4). However, even though polarity was more often invoked in biology than chemistry, a pairwise Pearson's χ^2 test showed no significant difference in engagement in CMR between these disciplines ($\chi^2 = 1.267$, $p = 0.531$, Cramér's $V = 0.067$) (see Supporting Information S3 for additional statistics comparing engagement in CMR across disciplines).

Space. While polarity could be used productively to explain the protein–ligand binding, some students in the cross-sectional cohort ($N = 48$) leveraged the idea of space to explain the preferential binding, an idea that probably emerges from the representation of the structure and that is not a productive approach to generating a causal mechanistic explanation of the phenomenon. We consider “space” to be a surface feature and, thus, a resource that does not support a causal mechanistic explanation, because a student who **only** recognized the available space would not have provided any evidence of considering a lower scalar level (the first requirement for CMR). Further, while space may be a productive approach to explaining binding activity when the ligand is exceptionally large, it is a negligible factor for the binding activity of small, spherical ions like Mg^{2+} (and even in that case, identifying space (or lack thereof) can be thought of as an explanatory black box³⁷ for the repulsive interactions that emerge when a binding site is “too hindered”).

In this set of responses, most frequently (67%) the idea of space showed up as a student noting “more room” or “less steric hindrance” in one of the protein versions (i.e., the spatial availability); however, explanations were also characterized by this bin if they discussed, for example, one version being a better “fit”. Unlike the use of polarity, we did not find a significant difference (Pearson's $\chi^2 = 4.490$, $p = 0.106$) in the proportion of students from each discipline (chemistry, biology, or biochemistry) using space to explain the protein–ligand binding.

Most (79%) of the students in this group selected version 1 (or were unclear in their selection) as the version which would better bind Mg^{2+} . For example, consider student 4000 from BCI who said, “*...I think Version 1 has a better binding site simply because there is less steric hindrance in Version 1.*” Similarly, student 1818 from OB said, “*...Protein 1 has a better binding site because there is less ‘clutter’ within the protein.*” Along with 58% of all explanations coded for space, this student also referenced electrostatics in their response by writing, “*...I show the Mg^{2+} binding to an oxygen because while there are no negative charges on these proteins, the oxygens within the proteins at least have a partially negative charge, which the Mg^{2+} is attracted to, causing it to bind.*” Thus, even though these students recognized the role of electrostatics in causing Mg^{2+} to bind, they leveraged the idea of space (instead of magnitude of

charge) to explain why Mg^{2+} binds preferentially to one site over another. A response such as this would be coded as partially CM, because they included the first two key ideas, but their invocation of space would not be enough for the “linking” code (third step in CMR), because according to Krist et al.,¹⁶ this step requires linking the underlying entities/interactions (in this case, the negative oxygens or polar amino acids) to the target phenomenon. Although the numbers are small, this is an important finding from a resource’s perspective, because while students leveraged appropriate resources (i.e., electrostatics), the format of the prompt may have also activated this alternative resource (i.e., space) for some students, indicating a need to further refine the instruction or activity. That being said, it is also important to note that we do not view the use of space as a misconception, since it can be a useful idea in other situations, such as macroscopic contexts. Rather, the use of space likely emerged as an artifact of (1) how we represented the binding sites and/or (2) the organization or framework of resources in students’ minds. For example, students who chose to leverage space when making their prediction may not have refined epistemic heuristics that support predicting phenomena of this complexity (i.e., choosing between competing resources).

RQ3: How Does Engagement in CMR Relate to Students’ Overall Predictions?

The most compelling evidence from our analysis of this data is the significant correlation between constructing a fully CM explanation and making a correct prediction. We used all responses from the cross-sectional cohort, as well as the 50 additional MB pre responses from the co-enrolled cohort ($N = 445$), to address this research question. Results indicated that nearly all (97%) of the students who constructed fully CM explanations also selected the correct protein version (version 2) (Figure 7). An overall Pearson’s χ^2 test of all three types of

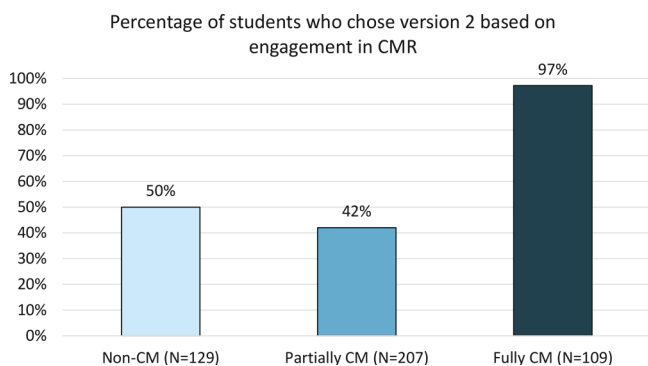


Figure 7. Relationship between engagement in CMR and the predictions that students made (i.e., which protein version they chose).

responses was significant; however, a pairwise comparison showed no significant difference between the non-CM and partially CM responses (Table 5). Thus, the fully CM group’s predictions must be driving the significance, which we see in the χ^2 values comparing fully CM predictions to both non-CM ($\chi^2 = 65.689$) and partially CM ($\chi^2 = 91.579$). With large effect sizes (Table 5), these results indicate the utility and power of engaging completely in CMR because doing so nearly always leads students to a correct prediction.

This analysis clearly indicates that CMR is a powerful tool for accurately predicting and explaining phenomena. Specifi-

Table 5. Pearson’s χ^2 Test Results Exploring the Relationship between Engagement in CMR and Binding Site Predictions

comparison	χ^2 value	p value	Cramér’s V^b
All	94.164	$p < 0.001$	0.460
Non-CM and Partially CM	1.847	$p^a = 0.174$	0.074
Non-CM and Fully CM	65.689	$p^a < 0.001$	0.525
Partially CM and Fully CM	91.579	$p^a < 0.001$	0.538

^aBonferroni-adjusted $\alpha = 0.017$ ^bSuggestions for interpreting Cramér’s V :³⁶ small, 0.1; medium, 0.3; large, 0.5

cally, for preferential protein–ligand binding, comparing the magnitude of charge/polarity between the two sites served as the critical idea for choosing the appropriate version. While many students identified the role of electrostatics in causing Mg^{2+} to bind, fewer explained why Mg^{2+} binds better to version 2; that is, these students explained how but not why this phenomenon occurs.

DISCUSSION

In this work, we examined students’ causal mechanistic reasoning (CMR) in an interdisciplinary context: preferential protein–ligand binding, a biologically relevant phenomenon that relies on understanding core chemistry ideas. While much of the work done regarding CMR is limited to students’ responses from one discipline, we have shared explanations provided by students across three chemistry, two biology, and two biochemistry courses for the same task. Research that engages students with tasks about interdisciplinary ideas is an important step toward addressing national calls for interdisciplinary education at all levels of science. Tasks such as this one are also important because they can engage students in and examine how students in different disciplines think about these ideas and areas they struggle or succeed.

While the results show varying types of student engagement in CMR across all courses, we provide evidence of nearly identical engagement in CMR between students co-enrolled in GCII and MB. These students maintained the level of engagement in CMR regardless of the discipline for which they completed the task, showing to our surprise that the course did not have a large impact on their explanations. This co-enrolled cohort may be leveraging ideas from both of their courses when constructing their explanations; however, additional work should be done to further support this, such as think-aloud interviews probing how and why students are including certain ideas in their explanations. Our results from RQ1 also show that students in more advanced chemistry courses used CMR more frequently and students in MB improved their explanations by the end of the semester, suggesting that students refined both their conceptual and epistemic resources over time.

Engagement in CMR did not differ significantly between students in chemistry (GCII and OCII) and biology (MB and OB); however, we did see notable differences in resource activation of charge or polarity based on discipline. The idea of polarity was leveraged far more frequently by students in biology than students in chemistry. This finding, while perhaps not surprising given the scale at which biology courses discuss molecular properties, highlights the importance of context in activating certain resources.¹⁷ From an instructor perspective, it is imperative that we acknowledge the ideas students leverage, even if those ideas are not the intent of the

assessment, because it may illuminate alternative appropriate explanations or, if incorrect, potential prompt revisions that better support students to use and apply relevant resources. For example, the handful of students who leveraged space might not have called upon that resource if a bulkier nonpolar amino acid was used in version 1 instead of alanine or if a more accurate representation of the structure (e.g., a space filling model) were presented. However, an important part of CMR involves choosing appropriate and explanatory resources, even amidst other (perhaps more salient) resources that, when used, provide competing explanations. In this case, invocation of space correlated with an incorrect prediction; however, that is not to say that these students have misconceptions about how/why ligands bind to proteins. Rather, the representation of the prompt activated a resource that may have been more accessible or more strongly connected to binding in their knowledge frameworks.

While the number of students constructing fully CM explanations was not particularly high across courses, the number of students engaging in some aspect of causal mechanistic reasoning (i.e., providing either a fully CM or partially CM response) is promising (75%). All these students leveraged electrostatics in some productive way, providing evidence of an understanding of at least one of the three key ideas needed for a fully CM explanation. Thus, this task and coding approach may serve as a means by which instructors can engage students in this type of deep reasoning, while highlighting key ideas that carry significance in both chemistry and biology. Additionally, using this coding approach allows instructors to identify the ideas present or absent in student responses, which can then serve as a guide for feedback and instruction.

Finally, the most significant finding in this work is the compelling evidence for the predictive power of CMR. The strong correlation between fully CM explanations and correct predictions (97%) shows the value of CMR: that is, considering relevant, lower-level entities, their behaviors and interactions, and how these key ideas link to the target phenomenon. We addressed this research question because, while constructing a fully CM response did not depend on correctness, making correct predictions using CMR shows that these students have developed coherent and sophisticated networks of knowledge (i.e., they have refined their epistemic and conceptual resources so that they can use and apply them to novel situations). For the phenomenon of preferential protein–ligand binding, most students identified the role of electrostatics in causing the Mg^{2+} to bind, but far fewer students linked the electrostatics to the preferential binding. This key idea, that a larger charge results in a stronger attractive force, proved to be the key factor in choosing the correct site, but because of the small number of students recognizing this idea, there should be additional work to support students in linking underlying factors to make predictions and explain phenomena.

It would be reasonable to predict that the partially CM explanations would more often choose the correct protein version when compared to the non-CM explanations; however, we saw that this did not happen. While there are six combinations of the three key ideas which could result in a partially CM code, we most often saw that these responses consisted of either the first idea alone, identifying the attraction of oppositely charged species ($N = 44$, 21.3%), or the first and second ideas, identifying this attraction and a lower-level

negative entity ($N = 152$, 73.4%). All other responses ($N = 11$, 5.3%) included the second idea alone, the second and third ideas, the first and third ideas, or the third idea alone. Thus, the partially CM responses rarely included the third idea: comparing the magnitude of charge between sites. That is, even though these students recognized the importance of electrostatics in causing Mg^{2+} to bind, they did not link these ideas to the phenomenon of preferential binding. In fact, a handful of students ($N = 42$) who constructed partially CM explanations used the idea of space to explain the preferential binding; the vast majority (90%) of these students explained that Mg^{2+} would more easily bind to the incorrect binding site (version 1) because there was “more room” or “less steric hindrance”.

The relationship between fully CM responses and correct predictions points to the importance of linking underlying factors to the phenomenon under consideration when constructing a causal mechanistic explanation. This third step in CMR about our phenomenon aligns with the third key idea which involves comparing the magnitude of the charge between sites to explain why Mg^{2+} binds preferentially to version 2: students who linked ideas about electrostatics to the phenomenon in this way nearly always (97% of the time) chose the correct binding site.

■ LIMITATIONS

There are a few notable limitations that should be addressed. The number of responses coded in the cross-sectional cohort represents a small number of students in the courses; for example, there were more than a thousand GCI responses, but we only coded 50. It is our intention to code additional responses in the future so that we may have a more complete understanding of each course as a whole; however, the smaller selection that we discussed in this paper provides a starting point in identifying patterns and themes among the student responses. Also, due to these small N values, we were not able to group students by previous coursework. For example, some of the OCII students did not complete GCII and took different GCI courses (e.g., residential college vs large enrollment). Additionally, each student was co-enrolled in several other courses and in different academic years, which undoubtedly impacted their prior knowledge. Typical higher education studies such as these focus on a phenomenon specific to the discipline in which it is administered, but in this case, the phenomenon is general and unfamiliar to students in most of the courses shown here, with perhaps the exception of the biochemistry courses, which discuss enzyme binding and active sites in more detail.

In addition to the sample selection and distribution, we note that the chemistry and biology courses in which this task was given were transformed. The GCI, GCII, and OCII courses use transformed curricula (CLUE and OCLUE),^{2,38} which emphasize CMR in frequent homework activities, recitation activities, and other assessments. Thus, these students may be better equipped than students in more traditional chemistry courses to explain this phenomenon. Likewise, the MB and OB courses at this institution encourage and provide opportunities for students to explain how and why phenomena occur. It is likely that students at a different institution would show different engagement in CMR.

Lastly, upon further analysis of student responses to parts 2 and 3 of this task, our colleagues found a small number ($N = 18$) of responses that were identical with one or more other

student responses (not all 18 were identical). Six of these responses were constructed by students in the co-enrolled cohort, meaning three students likely copied and pasted their response. Ten of these students seem to have worked with a classmate (or two) to construct identical responses, and the remaining two students had unique responses for the PL task, but their responses to parts 2 and/or 3 were identical with one or more other student responses. We did not reanalyze all of our data after discovering this, because the numbers were small in each course. With the activity being completed online, it is not surprising that a few students worked together or copied an explanation. Thus, while having negligible effects on our results, it is important to note that [such unobvious examples of] copied and pasted responses will likely play a role for future studies of open-ended responses in large-enrollment courses.

CONCLUSION AND FUTURE DIRECTIONS

Causal mechanistic reasoning is applicable across a range of phenomena and important for deeper and longer-lasting learning in science. This study emphasized CMR across the disciplinary boundary of chemistry and biology. Our task (addressing the phenomenon of preferential protein–ligand binding) requires an understanding of three key ideas: (1) the attraction of oppositely charged species, (2) the negative or polar nature of specific atoms or amino acids in a binding site, and (3) a larger charge (or more polar species) resulting in a stronger attractive force. These ideas are, ideally, learned early in an undergraduate science curriculum and used in future contexts such as protein–ligand binding in biology. However, causal mechanisms in earlier courses like chemistry might not be tied together explicitly within and across courses. However, each idea, when connected appropriately, can be used to construct a causal mechanistic explanation and productive prediction for the phenomenon.

The results of analysis from students' explanations indicate that most explanations were partially mechanistic, meaning that these students see the relevance of electrostatics in the phenomenon of protein–ligand binding but need additional guidance in explaining the *preferential* binding. Thus, we plan to develop feedback statements that might better support these students in developing and refining their resources involving electrostatics and structure–property relationships. However, providing feedback in a timely manner to a large population of students poses a hurdle we are working to overcome; we may investigate whether peer feedback or automated feedback could be used to navigate this hurdle. Additionally, the use of space, as a resource that competes with magnitude of charge to explain preferential binding, resulted in several rich discussions among our group. A revised activity that better supports students in choosing appropriate resources (without making it obvious that space is irrelevant) would be an interesting study. For example, adding a third binding site consisting of three alanine amino acids (and therefore the most “space” available) might provide additional insight into student use of this idea as it compares to their use of electrostatics. We encourage others in the field to further this investigation using related ideas or phenomena.

Our ultimate goal is to support students' interdisciplinary learning and mechanistic reasoning by engaging them in tasks that elicit CMR in the context of phenomena that span disciplinary boundaries. CMR, as evidenced by our results, is a useful way of thinking that spans scalar levels and results in powerful predictions and explanations. The work published

here is intended for both chemistry and biology instructors at the undergraduate level. We hope that these findings encourage our audience to engage students in CMR, because it will support them in their future lives as scientists and as citizens in an increasingly complex world.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at <https://pubs.acs.org/doi/10.1021/acs.jchemed.2c00737>.

Descriptions and details of each course in which the task was administered, independent *t* tests for the sample populations used for coding, student engagement in causal mechanistic reasoning across courses and disciplines, and GPA and SAT descriptive statistics for students in the three chemistry courses (PDF, DOCX)

AUTHOR INFORMATION

Corresponding Author

Clare G.-C. Franovic – Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States; orcid.org/0000-0003-1778-8407; Email: carls500@msu.edu

Authors

Keenan Noyes – Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States; orcid.org/0000-0002-8587-1694

Jon R. Stoltzfus – Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States

Christina V. Schwarz – College of Education, Michigan State University, East Lansing, Michigan 48824, United States

Tammy M. Long – Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, United States

Melanie M. Cooper – Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States; orcid.org/0000-0002-7050-8649

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jchemed.2c00737>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to acknowledge Juli Uhl and all the other current and past members of our research group for their thoughtful contributions to our group discussions. We also thank the instructors for allowing us to administer our task and the students for their participation. This work would not have been possible without them. This work was made possible through funding from the National Science Foundation under DUE 1725521.

REFERENCES

- (1) American Association for the Advancement of Science. *Vision and Change in Undergraduate Biology Education: A Call to Action*; Washington D.C., 2011; <https://live-visionandchange.pantheononline.io/wp-content/uploads/2011/03/Revised-Vision-and-Change-Final-Report.pdf> (accessed 2020-04-17).

- (2) Cooper, M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116–1122.
- (3) National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, DC, 2012.
- (4) Ledford, H. How to Solve the World's Biggest Problems. *Nature* **2015**, *525* (7569), 308–311.
- (5) Cooper, M. M.; Caballero, M. D.; Ebert-May, D.; Fata-Hartley, C. L.; Jardeleza, S. E.; Krajcik, J. S.; Laverty, J. T.; Matz, R. L.; Posey, L. A.; Underwood, S. M. Challenge Faculty to Transform STEM Learning. *Science* **2015**, *350* (6258), 281–282.
- (6) Luckie, D. B.; Bellon, R.; Sweeder, R. D. The BRAID: Experiments in Stitching Together Disciplines at a Big Ten University. *Journal of STEM Education: Innovations and Research* **2012**, *13* (2), 1.
- (7) Roche Allred, Z.; Shrode, A. D.; Gonzalez, J.; Rose, A.; Green, A. L.; Swamy, U.; Matz, R. L.; Underwood, S. M. Impact of Ocean Acidification on Shelled Organisms: Supporting Integration of Chemistry and Biology Knowledge through Multidisciplinary Activities. *J. Chem. Educ.* **2022**, *99* (5), 2182–2189.
- (8) Green, A. L.; Parent, K. N.; Underwood, S. M.; Matz, R. L. Connecting Ideas across Courses: Relating Energy, Bonds & How ATP Hydrolysis Powers a Molecular Motor. *American Biology Teacher* **2021**, *83* (5), 303–310.
- (9) Martinez, B. L.; Kararo, A. T.; Parent, K. N.; Underwood, S. M.; Matz, R. L. Creating and Testing an Activity with Interdisciplinary Connections: Entropy to Osmosis. *Chem. Educ. Res. Pract.* **2021**, *22* (3), 683–696.
- (10) Roche Allred, Z. D.; Farias, A. J.; Kararo, A. T.; Parent, K. N.; Matz, R. L.; Underwood, S. M. Students' Use of Chemistry Core Ideas to Explain the Structure and Stability of DNA. *Biochemistry and Molecular Biology Education* **2021**, *49* (1), 55–68.
- (11) Kohn, K. P.; Underwood, S. M.; Cooper, M. M. Energy Connections and Misconnections across Chemistry and Biology. *LSE* **2018**, *17* (1), ar3.
- (12) Kohn, K. P.; Underwood, S. M.; Cooper, M. M. Connecting Structure–Property and Structure–Function Relationships across the Disciplines of Chemistry and Biology: Exploring Student Perceptions. *LSE* **2018**, *17* (2), ar33.
- (13) Gayford, C. ATP: A Coherent View for School Advanced Level Studies in Biology. *Journal of Biological Education* **1986**, *20* (1), 27–32.
- (14) Halmø, S. M.; Sensibaugh, C. A.; Bhatia, K. S.; Howell, A.; Ferryanto, E. P.; Choe, B.; Kehoe, K.; Watson, M.; Lemons, P. P. Student Difficulties during Structure–Function Problem Solving. *Biochemistry and Molecular Biology Education* **2018**, *46* (5), 453–463.
- (15) Russ, R. S.; Scherr, R. E.; Hammer, D.; Mikeska, J. Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Science Education* **2008**, *92* (3), 499–525.
- (16) Krist, C.; Schwarz, C. V.; Reiser, B. J. Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *Journal of the Learning Sciences* **2019**, *28* (2), 160–205.
- (17) Hammer, D. Student Resources for Learning Introductory Physics. *American Journal of Physics* **2000**, *68* (S1), S52–S59.
- (18) Becker, N.; Noyes, K.; Cooper, M. Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *J. Chem. Educ.* **2016**, *93* (10), 1713–1724.
- (19) Cooper, M. M.; Williams, L. C.; Underwood, S. M. Student Understanding of Intermolecular Forces: A Multimodal Study. *J. Chem. Educ.* **2015**, *92* (8), 1288–1298.
- (20) Crandell, O. M.; Lockhart, M. A.; Cooper, M. M. Arrows on the Page Are Not a Good Gauge: Evidence for the Importance of Causal Mechanistic Explanations about Nucleophilic Substitution in Organic Chemistry. *J. Chem. Educ.* **2020**, *97* (2), 313–327.
- (21) Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, *96* (9), 1821–1832.
- (22) Crandell, O. M.; Kouyoumdjian, H.; Underwood, S. M.; Cooper, M. M. Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. *J. Chem. Educ.* **2019**, *96* (2), 213–226.
- (23) Caspari, I.; Kranz, D.; Graulich, N. Resolving the Complexity of Organic Chemistry Students' Reasoning through the Lens of a Mechanistic Framework. *Chemistry Education Research and Practice* **2018**, *19* (4), 1117–1141.
- (24) Graulich, N.; Caspari, I. Designing a Scaffold for Mechanistic Reasoning in Organic Chemistry. *Chemistry Teacher International* **2021**, *3* (1), 19–30.
- (25) Ralph, V. R.; Scharlott, L. J.; Schafer, A. G. L.; Deshayé, M. Y.; Becker, N. M.; Stowe, R. L. Advancing Equity in STEM: The Impact Assessment Design Has on Who Succeeds in Undergraduate Introductory Chemistry. *JACS Au* **2022**, *2*, 1869.
- (26) Schwarz, C.; Cooper, M.; Long, T.; Trujillo, C.; Noyes, K.; de Lima, J.; Kesh, J.; Stoltzfus, J. Mechanistic Explanations Across Undergraduate Chemistry and Biology Courses. In *ICLS 2020 Proceedings*; International Society of the Learning Sciences: Nashville, TN, USA, 2020; pp 625–628.
- (27) Yoho, R.; Foster, T.; Urban-Lurain, M.; Merrill, J.; Haudek, K. C. Interdisciplinary Insights from Instructor Interviews Reconciling “Structure and Function” in Biology, Biochemistry, and Chemistry through the Context of Enzyme Binding. *Disciplinary and Interdisciplinary Science Education Research* **2019**, *1* (1), 16.
- (28) Noyes, K.; Carlson, C. G.; Stoltzfus, J. R.; Schwarz, C. V.; Long, T. M.; Cooper, M. M. A Deep Look into Designing a Task and Coding Scheme through the Lens of Causal Mechanistic Reasoning. *J. Chem. Educ.* **2022**, *99* (2), 874–885.
- (29) Mislevy, R. J.; Almond, R. G.; Lukas, J. F. A Brief Introduction to Evidence-Centered Design. *ETS Research Report Series* **2003**, *2003* (1), i–29.
- (30) Harris, C. J.; Krajcik, J. S.; Pellegrino, J. W.; DeBarger, A. H. Designing Knowledge-In-Use Assessments to Promote Deeper Learning. *Educational Measurement: Issues and Practice* **2019**, *38* (2), 53–67.
- (31) Bryfczynski, S. *BeSocratic: An Intelligent Tutoring System for the Recognition, Evaluation, and Analysis of Free-Form Student Input*. Ph.D. Dissertation, Clemson University, 2012.
- (32) Cohen, J. A. Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **1960**, *20* (1), 37–46.
- (33) Landis, J. R.; Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33* (1), 159–174.
- (34) IBM. *SPSS Statistics*; 2020.
- (35) Green, S. B.; Salkind, N. J. Two-Way Contingency Table Analysis Using Crosstabs. In *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*; Pearson Education Inc.: Upper Saddle River, NJ, 2011; pp 366–376.
- (36) Cohen, J. A Power Primer. *Psychol. Bull.* **1992**, *112* (1), 155–159.
- (37) Haskel-Ittah, M. Explanatory Black Boxes and Mechanistic Reasoning. *Journal of Research in Science Teaching* **2023**, *60* (4), 915.
- (38) Cooper, M. M.; Stowe, R. L.; Crandell, O. M.; Klymkowsky, M. W. Organic Chemistry, Life, the Universe and Everything (OCLUE): A Transformed Organic Chemistry Curriculum. *J. Chem. Educ.* **2019**, *96* (9), 1858–1872.