Supporting Development of Mechanistic Reasoning Across Chemistry and Biology

Caleb M. Trujillo,[§†] Jenna Kesh,[‖] Melanie M. Cooper,[¶] Joelyn de Lima,[§‡] Tammy M. Long,[§] Keenan Noyes,[¶] Christina V. Schwarz,[#] and Jon R. Stoltzfus[††*]

[§]Plant Biology, Michigan State University, East Lansing, MI 49445; [‖]Neuroscience, Michigan State University, East Lansing, MI 49445; [¶]Chemistry, Michigan State University, East Lansing, MI 49445; [#]Teacher Education, Michigan State University, East Lansing, MI 49445; [††]Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 49445

ABSTRACT

While extensive research explores the difficulties undergraduate students face when learning chemistry or biology, less is known about how students learn across disciplines, especially when constructing mechanistic explanations. This study investigates (1) how undergraduate students explain differences in protein function, focusing on their use of conceptual resources and mechanistic reasoning, and (2) how task design influences their explanations. Using three different tasks with the same underlying mechanistic explanation, we analyzed responses from students who had completed a transformed cell and molecular biology course to determine what conceptual resources they use and how they connect those resources. Our findings show that, given a task with appropriate scaffolding, half of the students constructed mechanistic explanations. We also found that students rarely integrated ideas from both biology and chemistry, but when they did, they consistently constructed mechanistic explanations. Task design significantly influenced the conceptual resources students used and the frequency of mechanistic explanations. These findings highlight the need for careful instructional and curricular design to support cross-disciplinary learning and mechanistic reasoning in STEM education.

INTRODUCTION

Linkages between disciplines in undergraduate STEM curricula typically involve pre- or corequisites, for example, requiring that introductory chemistry be taken before, or at the same time as, introductory cell and molecular biology, which in turn serve as prerequisites for upper-level biology courses. However, explicit coordination across the curriculum is often lacking (Smith *et al.*, 2024), and more work is needed to understand whether and how introductory courses prepare students for future learning. To help address this gap, we explore how students who have completed transformed introductory chemistry and biology courses use mechanistic reasoning and integrate ideas across disciplines when asked to construct an explanation of why two proteins have different functions.

Mechanistic Explanations Across Biology and Chemistry

Mechanistic explanations require identifying entities that underlie a phenomenon and creating causal chains to explain that phenomenon (Krist *et al.*, 2019; Bachtiar *et al.*, 2022). They are highly valued in both chemistry and biology education because they help students make sense of and make predictions about complex systems (Talanquer, 2018; Bachtiar *et al.*, 2022). Mechanistic explanations of protein function are critical for understanding issues such as why bacteria develop antibiotic resistance and why we need new versions of the COVID vaccine. In STEM curricula, these explanations of protein function connect ideas taught in chemistry and biology and support learning in upper-level courses such as biochemistry

Paula P. Lemons, *Monitoring Editor*

Submitted Jul 21, 2025; Revised Jun 1, 2026; Accepted Jun 30, 2026

CBE Life Sci Educ September 1, 2026 25:ar34

DOI:10.1187/cbe.25-07-0149

[†]Present address: Interdisciplinary Arts and Sciences, University of Washington Bothell, Bothell, WA 98011

[‡]Present address: The Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

Author contributions: C.T., M.C., J.D., T.L., K.N., C.S., and J.S. conceived and designed the experiments; C.T. and J.S. performed the experiments; C.T. and J.S. analyzed the data; C.T. and J.S. drafted the article; C.T. and J.S. prepared the digital images; C.T. and J.S. led coding scheme development and coded student responses; J.K. assisted with coding scheme development and coded student responses; M.C., T.L., K.N., C.S., and J.D. provided input on coding scheme development, data analysis, and article revision.

Conflict of interest: The authors declare no competing financial interest.

*Address correspondence to: Jon R. Stoltzfus (stoltzfu@msu.edu).

© 2026 Trujillo *et al.* CBE—Life Sciences Education © 2026 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/3.0).
"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

(Villafañe *et al.*, 2011; Tansey *et al.*, 2013; Halmo *et al.*, 2018), cell biology/physiology (Mulnix, 2003; Agrawal, 2022), and physiology (Michael, 2021). However, there is a schism between the value educators place on mechanistic reasoning and students' abilities to construct mechanistic explanations (Bachtiar *et al.*, 2022).

Multiple studies using semistructured and think-aloud interviews have concluded that undergraduates have difficulty constructing mechanistic explanations of biological phenomena. Scott *et al.* (2018) found that undergraduate science majors used a causal mechanistic approach less than 10% of the time when explaining complex, real-world phenomena, including those related to biology. Flowers, Holder, and Rump *et al.* (2023) found that students included a mechanism in less than 15% of their explanations of cellular subsystems. In a separate study, Flowers, Holder, and Gardner (2023) found that only one of nine students interviewed constructed a connected mechanistic model of biofilm formation. Southard *et al.* (2016) found that students frequently conflated ideas across processes included in the central dogma and coded less than 20% of students' explanations as mechanistic. In a separate study, Southard *et al.* (2017) found that less than 30% of students created complete mechanistic explanations when explaining novel biological phenomena such as how bacteria sense and respond to the environment and how a plant generates some tissues that are poisonous and others that are not.

A significant body of research has produced similar results when undergraduate students explain chemical phenomena. The following studies illustrate common findings in this area. Becker *et al.* (2016) found complete mechanisms in less than 10% of students' drawings and written explanations of London Dispersion Forces. In semistructured interviews carried out by Macrie-Shuck and Talanquer (2020), the percentage of students who formed normative mechanistic explanations of phenomena involving energy transfer and transformation varied from just under 40% to less than 10% depending on the specific phenomena the students were explaining. Deng and Flynn (2021) also found differences in the prevalence of causal reasoning based on the specific phenomena being explained. Crandell *et al.* (2020) analyzed students' written responses and found that between 40% and 60% of students constructed causal mechanistic explanations of nucleophilic substitutions, with the difference attributed to the type of instruction students received. Scharlott *et al.* (2024) found that students use causal mechanistic reasoning to explain dissolution and boiling points in only a quarter to a third of their written explanations. Bodé *et al.* (2019) found that the vast majority of students did not provide significant granularity in their causal mechanistic arguments to meet course expectations for arguments about plausible reaction mechanisms. Haas *et al.* (2024) found that even with a highly scaffolded instructional task, some students needed additional support to develop fully causal mechanistic explanations.

Mechanistic explanations of why different proteins have different functions present an additional challenge. They require integration of ideas from chemistry into the explanation of biological phenomena, something we know is challenging. Analysis of students' explanations of their answers to a multiple-choice exam question in introductory biology indicated that, despite having completed the chemistry prerequisite, the students were unable to apply acid-base chemistry to biological contexts (Haudek *et al.*, 2012). Interviews of students who had completed both biology and chemistry revealed that many did not make connections between these courses when discussing energy-related phenomena (Kohn *et al.*, 2018b), but were able to make connections between structure-property relationships from chemistry and structure-function relationships from biology (Kohn *et al.*, 2018a). Analysis of an activity designed to bridge the application of electrostatic interactions and energy across chemistry and biology contexts found that many students who provided sophisticated responses in a chemistry context had difficulties integrating those same ideas into a biology context (Roche Allred *et al.*, 2021). A study looking at how students integrate ideas in an explanation that spanned chemistry, molecular biology, and organismal biology showed that students at all levels of a STEM curriculum had difficulty constructing a complete mechanistic explanation across these scales (Shiroda *et al.*, 2024). A study looking at protein structure–function problem solving found that introductory biology and upper-level biochemistry students frequently struggled to apply noncovalent interactions in this context (Halmo *et al.*, 2018).

On a positive note, studies also provide evidence that students who are not constructing complete mechanistic explanations often include aspects of mechanistic reasoning in their responses. For example, most students identified the underlying functional modules needed for a mechanistic explanation of how bacteria sense and respond to an environmental cue (Southard *et al.*, 2017). Most students included strong normative ideas in their explanations of phenomena involving energy transfer and transformation (Macrie-Shuck and Talanquer, 2020). And students who did not include fully developed causal-and-effect reasoning about reaction mechanisms often included elements needed for cause-and-effect reasoning (Haas *et al.*, 2024).

We also know that differences in instructional practices can impact students' use of mechanistic reasoning (Trujillo *et al.*, 2016; Crandell *et al.*, 2019; Houchlei *et al.*, 2021; Klymkowsky, 2021; Ralph *et al.*, 2022; Scharlott *et al.*, 2024; Talanquer, 2024) and the persistence of this type of reasoning (Crandell *et al.*, 2020). The amount and type of scaffolding can also influence students' construction of causal mechanistic explanations (Noyes and Cooper, 2019; Graulich and Caspari, 2020; Haas *et al.*, 2024) and the learning gains associated with reasoning activities (Kranz *et al.*, 2025). In addition, instructional practices aimed at reducing students' cognitive load may undermine the development of mechanistic reasoning. "Shortcutting" provides a simple rule or heuristic rather than a longer, more complex step-by-step discussion of the underlying mechanism, which allows instructors to reduce confusion and get through content but may create gaps in student understanding (Talanquer, 2024). "Black boxing" links cause and effect without describing the underlying connections. This is an important and necessary strategy in science communication, but a lack of clarity on whether, how, and when to use these "black boxes" can be problematic for students (Haskel-Ittah, 2022). "Bottoming out" ends an explanation without providing additional underlying mechanisms, which often exist but have been deemed irrelevant to the current discussion (Machamer *et al.*, 2000). This is

needed to keep instruction focused on course content, but a lack of clarity on when an explanation should "bottom out" in this way may create additional confusion for students as they navigate across biological scales (van Mil *et al.*, 2013). Clearly, supporting the development of mechanistic reasoning involves complex instructional choices.

Taken as a whole, these previous studies indicate that students often do not engage in the type of cross-disciplinary mechanistic reasoning expected in upper-level courses. It is possible that prerequisite courses have not prepared students to construct these explanations. It is also possible that the new course context and/or the context of the prompts failed to activate desired resources. This includes a failure to communicate to students what they were expected to do in the new context. These studies also indicate that carefully designed instruction can improve undergraduates' mechanistic reasoning and provide insights into how to design and implement this instruction for specific phenomena within a discipline.

Our work helps build understanding of whether, how, and in what context students at the end of a transformed introductory biology course connect ideas across biology and chemistry into mechanistic explanations, a relatively underexplored area of study. It also provides insights needed to design curricula that support mechanistic reasoning across courses. We draw on a resources perspective on student learning (Hammer, 2000; Hammer *et al.*, 2005) and a heuristic for analyzing mechanistic explanations (Krist *et al.*, 2019) to examine students' explanations of a phenomenon at the interface of chemistry and biology. We also explore how task details influence students' explanations.

A Resource Perspective on Student Learning and Curricula

We approach this study from a resources perspective that views learning as a transient cognitive state in which an individual activates a particular set of resources given the task at hand (Hammer, 2000; Hammer *et al.*, 2005). The resources perspective is a constructivist approach that seeks to build on students' existing knowledge rather than trying to confront and replace existing misconceptions. This is an extension of the "knowledge in pieces" approach to learning and instruction (diSessa, 1988; Smith III *et al.*, 1994). It argues that students rarely transfer an intact, coherent understanding of concepts from one context to another. Instead, details in instructional tasks, instructional messaging, assessment practices, and course design activate finer-grained resources that are highly context dependent (Hammer *et al.*, 2005). This suggests that when faced with instructional tasks, rather than bringing robust, fully coherent conceptions from previous units in a course, prerequisite courses, or personal experience, students activate a variety of finer-grained resources and assemble these into a response to the task. For example, undergraduate students completing a course in which they were taught about biological information flow frequently construct fragmented explanations of the processes involved, mixing ideas across transcription, translation, and replication rather than demonstrating a coherent and connected understanding of these processes and their cellular purposes (Southard *et al.*, 2016).

The resources perspective suggests that students bring both conceptual resources and epistemological resources to instructional tasks (Hammer, 2000). These resources are fine-grained and context-dependent, changing from task to task and even moment to moment within a task. Conceptual resources are the fragments of previous knowledge or intuitions students work with as they complete an instructional task. The exact grain size and nature of conceptual resources vary with discipline and research study. Examples include partial charges, electrostatic interactions, how bonds are formed and broken, the role of amino acids, and the effects of kicking a bowling ball (Redish and Hammer, 2009; Ralph *et al.*, 2022; Franovic *et al.*, 2023; Haas *et al.*, 2024; Shiroda *et al.*, 2024). Epistemological resources serve as in-the-moment, context-sensitive guides, determining how a student approaches a task and what conceptual resources they incorporate in their response (Hammer and Elby, 2003; Bing and Redish, 2009). Epistemological resources have to do with students' beliefs regarding how knowledge is generated and determine how a student approaches the task and what tools they use to complete it; for example, citing facts from authority or checking for logical coherence (Hammer and Elby, 2003; Bing and Redish, 2009). These approaches and tools are observable in students' responses to a task and provide insights into hidden epistemological beliefs. We call these observable elements in a student response epistemic resources (Krist *et al.*, 2019; Franovic *et al.*, 2023; Shiroda *et al.*, 2024).

The resources perspective defines framing as a student's interpretation of a task and how they should approach it based on its purpose and context. Instructional materials, instructional team messaging, and assessment practices can influence the epistemological frames students adopt by providing context and purpose to learning tasks (Gouvea *et al.*, 2022). When students complete instructional tasks, they can assume different epistemological frames that will determine the conceptual and epistemic resources they use (Hammer *et al.*, 2005; Bing and Redish, 2009) and alter the content of their responses. Researchers can infer students' epistemological framing by observing the resources used as students complete tasks. For example, students framing the task of writing a lab report as an opportunity to argue from evidence might carefully link observational data to claims, while students framing it as an expectation that they confirm already known facts might repeat key information provided in the lab manual (Gouvea *et al.*, 2022). Similarly, students using the epistemological frame that knowledge comes from authority may cite facts from a textbook, while a student using an alternative frame that knowledge is constructed might check for logical consistency (Bing and Redish, 2009). In our work, we use these observable elements in student responses, epistemic resources, to provide evidence of students' epistemological framing.

The idea that students use fine-grained, context-sensitive resources rather than fully formed, stable misconceptions when approaching instructional tasks has important instructional implications (diSessa, 1988; Smith III *et al.*, 1994; Hammer, 2000; Hammer and Elby, 2003). If an instructor believes students have stable misconceptions, instruction will focus on revealing the misconception to the student and convincing the student to replace the misconception with a scientifically accurate understanding. On the other hand, if an

instructor believes students bring fragmented knowledge and problem-solving approaches that are context-dependent but can be built into a coherent understanding, they will view the resources students bring as assets, and instruction will focus on understanding and repeatedly activating students' resources in productive ways.

This also has implications regarding the purpose of instruction and what one would hope to see following effective instruction (Bransford and Schwartz, 1999). Expertise is often associated with principle-oriented schemas that are activated when experts solve problems (Chi *et al.*, 1981). The resources perspective suggests that, with appropriate instruction, a particular pattern of fine-grained resource activation can become stable such that these resources are likely to be mutually activated and locally coherent within a given context, moving students toward more expert-like ways of thinking (Hammer *et al.*, 2005). Thus, exploring the resources students use to complete tasks following instruction provides insights into the impact of instruction. If the task presents students with contexts they will likely encounter in the future, it provides insights into how students might activate and use resources in future learning tasks, which has implications about their preparation for future learning (Bransford and Schwartz, 1999). If students who have completed a course are asked to explain a phenomenon they will likely encounter in a subsequent course, the resources they use provide insights into their preparation for learning in that subsequent course and have implications for curriculum design and student success.

Mechanistic Reasoning Framework

Distinct types of scientific explanations exist in discussions about science education (Machamer *et al.*, 2000; Duncan and Reiser, 2007; Russ *et al.*, 2008; Braaten and Windschitl, 2011; van Mil *et al.*, 2013; Trujillo *et al.*, 2015; Southard *et al.*, 2017; Talanquer, 2018; Alameh *et al.*, 2023). Here, we focus on an analytical heuristic for mechanistic reasoning described by Krist and colleagues to understand students' explanations (Krist *et al.*, 2019). In the Krist heuristic, mechanistic explanations rely on three key elements: 1) thinking across levels, 2) unpacking, which includes linking factors into causal chains, and 3) connecting the chains to explain the target phenomenon. Thinking across levels can involve identifying entities at scalar levels below the phenomenon and/or using non-material, abstract factors such as energy or forces. These three elements align with aspects of student responses frequently used to identify mechanistic reasoning in biology and chemistry education research: 1) considering scalar levels below the phenomenon, 2) using entities and their associated activities, and 3) a causal connection to the target phenomenon (Talanquer, 2018; Bachtiar *et al.*, 2022; Ralph *et al.*, 2022; Scharlott *et al.*, 2024). For example, a mechanistic explanation of why a ligand binds to a protein must move down a scale from the protein and ligand to identify atoms or functional groups. It must then determine the charges or partial charges (properties) and attractive noncovalent interactions (activities) of these atoms and functional groups. Finally, it must link these together to explain how and why this causes the ligand to bind to the protein (Noyes *et al.*, 2022).

We can apply the resources perspective to the Krist heuristic. Using this lens, the three elements of the Krist epistemic heuristic—moving across levels, unpacking by linking factors, and causal connection to the phenomenon—are epistemic resources and will be employed regardless of the phenomena being explained. Factors, on the other hand, serve as the conceptual resources needed to construct a specific mechanistic explanation and are context dependent. The specific entities, properties, and activities incorporated into an explanation will vary depending on the phenomena being explained.

The presence of specific epistemic resources in a student's response to a task can provide insights into how they are framing the task (Bing and Redish, 2009; Gouvea *et al.*, 2022). Using this approach, the inclusion of entities at scalar levels below the phenomen, linking factors, and connecting back to the phenomenon provide evidence that the student has framed the task as constructing a mechanistic explanation.

Scaffolding and Task Design

Scaffolding is the structured sequence of prompts in an instructional task required to help students complete a task for the first time and support new ways of thinking (McNeill *et al.*, 2006; Kranz *et al.*, 2025). It can also help students manage complex tasks and clarify key expectations (*How People Learn*, 2000; Pea, 2004; Caspari and Graulich, 2019). Details in the scaffolding help students determine how to frame the task and impact what conceptual resources they will use in their response. For example, scaffolded contrasting cases are often used to elicit reasoning about reaction mechanisms in organic chemistry (Graulich and Caspari, 2020; Haas *et al.*, 2024; Kranz *et al.*, 2025). Insufficient scaffolding can result in an underrepresentation of student knowledge. For example, changing the design of a prompt aimed at eliciting causal mechanistic reasoning about London dispersion forces to include three boxes for drawing connected by arrows rather than a blank drawing space significantly changed the frequency at which introductory chemistry students formulated causal mechanistic drawings (Noyes and Cooper, 2019). This same study showed differences in causal mechanistic reasoning between students' drawing and writing, a finding also seen in biology students' written and drawn explanations of the origin of variation (Bray Speth *et al.*, 2014). Other studies have shown that increased scaffolding does not always lead to stronger explanations (Crowder *et al.*, 2024; Kranz *et al.*, 2025) and the need to elicit student knowledge without making the expected answer so obvious that the task no longer measures student reasoning (Noyes *et al.*, 2022). Taken together, these studies show that while scaffolding is needed to support mechanistic reasoning, empirical testing is needed to ensure that scaffolding has the intended outcomes.

Developing Curricula that Support Mechanistic Reasoning and Connections Across Disciplines

Developing curricula that support mechanistic reasoning across courses will require appropriate and consistent approaches that align students' and instructors' expectations of success and repeatedly activate resources in productive ways (Gouvea *et al.*, 2013; Macrie-Shuck and Talanquer, 2020; Smith *et al.*, 2024). Here, we focus on three-dimensional learning (3DL) as a framework that supports consistent expectations when constructing cross-disciplinary instructional tasks that both facilitate student learning and reveal to

instructors and researchers what resources students are using (National Research Council, 2012; Laverty *et al.*, 2016; Bain *et al.*, 2020).

In 3DL, instruction and assessment focus on disciplinary core ideas, science practices, and cross-cutting concepts in well-defined and measurable ways (National Research Council, 2012; Laverty *et al.*, 2016; Bain *et al.*, 2020). This allows instructors to transform curriculum and prepare students to think like scientists and solve complex, real-world problems (Cooper *et al.*, 2015). 3DL provides a framework that can be used to construct assessments that require the use of disciplinary core ideas and reasoning in scientifically authentic ways (Laverty *et al.*, 2016). The framework also allows curricula to explicitly link introductory STEM courses by focusing on the same scientific practices and cross-cutting concepts across courses, helping provide consistent epistemological framing.

Phenomena that cross disciplinary boundaries can help students integrate knowledge across disciplines and provide opportunities for 3DL. For example, antibiotic resistance in bacteria sits at the interface between chemistry and molecular biology and lends itself to 3DL. Developing and using models or constructing explanations of how bacteria become antibiotic resistant engages students in a scientific practice that requires them to use a crosscutting concept, cause and effect, and connect core ideas from chemistry (atomic/molecular structure and properties and electrostatic and bonding interactions) and biology (information flow, structure and function, and evolution). From a resources perspective, these scientific practices require students to use epistemic resources related to mechanistic reasoning to integrate conceptual resources from the disciplines that include both physical entities (e.g., mRNA, amino acids, and proteins) and abstract concepts (e.g., forces and information). Using resources to construct a mechanistic explanation helps students create coherent and usable knowledge from fragmented pieces (Cooper, 2015). Designing curricula that support this kind of reasoning across courses requires a deeper understanding of how students integrate ideas taught in different courses when developing mechanistic explanations.

Research Questions

Building from these frameworks, we carried out an empirical study to explore how students who completed chemistry and biology courses that incorporate 3D instruction explain why two proteins have different functions. Our goal was to better understand the conceptual and epistemic resources students use and how task design shapes their responses. In this study, we analyze drawn and written student responses to three similar tasks to detect the use of resources and identify mechanistic explanations. We then integrate our findings with the 3DL framework to provide insights into instructional and curricular improvement. We focus on two research questions regarding these explanations.

1. What resources do students who have completed transformed introductory chemistry and cell and molecular biology courses include in explanations of protein function, and how are these resources linked within their explanations?
2. How does task design influence the types of conceptual resources these students include in explanations of protein function and the likelihood of producing a mechanistic explanation?

METHODS

Methodology

The goal of this exploratory research was to better understand the resources students use when asked to explain a phenomenon that crosses disciplinary boundaries and the impact of task details on resource activation. Therefore, we performed a qualitative study to describe and categorize the types of resources students use in written and drawn responses when asked to explain why two proteins function differently. To do this, we designed three tasks focused on protein function and created a codebook to analyze students' use of factors and causal relationships. We then collected and coded responses from students in a transformed introductory cell and molecular biology course and analyzed these data by mapping Krist's mechanistic criteria onto our coding scheme and clustering responses to examine how students linked resources and constructed mechanistic explanations.

Task and Coding Scheme Development

A detailed description of the process used in task development has already been published for a chemistry-focused companion task and coding scheme (Noyes *et al.*, 2022). The development of the molecular biology-focused tasks and coding scheme described here took place in parallel with that effort using similar approaches.

Context for Task Development. Task development followed this trajectory: 1) discussions among team members to identify key points of connection across introductory chemistry, cell and molecular biology, and organismal biology courses that were good candidates for developing a connected, mechanistic explanation; 2) development and piloting of an initial task that spanned all three disciplines followed by analysis of pilot data which resulted in the decision to simplify the tasks; 3) iterative development, piloting of tasks, and analysis using simplified tasks within specific disciplinary contexts to develop a set of tasks that linked across the disciplines (Noyes *et al.*, 2022; Franovic *et al.*, 2023; Shiroda *et al.*, 2024).

In developing our initial task, we used our knowledge from previous work on students' explanations in introductory chemistry and biology courses (Dauer *et al.*, 2013; Cooper *et al.*, 2015; Trujillo *et al.*, 2015; Bierema *et al.*, 2017) to explore content and ideas across courses that would be suitable for linked, high-quality mechanistic explanations. Based on this exploration, we chose to focus on two core ideas from chemistry—atomic/molecular structure and properties, and electrostatic and bonding interactions—and three from biology—information flow, structure and function, and evolution—using the following learning performances:

- Students should use molecular structures and properties to explain intermolecular forces, specifically why a protein binds to a specific ligand.

TABLE 1. Three task versions used in this study

Version		Task	Responses coded	Mechanistic explanations elicited
Task 1	Description	The task asks for an explanation of how two different proteins function and what causes those functions, allowing students to choose the functions of the proteins. The students then explain what causes the differences in function. The task provides no specific disciplinary cues.	40	18 (45%)
	Text	Each cell in your body produces thousands of different proteins with different functions. 1) Think of a function of a protein and construct a drawing and written explanation of how this protein functions and what causes it to have this specific function. 2) Think of a different function of a protein and construct a drawing and written explanation of how this protein functions and what causes it to have this specific function. 3) Provide a written explanation of what causes the proteins you drew to have different functions.		
Task 2	Description	The task structure is identical to task 1 but gives the biological function of the proteins, identifying them as transporters that move different sugars.	41	4 (10%)
	Text	Consider two different protein transporters in a cell. Transporter A is able to move sucrose across the membrane while transporter B is able to move lactose across the membrane. 1) Construct a drawing and written explanation of how transporter A moves sucrose across the membrane. 2) Construct a separate drawing and written explanation of how transporter B moves lactose across the membrane. 3) Provide a written explanation of what causes these transporters to move different sugars.		
Task 3	Description	The task provides the biological function of the proteins, receptors that differentially detect a hormone, and provides a specific chemistry cue by depicting functional groups (refer to Figure 1). The structure varies from tasks 1 and 2 in that the cause of differences in function is incorporated in part 2 rather than being prompted by a separate part 3.	50	25 (50%)
	Text	Consider two different protein receptors in a cell. Receptor A is able to detect the hormone depicted on the right (refer to Figure 1) while receptor B is not. 1) Construct a drawing and written explanation of how receptor A detects this hormone. 2) Construct a separate drawing and written explanation of why receptor B is unable to detect this hormone.		

- Students should use intermolecular/intramolecular forces and genetic information to explain protein function, specifically why a specific protein has a specific configuration of amino acids that allows that specific protein to bind a specific ligand.
- Students should use protein function to explain phenotypic variation, specifically why proteins that bind to different ligands can result in different characteristics at the organismal level.

Using these learning performances, we developed our initial task in the context of the evolution of fluoroquinolone resistance in *Mycobacterium tuberculosis*. This task was administered to students enrolled in an introductory cell and molecular biology course or an introductory organismal biology course. On reviewing student responses, it became clear that this pilot task did not elicit the kinds of explanations we were seeking, and the extensive variability in student explanations, coupled with complexities introduced by details of the context, led us to simplify the task and provide less specific context.

To decrease complexity and student confusion, we developed several simplified tasks that were focused within each discipline with the goal that they would later be linked into a larger, connected prompt. We administered each task to small sets of students and analyzed student responses for desired content and structure. Using this evidence allowed us to revise the tasks. This process eventually led to the development of the tasks used to examine mechanistic explanations across a variety of biology and chemistry courses (Shiroda *et al.*, 2024).

In this study, we explore student responses to three similar molecular biology-focused tasks that were developed as intermediate steps in creating the task used by Shiroda *et al.* (2024). All three tasks in this study focus on students' use of intermolecular/intramolecular forces and genetic information to explain protein function. Refer to Table 1 for a description and the text of each task and Figure 1 for the diagram included in Task 3.

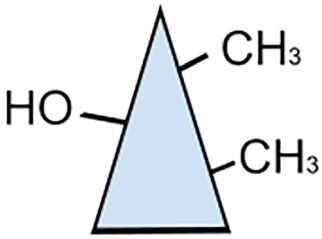


FIGURE 1. The diagrammatic hormone provided to students as part of Task 3 with the goal of activating chemistry resources related to structure-properties, bonding, and electrostatic interactions.

Rationale Underlying Task Development. Principles from evidence-centered design guided prompt development (Mislevy *et al.*, 2003; Noyes *et al.*, 2022). First, we defined the desired goal as mechanistic reasoning at the interface of chemistry and biology viewed from a resources perspective. We then determined that we wanted to observe: 1) the chemistry core ideas of atomic/molecular structure and properties and electrostatic and bonding interactions, 2) the biology core ideas of information flow or structure and function, and 3) the epistemic resources of moving down scalar levels, linking factors, and connecting to the phenomenon. The presence of these in explanations of how two proteins function and why they have different functions provides evidence that students have constructed a mechanistic explanation using ideas from both chemistry and biology.

Because our study aimed to understand what students know and can do at a specific point in a STEM curriculum (National Research Council, 2001), we targeted the level of scaffolding toward revealing understanding (Noyes *et al.*, 2022) rather than supporting learning (Graulich and Caspari, 2020). The goal is to provide enough scaffolding so that students who understand mechanistic reasoning will approach the problem using this epistemological frame, productively incorporating core ideas from both chemistry and biology to explain the phenomenon. At the same time, it is important not to provide so much scaffolding that students' explanations no longer reflect their own thinking but are obvious responses that require little understanding (Noyes *et al.*, 2022).

We used contrasting cases and multimodal representation in our task. Contrasting cases can be used as either learning tools or assessments and help elicit students' knowledge because they require students to think about critical similarities and differences between the two cases, slowing down their thinking to produce more deeply reasoned responses (Alfieri *et al.*, 2013; Graulich and Caspari, 2020). We also asked students to represent their thinking both in drawing and in writing. The ideas students bring forward in these different modalities sometimes vary (Bray Speth *et al.*, 2014; Noyes and Cooper, 2019), and thus a multimodal approach would likely capture a larger range of student thinking. In addition, there is evidence that drawing can support mechanistic reasoning (De Andrade *et al.*, 2022).

We also knew from our previous work the importance of asking students both how a process happened and how a molecule carries out its function and why this happens (Cooper, 2015; Becker *et al.*, 2016; Cooper *et al.*, 2016; Noyes and Cooper, 2019). Another way of phrasing the "why" question is what causes the process to happen. Thus, we ensured that our tasks asked the students both how and why. We are aware that our use of how and why differs from their use in evolutionary biology (Shan, 2025) and aligns with their use in molecular biology and chemistry (Machamer *et al.*, 2000; Haas *et al.*, 2024).

Each task in this study used contrasting cases by asking students to compare the functions of two proteins and multimodal representation by asking students to both draw and write (Table 1). Tasks 1 and 2 have parallel structures, each asking students to: 1) draw a diagram and write an explanation of how a protein functions and what causes it to have this function, 2) draw a diagram and write an explanation of how a second protein functions and what causes it to have this function, and 3) explain what causes the differences in function. They differ only in that Task 1 asks students to think about two different proteins and lets them choose any proteins they wish, while Task 2 asks students to compare transporters that move different sugars. Task 3 provides students with a diagram that includes functional groups and tells them that receptor A detects this molecule while receptor B does not. It then asks students to 1) construct a drawing and written explanation of how receptor A detects the hormone and 2) construct a drawing and written explanation of why receptor B does not detect the hormone. Our thinking during task construction was as follows. Task 1 provided little context and gave students more freedom to incorporate whatever ideas they wished. Task 2 provided a specific biological context that might help focus students' thinking on differences in protein–ligand interactions needed for selective transport. Task 3 also provided a specific biological function that might help focus students on protein–ligand binding needed for detection, but also included a clue that might activate chemistry resources. The hope was that the specific chemistry cue would prompt students to apply conceptual resources from chemistry to the biological context in a way that produces a deeper explanation (Gouvea *et al.*, 2013). Transporters and receptors were covered in the course, but these specific sugar transporters and the imaginary hormone were not explicitly described or discussed as part of the course materials.

One of the goals of these tasks was to help the research team develop a final, connected task that better elicited mechanistic reasoning across a range of scales and from students at different points in a STEM curriculum (Shiroda *et al.*, 2024). Thus, the research team felt the differences between these tasks would provide the insights needed to design the final prompt and attain this goal. Following data collection and analysis, it also became clear that the tasks resulted in significant differences in resource utilization.

Rationale Underlying Coding and Analysis. Based on discussions with the entire research team and guided by our learning performances, the mechanistic reasoning heuristic (Krist *et al.*, 2019), and our expectation of what factors students should identify and how they should link these factors together to explain why different proteins have different functions, authors JK, JRS, and CT developed an initial coding scheme. Following data collection with the three versions of

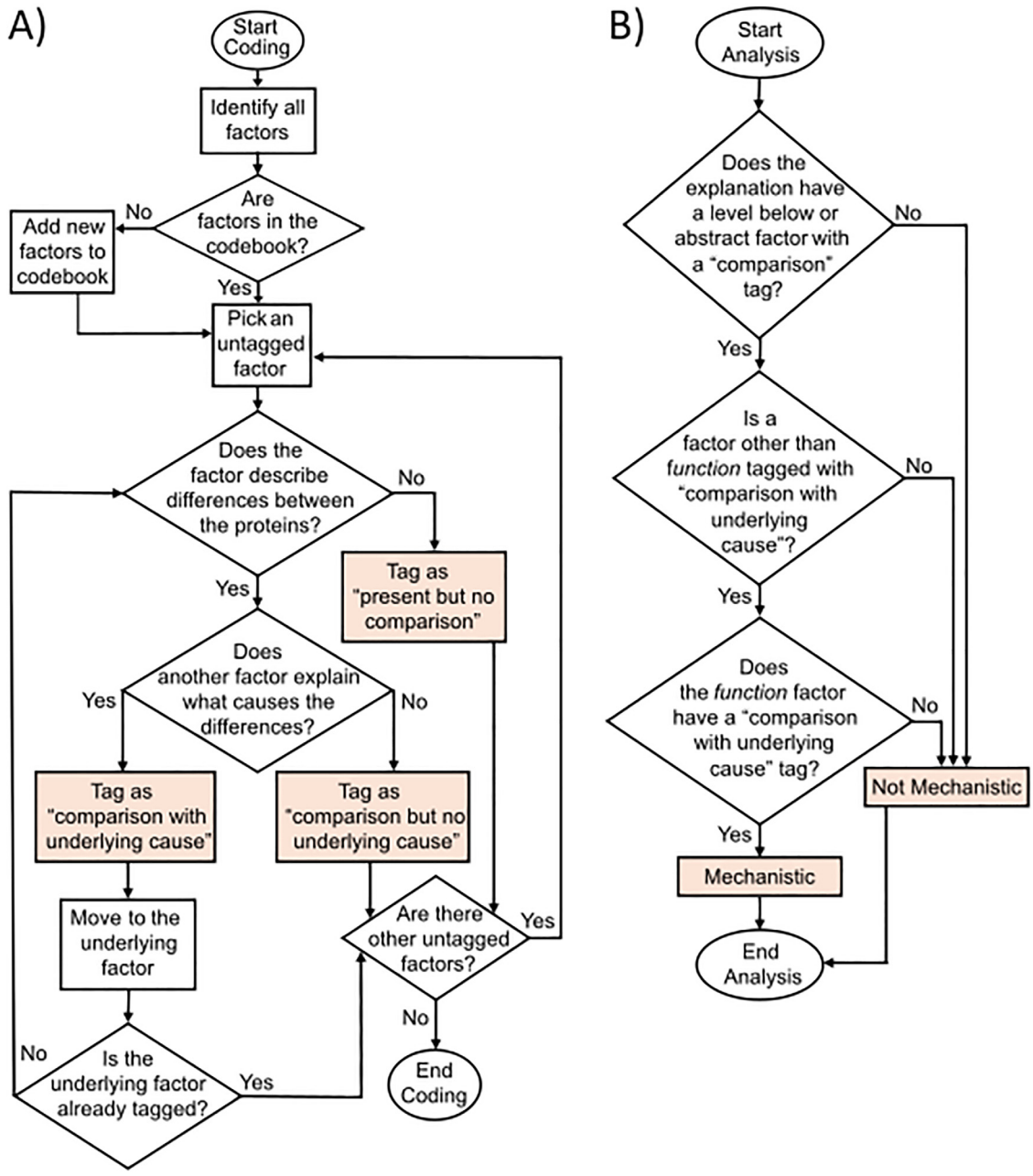


FIGURE 2. Flow charts showing the coding and subsequent analysis to determine if the response was a mechanistic explanation. (A) Each response was coded to identify the factors students included in their response, and each factor was given a linking tag to describe how it was connected to other factors (refer to Tables 1 and 2 for full details). (B) Following coding, Microsoft Excel was used to analyze the factor codes and linking tags applied to each response to determine if the response met the three mechanistic criteria described by Krist *et al.* (2019; refer to Table 4 for full details).

the protein function task, authors JK, JRS, and CT applied this initial coding scheme to student responses in iterative rounds involving independent coding of responses, group discussion of how well the coding scheme was functioning, and modification of the coding scheme before the next round of coding. Throughout this process, the entire interdisciplinary research team was consulted and provided input on the coding scheme as it developed.

Our analysis relies on students' use of three epistemic resources to identify mechanistic reasoning. The three epistemic resources are the inclusion of entities at scalar levels below the phenomenon or abstract entities, linking factors, and connecting back to the phenomenon (Krist *et al.*, 2019; Bachtiar *et al.*, 2022; Noyes *et al.*, 2022). Analysis consisted of two phases. During the first phase, we identified factors and the linkage of factors into causal chains (Figure 2A). During the second

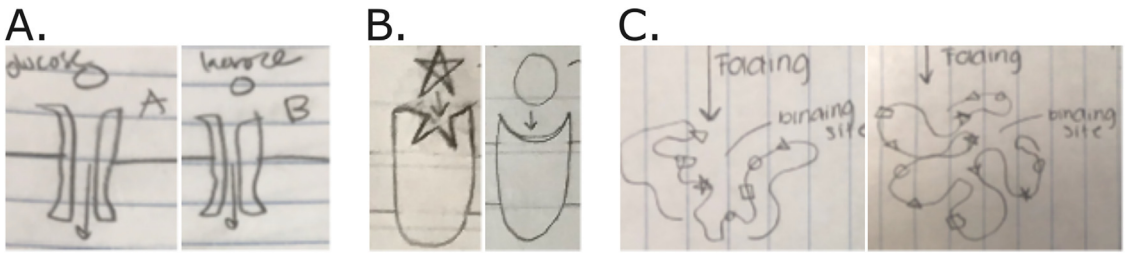


FIGURE 3. Example student drawings illustrating the application of linking tags to the *shape* factor. (A) Tagged as "present but no comparison"—the drawing shows shapes for each receptor, but no differences in shape are depicted (refer to Response TV2_49 in Supplemental Material). (B) Tagged as "comparison but no underlying cause"—the drawing shows a clear difference in shape affecting binding, but does not indicate why the receptor shapes differ (refer to Response TV2_45 in Supplemental Material). (C) Tagged as "comparison with underlying cause"—the drawing shows differences in the binding site, with varied symbols along the line representing amino acid differences that explain the shape differences (refer to Response TV1_17 in Supplemental Material).

phase, we used algorithmic analysis of this coding to determine if a response was a mechanistic explanation (Figure 2B).

Coding for Factors and Linking. The coding scheme first identifies the factors a student included in their response. According to the Krist heuristic, factors include 1) both concrete entities like amino acids or functional groups and abstract entities like information or forces, 2) system relationships like the binding of two entities, and 3) properties like charge or shape that influence behavior. Table 2 lists the factor codes used in this study. Throughout the manuscript, we have italicized factor codes to aid in clarity.

The coding scheme next determines if and how factors are linked to each other. If a factor is identified and coded, that factor is also given a linking tag that provides information on whether and how that factor was linked to other factors. Table 3 lists the linking tags. These tags indicate if that idea was used to compare proteins and if it was linked to other factors in the response. In our coding scheme, assigning an activity or a property to an entity results in the application of the "comparison with underlying cause" linking tag and reveals the presence of causal chains. We are enclosing linking tags in "quotation marks" for clarity. Figure 3 provides examples of linking tags applied to student drawings. Overall, this combinatorial approach allows detection of the many different ways students could link factors while still capturing the complexities present in student responses using a relatively small number of codes. Before analysis for the presence of mechanistic reasoning, the drawing and text factor codes for each explanation were combined to produce one final coding of the explanation. When the same factor code was applied to both the drawing and the text, but the linking tags differed, the highest-level linking tag was applied ("comparison with underlying cause" > "comparison but no underlying cause" > "present but no comparison").

Analysis for Mechanistic Reasoning. The coding described so far identifies factors and causal chains, but additional analysis is needed to determine if a response is a mechanistic explanation. To accomplish this, we mapped the three criteria from the Krist heuristic to the features of an ideal mechanistic explanation of why two proteins have different functions (Table 4). We then used formulas in Microsoft Excel to analyze the factors and linking present in a response to identify mechanistic reasoning (Figure 2B). Examples of coded responses illustrating the application of the coding scheme and analysis for mechanistic reasoning are presented in the results section.

To meet the first criterion for a mechanistic explanation, the response must consider entities at scalar levels below the phenomenon or use abstract factors. Three of our factor codes represent entities at lower scalar levels (*molecular structure*, *amino acids*, and *DNA*), and two factor codes represent abstract factors (*intermolecular forces* and *sequence*). The addition of a "comparison" linking tag ("comparison but no underlying cause" or "comparison with underlying cause") to these factors indicates that the factor was integrated into the explanation. Therefore, for a response to be considered a mechanistic explanation, it must include the factor code for *molecular structure*, *amino acids*, *DNA*, *intermolecular forces*, or *sequence* associated with one of the "comparison" linking tags.

To meet the second criterion for a mechanistic explanation, the response must unpack the properties and relationships of the factors. Unpacking is the intellectual work of characterizing the relevant properties, behaviors, and interactions of the factors (Krist *et al.*, 2019). An example of unpacking would be linking the structure of a functional group to a property such as polarity and then linking this with an activity such as interactions via intermolecular forces. The factors in our coding scheme include entities and properties that students link via activities. Thus, when students link factors, they are engaging in unpacking. To identify unpacking, our analysis uses linking tags to determine whether students link factors into causal chains. In our coding scheme, a causal chain is present when differences in one factor are caused by differences in a second factor. For example, the drawing could show that differences in *intermolecular forces* are caused by differences in *molecular structure*, or the text could describe differences in *intermolecular forces* caused by *molecular properties*. The linking tag "comparison with underlying cause" was only applied to a factor when it was linked to a second factor to which the "comparison but no underlying cause" or "comparison with underlying cause" linking tag was applied. This allowed the identification of factors that were connected to the differential function of the two proteins. Therefore, for a response to be considered a mechanistic explanation, it must include at least one factor other than *function* associated

TABLE 2. Factor codes representing the common physical entities and abstract concepts identified as conceptual resources in student responses

Factor code[a]	Description	Ideas from text	Ideas from drawing	Occurrences[b]	Factor cluster[c]
Function	Indicates what the protein is doing at the cellular or molecular level	binding, detecting, activating, moving, transporting, catalyzing	Depiction of arrows showing function such as transport through a protein across a membrane or a small molecule fitting into a binding site on the protein	**131** (100%)	*Function and Shape*
Shape	The shape of proteins	shape, size, tertiary structure	Depiction of protein as a shape (every drawing of a protein has a shape)	**130** (99%)	
Molecular structure	The specific arrangements of atoms	functional group, amino acid side chain, hydroxyl, methyl	Depictions of molecular structure with atoms linked together	**68** (52%)	*Chemical*
Molecular properties	A property associated with specific configurations of atoms	charge, polarity, hydrophobic, hydrophilic	Depictions of +, −, δ^+, δ^-	**42** (32%)	
Intermolecular forces	Interactions between molecules or within macromolecules	intermolecular forces, weak interactions, H-bonding, hydrophobic clustering	Depictions of + and − in close association, Depictions of functional groups that can H-bond connected by a dotted line, depicting methyl groups in close association	**40** (31%)	
Amino acids	The molecular building blocks used to build proteins	amino acid, aa	Symbols within a protein representing one or more building blocks from which the protein was made	**30** (23%)	*Biological Information*
DNA	The genetic information in the cell.	DNA, nucleic acids, genes, transcription	Depiction of a double helix	**21** (16%)	
Sequence	The order of monomers in a polymer	sequence, order of amino acids, primary structure	Depiction of amino acid symbols linked together	**31 (**24%)	
Generic characteristics	Vague/generic reference to a factor that could fall in one or more of the other categories, but not enough information is present to determine which category.	chemical makeup, molecular composition	This code was never applied to images.	**29** (22%)	*Generic*

[a]Factor codes are italicized for clarity.
[b]The total number and percentage of responses from all three tasks to which the factor code was applied ($N = 131$).
[c]Factor codes are grouped based on cluster analysis (refer to Supplemental Figure S1).

TABLE 3. Linking tags indicating if and how factor codes from Table 2 were linked in the response

Linking tag[a]	Operational definition	Example from text	Occurrences[b]
"Present but no comparison"	The factor was present in the response but there was no comparison of differences between the two proteins.	"Tertiary structure is the 3-D structure." (Response TV1_33)	80
"Comparison but no underlying cause"	Differences in this factor between the two proteins were indicated and used to compare the two proteins. However, the explanation gave no underlying causal linkage to another factor that explained why this factor differed between the two proteins.	"Transporters are very specific, only binding specific molecules due to their own unique binding site shape." (Response TV2_45)	247
"Comparison with underlying cause"	Differences in this factor between the two proteins were described and used to compare the two proteins. In addition, the explanation gave an underlying causal linkage to another factor that explained why this factor differed between the two proteins.	"The proteins function differently because of their different amino acids, causing it to have a different shape." (Response TV1_17)	195

[a]Linking tags are enclosed in "quotation marks" for clarity.
[b]Occurances indicate the total number of times each linking tag was associated with a factor.

with the "comparison with underlying cause" linking tag. The *function* factor was excluded because it is the phenomenon and therefore associated with the third criterion.

To meet the third criterion for a mechanistic explanation, the linked factors must be connected back to the phenomenon. In our coding scheme, this connection is represented by the *function* factor code associated with the "comparison with underlying cause" linking tag to indicate that other factors had been used to explain the differences in the functions of the proteins.

Example Application of Coding and Analysis. To illustrate how this coding scheme was applied to a student response, we will walk through coding and analysis using Figure 4 as an example. First, we identified the factors *function* and *shape* as present in the drawing, and the factors *function*, *molecular properties*, and a *generic* factor in the writing.

Next, we assigned a linking tag to each factor we identified. In the drawing, *function* was given the linking tag "comparison with underlying cause" because the proteins have shapes that are complementary to the sugars they transport, illustrating why the proteins transport different sugars. There is nothing in the drawing to illustrate why the proteins or sugars have differences in shape. However, the differences in shape are used to illustrate differences in function, so the linking tag "comparison but no underlying cause" and not the "present but no comparison" linking tag was applied to the *shape*.

In the writing, *function* was given the linking tag "comparison with underlying cause" because *molecular properties* and the *generic* factor of chemical makeup were used to explain why the transporters identify different molecules. There is nothing in the writing to explain what causes molecular properties or the chemical makeup. However, these factors are used to explain differences in function, so the linking tag

TABLE 4. Krist *et al.* (2019) heuristic criteria for mechanistic reasoning in the context of protein function

Three mechanistic criteria from Krist	Application to explain why different proteins have different functions.	Coding required to meet this criterion*
#1: Consider the level below the target phenomenon or abstract factors.	The phenomenon of protein function is at the macromolecular or cellular level. *Molecular structures* and *amino acids* are both entities at scalar levels below the phenomenon. Alternately, connecting to *DNA* in a comparison of protein function or using *sequence* involves the abstract factor of information. Inclusion of *intermolecular forces* involves the abstract factor of forces.	*Molecular structure*, *amino acid*, *DNA, sequence, or intermolecular forces* factor code with either the "comparison but no underlying cause" or "comparison with underlying cause" linking tag added
#2: Link factors into a causal chain (Unpacking).	Describing how one factor is causing differences in any of the factors other than function. This causal connection indicates students are considering behaviors or interactions of factors and meets the criterion for unpacking.	Any factor code other than *Function* that had a "comparison with underlying cause" linking tag added
#3: Connect how the lower-level interactions and behaviors give rise to the target phenomenon.	The phenomenon in this task is differences in protein function. Connecting function to another factor in a causal way meets this criterion.	*Function* factor code with a "comparison with underlying cause" linking tag added

[a]Refer to Table 2 for factor codes and Table 3 for linking tags.

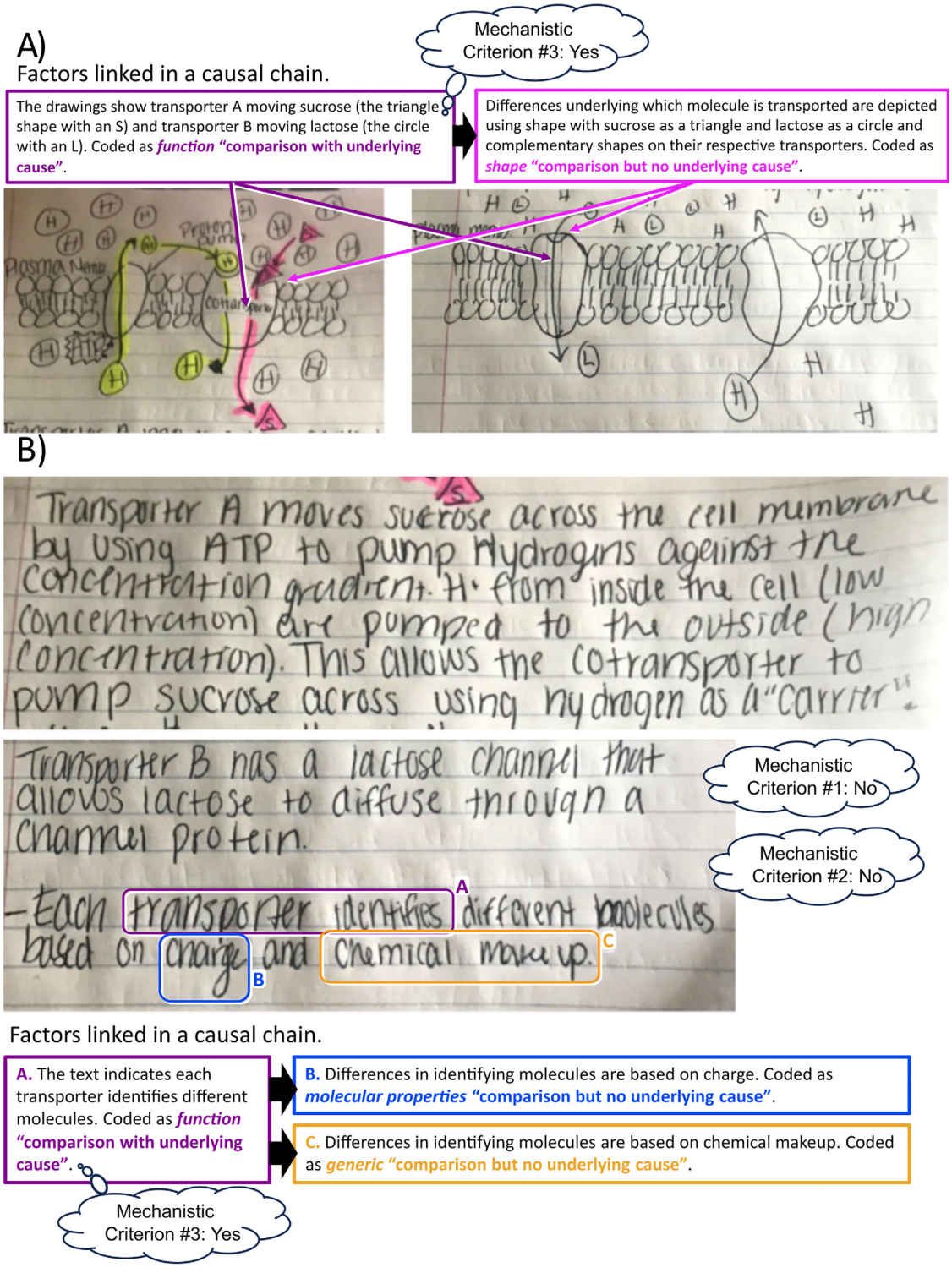


FIGURE 4. Coding of the drawings (A) and text (B) from Alex's response to Task 2 (response TV2_46 in Supplemental Material). Boxes with arrows for the drawings and lettered boxes for the writing show coding for factors and linking; cloud-shaped callouts show analysis for mechanistic criteria (refer to Figure 2 and Tables 2–4 for full details). This response links *shape* (drawing) and *molecular properties* and *generic* chemical makeup (writing) to differences in transport, which meets Criterion #3 (connection to function). However, because neither entities at a scalar level below the phenomenon nor abstract entities are identified, Criterion #1 (level below) is not met. In addition, no factors are linked to *shape, molecular properties,* or the *generic* chemical makeup, meaning Criterion #2 (linking factors) is not met. Therefore, this is not a mechanistic explanation. This response was placed in the *Chemical without IMFs Response Cluster* by hierarchical cluster analysis.

"comparison but no underlying cause" and not the "present but no comparison" linking tag was applied to both *molecular properties* and the *generic* factor.

We then combined the coding from the drawing and writing to get one set of codes from the entire response. Since *function* was present in both the drawing and writing, we compared the linking tags. Because they are the same, the combined coding is *function* "comparison with underlying cause." If the linking tags had been different, we would have used the higher-order tag. Thus, the final coding used for subsequent analysis was *function* "comparison with underlying cause", *shape* "comparison but no underlying cause", *molecular properties* "comparison but no underlying cause", and *generic* "comparison but no underlying cause."

To determine if the explanation was mechanistic, the combined codes were entered into Microsoft Excel and analyzed using a series of formulas to determine if the response met the three criteria needed to be classified as a mechanistic explanation (Table 4). First, the formula considered whether the response included a factor at a scalar level below the phenomenon or an abstract factor. Because it did not, it did not meet Criterion #1 (level below) and is not a mechanistic explanation. At this point in the analysis, the status of the other two criteria is a moot point, but we will consider them for clarity. The formula also considered whether a factor other than the function had the "comparison with underlying cause" tag. In this example, only *function* had this tag, so the response did not meet Criterion #2 (linking factors). Finally, the formula considered whether factors were linked to function. Because *function* has a "comparison with underlying cause" linking tag, the response does meet Criterion #3 (connection to function).

Study Participants

This study takes place in the context of a large-scale effort to incorporate evidence-based instruction into introductory STEM courses at a large public land-grant R1 university located in the Midwest (Cooper, Caballero *et al.*, 2015; Laverty *et al.*, 2016; Matz *et al.*, 2018; Bain *et al.*, 2020). The study analyzes explanations constructed in the spring of 2018 by students ($N = 131$) who had just completed a transformed introductory cell and molecular biology course that requires transformed chemistry as a pre- or corequisite. Both courses consistently incorporated 3DL instructional activities and assessments. In chemistry, students are frequently asked to use the core idea of atomic/molecular structure and properties and the cross-cutting concepts of structure and function and cause and effect to construct explanations of if, how, and why two molecules will interact (Cooper and Klymkowsky, 2013). In biology, students are frequently asked to use the core idea of information flow and the cross-cutting concepts of structure and function and cause and effect to construct explanations of if, how, and why a specific mutation would change the function of a protein. These ideas were a major component of the first unit in the course (refer to Supplemental Figures S1 and S2 for an activity focused on these ideas) and were revisited several times throughout the course, for example, as part of signal transduction and regulation of gene expression. Our results provide insights into whether and how students in these transformed courses link ideas from chemistry and biology and incorporate mechanistic reasoning when explaining protein function. Thus, this work provides insights that advance both our understanding of mechanistic reasoning and the need for a coordinated curriculum to support its development.

The institution's review board approved all research investigations (IRB# x17-793e). All responses were de-identified before coding, and names used in this manuscript are pseudonyms.

Generation of Student Explanations

Students enrolled in an introductory cell and molecular biology course were given the opportunity, during the week before final exams, to write an explanation of a phenomenon related to a topic they were told would appear on the final exam. Students were randomly divided into three groups using the group generation function in the learning management system, and each group was assigned one version of the task from Table 1. Students who completed the task earned extra credit that added 0.03% to their overall course grade. Following submission of their version of the task, students were given access to all three versions of the task as study tools for the final exam.

Students who completed different versions of the task did not differ in a statistically significant way based on incoming GPA (Kruskal-Wallis $\chi^2(2) = 0.210$, $p = 0.90$), course grade in the introductory cell and molecular biology course (Kruskal-Wallis $\chi^2(2) = 1.137$, $p = 0.57$), course grades for first semester of chemistry (Kruskal-Wallis $\chi^2(2) = 0.952$, $p = 0.62$), or institutional records of binary gender (Kruskal-Wallis $\chi^2(2) = 0.744$, $p = 0.69$).

Coding and Analysis

To analyze the student data, we performed a directed content analysis (Hsieh and Shannon, 2005). Coding was carried out using Dedoose (Dedoose, 2019). Initially, sets of three explanations, one from each prompt version, were coded independently by three members of the research team (JK, CT, and JRS), followed by meetings where we discussed differences in our coding and reached consensus. To measure reliability once our codebook was finalized, we calculated interrater reliability using percent agreement with and without tolerance, Cohen's κ, and Spearman's ρ for each code between the first (CT) and last (JRS) authors. These two authors completed the coding in sets of nine to 18 explanations containing a mixture of prompt versions. Consensus codes for all explanations were decided in weekly meetings in which we compared the coding results and, when necessary, discussed any disagreements to arrive at an agreed coding for each student's response.

All three versions of the task required students to both construct drawings and write an explanation. We independently applied the coding scheme to the drawing and the text. In cases where the text made clear reference to the drawing, this reference could be used to clarify and support coding decisions related to the drawing when ambiguity existed in the coding of the drawing. Refer to the Supplemental Material for the codebook and the Results section for coded examples of student responses. The drawing and text factor codes for each explanation were then combined to produce one final coding of the explanation. When the same factor code was applied to both the drawing and the text, but the linking tags differed,

the highest-level linking tag was applied. Thus, each explanation's final coding consisted of all the factors identified in both the image and the text, with one linking tag associated with each factor. As shown in Table 4, explanations that met three criteria related to mechanistic reasoning were considered mechanistic explanations. First, the explanation must include at least one entity at scalar levels below the phenomenon or one abstract factor. Second, it must include at least one additional factor with the "comparison with underlying cause" linking tag (refer to Table 3). Finally, it must include the *function* code with the "comparison with underlying cause" linking tag.

Our coding did not focus on absolute accuracy in students' understanding of factors. We disregarded errors in detail if the student incorporated factors in meaningful and appropriate ways. This is illustrated by the following examples: indicating that both the hydrophobic clustering and hydrogen bonding were on the left side in a drawing when the hydrophobic clustering was depicted on the left and the hydrogen bonding was depicted on the right of the drawing; saying that hydrophobic clustering is caused by interactions between the functional groups when this intermolecular force is really driven by entropy; drawing hydrogen bonding taking place between two hydrogen atoms rather than a hydrogen atom and an electronegative atom with an unshared electron pair. When students provided conflicting details, if one of the statements was correct, we disregarded the incorrect statement, for example, a student saying that proteins are made of DNA but also saying that DNA encodes proteins. Refer to figures with coded student responses in the *Results* section for in situ examples.

Interrater Reliability. Interrater reliability was at acceptable levels for our purposes. For coding of students' responses, the overall percent agreement between authors (CT and JS) was 83% without tolerance and 94% with a tolerance of one level difference in linking tags, and Cohen's κ was 0.71. The κ value and levels of agreement were considered suitable for our purpose (Gamer *et al.*, 2012; McHugh, 2012). After initial coding and calculation of interrater reliability, all responses were discussed until consensus was reached. The analysis presented below reports the consensus coding for the combined codes from the textual and image analyses.

Cluster Analysis. To analyze the coded explanations for grouping among the factors, we exported our coded dataset and conducted a two-dimensional hierarchical cluster analysis in RStudio (Galili, 2015; Galili *et al.*, 2018; RStudio Team, 2020). Factors were recoded to values as follows: "absent" = 0; "present but no comparison" = 1; "comparison but no underlying cause" = 2; "comparison with underlying cause" = 3. We used the Euclidean distance measure to evaluate dissimilarity between students and between codes, then used Ward's minimum variance method to create hierarchical clusters (Murtagh and Legendre, 2014). Ward's approach was appropriate to examine how students' ideas grouped together rather than to examine the most similar or distant relationships using single or complete linkage methods, respectively. These methods align with validation studies of the appropriate use of cluster analysis for ordinal data (Žiberna *et al.*, 2004) and can be displayed as a heatmap. Visualization was carried out using heatmap visualization packages in R Studio (Galili, 2015; Galili *et al.*, 2018). The two-dimensional clustering approach resulted in two types of clusters. Factor clusters (dimension 1) show factors that occur together across responses. Response clusters (dimension 2) show student responses that include similar factors.

Statistical Analyses. Statistical analyses were performed using RStudio (RStudio Team, 2020). We used the binomial test (Wagner-Menghin, 2005) to determine if explanations that included two or more factors differed significantly from a random distribution between explanations that contained two or more factors from only one discipline versus explanations that contained factors from both biology and chemistry.

To test the null hypotheses that: 1) mechanistic explanations were independent of task version, 2) explanation clusters were independent of task version, and 3) mechanistic explanations were independent of clusters, we performed Chi-squared tests of independence including calculating the adjusted Cramér's *V*. While 28% of expected values used in (2) and 17% of expected values used in (3) were less than 5, each table used in Chi-squared calculations had a total count greater than or equal to five times the total number of cells indicating the Chi-squared test results are reliable (Alberti, 2024). To determine which associations were driving the significance in the Chi-squared test, we compared the adjusted Pearson residuals for each cell to the Bonferroni-corrected critical value for significance at the $\alpha = 0.05$ level (Naioti and Mudrak, 2022).

Post hoc Analysis of Additional Resources As coding progressed, we noticed less relevant but recurring ideas that had not emerged as part of the initial codebook development. One set of ideas was related to concentration gradients, primary and secondary transport, and passive and active transport. Another was related to cellular signal transduction pathways. While these ideas were not captured in the coding scheme, a *post hoc* analysis was done to identify responses that included these ideas.

RESULTS

In this section, we answer our research questions using an analytical coding scheme to identify the conceptual and epistemic resources students used in their explanations. The scheme follows a combinatorial structure in which coding for factors (e.g., *amino acids* and *DNA*) is paired with linking tags (e.g., "comparison with underlying cause") to reflect if and how that factor was linked in the explanation. We share the results for research question 1 (RQ1) using our coding of factors to understand conceptual resource use, analysis of the combinatorial coding to determine if students were using mechanistic reasoning, and cluster analysis to determine ways students link resources and construct mechanistic explanations. For research question 2 (RQ2), we explore how the task version influenced the resources students used in their explanations.

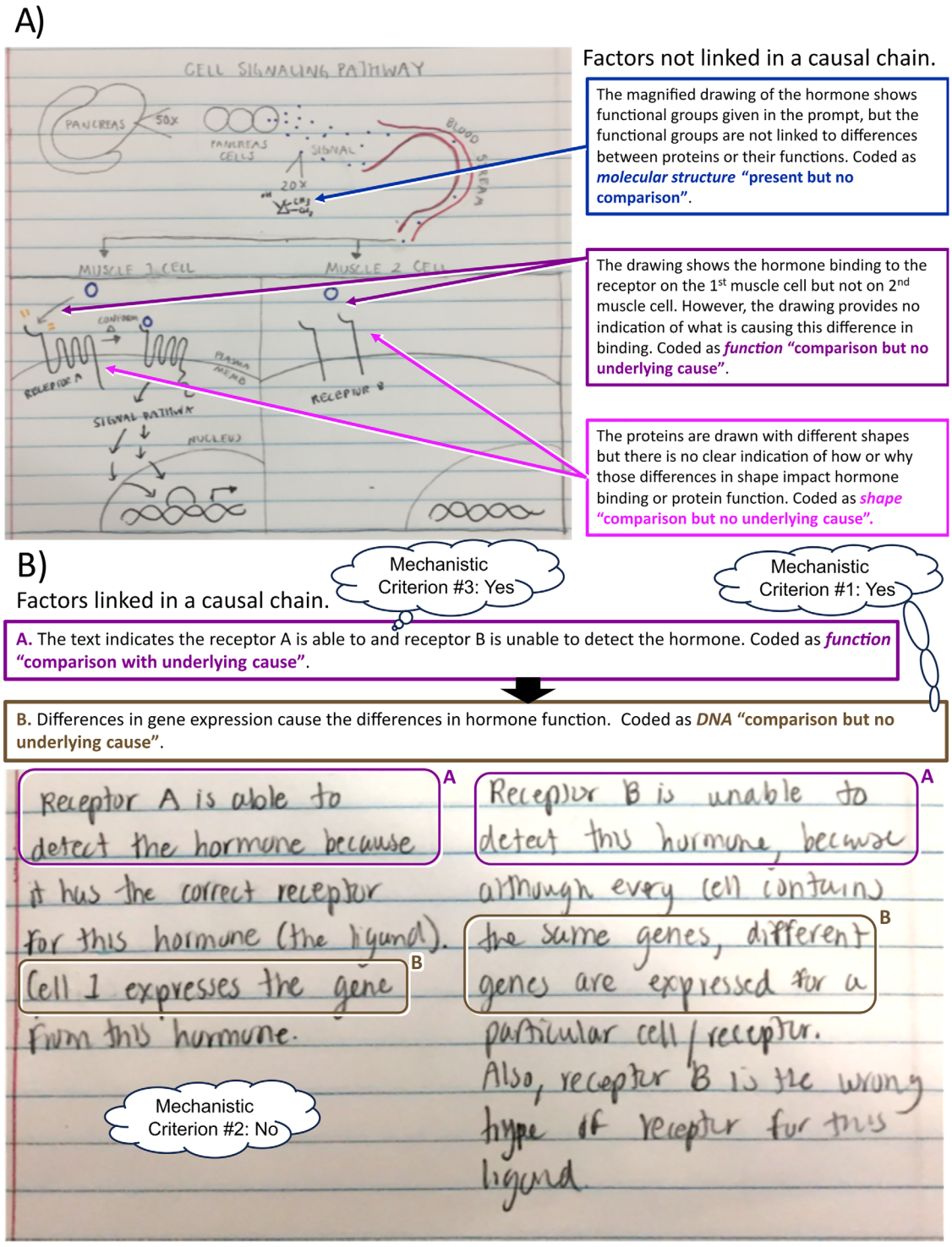


FIGURE 5. Coding of the drawing (A) and text (B) from Tracy's response to Task 3 (response TV3_131 in Supplemental Material). Boxes with arrows for the drawings and lettered boxes for the writing show coding for factors and linking; cloud-shaped callouts show analysis for mechanistic criteria (refer to Figure 2 and Tables 2–4 for full details). This written response identifies *DNA,* which is an entity at a scalar level below the phenomena and meets Criterion #1 (scale below). It links *DNA* to differences in hormone detection to meet Criterion #3 (connection to function). However, the response does not link to any other factors and therefore does not meet Criterion #2 (linking factors) and is not a mechanistic explanation. The explanation contains one factor associated with biological information and one factor associated with chemistry, but these factors are not linked. This explanation was placed in the *Shape Response Cluster* by hierarchical cluster analysis. Note: as described in the *Methods* section, both the image and text were coded as *function* but were given different linking tags. The overall coding for the explanation was *function* "comparison with underlying cause" as this represents the higher-order linking tag.

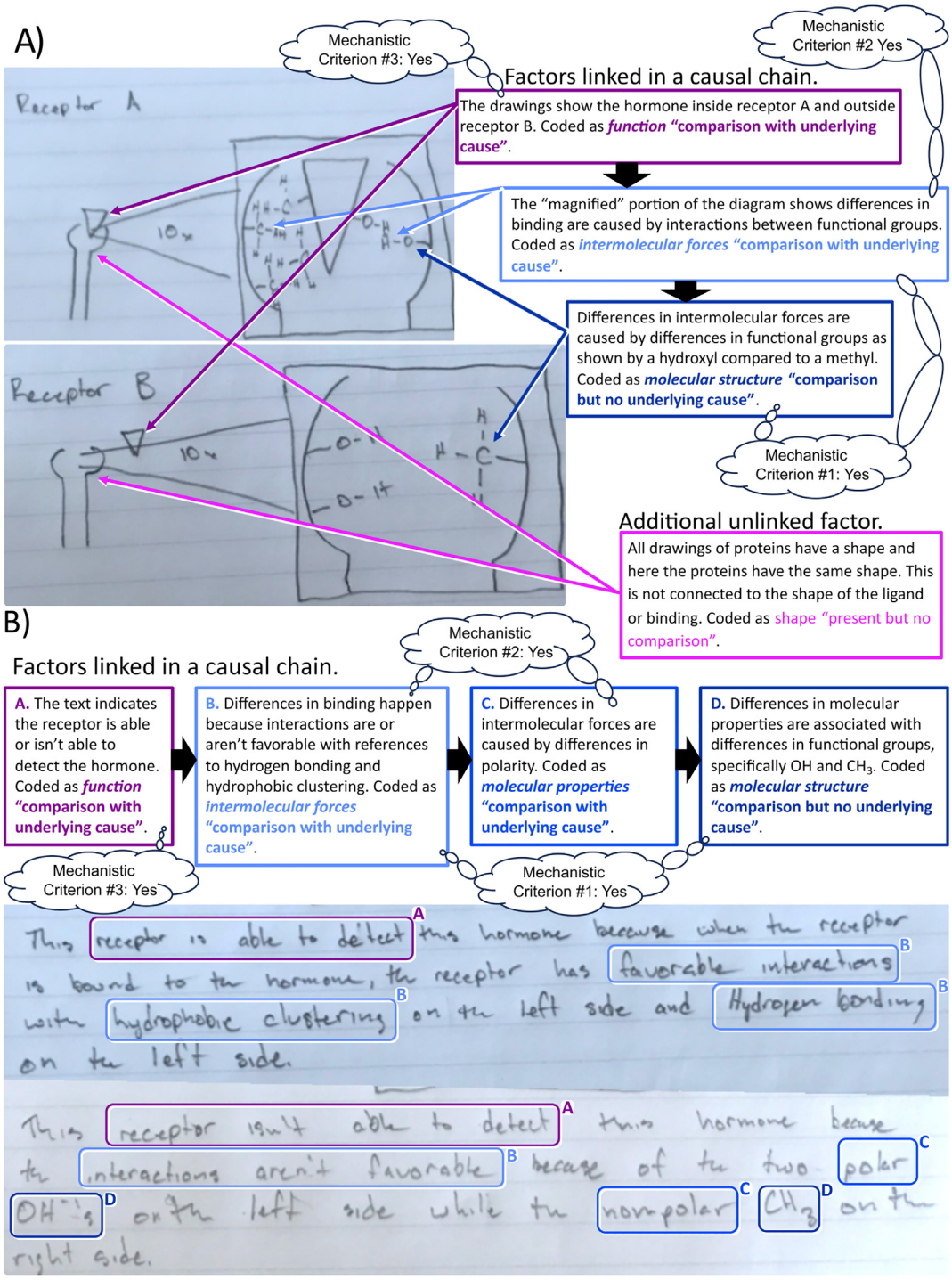


FIGURE 6. Coding of the drawing (A) and text (B) from Skyler's response to Task 3 (response TV3_118 in Supplemental Material). Boxes with arrows for the drawings and lettered boxes for the writing show coding for factors and linking; cloud-shaped callouts show analysis for mechanistic criteria (refer to Figure 2 and Tables 2–4 for full details). This response meets Criterion #1 (level below) by identifying specific *molecular structures* and *intermolecular forces*. It links the structures and forces to *molecular properties,* which meets Criterion #2 (linking factors). The response uses this causal chain to explain how and why the hormone binds to one receptor and not the other. This meets Criterion #3 (connection to function) and makes this a mechanistic explanation. The response relies entirely on ideas from chemistry and was placed in the *Chemical with IMFs Response Cluster* by hierarchical cluster analysis. Note: as described in the *Methods* section, we overlooked errors such as saying both the hydrophobic clustering and hydrogen bonding are on the left side instead of one on the left and the other on the right side, saying hydrophobic clustering is caused by interactions between the functional groups while it is actually driven by entropy, and showing hydrogen bonding taking place between two hydrogen atoms.

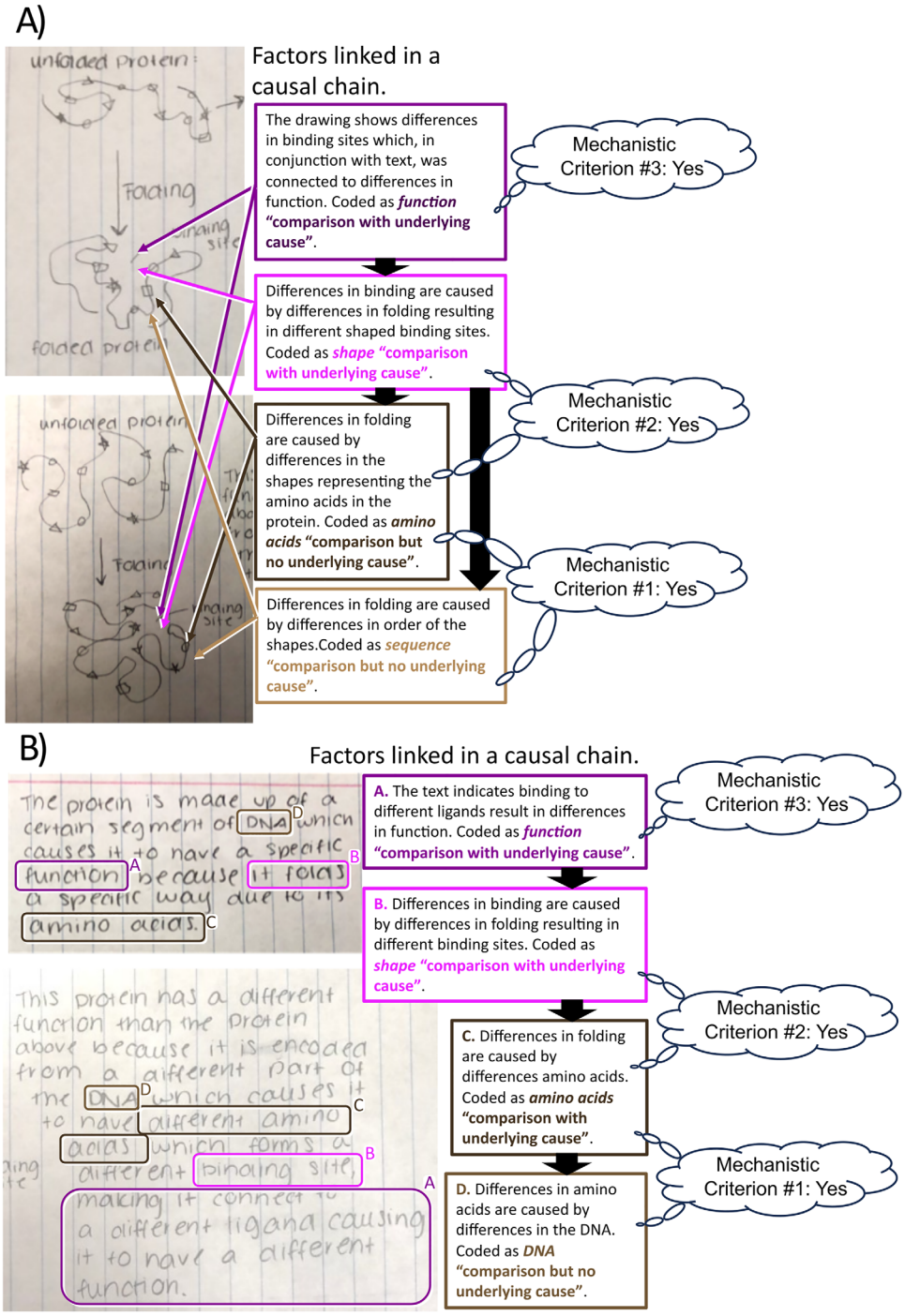


FIGURE 7. Coding of the drawing (A) and text (B) from Finley's response to Task 1 (response TV1_17 in Supplemental Material). Boxes with arrows for the drawings and lettered boxes for the writing show coding for factors and linking; cloud-shaped callouts show analysis for mechanistic criteria (refer to Figure 2 and Tables 2–4 for full details). The response meets Criterion #1 (level below) by identifying *amino acids*, *DNA*, and *sequence*. It links factors by indicating that the different *amino acids* in different proteins are caused by differences in the information encoded in the *DNA sequence,* which meets Criterion #2 (linking factors). This is connected to differences in protein shape and function, which meet Criterion #3 (connection to function). Since it meets all three mechanistic criteria, this is a mechanistic explanation. The response relies entirely on ideas from biology and was placed in the *Biological Information Response Cluster* by hierarchical cluster analysis. Note: as described in the *Methods* section, we overlooked errors such as saying the protein is made up of DNA when other parts of the explanation correctly explained this idea by saying proteins are encoded by DNA.

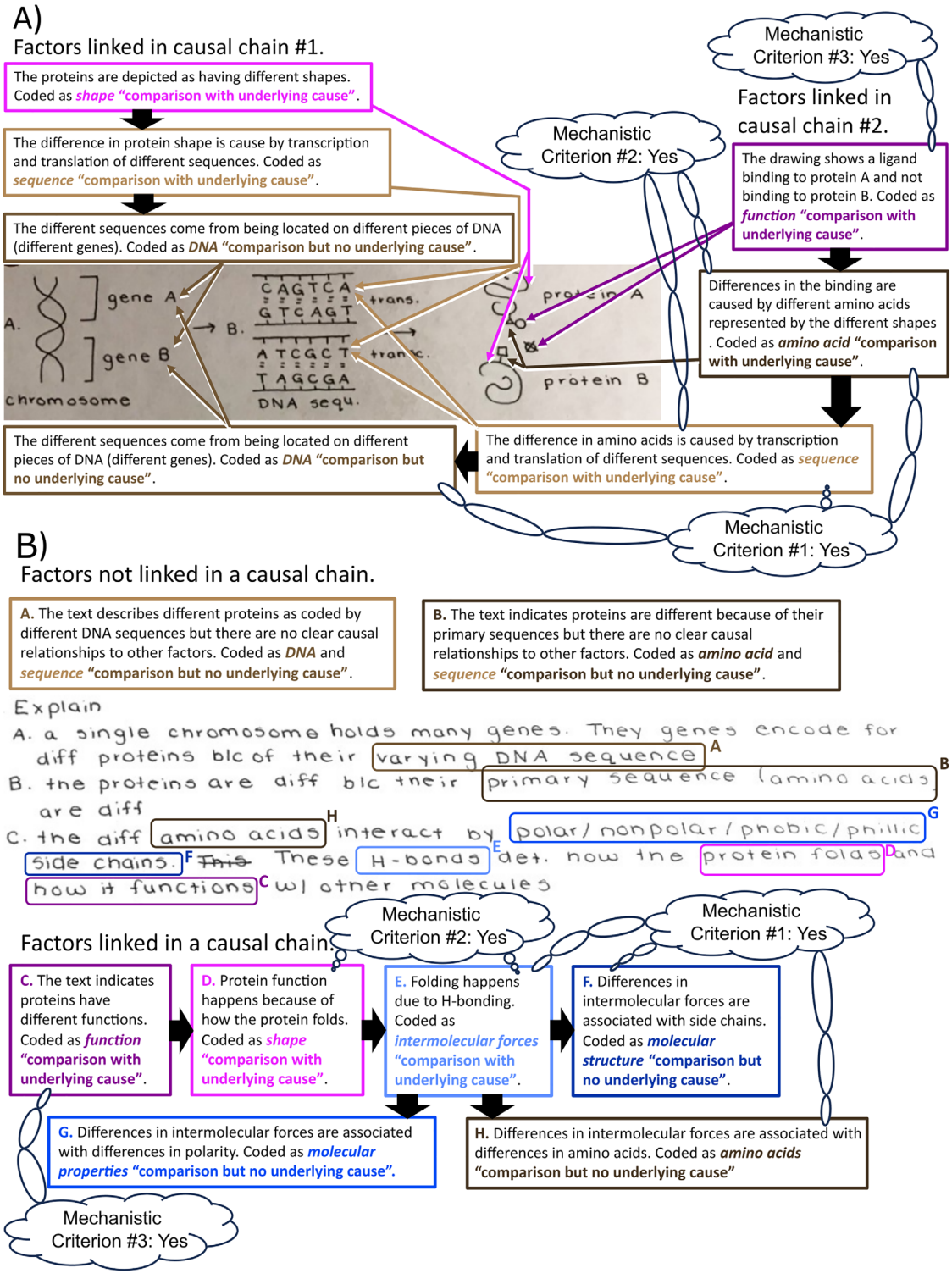


FIGURE 8. Coding of the drawing (A) and text (B) from Emerson's response to Task 1 (response TV1_30 in Supplemental Material). Boxes with arrows for the drawings and lettered boxes for the writing show coding for factors and linking; cloud-shaped callouts show analysis for mechanistic criteria (refer to Figure 2 and Tables 2–4 for full details). This response meets Criterion #1 (level below) by identifying *molecular structure*, *DNA*, and *amino acids*, as well as *intermolecular forces* and *sequence*. These factors are linked in ways that meet Criterion #2 (linking factors) and are used to explain how and why this causes different proteins to have different functions, meeting Criterion #3 (connection to function), making this a mechanistic explanation. The explanation includes ideas about both biological information and chemistry, and was placed in the *Chemical and Biological Response Cluster* by hierarchical cluster analysis.

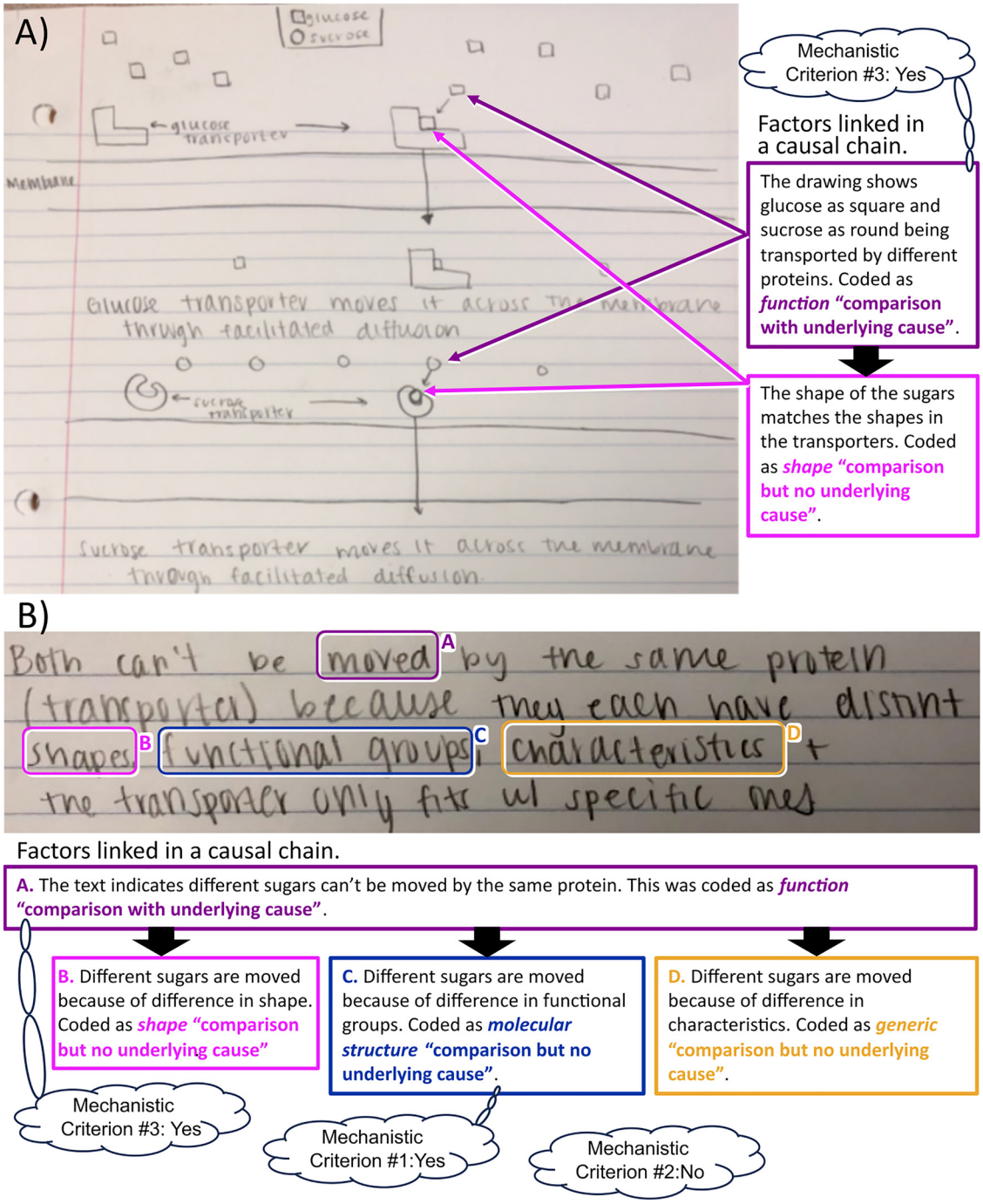


FIGURE 9. Coding of the drawing (A) and text (B) from Jessie's response to Task 2 (response TV2_57 in Supplemental Material). Boxes with arrows for the drawings and lettered boxes for the writing show coding for factors and linking; cloud-shaped callouts show analysis for mechanistic criteria (refer to Figure 2 and Tables 2–4 for full details). This response meets Criterion #1 (level below) by identifying *molecular structure*. It indicates that the proteins function differently because of differences in *shape*, *molecular structure*, and *generic* characteristics, satisfying Criterion #3 (connection to function). However, because no factors are linked to differences in *shape*, *molecular structure*, or *generic* characteristics, the response does not meet Criterion #2 (linking factors). Therefore, this is not a mechanistic explanation. This response was placed in the *Generic Response Cluster* by hierarchical cluster analysis. Note: as described in the *Methods* section, we overlooked errors such as using glucose rather than lactose, the sugar asked for in the task.

RQ1: What Resources do Students Draw upon when Explaining Protein Function, and how are these Resources Linked within their Explanations?

Resource use in Explanations of Protein Function

Conceptual Resources. The factors in our coding scheme are conceptual resources. Figures 4–9 provide examples of how student responses were coded. Our coding for factors revealed that all students included the function factor in their responses, and all but one student included the shape factor in their responses, typically by drawing proteins with shapes. Students included other factors at rates ranging from 16% to 52%. Table 2 shows how frequently students included these factors in their explanations.

Post hoc analysis also identified explanations that depicted or discussed transport energetics (n = 14, 11%) or signal transduction (n = 7, 5%), factors not included in the original coding scheme. Factors included in discussions of transport energetics were concentration gradients, primary/secondary transport, and passive/active transport. Alex's response (Figure 4) provides an example of a response that included factors related to transport energetics. Tracy's response (Figure 5) provides an example of a response that included factors related to signal transduction.

Epistemic Resources. Our analysis of 131 responses across the three tasks revealed that 47 students (36%) created mechanistic explanations by including relevant factors at scalar levels below the phenomenon or relevant abstract factors, linking these factors, and then connecting these factors to protein function (Table 4). These students used a variety of relevant factors and made different types of connections between the factors. For example, Skyler relied on conceptual resources from chemistry, including an entity at a scalar level below the phenomenon (*molecular structure*), and used an epistemological frame in which they linked the factors into a mechanistic explanation (Figure 6). Findley (Figure 7) provides an example of a mechanistic explanation based solely on biological resources, while Emerson's mechanistic explanation (Figure 8) draws on both biology and chemistry.

Students who did not construct mechanistic explanations often included elements of a mechanism. For example, Tracy's response (Figure 5) included conceptual resources from both chemistry and biology, including *molecular structure* and *DNA*, both factors at a scalar level below the phenomenon, but did not link the factors in meaningful ways. Therefore, Tracy did not construct a mechanistic explanation. Figures 4 and 9 provide additional examples of responses that are not mechanistic explanations. Of the responses that did not meet all three criteria for a mechanistic explanation, 27 identified a factor at scalar levels below the phenomenon and 12 included an abstract factor (Criteroin #1), one linked at least two factors into a causal chain (Criterion #2), and 60 connected a factor to protein function (Criterion #3). Overall, only 21 students (16%) did not include any of the elements of a mechanistic explanation in their response. In total, 83 (63%) students did not link factors into causal chains, making this the most underutilized epistemic resource, with 28 (21%) students missing only this element of mechanistic reasoning. The only student who linked factors in a causal chain but did not construct a mechanistic explanation did not identify entities at a level below the phenomenon or include an abstract entity (refer to TV2_82 in the Supplemental Material).

How Students Link Resources when Explaining Protein Function. To better understand the patterns of factors and linking found in students' explanations, we carried out hierarchical cluster analysis in two dimensions. The first dimension in the cluster analysis grouped factors by their tendency to occur together across responses (*Factor Clusters)*, and the second dimension grouped responses by patterns of factors found in the responses (*Response Clusters*). Refer to Supplemental Figure S3 for a heatmap of the cluster analysis.

Hierarchical Clustering Dimension #1: When Students use Multiple Factors, the Factors are Commonly from the same Discipline. The hierarchical clustering of factors based on their tendency to occur together across responses identified four meaningful factor clusters, which we named based on the types of factors found in each cluster (refer to Table 2 and Supplemental Figure S3). The first factor cluster includes the *function* and *shape* factors that were present in essentially every response and was named the Function-shape Factor Cluster. The second factor cluster contains the *molecular structure*, *molecular properties*, and *intermolecular forces* factors, and was named the Chemical Factor Cluster as these factors are typically associated with instruction in chemistry courses. The third factor cluster contains the *amino acids*, *sequence*, and *DNA* factors and was named the Biological Information Factor Cluster, as these factors are typically associated with instruction in biology courses. The final factor cluster included only the *generic* characteristics factor and was named the Generic Factor Cluster.

The fact that chemistry factors clustered together and biology factors clustered together indicates that when students include one factor from a particular discipline, for example, chemistry, the student is more likely to include a second factor from that same discipline than to include a factor from the other discipline. A binomial test, which checks for statistically significant deviation from random selection between binary outcomes (Wagner-Menghin, 2005), showed that students who used two or more factors are more likely to use factors from the same discipline than from different disciplines ($p = 0.04296$; probability of using factors from the same category is 0.635). Overall, 28 students did not include any chemistry or biology factors, 40 students included only one chemistry factor or one biology factor, 40 students included two or more factors from either chemistry or biology, and 23 students included two or more factors with factors from both chemistry and biology. Of the students in the latter two categories, 68% constructed mechanistic explanations.

Hierarchical Clustering Dimension #2: Students Include Conceptual Resources from Chemistry or Biology, but Rarely Both. The hierarchical clustering of student responses based on the patterns of factors included within their responses identified six meaningful clusters which we named by resources commonly used in those responses: (1) the *Chemical and Biological Response Cluster*, (2) the *Chemical with IMFs Response Cluster*, (3) the *Biological Information Response Cluster*, (4) the *Chemical*

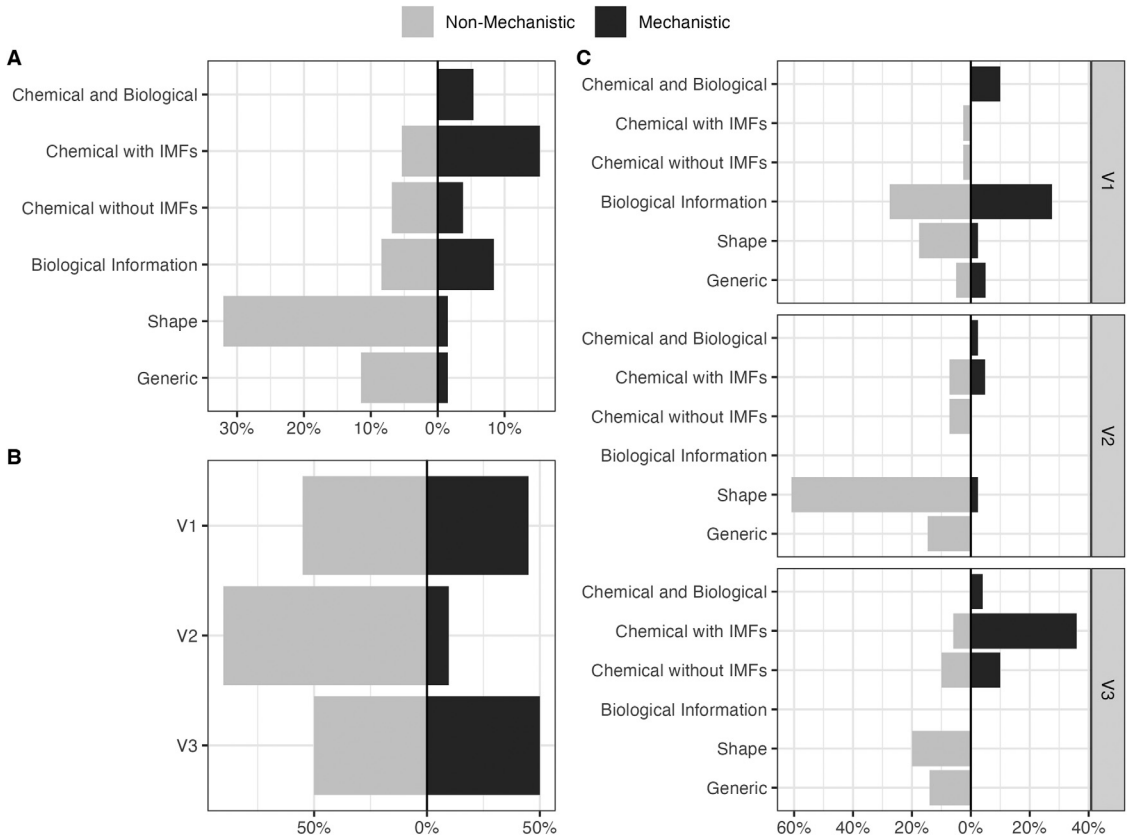


FIGURE 10. Percentages of explanations classified as mechanistic or non-mechanistic. Gray bars to the left of the center line (0%) indicate explanations that were not coded as mechanistic, and black bars to the right indicate explanations that were coded as mechanistic. Panel A shows the percentage of all responses ($N = 131$) that were nonmechanistic and mechanistic, divided into the six *Response Clusters* identified by hierarchical cluster analysis. Panel B shows the percentage of responses within each task that were nonmechanistic and mechanistic (V1 $n = 40$, V2 $n = 41$, V3 $n = 50$). Panel C shows the percentage of responses from each task that were nonmechanistic and mechanistic, broken down by *Response Cluster*.

without IMFs Response Cluster, (5) the *Shape Response Cluster*, and (6) the *Generic Response Cluster* (refer to Supplemental Figure S3). The percentage of responses placed in each cluster, along with their status as mechanistic or not mechanistic, is shown in Figure 10A and Supplemental Table S1. While the specific factors and their linking varied somewhat for responses within a cluster, the following patterns emerged.

The *Chemical and Biological Response Cluster* was the smallest cluster and contained only 5% of responses. This cluster included responses that incorporated multiple factors from both chemistry and biology, with meaningful linking occurring for factors from both disciplines. Emerson's response (Figure 8) illustrates the type of response seen in this cluster.

The *Chemical with IMFs Response Cluster* contained 21% of responses and included those that incorporated intermolecular forces and at least one other factor from chemistry, frequently linking these in meaningful ways. Most of these responses did not include any factors from biology. When they did, these factors were typically coded as "present but no comparison" or "comparison but no underlying cause" and typically were not linked into causal chains. Skyler's response (Figure 6) illustrates the type of responses found in this cluster.

The *Biological Information Response Cluster* contained 17% of responses and included those that incorporated at least two factors from biology, frequently linking these in meaningful ways. Some responses in this cluster included factors from chemistry, but rarely linked these factors. Finley's response (Figure 7) illustrates the type of response found in this cluster.

The *Chemical without IMFs Response Cluster* contained 11% of responses and included those that incorporated molecular properties with a linking tag of either "comparison but no underlying cause" or "comparison with underlying cause" present (Table 3), but did not include specific references to intermolecular forces or biological factors. Most responses in this cluster also incorporated *molecular structure* or a *generic* factor. Alex's response (Figure 4) illustrates the type of response found in this cluster.

The final two response clusters, the *Generic Response Cluster* and the *Shape Response Cluster*, contained 13% and 34% of responses and included responses that were largely devoid of chemical or biological factor codes. When biology or chemistry factors were included in the response, they were typically coded as "present but no comparison" or "comparison but no underlying cause" and typically were not linked together into causal chains. Jessie's response (Figure 9) and Tracy's

response (Figure 5) provide examples of responses from the *Generic Response Cluster* and the *Shape Response Cluster*, respectively.

Factors Associated with Mechanistic Explanations. The proportion of responses that were mechanistic explanations varied across the response clusters (Figure 10A and Supplemental Table S1). The *Chemical and Biological Response Cluster*, the *Chemical with IMFs Response Cluster*, and the *Biological Information Response Cluster* had 100%, 74%, and 50% mechanistic explanations, while the *Chemical without IMFs Response Cluster*, the *Generic Response Cluster*, and the *Shape Response Cluster* had 36%, 12%, and 5% mechanistic explanations. A Chi-squared test of independence of the response clusters and mechanistic explanations supports the observations that constructing mechanistic explanations appears to be dependent on what factors were included in the explanation ($\chi^2(5) = 54.613$, $p = < 0.001$, Cramér's $V = 0.62$). Analysis of the adjusted Pearson residuals using the Bonferroni-corrected critical value for significance at the $\alpha = 0.05$ level indicates that there is an overrepresentation of mechanistic explanations in both the *Chemical and Biological Response Cluster* and the *Chemical with IMFs Response Cluster*, as well as an underrepresentation of mechanistic explanations in the *Shape Response Cluster*.

Non-Productive Conceptual Resources. Post hoc analysis found responses that depicted or discussed transport energetics or signal transduction, ideas not captured in the original coding scheme. Of the *14* responses that included ideas related to transport energetics, nine of these responses were clustered in the *shape* cluster, three in the *generic* cluster, and two in the *chemical without IMFs* cluster. Of the seven responses that included ideas related to cellular signal transduction pathways, four were clustered in the *biological information* cluster, two in the *shape* cluster, and one in the *chemical without IMFs* cluster. Because conceptual resources related to transport energetics and signaling cannot be used to explain how and why proteins bind to different ligands, these factors cannot be used to construct a mechanistic explanation of this phenomenon. Not surprisingly, none of the responses that included transport energetics factors and only one of the responses that included signal transduction factors included and linked other relevant factors from the coding scheme into a mechanistic explanation of how and why proteins bind different ligands.

RQ2: How do Details of the Task Influence the Types of Conceptual Resources Students Include and the Likelihood of Producing a Mechanistic Explanation?

Impact of Task Version on use of Conceptual Resources. Different task versions elicited conceptual resources at different frequencies. As seen in Figure 10C, the responses that rely on ideas from biology (*Chemical and Biological Response Cluster* and *Biological Information Response Cluster*) were predominantly from Task 1, while responses that rely heavily on ideas from chemistry (*Chemistry with IMFs Response Cluster* and *Chemistry without IMFs Response Cluster*) were predominantly from Task 3. While all three tasks were well represented in the *shape* cluster, Task 2 was overrepresented. A Chi-squared test of independence revealed a significant association between response clusters and prompt version ($\chi^2(10) = 94.572$, $p < 0.001$, Cramér's $V = 0.57$), indicating that the distribution of ideas was not independent of the prompt given. Analysis of the adjusted Pearson residuals using the Bonferroni-corrected critical value for significance at the $\alpha = 0.05$ level confirmed that the patterns observed in the graph were significant. There was an underrepresentation of explanations from Task 1 and an overrepresentation of explanations from Task 3 in the *Chemical with IMF Response Cluster*, an underrepresentation of explanations from Task 2 and Task 3 and an overrepresentation of explanations from Task 1 in the *Biological Information Response Cluster*, and an overrepresentation of explanations from Task 2 in the *Shape Response Cluster*.

Post-hoc coding found responses that depicted or discussed transport energetics or signal transduction, ideas not captured in the original coding scheme but that may provide meaningful insights into resource activation by the tasks. All 14 responses that included ideas related to transport energetics were developed by students given Task 2. Of the seven responses that included ideas related to cellular signal transduction pathways, five were from Task 1, and two were from Task 3.

Impact of Task Version on Construction of Mechanistic Explanations. The linking of resources into mechanistic explanations occurred at different frequencies when students were given different task versions. As seen in Table 1 and Figure 10B, 45% and 50% of responses from Task 1 and Task 3, respectively, were mechanistic explanations, while only 10% of responses from Task 2 were mechanistic explanations. Development of mechanistic explanations was significantly associated with task version ($\chi^2(2) = 17.924$, $p < 0.001$, Cramér's $V = 0.35$). *Post-hoc* analysis of adjusted Pearson residuals (Bonferroni corrected at $\alpha = 0.05$ level) indicated that there was an overrepresentation of non-mechanistic explanations in Task 2 and an overrepresentation of mechanistic explanations in Task 3.

Limitations

Student responses in this study were collected in a course transformed using the 3DL framework (National Research Council, 2012), and it is probable that responses from students learning biology in different contexts would differ from those in this study. In addition, responses were only collected during one semester, with fewer than 50 student responses analyzed for each version of the task, so, likely, this study does not capture the breadth of resources activated by these prompts. However, despite these limitations, we believe this report fulfills its intended purpose of illustrating common types of student responses, the impact of task details on the responses, and providing insights into students' preparation for future learning.

Inferring framing using the presence of epistemic resources is just that, an inference. It is possible that students are including these resources while mimicking instruction without actually framing the activity as developing a mechanistic explanation. Disentangling this would require additional data, such as surveys or interviews, which are not available in this study. Regardless, the presence of epistemic resources in student responses provides insights into resources students might use

when they encounter similar phenomena in similar contexts in subsequent courses and their careers.

DISCUSSION

Overview

From a resources perspective, instruction should repeatedly activate both conceptual and epistemic resources in productive ways across different contexts and courses. This will facilitate stabilization of productive resource configurations that, over time, will allow students to become more expert-like in how they use their disciplinary knowledge to address complex problems and, ideally, prepare students for future learning in subsequent courses and their careers (Bransford and Schwartz, 1999; Hammer *et al.*, 2005; Houchlei *et al.*, 2021; Edelsbrunner *et al.*, 2024). In this model of curricula, assessment data that show how tasks activate resources following instruction in one course can and should inform instructional design in subsequent courses. Here, we explore resource activation in students who have completed a transformed biology course that requires a transformed chemistry course as a corequisite. These courses focus on mechanistic reasoning and are prerequisites for most upper-level biology courses in life-science curricula at our university.

We identified the three epistemic resources from Krist's heuristic for mechanistic explanations (Table 4) in 45% of responses to Task 1 and 50% of responses to Task 3, but only 10% of responses to Task 2. This provides evidence that many students framed the task as creating a mechanistic explanation, understood that these tasks required mechanistic reasoning, and had the epistemic and conceptual resources needed to meet our expectations for a mechanistic explanation (Table 1 and Figure 10B). The results from Tasks 1 and 3 align with the approximately 45%–60% of students who constructed mechanistic explanations in chemistry courses that focus on causal mechanistic reasoning (Crandell *et al.*, 2019, 2020; Ralph *et al.*, 2022). It is also greater than the 10%–40% of students who constructed mechanistic explanations in other studies (Becker *et al.*, 2016; Southard et al., 2016, 2017; Scott *et al.*, 2018; Macrie-Shuck and Talanquer, 2020; Flowers, Holder, and Gardner, 2023; Flowers, Holder, Rump *et al.*, 2023) and is consistent with findings showing that instruction focused on developing students' mechanistic reasoning increases the proportion of students who construct them (Crandell et al., 2019, 2020; Houchlei *et al.*, 2021; Ralph *et al.*, 2022; Scharlott *et al.*, 2024). While the phenomena students explained, the methods used to identify mechanistic reasoning, and the specific student populations varied widely across these studies, several themes emerge regarding challenges students face in constructing mechanistic explanations.

Common Barriers to Mechanistic Reasoning

Students in our study struggled with the same barriers to mechanistic reasoning identified in a recent review: using an appropriate number of entities and assigning activities to entities (Bachtiar *et al.*, 2022). In addition, Russ *et al.* (2008) asserted that chaining, an important aspect of unpacking, is the most difficult and sophisticated of the nine aspects of mechanistic reasoning they identified. Our findings on students' explanations of protein function support this. Across all three prompts in our study, 63% of students did not link factors into causal chains that explain protein function, with 21% of students missing only this element. Refer to responses from Alex (Figure 4), Tracy (Figure 5), and Jessie (Figure 9) for examples of explanations that lacked linking of factors (Criterion #2) related to protein function. Only one response that included linked factors was not fully mechanistic, as it included neither an entity at scalar levels below the phenomenon nor an abstract entity (refer to TV2_82 in Supplemental Material). Here, we summarize three major areas where students in our study need additional support in constructing mechanistic explanations.

Identifying Sufficient Relevant Factors. The first area where support is needed is in identifying multiple relevant factors. Mechanistic reasoning requires connecting multiple factors. Without the relevant conceptual resources, it is impossible to construct satisfactory mechanistic explanations because the epistemic resource of linking factors relies on the presence of conceptual resources. In our study, students in the *Shape Response Cluster* typically included only one additional relevant factor beyond shape and rarely constructed mechanistic explanations (Figure 10A; Supplemental Figure S3). On the other hand, when students included ideas from both chemistry and biology (*Chemical and Biological Response Cluster*) or intermolecular forces (*Chemical with IMF Response Cluster*), they were more likely to construct mechanistic explanations (Figures 6, 8, and 10A). In fact, 68% of students who included two or more ideas from the *Chemical Factor Cluster* and/or the *Biological Factor Cluster* constructed mechanistic explanations. Thus, most students in the shape cluster would benefit from support in identifying additional factors.

Beyond just the number of factors, support in this area could also help students who did not identify factors across scales or abstract factors. This was the third barrier to mechanistic reasoning identified by Bachtiar *et al.* (2022). For example, Alex's response (Figure 4) would benefit from additional specificity regarding "chemical makeup," which could lead to an explanation that moves across scales.

It is also important that students choose factors that are relevant to the mechanism underlying the phenomenon being explained. An important tenet of the resources framework is identifying resources students bring when approaching a problem and helping them build on those resources. This includes helping students identify which of the resources activated by a prompt are relevant to the problem at hand. Work in organic chemistry found that some students identify features of molecules that could not be used to judge the plausibility of a reaction mechanism (Bodé *et al.*, 2019). We also found examples where students chose factors that were not relevant. For example, Alex's response (Figure 4) spent significant time discussing transport energetics and identified factors relevant for explaining the direction in which solutes move during transport. However, these factors cannot be used to explain differential protein function. Thus, Alex needed additional support in recognizing relevant factors.

Responses that lack relevant resources also impact our recognition of mechanistic reasoning. Because our coding scheme relied on the incorporation of relevant factors to identify mechanistic explanations, it is possible that some students

framed the task as requiring mechanistic reasoning even though this was not identified in their response. For example, response TV2_73 (Supplemental Material) included, "The transporters are able to move different sugars due to the active and passive transport functions they can do." The student goes on to create a causal chain that includes conformational changes and concentration gradients, important factors in transport. This response does try to explain differences in protein function in a causal manner, and the student may have been working in a mechanistic frame. However, because the student was working with conceptual resources from transport energetics rather than structure–function relationships, the response was not identified as a mechanistic explanation in our coding scheme. From a resources perspective, it is important that instructors recognize the valuable epistemic resources a student brings and help them build on these resources even if they are not using relevant conceptual resources.

Linking Factors into Causal Chains. A second area of support required by a significant number of students in our study is unpacking entities by identifying activities and properties and linking these into causal chains. Mechanistic reasoning requires more than just identifying enough relevant factors; it also requires linking the factors. Our data show that students can do this. For example, Emerson identifies that intermolecular forces are important for folding proteins, which give the protein its shape. Emerson then associates this with properties such as polarity (Figure 8). However, many students who identified multiple factors needed additional support to link these factors, for example, by assigning activities or properties to entities. Responses from Alex (Figure 4), Tracy (Figure 5), and Jessie (Figure 9) all include multiple relevant factors but do not link these into coherent, causal chains. Thus, these students needed additional support linking factors into causal chains that show how the properties and activities of the entities produce differences in protein function.

Integration Across Disciplines. A third key area where students need support is integrating ideas from different domains into their mechanistic explanations. On the surface, this may seem similar to the previously discussed need to identify multiple factors. However, we posit that this differs because it requires promoting students' abilities to identify factors from disparate core ideas, courses, or disciplines that are not frequently associated during instruction. For example, if a student has already included charge in their response, as Alex did in Figure 4, linking this to another factor such as IMFs requires adding a resource that is frequently and tightly linked to the idea of charge during instruction. This differs from integrating resources from different core ideas, such as energetics and electrostatics, or information flow and structure and function, as these resources are less frequently linked during instruction. As an example, Skyler's response in Figure 6 links multiple factors related to electrostatics into a mechanistic explanation. Adding resources related to biological information flow to this explanation would be substantively different than adding in an additional closely related resource from chemistry.

Previous work has revealed challenges students face when integrating ideas across disciplines (Haudek *et al.*, 2012; Nagel and Lindsey, 2015; Kohn et al., 2018a, 2018b; Roche Allred *et al.*, 2021). Our work supports this, with only 5% of responses in this study found in the *Chemical and Biological Response Cluster*. This indicates that few students were productively linking conceptual resources from both biology and chemistry and that students will need additional support when asked to do this in upper-level courses.

One possible explanation for students' infrequent integration of ideas from both chemistry and biology is their use of one-reason decision-making, which reduces cognitive load and is a common strategy when facing day-to-day decisions (Talanquer, 2018). As a result, chemistry education researchers have found that students rarely consider multiple variables as part of mechanistic explanations, even when multiple variables influence the outcome. Instead, students tend to pick one variable and use only that variable in their explanation (Talanquer, 2018; Caspari and Graulich, 2019).

Another possibility is that students believe their responses met the expectations of the task. When disciplinary experts construct mechanistic explanations, they typically limit the number of levels below the phenomenon and/or the types of entities and activities they include in their explanation, not because additional mechanisms do not exist, but because they deem the explanation sufficient for its purpose (Machamer *et al.*, 2000; Bodé *et al.*, 2019). The task may have activated a range of resources that a student did not include in their explanation because they believed their explanation sufficiently completed the task. For example, when students write individual explanations following group work, the students typically focus on a single variable they believe most directly supports their claim, despite having previously explored a variety of variables as part of a group activity (Haas *et al.*, 2024).

It is also possible that some students in our study did not connect ideas across biology and chemistry because they had not yet mastered the required disciplinary content, or students had mastered the disciplinary content and the task failed to activate relevant resources.

Instructional and Curricular Implications

Regardless of the underlying reasons that students struggle in the areas described above, the instructional and curricular implications are the same. Students in our courses, even in transformed versions, need more opportunities and support to activate and incorporate relevant resources into mechanistic explanations in ways that build stable and usable connections. This is especially true for explanations that require integrating resources typically taught in different courses. Both the existing literature and this study provide us with paths toward achieving this goal.

Based on existing literature and our results, an introductory cell and molecular biology course designed to foster mechanistic reasoning across chemistry and biology would consider multiple aspects of context in order to influence resource activation and students' understanding of what they should be doing (Gouvea *et al.*, 2013; Macrie-Shuck and Talanquer, 2020; Smith *et al.*, 2024). Instruction would explicitly discuss mechanisms more frequently and would thoughtfully incorporate and discuss "shortcutting," "black boxing," and "bottoming out" when explaining phenomena (Machamer *et al.*, 2000; Haskel-Ittah, 2022; Scharlott *et al.*, 2024;

Shiroda *et al.*, 2024; Talanquer, 2024). Students would engage in scaffolded in-class activities that activate productive configurations of resources, both conceptual and epistemic (Graulich and Caspari, 2020; Haas *et al.*, 2024; Supplemental Figures S1 and S2). These activities would explicitly use vocabulary, figures, and examples from the linked chemistry course to provide contexts that activate conceptual resources from chemistry alongside those from biology and would connect these to real-world problems that students will encounter in their lives and careers. Instructional team messaging in this class would prioritize and reward engagement in mechanistic reasoning, for example, during conversation with groups as they engage in these in-class activities. The course would frequently engage students in these activities and ideally connect to similar activities in other courses (Crandell *et al.*, 2019; Ralph *et al.*, 2022; Kranz *et al.*, 2025).

An important element of this vision is designing unique scaffolds for each set of disciplinary core ideas we wish to co-activate and iteratively improving the scaffolding based on careful analyses of student responses. Our data provide some insights into effective and ineffective scaffolding and the challenges instructors may face when designing tasks requiring students to construct mechanistic explanations.

Task Design and Resource Activation

We observed that Tasks 1 and 3 resulted in mechanistic explanations that explained why two proteins have different functions more frequently than Task 2 (Table 1 and Figure 10B). We think much of this difference is explained by differences in the conceptual resources activated by these tasks. The percentage of responses assigned to the *Chemical and Biological*, *Chemical with IMFs,* or *Chemical without IMFs* Response Clusters was less than 20% for Task 1 but over 60% for Task 3 (Figure 10C and Supplemental Table S1). We think the diagram depicting functional groups in Task 3 (Figure 1) likely cued students to include and link two chemistry-related factors in their explanations, functional groups and intermolecular forces, which likely contributed to the higher frequency of Task 3 responses in the *chemical with IMF cluster* and the high frequency of mechanistic explanations.

On the other hand, task details can also lead to decreased mechanistic reasoning about the target phenomenon by activating non-productive resources. We think this was an important factor contributing to the lack of mechanistic explanations in response to Task 2. Our *post hoc* analysis revealed that 34% of Task 2 responses included ideas related to transport energetics, while none of the responses to Task 1 or Task 3 included these ideas. Task 1 contains a few specific details, while Task 2 asks how and why two transporters move different sugars across the membrane. In hindsight, it is not surprising that responses to Task 2 included resources related to the energetics of transport, something students in the course spent significant time learning about and working with when studying membrane transport.

Several pieces of evidence indicate that the focus on transport in Task 2 reduced the construction of mechanistic explanations. First, responses to Task 2 are underrepresented in the *Biological Information Response Cluster* and overrepresented in the *Shape Response Cluster*, suggesting that resources related to transport energetics replaced the more productive biology resources activated by Task 1. Next, Task 1 is overrepresented in the *Biological Information Response Cluster*, showing that, without the transport cue in the task, students were activating the appropriate conceptual resources needed to construct a mechanistic explanation about protein function. Finally, none of the responses that included ideas related to transport energetics were mechanistic explanations. Therefore, it appears that activation of resources related to transport energetics led to reduced activation of the conceptual resources needed to construct a mechanistic explanation about differential protein function.

While there appears to be a connection between task details and activation of non-productive resources for Task 2, it is not always clear what prompts students to include non-productive ideas. Our *post hoc* analysis identified seven students who focused on cellular signal transduction, only one of whom constructed a mechanistic explanation. Tracy's response to Task 3 (Figure 5) is an example of a response focused on cellular signal transduction rather than molecular-level structure or interactions and is not a mechanistic explanation of why two proteins have different functions. Although Tracy included relevant biological concepts, the explanation did not link these factors in a way that met the criteria for a mechanistic explanation of differential protein function (Table 4). This illustrates how activating non-productive resources—despite surface relevance—can divert students from addressing the core question. Two of these signal transduction-focused responses were from Task 3, which included receptors as a detail in the task and may have cued these students toward the unintended phenomenon. The remaining five responses that focused on signal transduction came from Task 1, which provided minimal context and did not reference signal transduction or its components. This suggests that for Task 1, signal transduction resources were not activated in response to a specific detail in the task. Students may have drawn on signal transduction because it was a recently covered topic in the course and more accessible in memory. More research is needed to better understand the conditions that promote or inhibit productive resource activation.

Co-Activating Chemistry and Biology Resources

Our data also provide insights into challenges faced in creating tasks that activate resources across courses or disciplines. Our study showed a statistically significant influence of task version on students' use of resources (Figure 10C). However, students who completed the different tasks had similar grades in biology and chemistry, suggesting the resources students included were influenced more strongly by the task than by their underlying disciplinary knowledge. Based on responses across the tasks, it is highly likely that more than 5% of the students in our study had the epistemic and conceptual resources needed to create a mechanistic explanation that spans the disciplines, but our tasks and the context in which they were completed failed to co-activate these resources.

As a specific example of this, student responses to Task 3 focused on conceptual resources from chemistry more frequently than responses to Task 1. However, it appears that including chemistry-specific cues came at the cost of reducing the inclusion of biology resources. Responses from Task 3 are underrepresented in the *Biological Information Response*

Cluster, while responses from Task 1 are overrepresented in this cluster. Over two-thirds of Task 1 responses fell in either the *Chemical and Biological* or the *Biological Information Response Clusters*. In comparison, less than 5% of Task 3 responses were assigned to the *Chemical and Biological Response Cluster,* and none of them to the *Biological Information Response Cluster*. This trade-off suggests that cueing one discipline may suppress activation or inclusion of resources from another, underscoring the challenge of designing tasks with sufficient scaffolding to explicitly activate disciplinary resources from both biology and chemistry. Other work by our group has shown that undergraduate students from a range of STEM courses rarely make explicit connections between biology and chemistry when explaining protein function, even with tasks specifically focused on eliciting ideas from both disciplines (Shiroda *et al.*, 2024). Additional work is needed to design scaffolds that support activation and integration of conceptual resources from different courses into mechanistic explanations.

Conclusions

Students in a transformed biology course received scaffolded instruction designed to support mechanistic reasoning that incorporates ideas from both chemistry and biology as they explain protein structure and function. The biology course has a transformed chemistry corequisite that focuses on mechanistic reasoning about structure and properties. Despite the transformed nature of both courses and the intentionality of instructional design, we found that students rarely integrated resources from both biology and chemistry and frequently omitted elements required for a mechanistic explanation. This has important implications for instruction in subsequent biology courses that expect phenomena related to protein structure and function to activate resources from these prerequisite courses: most students will need additional support to do this.

From a resources perspective on learning, instructors should carefully scaffold instructional tasks to activate the desired resources, monitor what resources instructional tasks are activating, and tailor instruction based on what they see. They must ensure the curriculum is frequently activating both conceptual and epistemic resources needed to construct mechanistic explanations. And they need to connect instruction and assessments with the real-world problems the students will encounter in their lives and careers. Designing instructional tasks that co-activate resources from different disciplines is challenging and will require empirical testing. Effective implementation of these tasks will require dialog between students and instructional teams during instruction and feedback following instruction and task completion. This requires appropriately trained faculty with enough time and resources to read the appropriate literature, design and administer carefully scaffolded instruction and tasks, evaluate student responses and iteratively improve instruction and tasks, and work across courses to support students throughout the curriculum (Gouvea *et al.*, 2013; Macrie-Shuck and Talanquer, 2020; Smith *et al.*, 2024). Administrators and academic leaders can support this type of integrated curriculum by encouraging intergenerational and interdisciplinary instructional teams, providing access to professional development, and developing structural incentives that promote 3DL or similar evidence-based instruction throughout the curriculum. Preparing students to explain complex biological phenomena using integrated disciplinary knowledge requires not just better tasks and transformed courses, but deliberate, structural connections between courses.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1725521. We thank Juli Uhl, Clare Franovic, and Seth Hunt for providing valuable feedback on the manuscript. ChatGPT was used to provide feedback on clarity and flow of text provided to it by the authors and to edit and troubleshoot R-code used in statistical analysis. Some literature searches were performed using the generative AI-assisted Elicit platform (https://elicit.com/).

REFERENCES

Agrawal, S. (2022). When molecules come to life – using augmented reality for studying protein structure in Cell Biology class. *The FASEB Journal*, *36*(S1). https://doi.org/10.1096/fasebj.2022.36.S1.L7980

Alameh, S., Abd-El-Khalick, F., & Brown, D. (2023). The nature of scientific explanation: Examining the perceptions of the nature, quality, and "goodness" of explanation among college students, science teachers, and scientists. *Journal of Research in Science Teaching*, *60*(1), 100–135. https://doi.org/10.1002/tea.21792

Alberti, G. (2024). *From Data to Insights: A Beginner's Guide to Basic Cross-tabulation Analysis*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781032726328

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Bachtiar, R. W., Meulenbroeks, R. F. G., Van Joolingen, W. R. (2022). Mechanistic reasoning in science education: A literature review. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(11), em2178. https://doi.org/10.29333/ejmste/12512

Bain, K., Bender, L., Bergeron, P., Caballero, M. D., Carmel, J. H., Duffy, E. M., ... Cooper, M. M. (2020). Characterizing college science instruction: The Three-Dimensional Learning Observation Protocol. *PLoS ONE*, *15*(6), e0234640. https://doi.org/10.1371/journal.pone.0234640

Becker, N., Noyes, K., & Cooper, M. (2016). Characterizing students' mechanistic reasoning about London Dispersion Forces. *Journal of Chemical Education*, *93*(10), 1713–1724. https://doi.org/10.1021/acs.jchemed.6b00298

Bierema, A. M.-K., Schwarz, C. V., & Stoltzfus, J. R. (2017). Engaging undergraduate biology students in scientific modeling: Analysis of group interactions, sense-making, and justification. *Cell Biology Education*, *16*(4), ar68. https://doi.org/10.1187/cbe.17-01-0023

Bing, T. J., & Redish, E. F. (2009). Analyzing problem solving using math in physics: Epistemological framing via warrants. *Physical Review Special Topics - Physics Education Research*, *5*(2), 020108. https://doi.org/10.1103/PhysRevSTPER.5.020108

Bodé, N. E., Deng, J. M., & Flynn, A. B. (2019). Getting past the rules and to the WHY: Causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *Journal of Chemical Education*, *96*(6), 1068–1082. https://doi.org/10.1021/acs.jchemed.8b00719

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, *95*(4), 639–669. https://doi.org/10.1002/sce.20449

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61–100. https://doi.org/10.2307/1167267

Bray Speth, E., Shaw, N., Momsen, J., Reinagel, A., Le, P., Taqieddin, R., & Long, T. (2014). Introductory biology students' conceptual models and explanations of the origin of variation. *CBE-Life Sciences Education*, *13*(3), 529–539. https://doi.org/10.1187/cbe.14-02-0020

Caspari, I., & Graulich, N. (2019). Scaffolding the structure of organic chemistry students' multivariate comparative mechanistic reasoning. *International Journal of Physics and Cemistry Education*, *11*(2), 31–43. https://doi.org/10.13140/RG.2.2.35136.48646

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices*. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Cooper, M. M. (2015). Why ask why? *Journal of Chemical Education*, *92*(8), 1273–1279. https://doi.org/10.1021/acs.jchemed.5b00203

Cooper, M. M., Caballero, M. D., Ebert-May, D., Fata-Hartley, C. L., Jardeleza, S. E., Krajcik, J. S., … Underwood, S. M. (2015). Challenge faculty to transform STEM learning. *Science*, *350*(6258), 281–282. https://doi.org/10.1126/science.aab0933

Cooper, M. M., & Klymkowsky, M. (2013). Chemistry, life, the universe, and everything: A new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical Education*, *90*(9), 1116–1122. https://doi.org/10.1021/ed300456y

Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating students' reasoning about acid–base reactions. *Journal of Chemical Education*, *93*(10), 1703–1712. https://doi.org/10.1021/acs.jchemed.6b00417

Cooper, M. M., Williams, L. C., & Underwood, S. M. (2015). Student understanding of intermolecular forces: A multimodal study. *Journal of Chemical Education*, *92*(8), 1288–1298. https://doi.org/10.1021/acs.jchemed.5b00169

Crandell, O. M., Kouyoumdjian, H., Underwood, S. M., & Cooper, M. M. (2019). Reasoning about reactions in organic chemistry: Starting it in general chemistry. *Journal of Chemical Education*, *96*(2), 213–226. https://doi.org/10.1021/acs.jchemed.8b00784

Crandell, O. M., Lockhart, M. A., & Cooper, M. M. (2020). Arrows on the page are not a good gauge: Evidence for the importance of causal mechanistic explanations about nucleophilic substitution in organic chemistry. *Journal of Chemical Education*, *97*(2), 313–327. https://doi.org/10.1021/acs.jchemed.9b00815

Crowder, C. J., Yik, B. J., Frost, S. J. H., Cruz-Ramírez de Arellano, D., & Raker, J. R. (2024). Impact of prompt cueing on level of explanation sophistication for organic reaction mechanisms. *Journal of Chemical Education*, *101*(2), 398–410. https://doi.org/10.1021/acs.jchemed.3c00710

Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, *50*(6), 639–659. https://doi.org/10.1002/tea.21094

De Andrade, V., Shwartz, Y., Freire, S., & Baptista, M. (2022). Students' mechanistic reasoning in practice: Enabling functions of drawing, gestures and talk. *Science Education*, *106*(1), 199–225. https://doi.org/10.1002/sce.21685

Deng, J. M., & Flynn, A. B. (2021). Reasoning, ganularity, and comparisons in students' arguments on two organic chemistry items. *Chemistry Education Research and Practice*, *22*(3), 749–771. https://doi.org/10.1039/d0rp00320dDedoose

Dedoose. (2019). *Web Application for Managing, Analyzing, and Presenting Qualitative and Mixed Method Research Data* (Version 7.0.18) [Computer software]. SocioCultural Research Consultants, LCC.

diSessa, A. A. (1988). Knowledge in pieces. In: *Constructivism in the Computer Age*, ed. G Formanand P Pufall, New Jersey: Lawrence Erlbaum Publishers, 49–70.

Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. *Journal of Research in Science Teaching*, *44*(7), 938–959. https://doi.org/10.1002/tea.20186

Edelsbrunner, P., Schumacher, R., Hänger-Surer, B., Schalk, L., & Stern, E. (2024). Preparation for future conceptual learning: Content-specific long-term effects of early physics instruction. *Journal of Educational Psychology*, *116*, 1479–1499. https://doi.org/10.3929/ETHZ-B-000703158

Flowers, S., Holder, K. H., & Gardner, S. M. (2023). Here is the biology, now what is the mechanism? Investigating biology undergraduates' mechanistic reasoning within the context of biofilm development. *Journal of Microbiology and Biology Education*, *24*(2), e00221–22. https://doi.org/10.1128/jmbe.00221-22

Flowers, S., Holder, K. H., Rump, G. K., & Gardner, S. M. (2023). Missed connections: Exploring features of undergraduate biology students' knowledge networks relating gene regulation, cell–cell communication, and phenotypic expression. *CBE—Life Sciences Education*, *22*(4), ar44. https://doi.org/10.1187/cbe.22-03-0041

Franovic, C. G.-C., Noyes, K., Stoltzfus, J. R., Schwarz, C. V., Long, T. M., & Cooper, M. M. (2023). Undergraduate chemistry and biology students' use of causal mechanistic reasoning to explain and predict preferential protein–ligand binding activity. *Journal of Chemical Education*, *100*(5), 1716–1727. https://doi.org/10.1021/acs.jchemed.2c00737

Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718–3720. https://doi.org/10.1093/bioinformatics/btv428

Galili, T., O'Callaghan, A., Sidi, J., & Sievert, C. (2018). heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, *34*(9), 1600–1602. https://doi.org/10.1093/bioinformatics/btx657

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). *Various Coefficients of Interrater Reliability and Agreement*. https://www.r-project.org

Gouvea, J., Appleby, L., Fu, L., & Wagh, A. (2022). Motivating and shaping scientific argumentation in lab reports. *CBE—Life Sciences Education*, *21*(4), ar71. https://doi.org/10.1187/cbe.21-11-0316

Gouvea, J. S., Sawtelle, V., Geller, B. D., & Turpen, C. (2013). A framework for analyzing interdisciplinary tasks: Implications for student learning and curricular design. *CBE—Life Sciences Education*, *12*(2), 187–205. https://doi.org/10.1187/cbe.12-08-0135

Graulich, N., & Caspari, I. (2020). Designing a scaffold for mechanistic reasoning in organic chemistry. *Chemistry Teacher International*, *3*(1), 20200001. https://doi.org/10.1515/cti-2020-0001

Haas, D. B., Watts, F. M., Dood, A. J., & Shultz, G. V. (2024). Analysis of organic chemistry students' developing reasoning elicited by a scaffolded case comparison activity. *Chemistry Education Research and Practice*, *25*(3), 742–759. https://doi.org/10.1039/d4rp00021h

Halmo, S. M., Sensibaugh, C. A., Bhatia, K. S., Howell, A., Ferryanto, E. P., Choe, B., … Lemons, P. P. (2018). Student difficulties during structure–function problem solving. *Biochemistry and Molecular Biology Education*, *46*(5), 453–463. https://doi.org/10.1002/bmb.21166

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics*, *68*(S1), S52–S59. https://doi.org/10.1119/1.19520

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *Journal of the Learning Sciences*, *12*(1), 53–90. https://doi.org/10.1207/S15327809JLS1201_3

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of Learning from a Modern Multidisciplinary Perspective* (pp. 89–119). Greenwich, CT: IAP.

Haskel-Ittah, M. (2022). Explanatory black boxes and mechanistic reasoning. *Journal of Research in Science Teaching*, *60*(4), 915–933. https://doi.org/10.1002/tea.21817

Haudek, K. C., Prevost, L. B., Moscarella, R. A., Merrill, J., & Urban-Lurain, M. (2012). What are they thinking? Automated analysis of student writing about acid–base chemistry in introductory biology. *CBE-Life Sciences Education*, *11*(3), 283–293. https://doi.org/10.1187/cbe.11-08-0084

Houchlei, S. K., Bloch, R. R., & Cooper, M. M. (2021). Mechanisms, models, and explanations: Analyzing the mechanistic paths students take to reach a product for familiar and unfamiliar organic reactions. *Journal of Chemical Education*, *98*(9), 2751–2764. https://doi.org/10.1021/acs.jchemed.1c00099

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Klymkowsky, M. W. (2021). Making mechanistic sense: Are we teaching students what they need to know? *Developmental Biology*, *476*, 308–313. https://doi.org/10.1016/j.ydbio.2021.04.004

Kohn, K. P., Underwood, S. M., & Cooper, M. M. (2018a). Connecting structure–property and structure–function relationships across the disciplines of chemistry and biology: Exploring student perceptions. *CBE—Life Sciences Education*, *17*(2), ar33. https://doi.org/10.1187/cbe.18-01-0004

Kohn, K. P., Underwood, S. M., & Cooper, M. M. (2018b). Energy connections and misconnections across chemistry and biology. *CBE—Life Sciences Education*, *17*(1), ar3. https://doi.org/10.1187/cbe.17-08-0169

Kranz, D., Martin, P. P., Schween, M., & Graulich, N. (2025). Should we scaffold it? Analysing the effect of task format and scaffolding on students' learning gain. *Chemistry Education Research and Practice*, *27*, 151–171. https://doi.org/10.1039/d4rp00241e

Krist, C., Schwarz, C. V., & Reiser, B. J. (2019). Identifying essential epistemic heuristics for guiding mechanistic reasoning in science learning. *Journal of the Learning Sciences*, *28*(2), 160–205. https://doi.org/10.1080/10508406.2018.1510404

Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., … Cooper, M. M. (2016). Characterizing college science

assessments: The three-dimensional learning assessment protocol. *PLoS ONE*, *11*(9), e0162333. https://doi.org/10.1371/journal.pone.0162333
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759
Macrie-Shuck, M., & Talanquer, V. (2020). Exploring students' explanations of energy transfer and transformation. *Journal of Chemical Education*, *97*(12), 4225–4234. https://doi.org/10.1021/acs.jchemed.0c00984
Matz, R. L., Fata-Hartley, C. L., Posey, L. A., Laverty, J. T., Underwood, S. M., Carmel, J. H., … Cooper, M. M. (2018). Evaluating the extent of a large-scale transformation in gateway science courses. *Science Advances*, *4*(10), eaau0554. https://doi.org/10.1126/sciadv.aau0554
McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031
McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1
Michael, J. (2021). What do we mean when we talk about "structure/function" relationships? *Advances in Physiology Education*, *45*(4), 880–885. https://doi.org/10.1152/advan.00108.2021
Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research & Perspective*, *1*(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02
Mulnix, A. B. (2003). Investigations of protein structure and function using the scientific literature: An assignment for an undergraduate cell physiology course. *Cell Biology Education*, *2*(4), 248–255. https://doi.org/10.1187/cbe.03-06-0025
Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z
Nagel, M. L., & Lindsey, B. A. (2015). Student use of energy concepts from physics in chemistry courses. *Chemistry Education Research and Practice*, *16*(1), 67–81. https://doi.org/10.1039/C4RP00184B
Naioti, E., & Mudrak, E. (2022). *Using Adjusted Standardized Residuals for Interpreting Contingency Tables*. Cornell Statistical Consulting Unit. Retrieved from https://cscu.cornell.edu/wp-content/uploads/conttableresid.pdf
National Research Council. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. The National Academies Press. Retrieved from http://www.nap.edu/openbook.php?record_id=10019
National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press. Retrieved from http://www.nap.edu/openbook.php?record_id=13165
National Research Council. (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition (p. 9853)*. Washington, DC: The National Academies Press. https://doi.org/10.17226/9853
Noyes, K., Carlson, C. G., Stoltzfus, J. R., Schwarz, C. V., Long, T. M., & Cooper, M. M. (2022). A deep look into designing a task and coding scheme through the lens of causal mechanistic reasoning. *Journal of Chemical Education*, *99*(2), 874–885. https://doi.org/10.1021/acs.jchemed.1c00959
Noyes, K., & Cooper, M. M. (2019). Investigating student understanding of London Dispersion Forces: A longitudinal study. *Journal of Chemical Education*, *96*(9), 1821–1832. https://doi.org/10.1021/acs.jchemed.9b00455
Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, *13*(3), 423–451. https://doi.org/10.1207/s15327809jls1303_6
Ralph, V. R., Scharlott, L. J., Schwarz, C. E., Becker, N. M., & Stowe, R. L. (2022). Beyond instructional practices: Characterizing learning environments that support students in explaining chemical phenomena. *Journal of Research in Science Teaching*, *59*(5), 841–875. https://doi.org/10.1002/tea.21746
Redish, E. F., & Hammer, D. (2009). Reinventing college physics for biologists: Explicating an epistemological curriculum. *American Journal of Physics*, *77*(7), 629–642. https://doi.org/10.1119/1.3119150
Roche Allred, Z. D., Farias, A. J., Kararo, A. T., Parent, K. N., Matz, R. L., & Underwood, S. M. (2021). Students' use of chemistry core ideas to explain the structure and stability of DNA. *Biochemistry and Molecular Biology Education*, *49*(1), 55–68. https://doi.org/10.1002/bmb.21391
RStudio Team. (2020). *RStudio: Integrated Development Environment for R. RStudio*, Boston, MA: PBC. Retrieved from http://www.rstudio.com/
Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499–525. https://doi.org/10.1002/sce.20264
Scharlott, L. J., Rippey, D. W., Rosa, V., & Becker, N. M. (2024). Progression toward causal mechanistic reasoning through phenomenon-based learning in introductory chemistry. *Journal of Chemical Education*, *101*(3), 777–788. https://doi.org/10.1021/acs.jchemed.3c00517
Scott, E. E., Anderson, C. W., Mashood, K. K., Matz, R. L., Underwood, S. M., & Sawtelle, V. (2018). Developing an analytical framework to characterize student reasoning about complex processes. *CBE-Life Sciences Education*, *17*(3), ar49. https://doi.org/10.1187/cbe.17-10-0225
Shan, Y. (2025). The debate over proximate and ultimate causation in biology. *Synthese*, *205*(1), 37. https://doi.org/10.1007/s11229-024-04890-8
Shiroda, M., Franovic, C. G.-C., De Lima, J., Noyes, K., Babi, D., Beltran-Flores, E., … Stoltzfus, J. R. (2024). Examining and supporting mechanistic explanations across chemistry and biology courses. *CBE—Life Sciences Education*, *23*(3), ar38. https://doi.org/10.1187/cbe.23-08-0157
Smith, D., Finkelstein, N., Horii, C. V., Miller, E., & York, T. T. (2024). *Levers for change: Enacting a national agenda for undergraduate STEM education*. ed. L. Schererand T. Veague, Washington, D.C.: American Association for the Advancement of Science. https://aaas-iuse.org/wp-content/uploads/2024/06/AAAS-IUSE-Report_Levers_for_Change.pdf.
Smith, III, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, *3*(2), 115–163. https://doi.org/10.1207/s15327809jls0302_1
Southard, K. M., Espindola, M. R., Zaepfel, S. D., & Bolger, M. S. (2017). Generative mechanistic explanation building in undergraduate molecular and cellular biology. *International Journal of Science Education*, *39*(13), 1795–1829. https://doi.org/10.1080/09500693.2017.1353713
Southard, K. M., Wince, T., Meddleton, S., & Bolger, M. S. (2016). Features of knowledge building in biology: Understanding undergraduate students ideas about molecular mechanisms. *CBE-Life Sciences Education*, *15*(1), ar7. https://doi.org/10.1187/cbe.15-05-0114
Talanquer, V. (2018). Exploring mechanistic reasoning in chemistry. In: *Science Education Research and Practice in Asia-Pacific and Beyond* ed. J. Yeo, T. W. Teo, and K.-S. Tang, Singapore: Springer, 39–52. https://doi.org/10.1007/978-981-10-5149-4_3
Talanquer, V. (2024). How we simplify, circumvent, or distort causal mechanistic reasoning in chemistry. *Journal of Chemical Education*, *101*(5), 1793–1797. https://doi.org/10.1021/acs.jchemed.4c00281
Tansey, J. T., Baird, T., Cox, M. M., Fox, K. M., Knight, J., Sears, D., & Bell, E. (2013). Foundational concepts and underlying theories for majors in "biochemistry and molecular biology". *Biochemistry and Molecular Biology Education*, *41*(5), 289–296. https://doi.org/10.1002/bmb.20727
Trujillo, C. M., Anderson, T. R., & Pelaez, N. J. (2015). A model of how different biology experts explain molecular and cellular mechanisms. *Cell Biology Education*, *14*(2), ar20–ar20. https://doi.org/10.1187/cbe.14-12-0229
Trujillo, C. M., Anderson, T. R., & Pelaez, N. J. (2016). An instructional design process based on expert knowledge for teaching students how mechanisms are explained. *Advances in Physiology Education*, *40*(2), 265–273. https://doi.org/10.1152/advan.00077.2015
van Mil, M. H. W., Boerwinkel, D. J., & Waarlo, A. J. (2013). Modelling molecular mechanisms: A framework of scientific reasoning to construct molecular-level explanations for cellular behaviour. *Science & Education*, *22*(1), 93–118. https://doi.org/10.1007/s11191-011-9379-7
Villafañe, S. M., Loertscher, J., Minderhout, V., & Lewis, J. E. (2011). Uncovering students' incorrect ideas about foundational concepts for biochemistry. *Chemistry Education Research and Practice*, *12*(2), 210–218. https://doi.org/10.1039/C1RP90026A
Wagner-Menghin, M. M. (2005). Binomial test. In: *Encyclopedia of Statistics in Behavioral Science (Vol. 1, pp. 158-163)*, ed. B. S. Everittand D. C. Howell, 1st ed., Hoboken, NJ: Wiley. https://doi.org/10.1002/0470013192.bsa743
Žiberna, A., Kejzar, N., & Golob, P. (2004). A comparison of different approaches to hierarchical clustering of ordinal data. *Metodoloski Zvezki*, *1*, 57–73.