

EXPLORING STUDENTS' UNDERSTANDING OF INTERACTIONS AND ENERGY ACROSS CHEMISTRY AND  
BIOLOGY

By

Keenan Chun Hong Lee Noyes

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Chemistry – Doctor of Philosophy

2022

## **ABSTRACT**

### **EXPLORING STUDENTS' UNDERSTANDING OF INTERACTIONS AND ENERGY ACROSS CHEMISTRY AND BIOLOGY**

By

Keenan Chun Hong Lee Noyes

One of the goals of science education is to help students make sense of the world around them. To that end, it is critical that students understand the central ideas in each discipline like, in chemistry, energy and interactions. These ideas are of particular importance because they are directly related to one another and are relevant across other science disciplines. Unfortunately, researchers have found that students often struggle to develop a deep understanding of these ideas. To uncover better ways to support students' learning, I explored how students understand interactions and energy in both chemistry and biology.

In this dissertation, I focused on London dispersion forces (LDFs), a type of intermolecular force (IMF) which occurs between all atoms and molecules. Specifically, I used the lens of causal mechanistic reasoning to think about students' knowledge. That is, how students connect the properties and behaviors of the underlying entities to the overall phenomenon. If we can help students to develop this type of understanding, they may be able to make powerful predictions about new, unfamiliar phenomena in which IMFs play an important role. Additionally, I explored how students thought about the energy changes which result from the formation of LDFs. Lastly, I designed assessments to elicit and characterize explanations of protein-ligand binding, a biological phenomenon governed by IMFs.

To explore these questions, I used a mix of qualitative and quantitative techniques. I designed tasks to elicit causal mechanistic responses from students, using students' responses to refine the task design. I also developed coding schemes to characterize students' engagement in causal mechanistic reasoning. Furthermore, I developed and used automated resources to analyze thousands of responses

in a matter of minutes. In these studies, I focused primarily on undergraduate students enrolled in Chemistry, Life, the Universe, and Everything (CLUE), a transformed, core-idea centered general chemistry curriculum.

From these studies, I found that the majority of CLUE students could leverage electrostatic ideas to explain LDFs, and that a meaningful proportion of those students could provide a full causal mechanistic account. This highlights the importance of emphasizing these interactions, and the mechanism by which they form, throughout the general chemistry course sequence. Additionally, students who used causal mechanistic reasoning to discuss LDFs were more likely to use that same reasoning in the context of the associated changes in potential energy. However, this relationship was weaker among those providing a partially causal mechanistic response. This suggests that more work needs to be done to find ways of supporting students to connect the ideas of interactions and energy. Additionally, in this thesis, I describe the process by which I used iterative design to develop a task eliciting causal mechanistic explanations of a biological phenomenon. In future work, these materials can be used to explore how broader groups of students engage with this task in an effort to foster interdisciplinary coherence.

I dedicate this dissertation to my mom, Yu Man Lee, who taught me to love the natural world and the joy of sharing it with others.



## ACKNOWLEDGEMENTS

I could write an entire second dissertation thanking everyone who has helped me reach where I am today, but I will try to keep this brief. I would like to thank Drs. Becky Matz, J.T. Lavery, Sarah Jardeleza, and Lynmarie Posey for introducing me to the world of education research. I would also like to thank my graduate committee: Drs. Heedeok Hong, Mark Urban-Lurain, and Tammy Long. I have been incredibly fortunate to know them all outside of this committee, and I am grateful for all their guidance. Throughout both undergraduate and graduate experiences, there has been one constant: my advisor, Dr. Melanie Cooper. There are very few people who have had such a positive impact on my life as she has. I cannot thank her enough for all she has done for me.

I would like to thank all the Cooper group members I have been fortunate enough to work with over the years. There are so many wonderful friends and colleagues I have made along the way, but I would like to recognize three individuals in particular. First, I would like to thank Dr. Nicole Becker for taking me under her wing and sparking my love for chemical education research. Second, I would like to thank soon-to-be Dr. Samantha Houchlei; graduate school is hard, but it becomes a lot more enjoyable when you get to do it with a friend like her. Third, future Dr. Clare Carlson is the only other person who has come into this group who loves biology as much as I do, and it was a privilege to work alongside her. I am so proud of this supportive community we have all built together—I will miss this group terribly.

I would like to recognize the Olin Health Center and their Counseling and Psychiatric Services specifically. Asking for help is hard, and that is especially true when it comes to mental health. I would not have been able to finish this dissertation without these resources and, in particular, the help of Dr. Yvonne Connelly. I am forever grateful.

Last but not least, I would like to thank my family. My parents and siblings have given me endless support over the years, and I feel this love every day—it makes everything worthwhile. Finally, I would like to thank my wonderful partner, Taylor. The last ten years have been filled with many ups and downs, but she has supported me at every turn. This dissertation is the product of her hard work as much as it is mine, and I am so lucky to have her by my side.

## TABLE OF CONTENTS

LIST OF TABLES.....	xii
LIST OF FIGURES.....	xiv
CHAPTER I – INTRODUCTION .....	1
Study goals and research questions.....	3
Study 1: Investigating student understanding of London dispersion forces: a longitudinal study.....	3
Study 2: Developing computer resources to automate analysis of students’ explanations of London dispersion forces .....	3
Study 3: Using machine learning resources to explore the long-term impact of CLUE on undergraduate students’ explanations of London dispersion forces .....	4
Study 4: Exploring connections between students’ explanations of London dispersion forces and potential energy .....	5
Study 5: A deep look into designing a task and coding scheme through the lens of causal mechanistic reasoning .....	5
CHAPTER II - THEORETICAL FRAMEWORKS .....	7
How people learn: the resources perspective .....	7
Constructivism.....	7
Misconceptions and conceptual change.....	7
Resources perspective .....	8
Using assessments to collect evidence .....	10
Causal mechanistic reasoning .....	11
CHAPTER III – LITERATURE REVIEW .....	13
Assessment scaffolding and iterative design .....	13
Eliciting and characterizing causal mechanistic reasoning .....	15
The importance of forces and interactions.....	18
CLUE and the impact on forces and interactions.....	20
Prior research on CM reasoning and LDFs.....	21
CHAPTER IV - INVESTIGATING STUDENT UNDERSTANDING OF LONDON DISPERSION FORCES: A LONGITUDINAL STUDY .....	23
Preface .....	23
Introduction .....	23
Causal Mechanistic Reasoning and LDFs.....	25
Designing the Task .....	27
Research questions .....	30
Methods.....	31
Participants .....	31
Data Collection.....	32
Data Analysis.....	37
Results and discussion .....	40
Study 1: What effect does the level of scaffolding have on student drawings and explanations.....	40

Study 2: How do students' explanations and models of LDFs change over a two-semester sequence .....	42
Finding 1: A large percentage of students construct a full CM drawing on the first summative exam after being taught the material .....	42
Finding 2: Removal of the scaffold appears to produce a decline in causal mechanistic drawings on unstructured posttest assessments (Items 3 and 4) .....	43
Finding 3: The initial scaffolded drawing prompt does have an effect on the final number of CM drawings .....	44
Finding 4: The text responses do not vary as widely over the four time-points, but there is a general decrease in CM written explanations .....	45
Study 3: How do students written explanations and drawn models relate to one another .....	45
Finding 1: There is generally an association between text and drawing responses except for EC responses .....	45
Finding 2: If we use the "best" response - drawing or text - we see a significant increase in CM from item 1 to 2, a decrease from 2 to 3 and then little change from 3 to 4 .....	48
Summary .....	49
Implications For Teaching .....	50
The use of scaffolding in formative assessments .....	50
Formative and summative assessments .....	51
The importance of long-term post-tests: what do students retain .....	52
Limitations .....	52
APPENDICES .....	53
APPENDIX A: Permissions .....	54
APPENDIX B: Studies 1-3 participant demographics and interrater reliability .....	55
APPENDIX C: Full activity prompts and additional information .....	57
APPENDIX D: Additional contingency tables .....	61
 CHAPTER V - DEVELOPING COMPUTER RESOURCES TO AUTOMATE ANALYSIS OF STUDENTS' EXPLANATIONS OF LONDON DISPERSION FORCES .....	62
Preface .....	62
Introduction .....	62
The importance of assessments .....	63
Designing assessments .....	64
The role of formative, or low-stakes, assessments .....	66
The case for machine learning .....	66
Overview of the Constructed Response Classifier .....	67
Challenges with machine learning and potential solutions .....	68
Research questions .....	70
Methods .....	70
Strategy for developing machine learning resources .....	70
Participants .....	71
Group 1A .....	71
Group 1B .....	72
Group 2 .....	72
Group 3 .....	73
Question prompt .....	73
Coding scheme .....	75
Human coding of responses .....	76

General overview of how the CRC works.....	79
Results and discussion .....	82
Developing an initial model to characterize LDF responses .....	82
Expanding our initial model with responses from 3 other groups .....	85
Testing the model performance .....	86
Detecting signal at the group level .....	89
Limitations.....	92
Conclusion and implications for teaching and research .....	94
APPENDICES .....	96
APPENDIX A: Permissions.....	97
APPENDIX B: Additional demographic information about the participants .....	98
APPENDIX C: Additional information about the nature of the responses.....	100
APPENDIX D: Additional results of groups 1B, 2, and 3 test set coding.....	113
 CHAPTER VI - USING MACHINE LEARNING RESOURCES TO EXPLORE THE LONG-TERM IMPACT OF CLUE ON UNDERGRADUATE STUDENTS' EXPLANATIONS OF LONDON DISPERSION FORCES .....	 116
Introduction .....	116
CLUE and intermolecular forces.....	116
Causal mechanistic reasoning and LDFs.....	117
Machine learning and assessments .....	118
Online learning and the global pandemic .....	119
Research questions .....	120
Methods.....	120
LDF task .....	120
Participants .....	120
Coding scheme .....	123
Automated analysis tools.....	123
Statistical tests .....	125
Results and discussion .....	125
Study 1: How has the long-term adoption of CLUE impacted how students use causal mechanistic reasoning to explain the attraction between neutral species .....	125
Study 2: How do “on” and “off” sequence students compare in their explanations of LDFs.....	127
Study 3: How has the emergence of the pandemic impacted the students' responses .....	129
Limitations.....	131
Conclusions .....	131
 CHAPTER VII - EXPLORING CONNECTIONS BETWEEN STUDENTS' EXPLANATIONS OF LONDON DISPERSION FORCES AND POTENTIAL ENERGY .....	 133
Introduction .....	133
Prior research on understanding of potential energy .....	133
Prior research on understanding of intermolecular forces .....	134
Causal mechanistic reasoning .....	135
Resources perspective .....	136
Research questions .....	137
Methods.....	138
Participants .....	138
Data collection .....	139
LDF text and drawing tasks .....	139

PE task .....	140
Coding schemes .....	141
LDF coding scheme.....	141
PE coding scheme.....	142
Data analysis .....	144
Fall 2018 data .....	144
Fall 2020 data .....	145
Statistical tests .....	145
Results and discussion .....	146
Study 1: How do students use causal mechanistic reasoning to explain differences in PE minima of interacting neutral species.....	146
Results of PE task version 1.....	146
Results of PE task version 2.....	148
Study 2: How does students' causal mechanistic explanations of the formation of LDFs compare to their explanations of differences in PE minima .....	150
Study 3: How do students' causal mechanistic explanations about LDFs and the depth of PE wells impact students' responses about associated macroscopic phenomena .....	153
Limitations.....	155
Conclusion.....	156
 CHAPTER VIII - A DEEP LOOK INTO DESIGNING A TASK AND CODING SCHEME THROUGH THE LENS OF CAUSAL MECHANISTIC REASONING.....	159
Preface .....	159
Introduction .....	159
Causal mechanistic reasoning.....	160
Resources perspective of student learning.....	160
Assessment design .....	161
Research questions .....	165
Methods.....	165
Overview of the rationale for the design of the PL task and associated coding scheme .....	165
Participants .....	166
Analysis guiding task development.....	167
Analysis guiding coding scheme development .....	168
Results and discussion .....	170
PL task development.....	170
PL task version 1.....	170
PL task version 2.....	172
PL task version 3.....	174
PL task final version.....	176
PL coding scheme development .....	177
Developing the analytic rubric .....	178
The holistic scheme.....	179
Limitations.....	183
Conclusions and future work .....	184
APPENDICES .....	188
APPENDIX A: Permissions.....	189
APPENDIX B: Analytic rubric development and application .....	190
APPENDIX C: Alternate task administration details .....	197

APPENDIX D: Deidentified student responses used in task development.....	198
APPENDIX E: Screenshots of each activity version .....	254
APPENDIX F: Example responses mentioning hydrogen bonding .....	268
CHAPTER IX - CONCLUSIONS, IMPLICATIONS, AND FUTURE RESEARCH .....	269
Conclusions .....	269
CLUE helps students develop a deeper understanding of LDFs.....	269
Scaffolding and iterative design are useful in eliciting causal mechanistic reasoning .....	270
The “messy middle” of causal mechanistic reasoning.....	271
Implications.....	272
The importance of long-term educational initiatives .....	272
Automated resources can facilitate the use of explanation questions .....	272
Supporting students in the “messy middle” .....	273
Future directions.....	273
REFERENCES.....	275

## LIST OF TABLES

Table 4.1. Information about the LDF prompts used throughout all four studies. ....	33
Table 4.2. Coding scheme for text LDF responses. ....	37
Table 4.3. Coding scheme for LDF drawings. Student drawings used with permission. ....	39
Table 4.4. Pearson's $\chi^2$ test results exploring the relationship between the text and drawing LDF Responses. ....	46
Table 4.5. Contingency table for the item 4 (exam posttest assessment) text and drawing LDF responses. In each cell the adjusted residual value is reported along with the observed and expected values. Adjusted residuals larger than the Bonferroni adjusted critical value ( $\pm 2.78$ for 9-cell table) are bolded. To visualize the sign and magnitude of the adjusted residuals, the cells are color coded from dark red (most negative) to dark blue (most positive). ....	47
Table 4.6. The contingency table for the item 1 text and drawing LDF responses. In each cell the adjusted residual value is reported along with the observed value and the value we would expect if there was no relationship between the text and drawing responses (expected value). Adjusted residuals larger than the Bonferroni adjusted critical value ( $\pm 2.78$ for 9-cell table) are bolded. To visualize the sign and magnitude of the adjusted residuals, the cells are color coded from dark red (most negative) to dark blue (most positive). ....	61
Table 4.7. The contingency table for the item 3 text and drawing LDF responses. In each cell the adjusted residual value is reported along with the observed value and the value we would expect if there was no relationship between the text and drawing responses (expected value). Adjusted residuals larger than the Bonferroni adjusted critical value ( $\pm 2.78$ for 9-cell table) are bolded. To visualize the sign and magnitude of the adjusted residuals, the cells are color coded from dark red (most negative) to dark blue (most positive). ....	61
Table 5.1. Overview of coding scheme from Noyes and Cooper. <sup>75</sup> ....	76
Table 5.2. The initial level of agreement for each of three cohorts between authors RM and MN. ....	79
Table 5.3. Crosstab relating the number (and percentage of total) of reference human scores to the predicted computer scores for the training set of the initial model. ....	84
Table 5.4. Crosstab relating the number (and percentage of total) of reference human scores to the predicted computer scores for the training set of the combined model. ....	86
Table 5.5. Cohen's kappa value and percent agreement between human coders and with the computer models for groups 1B, 2 and 3. ....	87
Table 5.6. Example LDF responses. ....	102
Table 5.7. Crosstab relating the number of reference consensus human scores to the predicted computer scores using the initial model for the Group 1B test set. ....	113



Table 5.8. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the combined model for the Group 1B test set. ....	113
Table 5.9. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the initial model for the Group 2 test set.....	113
Table 5.10. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the combined model for the Group 2 test set. ....	114
Table 5.11. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the initial model for the Group 3 test set.....	114
Table 5.12. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the combined model for the Group 3 test set. ....	114
Table 5.13. Results of Pearson's $\chi^2$ tests comparing the human consensus and combined model codes. ....	115
Table 6.1. Number of student responses collected by time and course location. ....	121
Table 6.2. Confusion matrix for the computer model used to code the responses in this study. ....	124
Table 6.3. Results of pairwise Pearson's $\chi^2$ tests comparing fall and spring semesters each academic year. ....	128
Table 7.1. Descriptions and examples of the coding categories for characterizing responses to the PE Prompt. ....	143
Table 8.1. Overview of administration of each version of the PL task. ....	167
Table 8.2. Details of the administration of the original and alternate versions of the PL task. ....	197

## LIST OF FIGURES

Figure 4.1. Relative timing of the four assessment items relative to the two semesters in the general chemistry sequence in the 2015-2016 academic year. The four assessments consist of a formative homework assessment (item 1), a summative exam assessment (item 2), a homework posttest assessment (item 3), and an exam posttest assessment (item 4).	33
Figure 4.2. Item 2 drawing prompt including the scaffolded 3 box drawing prompt to encourage students to draw the process by which the interaction formed.	36
Figure 4.3. Distribution of text, drawing, and best (most sophisticated code for drawing or text) responses for the different formative assessments given during fall 2014 (N = 129) and fall 2015 (item 1, N = 150).	41
Figure 4.4. Percentage of 15-16 cohort drawing responses (n = 150) in each of the LDF categories for items 1-4 (see Table 4.1 for item descriptions).	43
Figure 4.5. Percentage of 15-16 cohort text responses (n = 150) in each of the categories for items 1-4 (see Table 4.1 for item descriptions).	45
Figure 4.7. Percentage of 15-16 cohort's (n = 150) best responses (based on each students most highest level text or drawing response) in each of the categories for items 1-4 (see Table 1 for item descriptions).	48
Figure 4.8. Permissions to reproduce manuscript in its entirety.	54
Figure 4.9. Formative LDF text prompt administered to general chemistry 1 students via beSocratic in fall of 2014.	57
Figure 4.10. Formative LDF drawing prompt administered to general chemistry 1 students via beSocratic in fall of 2014.	57
Figure 4.11. Item 1 (formative) LDF text prompt administered to the 15-16 cohort in fall of 2015 during general chemistry 1 via beSocratic.	57
Figure 4.12. Item 1 (formative) LDF drawing prompt administered to the 15-16 cohort in fall of 2015 during general chemistry 1 via beSocratic.	58
Figure 4.13. Item 2 (summative) LDF text and drawing prompt administered to the 15-16 cohort in fall of 2015 during general chemistry 1 on the first midterm exam.	58
Figure 4.14. Item 3 (homework posttest) LDF text prompt administered to the 15-16 cohort in spring of 2016 during general chemistry 2 via beSocratic.	59
Figure 4.15. Item 3 (homework posttest) LDF drawing prompt administered to the 15-16 cohort in spring of 2016 during general chemistry 2 via beSocratic.	59
Figure 4.16. Item 4 (exam posttest) LDF text and drawing prompt administered to the 15-16 cohort in spring of 2016 during general chemistry 2 on the last midterm exam.	60

Figure 5.1. LDF prompt. ....	74
Figure 5.2. Summary of the human coding reported in this study with the distinction as to which data was reported in our previous study, Noyes and Cooper. <sup>75</sup> .....	77
Figure 5.3. General overview of the automated coding process using several hypothetical student responses highlighting the creation of a document-term matrix, the training of the computer model, and the process by which the model then codes new responses. ....	80
Figure 5.4. Overview of 10-fold cross-validation procedure for assessing the accuracy of the developed computer model. ....	81
Figure 5.5. Agreement between the human and computer described by the Cohen's kappa value calculated in the 10-fold cross-validation as a function of the size of the number of responses used to train (and also validate) the computer model. The dashed line at 0.78 indicates the Cohen's kappa value for the human-human IRR with the responses.....	83
Figure 5.6. Distribution for LDF codes for the subsets of 100 responses from groups 1B, 2, and 3 as coded by humans (consensus score) and the computer (combined model). The human consensus score represents the agreed upon codes of authors R.L.M. and M.N. after discussion. The computer scores were coded by the combined model, which was developed using responses from groups 1A, 1B, 2, and 3. ....	91
Figure 5.7. Permissions to reproduce manuscript in its entirety.....	97
Figure 5.8. The racial and/or ethnic identities of the students from each of the four groups used in this study. We did not collect the racial and/or ethnic identities from the students of Institution 2, so instead we report the information for the entire undergraduate body for that academic year from the registrar's website. For Groups 1A, 1B, and 3 some students did not report their demographic information which is why the number of students described in this figure is slightly less than the number of students coded. Due to the racial/ethnic identity categories used by this institution's registrar, for Group 3 the category for Asian students also includes Pacific Islander students. ....	98
Figure 5.9. The percentage of male and female students in each of the four groups used in this study. .	98
Figure 5.10. The percentage of students in each group broken down by their grade level.....	99
Figure 5.11. Histogram of the character lengths of all the student responses in Groups 1A, 1B, 2, and 3 (N = 2,030). ....	100
Figure 5.12. Histogram of the word lengths of all the student responses in Groups 1A, 1B, 2, and 3 (N = 2,030). ....	101
Figure 6.1. The distribution of non-electrostatic (NE), electrostatic causal (EC), and causal mechanistic (CM) explanations to the LDF task from students enrolled in the fall semester of general chemistry (2015-2019).....	126
Figure 6.2. The fall and spring response for the in-person academic years 2015-2016, 2017-2018, 2018-2019, and 2019-2020. The fall semesters are color coded with a light brown background while the spring semesters have a yellow background.....	127

Figure 6.3. The distribution of LDF responses for the in-person and online learning environments (before and after March 2020). .....	130
Figure 7.1. The LDF text and drawing tasks given in fall 2018. We modified the format of these questions to reduce the size for this manuscript. ....	140
Figure 7.2. Both versions of the PE task. The LDF tasks (Figure 7.1) are question 1 in this figure. We modified the format of this question to reduce the size for this manuscript. ....	141
Figure 7.3. Distribution of fall 2018 students' responses to the LDF text task, LDF drawing task, and PE task version 1. ....	147
Figure 7.4: Distribution of fall 2020 students' responses to the LDF text task, LDF drawing task, and PE task version 2. ....	149
Figure 7.5. The adjusted standardized residuals of the cross table comparing the PE and LDF text responses as well as the PE and LDF drawing responses. The cells are color coded according to the scale on the right. Cells with more positive adjusted residuals (blue) have more observed instances than would be expected if there was no relationship between the variables. Cells with more negative adjusted residuals (red) have less observed instances than expected. Bolded adjusted standardized residuals have a magnitude greater than 2.78 (yellow line on the scale) and are considered significant drivers of the initial $\chi^2$ result. ....	152
Figure 7.6. The distribution of correct and incorrect students' predictions about the relative depth of the PE minima (left of black divider) and boiling point (right of black divider). Note that in fall 2018 we did not ask the students to compare the boiling points. ....	154
Figure 8.1. Our iterative process of designing a well-scaffolded assessment task. ....	164
Figure 8.2. The response from Isabella to PL task version 1. ....	171
Figure 8.3. The response from Katrina to PL task version 2. ....	173
Figure 8.4. The response from Conor to PL task version 3 alternate. ....	175
Figure 8.5. The final version of the PL task. ....	177
Figure 8.6. The process of using the analytic rubric to assign a holistic code for each response. ....	180
Figure 8.7. Examples of student engagement in causal mechanistic reasoning. ....	181
Figure 8.8. Wayne's drawing and text responses. ....	182
Figure 8.9. Permissions to reproduce manuscript in its entirety. ....	189
Figure 8.10 Drawing response from Penny. ....	195
Figure 8.11 Drawing response from David. ....	195
Figure 8.12. Response from student V1_OC1_101. ....	199

Figure 8.13. Response from student V1_OC1_102.....	199
Figure 8.14. Response from student V1_OC1_103.....	200
Figure 8.15. Response from student V1_OC1_104.....	200
Figure 8.16. Response from student V1_OC1_105.....	201
Figure 8.17. Response from student V1_OC1_106.....	201
Figure 8.18. Response from student V1_OC1_107.....	202
Figure 8.19. Response from student V1_OC1_108.....	202
Figure 8.20. Response from student V1_OC1_109.....	203
Figure 8.21. Response from student V1_OC1_110.....	203
Figure 8.22. Response from student V1_OC1_111 (Isabella). ....	204
Figure 8.23. Response from student V1_OC1_112.....	204
Figure 8.24. Response from student V1_OC1_113.....	205
Figure 8.25. Response from student V1_OC1_114.....	205
Figure 8.26. Response from student V1_OC1_115.....	206
Figure 8.27. Response from student V1_OC1_116.....	206
Figure 8.28. Response from student V1_OC1_117.....	207
Figure 8.29. Response from student V1_OC1_118.....	207
Figure 8.30. Response from student V1_OC1_119.....	208
Figure 8.31. Response from student V1_OC1_120.....	208
Figure 8.32. Response from student V1_MB_101.....	209
Figure 8.33. Response from student V1_MB_102.....	209
Figure 8.34. Response from student V1_MB_103.....	210
Figure 8.35. Response from student V1_MB_104.....	210
Figure 8.36. Response from student V1_MB_105.....	211
Figure 8.37. Response from student V1_MB_106.....	211
Figure 8.38. Response from student V1_MB_107.....	212

Figure 8.39. Response from student V1_MB_108.....	212
Figure 8.40. Response from student V1_MB_109.....	213
Figure 8.41. Response from student V1_MB_110.....	213
Figure 8.42. Response from student V1_MB_111.....	214
Figure 8.43. Response from student V1_MB_112.....	214
Figure 8.44. Response from student V1_MB_113.....	215
Figure 8.45. Response from student V1_MB_114.....	215
Figure 8.46. Response from student V1_MB_115.....	216
Figure 8.47. Response from student V1_MB_116.....	216
Figure 8.48. Response from student V1_MB_117.....	217
Figure 8.49. Response from student V1_MB_118.....	217
Figure 8.50. Response from student V1_MB_119.....	218
Figure 8.51. Response from student V1_MB_120.....	218
Figure 8.52. Response from student V2Alt_OC2_101.....	219
Figure 8.53. Response from student V2Alt_OC2_102.....	219
Figure 8.54. Response from student V2_OC2_103 (Katrina).....	220
Figure 8.55. Response from student V2_OC2_104.....	220
Figure 8.56. Response from student V2_OC2_105.....	221
Figure 8.57. Response from student V2_OC2_106.....	221
Figure 8.58. Response from student V2Alt_OC2_107.....	222
Figure 8.59. Response from student V2_OC2_108.....	222
Figure 8.60. Response from student V2_OC2_109.....	223
Figure 8.61. Response from student V2Alt_OC2_110.....	223
Figure 8.62. Response from student V2_OC2_111.....	224
Figure 8.63. Response from student V2_OC2_112.....	225
Figure 8.64. Response from student V2_OC2_113.....	225

Figure 8.65. Response from student V2_OC2_114.....	226
Figure 8.66. Response from student V2Alt_OC2_115. ....	226
Figure 8.67. Response from student V2Alt_OC2_116. ....	227
Figure 8.68. Response from student V2Alt_OC2_117. ....	227
Figure 8.69. Response from student V2Alt_OC2_118. ....	228
Figure 8.70. Response from student V2_OC2_119.....	228
Figure 8.71. Response from student V2Alt_OC2_120. ....	229
Figure 8.72. Response from student V2Alt_MB_101. ....	229
Figure 8.73. Response from student V2_MB_102 (Liam).....	230
Figure 8.74. Response from student V2_MB_103.....	230
Figure 8.75. Response from student V2_MB_104 (Paul).....	231
Figure 8.76. Response from student V2_MB_105.....	231
Figure 8.77. Response from student V2_MB_106.....	232
Figure 8.78. Response from student V2_MB_107.....	232
Figure 8.79. Response from student V2_MB_108. Note that the student provided a drawing rather than an explanation in response to “Explain what causes glucose to bind to the protein.” .....	233
Figure 8.80. Response from student V2_MB_109.....	233
Figure 8.81. Response from student V2Alt_MB_110. ....	234
Figure 8.82. Response from student V2_MB_111.....	234
Figure 8.83. Response from student V2_MB_112. Note that the student provided a drawing rather than an explanation in response to “Explain what causes glucose to bind to the protein.” .....	235
Figure 8.84. Response from student V2_MB_113.....	235
Figure 8.85. Response from student V2_MB_114.....	236
Figure 8.86. Response from student V2Alt_MB_115. ....	236
Figure 8.87. Response from student V2Alt_MB_116. ....	237
Figure 8.88. Response from student V2Alt_MB_117. ....	237
Figure 8.89. Response from student V2Alt_MB_118. ....	238

Figure 8.90. Response from student V2_MB_119.....	238
Figure 8.91. Response from student V2Alt_MB_120. ....	239
Figure 8.92. Response from student V3Alt_GC2_101. ....	239
Figure 8.93. Response from student V3Alt_GC2_102 (Claudia). ....	240
Figure 8.94. Response from student V3_GC2_103.....	240
Figure 8.95. Response from student V3Alt_GC2_104. ....	241
Figure 8.96. Response from student V3Alt_GC2_105 (Conor). ....	241
Figure 8.97. Response from student V3Alt_GC2_106. ....	242
Figure 8.98. Response from student V3Alt_GC2_107. ....	242
Figure 8.99. Response from student V3_GC2_108.....	243
Figure 8.100. Response from student V3_GC2_109.....	243
Figure 8.101. Response from student V3Alt_GC2_110. ....	244
Figure 8.102. Response from student V3Alt_GC2_111. ....	244
Figure 8.103. Response from student V3_GC2_112.....	245
Figure 8.104. Response from student V3_GC2_113.....	245
Figure 8.105. Response from student V3_GC2_114.....	246
Figure 8.106. Response from student V3Alt_GC2_115. ....	246
Figure 8.107. Response from student V3Alt_GC2_116. ....	247
Figure 8.108. Response from student V3_GC2_117.....	247
Figure 8.109. Response from student V3Alt_GC2_118. ....	248
Figure 8.110. Response from student V3_GC2_119.....	248
Figure 8.111. Response from student V3_GC2_120.....	249
Figure 8.112. Response from student V3_GC2_121.....	249
Figure 8.113. Response from student V3_GC2_122.....	250
Figure 8.114. Response from student V3_GC2_123.....	250
Figure 8.115. Response from student V3Alt_GC2_124. ....	251



Figure 8.116. Response from student V3_GC2_125.....	251
Figure 8.117. Response from student V3_GC2_126.....	252
Figure 8.118. Response from student V3_GC2_127.....	252
Figure 8.119. Response from student V3Alt_GC2_128. ....	253
Figure 8.120. Response from student V3Alt_GC2_129. ....	253
Figure 8.121. PL task version 1 slide one of three. ....	254
Figure 8.122. PL task version 1 slide two of three. ....	255
Figure 8.123. PL task version 1 slide three of three.....	256
Figure 8.124. PL task version 2 (original and alternate) slide one of four. ....	257
Figure 8.125. PL task version 2 (original and alternate) slide two of four. ....	258
Figure 8.126. PL task version 2 (original) slide three of four. ....	259
Figure 8.127. PL task version 2 (original) slide four of four. ....	260
Figure 8.128. PL task version 2 (alternate) slide three of four. ....	261
Figure 8.129. PL task version 2 (alternate) slide four of four. ....	262
Figure 8.130. PL task version 3 (original) slide one of two. ....	263
Figure 8.131. PL task version 3 (original) slide two of two. ....	264
Figure 8.132. PL task version 3 (alternate) slide one of two.....	265
Figure 8.133. PL task version 3 (alternate) slide two of two. Note: we accidentally included protein B from the original version of this slide. ....	266
Figure 8.134. PL task final version slide one of two.....	267
Figure 8.135. PL task final version slide two of two.....	267
Figure 8.136. Two MB students' responses to PL task version 2 who explain hydrogen bonding as the cause for the protein-ligand binding but attribute the interaction to two different groups and do not explicitly connect to charge or polarity. ....	268

## CHAPTER I – INTRODUCTION

Intermolecular forces (or non-covalent interactions) play an important role in a wide variety of both chemical and biological phenomena.<sup>1</sup> For example, these interactions are key to explaining why ethanol and water boil at different temperatures and how DNA can be replicated while maintaining its genetic code. Additionally, forming and breaking these interactions is associated with changes in energy, another core idea across the two disciplines.<sup>1,2</sup> Unfortunately, despite their importance, prior research has found that students have difficulties understanding intermolecular forces (IMFs).<sup>3,4</sup> That is why, in this thesis, I set out to explore how students use causal mechanistic reasoning to explain IMFs in the context of both chemical and biological phenomena.

Across these studies, I focus primarily on London dispersion forces (LDFs), a type of IMF that occurs between all molecules, though I explore students' understanding of other IMFs in the final study. I used the lens of causal mechanistic reasoning throughout this work to make sense of the students' responses. Causal mechanistic reasoning is a powerful form of thinking in which the learner leverages the properties and behaviors of the entities at a scalar level below a phenomenon to explain *why* and *how* the phenomenon occurs.<sup>5,6</sup> In the context of IMFs, this would require a discussion of the electrostatic nature of the entities at the subatomic or atomic level and how those charges would lead to attractions and repulsions producing the overall phenomenon.

The first three studies focus on the use of a prompt and coding scheme we previously developed to elicit and characterize causal mechanistic explanations of LDFs.<sup>7</sup> In this prompt, we asked students what happens when two neutral noble gas atoms (e.g., helium, argon) approach one another. That is, an explanation for why and how LDFs form. An ideal causal mechanistic explanation would step down a scalar level and explain how random and induced fluctuations in electron density produce temporarily charged poles; the oppositely charged ends of two atoms then attract one another. We developed a

corresponding coding scheme to characterize the degree to which students engaged in causal mechanistic reasoning.<sup>7</sup> In the first study, I used these materials to explore how the scaffolding of the prompt impacted the students' responses and how the students' responses changed during the general chemistry course.

The next two studies focused on automating the analysis of students' explanations of LDFs. In the second study, we collaborated with a research group specializing in automated analysis of students' written responses to develop a computer model capable of analyzing students' responses like a human coder. After training this model with over 1,500 human-coded responses, we found this resource could code new responses accurately (compared to human coding). We then used this resource in the third study to analyze tens of thousands of responses to answer questions about the long-term impact of the curriculum on students' understanding of LDFs and the impact of the emergency shift to online learning caused by the COVID-19 pandemic.

The final two studies move beyond an exploration of students' understanding of IMFs in chemistry. In study four, we explore the connection between LDFs and the corresponding changes in potential energy. That is, how does having a causal mechanistic understanding of this IMF impact students' explanations of the depth of the potential energy minimum that results from the formation of LDFs. In the final study, we shift from a chemical phenomenon to a biological phenomenon that is also governed by IMFs: protein-ligand binding. Unlike the LDF prompt, which we initially developed prior to this thesis, this task was entirely new. That is why in this final study we provide an in-depth look into the iterative task development process we used to produce a prompt capable of eliciting causal mechanistic explanations of protein-ligand binding and how we designed the corresponding coding scheme to make sense of the students' responses.

Taken as a whole, these studies illuminate how students use causal mechanistic reasoning to think about IMFs. By understanding the ways in which students reason through these phenomenon, we can create learning opportunities for students to activate productive resources associated with IMFs. Additionally, this can reveal ways in which we can help students connect their understanding of IMFs to other related important ideas like energy and electrostatics. Finally, given the importance of IMFs across disciplines, this work can help to support efforts to foster interdisciplinary coherence between chemistry and biology.

### **Study goals and research questions**

#### ***Study 1: Investigating student understanding of London dispersion forces: a longitudinal study***

In the first study, we asked a cohort of students (N = 150) to explain how and why LDFs form between two neutral atoms at four timepoints over the general chemistry course sequence. In addition to the varying lengths of time between initial instruction and the administration of the task, at each timepoint we varied the scaffolding of the prompt as well as the stakes of the assessment (low vs high). This longitudinal study explored how students were using this causal mechanistic reasoning to explain LDFs throughout the course, revealing how their explanations changed over time. Additionally, our analysis uncovered how the prompt scaffolding and nature of the assessment affected the students' responses. Our specific research questions were:

1. How does scaffolding impact student responses to prompts about LDFs?
2. How do students' explanations and models of LDFs change over a two-semester sequence?
3. How do students' written explanations and drawn models correlate with each other?

#### ***Study 2: Developing computer resources to automate analysis of students' explanations of London dispersion forces***

In this study, we adapted the coding scheme used in study 1 for use with automated analysis technology. We collaborated with a research group specializing in machine learning research and used

their Constructed Response Classifier (CRC) tool to develop computer models capable of analyzing students written work like a human would. To “train” the computer model, we coded over 1,500 students’ responses, collected from three institutions, so the model could detect the lexical patterns associated with each of the codes characterizing the use of causal mechanistic reasoning. We then used this model to code new students’ responses and evaluate the overall accuracy between the human and computer-predicted codes. The research questions of this study were:

1. How does the machine coding for causal mechanistic explanations compare to humans?
2. How does the machine coding for different groups of students compare?

***Study 3: Using machine learning resources to explore the long-term impact of CLUE on undergraduate students’ explanations of London dispersion forces***

Using the model developed in study two, here we analyzed over 20,000 students’ responses collected across eight years to quickly and accurately assess how students were using causal mechanistic reasoning to explain the formation of LDFs. Analyzing such a large data set allowed us to explore questions that would be more difficult and resource intensive to conduct using human coding methods. Specifically, we evaluated the long-term impact of the adoption of the transformed general chemistry curriculum Chemistry, Life, the Universe, and Everything (CLUE) by exploring the consistency of the students’ responses over time. Additionally, by collecting thousands of students’ responses from both the “on” and “off” sequence general chemistry course offerings, we explored potential differences between those two groups of students. Finally, we explored the effect of the emergency shift to online learning in response to the emergence of COVID-19. The research questions were:

1. How has the long-term adoption of CLUE impacted students use of causal mechanistic reasoning to explain the attraction between neutral species (i.e., LDFs)?
2. How do “on” and “off” sequence students compare in their explanations of LDFs?

3. How has the pandemic-induced shift to online learning impacted students' explanations of LDFs?

***Study 4: Exploring connections between students' explanations of London dispersion forces and potential energy***

In study four, we explored the connections between how students think about interactions and subsequent changes in potential energy (PE) that result from the formation of these interactions. To do this, we used our previously developed LDF prompt and coding scheme to explore how students used causal mechanistic reasoning to reason through the formation of this interaction. We then developed a new task and coding scheme capable of eliciting and characterizing causal mechanistic explanations of the potential energy changes which result from the formation of LDFs. Specifically, how the potential energy minimum differs for two pairs of interacting atoms experiencing LDFs of different strengths. From these two tasks, we could compare how the same students used causal mechanistic reasoning to explain both phenomena. The research questions for this study were:

1. How do students use causal mechanistic reasoning to explain differences in PE minima of interacting neutral species?
2. How do students' causal mechanistic explanations of the formation of LDFs compare to their explanations of differences in PE minima?
3. How do students' causal mechanistic explanations about LDFs and the depth of PE wells impact students' responses about associated macroscopic phenomena?

***Study 5: A deep look into designing a task and coding scheme through the lens of causal mechanistic reasoning***

One reason why IMFs are so important is the key role they play in phenomenon across disciplines. While the other four studies focused on chemical phenomena, in this final study we set out to design a task and corresponding coding scheme situated in a biological phenomenon: protein-ligand binding. While in studies 1-4, we used a previously developed task, here we developed an entirely new

task. To design this task, we used a modified evidence-centered design approach in which we initially defined what evidence we would accept for causal mechanistic understanding of this binding (specifically, the electrostatic interactions between the charged atoms of the protein and ligand), but then we iteratively refined the task based on how the students responded. We explored different scaffolding techniques, like contrasting cases, to design an appropriate prompt. We then used the responses from the final prompt and our understanding of causal mechanistic reasoning to develop a scheme to characterize the degree to which the student used causal mechanistic reasoning in their response. In this process, we highlight effect strategies for iteratively developing effective assessments while also producing materials we can later use to explore students understanding of protein-ligand binding. The research questions were:

1. What is the impact of different types of scaffolding on the resources students use to respond to the task?
2. In what ways can we characterize the degree to which students are engaging in CMR to explain this phenomenon?

## CHAPTER II - THEORETICAL FRAMEWORKS

### How people learn: the resources perspective

#### *Constructivism*

In order to support students' education, it is vital that we have a working model for how people learn. Our research is, at its core, based on constructivism, which proposes that new knowledge must be constructed from old knowledge and that this process must be carried out by the learner.<sup>8,9</sup> Jean Piaget is credited with developing this theory at a time when the dominant ideology was stimulus-response theory.<sup>8,10</sup> That is, only through rote association between a specific event and the corresponding response could people "learn". Piaget noted that this theory excluded the role of the learner in the process, even though it was the learner who interpreted the stimulus and produced the response.<sup>9</sup> Piaget's constructivism placed the learner at the center of this model as they engaged in two key steps through which learning occurred: assimilation and accommodation. Assimilation describes the process by which a child applies their existing knowledge structures to a new situation.<sup>8</sup> Accommodation is then when the child modifies their existing knowledge structures to incorporate the information that was just assimilated.<sup>8</sup> While some of Piaget's other ideas, like his stages of development, have become less popular, the idea that we must attend to the knowledge that students bring with them has become widely accepted.

#### *Misconceptions and conceptual change*

While constructivism suggests that knowledge is constructed by the learner, it does not elaborate on *how* that knowledge is constructed. More specifically, it does not provide a mechanism for how wrong ideas, or misconceptions, are corrected. In this space, Posner et al.'s "theory theory" framework emerged as one of the most popular ways to "fix" misconceptions.<sup>10,11</sup> In their view, incorrect student ideas were relatively stable and needed to be replaced by more correct ones.<sup>10,11</sup> To



bring about this change, students had to encounter a situation in which their incorrect ideas would lead them astray, therefore requiring them to replace their incorrect ideas with more correct ones.<sup>10,11</sup>

While this theory is aligned with constructivism in that the student and their existing ideas are still central to the conceptual change process, instruction is primarily focused on creating dissatisfaction with the students' existing ideas. In this sense, "theory theory" takes a deficit approach on student learning, viewing their knowledge as something just to be replaced.<sup>12</sup> Additionally, "theory theory" does not explore why students might be struggling with these new ideas, limiting the ways in which instruction of these new ideas can be supported.<sup>12</sup> In response to these shortcomings, other researchers proposed alternative models for how conceptual change occurs. A common theme across the theories proposed by Minstrell, diSessa, Hammer, and Elby is the fragmented nature of knowledge.<sup>10,12–14</sup> While their theories do differ in some ways, all suggest that students bring with them a set of ideas (facets, raw intuitions, p-prims, resources) which they use to reason through the situation at hand.<sup>10,12–14</sup> What is key across all these theories is that just because an idea is used in an unproductive manner, it does not mean that the idea itself is not valuable.<sup>10,12–14</sup> For example, in many contexts the rule of thumb "more means more" is quite useful—if you wish for something to move further, you push it harder. However, such heuristics do not always apply at the molecular level (e.g., the quantization of energy). This does not mean that the intuitive idea of "more means more" is wrong, it was just used in an inappropriate context. Of all these theories, we have found the resources perspective to be the most useful in understanding how students learn.<sup>12,14</sup>

### ***Resources perspective***

Hammer's resource perspective suggests that students' have resources in their minds which they use to reason through the task at hand. These resources can either be conceptual (focused on the content) or epistemological (relevant to the nature of knowledge).<sup>12</sup> A key feature of this perspective is that these resources are context dependent. As a consequence, resources cannot be inherently right or

wrong, but rather productive or unproductive, depending on the situation.<sup>12</sup> Through this lens, the goal of education is to help students activate productive resources in appropriate contexts, helping the students to build a well-connected framework of knowledge, an important characteristic of expertise.<sup>15</sup>

As an example, consider fire. The context in which one encounters fire dramatically changes which resources are activated. If it is in the setting of a small campfire, this might activate resources which would be useful in cooking foods over an open flame or techniques of maintaining a fire. If instead the fire grows to engulf an entire forest, this will activate entirely different resources—likely, methods of extinguishing it. All these resources are valid, but depending on the context, some are far more appropriate than others. Additionally, even if the context is identical, the same resources will not be activated for everyone. For example, forest fires might lead most of us scrambling for ways to deliver water to the area, but expert fire fighters might incorporate other techniques, like cutting down trees or digging trenches to create fire barriers to manage its spread.

How then can we impact the resources students have and when they are activated? One way is through repeated use in multiple contexts. Over time, these connections build upon one another, so it becomes automatic for those productive resources to be activated in those contexts.<sup>14</sup> Additionally, when one resource is activated, closely linked resources are also activated, similar to diSessa's concept of coordination classes.<sup>10,12</sup> Again, those connections between resources develop through use over time. This perspective has important and actionable implications for instruction and learning. Instruction should identify what students know (i.e., the resources they have) and help them to connect their knowledge in productive ways. This highlights the advantage of using this theory of learning—now we are not solely focused on what students *cannot* do, but instead on what they *can* do.

### Using assessments to collect evidence

Assessments play a critical role in supporting students learning. Specifically, they allow us to understand what students know and can do. Key to this argument is the idea that students' responses act as evidence for what is going on in students' heads.<sup>16</sup> Through the resources perspective, this is evidence of the resources they have, how they are connected, and in what contexts they are activated. Knowing what students know is important for several reasons. First, the theory of constructivism states that new student knowledge is built upon older knowledge, so therefore we must know what resources students have so we can design instruction to create opportunities for students to connect those resources in productive ways.<sup>8,12</sup> Understanding what students know is also important following instruction as well, as this evidence can be used to evaluate the current learning environment and identify areas of improvement. That is, assessments help us make informed decisions to support student learning.

However, all of this is built upon the assumption that the evidence we have collected is an accurate representation of what students know and can do. Through the lens of resources, the assessment is the context which determines which resources are activated. Therefore, it is critical that the assessment activates the appropriate resources necessary to accomplish the task at hand; otherwise, we might falsely assume that the student does not have a particular idea when it may be that they do have productive resources that were not activated. In this endeavor, scaffolding is useful.

Scaffolding is the process in which someone more expert helps someone less knowledgeable accomplish a task they would otherwise be unable to do on their own.<sup>17</sup> Scaffolding is also very similar to Vygotsky's Zone of Proximal Development.<sup>18,19</sup> According to Vygotsky, we can think of three "locations" of student learning. The students' *actual development level* and *potential development level* represent what the student can do on their own and what is beyond their capabilities.<sup>18</sup> Between these two levels is the *zone of proximal development* which represents what the student can do with some

assistance.<sup>18</sup> It is in this zone where learning takes place. Eventually, the students' potential development level becomes their actual development level, and the process repeats. While issues of translation and active repression from the Russian government resulted in the development of these two theories occurring independent of one another, they work quite well together.<sup>19,20</sup> In a sense, scaffolding is the bridge which helps students cross the zone of proximal development.

Methods of scaffolding and targeting the zone of proximal development can be useful in designing assessments that activate the appropriate resources. If we do not provide enough scaffolding, the student might not be clear what they are being asked to do, and we may not get an accurate picture of their knowledge. On the other hand, if we provide too much assistance, the task becomes rote, and their responses will also not be good indicator of their abilities.

### **Causal mechanistic reasoning**

Causal mechanistic reasoning is a type of thinking that is highly valued in science. Reasoning about a phenomenon in this way requires that the learner identifies the entities at a scalar level below the phenomenon, unpacks their properties and behaviors, and then connects those behaviors back up to the overall phenomenon to explain both how and why it occurs.<sup>5,6</sup> This type of reasoning is powerful because it enables the learner to produce rich explanations and make robust predictions. The resources lens helps to reveal why causal mechanistic reasoning is so useful; this type of reasoning is all about making connections—connections between the entities and their properties, the properties and the behaviors that result, and across scalar levels. If students can make those connections, then they can better predict what might happen if the properties of some lower-level entity is changed or how entities with similar properties might behave in a different system.

For example, consider a mechanical camera. Developing a (simple) causal mechanistic understanding how this device works would require knowledge of two key components within the

camera: the aperture and the shutter. When the button is pressed to take a picture, the shutter opens, allowing light through aperture (the hole in the lens) where it is then focused on light-sensitive film which “captures” the image. The aperture size (how big is the hole the light passes through) and the shutter speed (how long light passes through the hole) can both be modified by the user to control certain aspects of the photo, like brightness. Students with a causal mechanistic understanding of how such a camera works would be equipped to predict what would happen if they took a photo with a large aperture size and long shutter speed (very bright, over-exposed photo as lots of light is let through). Similarly, they could predict what actions would be required to take a photo with the appropriate level of brightness (reduce aperture size or shutter speed).

This example highlights how causal mechanistic reasoning is important both in science and everyday life. It is an epistemic heuristic, a rule of thumb that one can use to organize their knowledge in productive ways.<sup>5</sup> Furthermore, it is a decision the learner needs to make—to probe the underlying parts to understand how and why something happens.<sup>5</sup> Emphasizing this epistemic heuristic in our courses is important for two reasons. First, if we help students to engage in this type of reasoning often, they may be able to rely on this resource in their future science classes and beyond. Second, making this type of reasoning explicit is an important part of developing equitable learning environments. As scientists, we implicitly understand the value of mechanisms and can, through years of experience, engage in this process instinctively. It is therefore important that we, as science educators, make epistemic heuristics like this one explicit, so students understand what is expected and valued in these communities. This can help make our courses, and science as a whole, more accessible to all learners.

## CHAPTER III – LITERATURE REVIEW

### Assessment scaffolding and iterative design

Assessments play an important role in students learning. They send a message to the students about what is valued in the classroom.<sup>21</sup> When used in a formative manner, they provide an opportunity for students to connect their ideas in productive ways.<sup>21</sup> They also help us (e.g., researchers, instructors) elicit evidence of what students know and can do.<sup>22</sup> In this section, we will review some literature regarding the role of assessments in collecting evidence. Specifically, how the structure and design of the assessment impact the quality and accuracy of the evidence collected. Two areas of research which researchers have drawn on in this domain are scaffolding and iterative design.

Scaffolding, in brief, is the support which allows a learner to accomplish a task that they would otherwise be unable to complete on their own.<sup>17</sup> Wood et al.<sup>17</sup> first proposed scaffolding in the context of tutoring young children, helping them to solve a puzzle involving a series of wooden blocks. Based on their observations, they identified six strategies that the tutor used to scaffold their experience: *recruitment, reduction in degrees of freedom, direction maintenance, marking critical features, frustration control, demonstration*.<sup>17</sup> Some of these roles are more applicable for younger children, like *recruitment, direction maintenance, and frustration control*, but the other steps are quite relevant for undergraduate students. *Reduction in degrees of freedom* involves the tutor limiting what the student is responsible for so that they are not overwhelmed and can focus on the parts of the task which they can meaningfully work through (while the tutor deals with the more complex parts). Similarly, when the tutor *marks critical features* they focus the students' attention to the most productive aspects of the task. Finally, *demonstration* is a tool the tutor can use to exemplify certain behaviors and processes.

While Wood et al.<sup>17</sup> focused on in-person tutoring, these techniques have also been used to guide assessment design.<sup>20</sup> For example, Reiser<sup>23</sup> described how an online learning software provided

scaffolding to support students constructing scientific explanations. This software acted as a “digital journal”, breaking down the complex process of building an explanation into a series of components.<sup>23</sup> In this way, they sought to *reduce the degrees of freedom* and shrink students’ cognitive load so they could better engage with the meaningful parts of the activity. Similarly, McNeil et al.<sup>24</sup> constructed a scaffolded activity to help students construct explanations by explicitly prompting for claim, evidence, and reasoning. The researchers found the students’ reasoning significantly improved when they provided the students with additional support via the scaffold.

However, figuring out how much scaffolding to provide can be a difficult task.<sup>23</sup> The scaffolding should provide *just* enough structure to activate the appropriate resources so the student can accomplish the task at hand, but not so much that they can complete the task without thoughtful effort. Using students’ responses to the activity can provide useful information to guide iterative refinement of the scaffolding to strike this balance. Even beyond scaffolding, iterative refinement is a useful tool for all assessment design.<sup>25</sup> For example, Brandriet et al.<sup>26</sup> designed a task in which students had to use hypothetical data from kinetic experiments to construct a rate law, modeling the relationship between the rate of the reaction and the concentration of the reactants. In their prior study<sup>27</sup> they found that students struggled when the experimental data did not fit the model exactly, so they used data which did not yield an integer exponent.<sup>26</sup> By using numbers that did not “work out” cleanly, they could elicit nuanced explanations of students’ understanding of what the rate law represents. Similarly, Bishop and Anderson<sup>28</sup> designed an assessment to probe students understanding of natural selection. They gave their assessment to a group of students, analyze the responses, and then modified the assessment accordingly; for example, by decreasing the number of constructed response questions and increasing the number of multiple-choice questions.<sup>28</sup> Then, to evaluate those new changes, the process was then repeated with a new group of students and new assessment modifications until their assessment had been optimized. Others have also used this approach as well to design assessments about acid base

chemistry and carbon cycles.<sup>29,30</sup> This process highlights the utility of iterative refinement. It is a fact that we do not know how students will interact with our assessments until we try them out and evaluated the responses.

### **Eliciting and characterizing causal mechanistic reasoning**

Causal mechanistic reasoning requires the learner to address *how* and *why* a phenomenon occurs by leveraging the entities a scalar level below the phenomenon.<sup>5,6</sup> In this thesis, we rely primarily upon the definition of causal mechanistic reasoning put forth by Krist et al.<sup>5</sup> Their framework, which is adapted from the work of Russ et al.,<sup>6</sup> identifies three key features of this type of reasoning.<sup>5</sup> First, the learner must step down to the scalar level below the phenomenon to identify the relevant entities at that level. Next, they must identify the properties of those entities and their resulting behaviors. Finally, they must link those behaviors of these entities back up the phenomenon. In order to support students' engagement in causal mechanistic reasoning, it is important that we can design assessments which elicit causal mechanistic reasoning. This would allow us to gather evidence about what practices best support this type of reasoning.

Engaging in causal mechanistic reasoning is a difficult task for students. Bodé et al.<sup>31</sup> explored how students leveraged reaction coordinate diagrams, comparing two first-order nucleophilic substitution reactions, to predict which reaction would occur. They hoped that students would compare the reactants (primary vs tertiary alkyl halide) and use ideas related to the transition state, rate determining step, and hyperconjugation to explain why the reaction with the tertiary alkyl halide would be more likely to proceed.<sup>31</sup> However, they found that few students addressed the “granularity” (their term for the appropriate scalar level) necessary to support their prediction.<sup>31</sup> Additionally, none of the students provided the highest level of reasoning, multicomponent causal—considering the dynamic nature of the system, in their explanation.<sup>31</sup> Macrie-Shuck and Talanquer<sup>32</sup> also found that students struggled to use causal mechanistic reasoning. They conducted semi-structured interviews with



students, exploring their understanding of energy transfer and transformation in chemical and physical processes.<sup>32</sup> They found that while students could identify some correct associations between ideas (for example, that potential energy decreases when bonds form) they could not explain *how* or *why* the phenomena occurred.<sup>32</sup>

Even when students can engage in this type of reasoning, eliciting these kinds of responses is no small task. One of the biggest challenges is encouraging students to go beyond providing just a description of the phenomenon and addressing *how* and *why* it occurs. Cooper et al.<sup>29</sup> discussed the iterative process by which they developed a prompt capable of eliciting a causal mechanistic explanation of the reaction between water and hydrochloric acid (i.e., an acid base reaction). Often, these types of reactions are only described as the transfer of a hydrogen from the acid to the base (i.e., using the Bronsted-Lowry theory). However, in order to explain both *how* and *why* this reaction occurs, the student must discuss the movement of the electrons from the base to the acidic proton, and that this is caused by an electrostatic attraction between opposite charges. Based on pilot work with this task, the researchers learned that students were less likely to discuss the electrons if the structures of water and hydrochloric acid were condensed in the prompt.<sup>29</sup> To encourage students to think about the role of the electrons, they used the full Lewis structures in which the lone pairs are explicit.<sup>29</sup> After administering this first version of the activity, they found that students were primarily *describing* what was occurring instead of engaging in causal mechanistic reasoning like they had hoped.<sup>29</sup> In response, the researchers modified the questions so that students were asked to both explain *what* was occurring and also *why* the reaction occurred in two separate questions.<sup>29</sup> They found that by asking *what* and *why* in separate questions, they could elicit about 5 times the number of fully causal mechanistic explanations.<sup>29</sup> That is, through iterative design they designed a prompt that elicited not just *what* occurred, but *why* and *how* it occurred as well.

Other researchers have identified additional scaffolds that can support students causal mechanistic reasoning. One such technique is the use of contrasting cases. This has been used by researchers to help students reason through a variety of phenomenon including phase changes, structure-property relationships, and substituent effects on reaction speed.<sup>33–35</sup> In this format, the contrasting cases and associated questions are used to help students (1) identify the differences between the two contrasting cases, (2) reflect on the impact of that difference, and (3) link the effect of the difference to some overall change.<sup>34</sup> This aligns well with the steps of causal mechanistic reasoning as students are encouraged to think about the properties of the underlying entities (i.e., the point of difference between the two cases).

Using these types of scaffolds provide useful ways to elicit causal mechanistic reasoning, but we need to be able to make sense of these responses as well. That is, we need to characterize the degree to which students engage with causal mechanistic reasoning. Moreira et al.<sup>36</sup> used a prompt asking students to explain why water does not freeze when alcohol is added. While they do not delve into much detail about the development of this question, they do discuss how they characterized the responses. In their work, they primarily leaned on the definition of causal mechanistic reasoning laid out by Russ et al.<sup>6</sup> First, they tried to identify the components of causal mechanistic reasoning that students were using in their explanations. They identified the categories of *entities*, *properties*, *activities* (behaviors which cause changes to entities or the phenomenon), and *organisation* [sic] (when/where the entities are during an activity).<sup>36</sup> They defined the types of relationships between the categories and used a series of levels to characterize their reasoning. At the lower levels, only *entities* and/or their *properties* are included in the explanation.<sup>36</sup> However, at the upper levels, students addressed the *activities* and/or *organisation* [sic] of the entities.<sup>36</sup> That is, these researchers identified those students who connected how the entities and their properties produce the observed phenomenon as more fully engaging in this practice.

This is a type of thinking that is highly valued in science because it enables the learner to predict and explain phenomenon.<sup>5,6</sup> However, figuring out how to elicit this type of reasoning and make sense of students' responses is not trivial. Recent research in this field has uncovered productive ways of engaging students in causal mechanistic through assessment scaffolding and careful considerations of the connections students make between the underlying entities, their properties, and the overall phenomenon.

### **The importance of forces and interactions**

Forces and interactions are a core idea in chemistry. In conjunction with the other core ideas of structure and properties, energy, and change and stability, they explain how and why atoms and molecules behave in the ways that they do, and how these behaviors in turn give rise to many chemical phenomena.<sup>37,38</sup> In the *Framework for K-12 Science Education*,<sup>1</sup> the National Research Council also includes forces and interactions as one of the four core ideas in the physical sciences. While this core idea encapsulates both covalent (bonds) and non-covalent interactions (intermolecular forces), we will be focusing on non-covalent interactions in this thesis.

Intermolecular forces (IMFs) are the pushes and pulls which occur between electrostatically charged species—opposite charges attract, while like charges repel. Even though this a fundamental idea in chemistry, studies have shown that IMFs are difficult for students to understand. For example, Henderleiter et al.<sup>3</sup> interviewed undergraduate students at the end organic chemistry and found that despite all their chemistry instruction, these students still struggled with hydrogen bonding interactions, a commonly discussed IMF. Students were confused about where this interaction occurred, possibly due to the “bonding” name, or why these interactions occurred between particular atoms.<sup>3</sup> The researchers found that most students were not using chemical ideas like charge or electronegativity to determine where and why these interactions formed; instead, the students primarily relied on memorization.<sup>3</sup>

Other studies have also found that students struggle with IMFs, specifically in regard to their occurrence between, rather than within, molecules.<sup>4,39</sup>

This issue extends beyond chemistry; ideas about forces and interactions can create difficulties in biology as well. One of the most well-documented “misconceptions” in biology is the idea that breaking bonds, specifically the phosphate bonds in adenosine triphosphate (ATP), releases energy.<sup>40–42</sup> While this simplification of the role of ATP is productive in biology, it contradicts how the relationship between interactions and energy is discussed in chemistry, where breaking bonds or interactions *requires* energy. Kohn et al.<sup>43</sup> explored this cross-discipline connection (or lack thereof) by interviewing undergraduate students co-enrolled in both general chemistry and molecular biology about their perceptions of the “big ideas” in each course. One of the main ideas that students discussed was the energy changes associated with the forming and breaking of bonds and interactions.<sup>43</sup> Kohn et al.<sup>43</sup> found that the majority of the students who discussed such energy changes in the context of biology felt that these ideas were in conflict with what they were taught in chemistry. For example, one student explained that “...ATP, when the bond’s broken, energy is released...when one of the phosphates is broken off, it releases energy. That’s, I think, what’s getting me. Because when the bond’s broken, it should absorb energy. So I’m getting very confused.”<sup>43</sup>

The confusion surrounding forces and interactions is troublesome, but the issue becomes much more dire when you consider the connection with energy and disconnect in how these ideas are discussed across disciplines. The goal of science education is to help students make sense of world around them, and this understanding should be coherent and consistent across the science disciplines. While different disciplines may focus on different aspects of a system, they all still describe the same system. Another student from Kohn et al.’s<sup>43</sup> study described the danger of this disconnect quite well: “I feel like I can ration [sic] it out both ways, so then, I understand why someone would say either/or, but then I know for biology what [the instructor] wants us to say and then for chemistry what we have to

say". It appears that inconsistencies in regarding how force and interactions are discussed (and the associated changes in energy) actively prevent students from developing an interdisciplinary understanding of science. How can we remedy this? That is, how can we help students develop a deeper understanding of these interactions? Recent studies suggest that the transformed general chemistry curriculum Chemistry, Life, the Universe, and Everything (CLUE) is helping students better understand forces and interactions.<sup>37</sup>

### **CLUE and the impact on forces and interactions**

As our knowledge of chemistry has increased, so has the list of topics traditionally covered in chemistry courses. The result is that general chemistry has become "a mile wide and an inch-deep".<sup>44,45</sup> Attempting to cover the maximum number of topics in the course inherently requires that each idea is covered in less depth. The trouble is that becoming an expert, building meaningful connections between important disciplinary ideas, requires robust, in-depth instruction.<sup>15</sup> To help our students develop a deep understanding of chemistry, researchers have proposed a few specific core ideas which run throughout chemistry, as discussed earlier. All new topics covered in the course can then relate back to these core ideas, providing students with a way to organize their knowledge. After all, one of the major differences between experts and novices is that experts have well-organized knowledge.<sup>15</sup> CLUE is based around these core ideas, thereby specifically focusing on bonding and interactions which, as discussed above, can be a difficult area for students.

Studies have shown that enrollment in CLUE has positive outcomes for students, especially with regard to their understanding of bonding and interactions. For example, Williams et al.<sup>46</sup> asked both traditional and CLUE general chemistry students where IMFs (hydrogen bonding, dipole-dipole, and LDFs) occurred. They found that the majority of CLUE students correctly indicated that such interactions occurred between molecules while the majority of traditional students said the opposite, that these IMFs occurred within molecules.<sup>46</sup> Furthermore, when they followed a subgroup of students

through organic chemistry they found that these results were maintained.<sup>46</sup> That is, even after four semesters of introductory chemistry, the CLUE students could still identify where these interactions occurred while the traditional students could not.

In another study, researchers adapted CLUE for the high school environment and studied how students in the high school CLUE course compared to students in a traditional course and an alternative “modeling” course (a transformed curriculum centered around different models of the atom and their evolution over time).<sup>47</sup> They administered an activity to the students in which students had to compare the Lewis structures of ethanol and dimethyl ether and draw the hydrogen bonds between several ethanol molecules.<sup>47</sup> Then, they were asked to explain why the presence of hydrogen bonding between the ethanol molecules results in ethanol having a higher boiling point. They found that about half of the CLUE and modeling students could draw the hydrogen bonds as between molecules compared to just 16% of the traditional students.<sup>47</sup> However, when it came time to link those hydrogen bonds to the boiling point, a third of the CLUE students leveraged the strength of the interactions to this phenomenon and another third brought in the role of energy as well.<sup>47</sup> For the traditional and modeling students though, fewer than 10% of those students fell into either of those two categories.<sup>47</sup>

These studies suggest that CLUE can have a meaningful impact on students understanding of interactions. However, these studies have primarily focused on the location of the interactions, not how or why they form. That is, the causal mechanism by which these interactions occur. Prior to this thesis, we explored how CLUE students used causal mechanistic reasoning to think about the formation of the London dispersion forces.<sup>7</sup>

#### **Prior research on CM reasoning and LDFs**

Previously, we conducted a study exploring how students used causal mechanistic reasoning to think about one intermolecular force in-depth: the London dispersion force or LDF.<sup>7</sup> We focused on how

this interaction arises in systems of neutral atoms (e.g., helium, argon) because this interaction occurs between all atoms, and the strength of this interaction is only dependent on the number of charged subatomic particles. We began by interviewing twelve CLUE general chemistry students, building upon the work conducted by Williams et al.<sup>46</sup> In those interviews, we asked students not just to draw where these interactions occurred (which Williams et al.<sup>46</sup> investigated), but to also explain how and why these interactions occurred. Based on their responses, we generated a set of codes characterizing the degree to which students engaged in causal mechanistic reasoning to explain or draw the formation of LDFs.<sup>7</sup>

Based on this characterization, we designed a homework activity which we administered to an entire CLUE general chemistry class to get a sense of how a broader group of students would explain the formation of this interaction. We then analyzed the responses using the approach developed from the interview portion. Based on this analysis, we found that over half of the students identified the electrostatic cause to this attraction, while another 30% discussed how the temporary shift in electron distribution led to the charged poles (i.e., a fully causal mechanistic explanation).<sup>7</sup> However, the students' drawings told a slightly different story. In the drawings, practically none of the students showed how the movement of electrons resulted in temporary dipoles; instead, while about 80% of the drawings indicated the presence of an electrostatic attraction, few showed the mechanism by which neutral atoms developed these charges.<sup>7</sup>

Overall, this study showed that CLUE can lead students to think deeply and mechanistically about LDFs. However, it also showed that there are still many more questions to answer (e.g., why were there no causal mechanistic drawings?). This study was the jumping off point for all the studies included in this thesis. As we explore how students use causal mechanistic reasoning to think about forces and interactions, we hope the information we uncover can better support students' learning going forward.

## CHAPTER IV - INVESTIGATING STUDENT UNDERSTANDING OF LONDON DISPERSION FORCES: A LONGITUDINAL STUDY

### Preface

In this longitudinal study, we investigate how students use causal mechanistic reasoning to explain and illustrate the origins of LDFs. This research has been previously published in the Journal of Chemical Education and is reprinted with permission from Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, 96 (9), 1821–1832. Copyright 2019 American Chemical Society.

A copy of permissions obtained is included in the Appendices. Supporting Information for this manuscript is included in the Appendices.

### Introduction

Understanding the causes and consequences of intermolecular forces (IMFs) (or more generally noncovalent interactions) is crucial to an understanding of the macroscopic behavior of molecular species. From phase changes to the behavior of complex biological systems, the interactions and concomitant energy changes that arise from transient or permanent electron density differences play a central role in predicting and explaining how molecular level systems behave. Unfortunately, there is ample evidence that the idea that molecules (or parts of molecules) can “stick together” without being permanently bonded appears to be difficult for students.<sup>3,4</sup> In prior work we found that students in traditional general chemistry courses rarely were able to draw and explain the types of IMFs typically taught in general chemistry in a coherent fashion.<sup>39,46</sup> While students could describe IMFs in textbook terms, they were unable to translate these ideas into a visual representation. That is, many students drew not only hydrogen bonding, but also dipole-dipole interactions and London dispersion forces (LDFs), as being located within a (small) molecule, rather than between molecules.<sup>39</sup> This may well be one source of the problematic ideas about how substances undergo phase changes. For example, if



students are told that when water boils hydrogen bonds are broken, then we cannot be surprised when they interpret phase changes as breaking covalent bonds.<sup>48–50</sup>

While the mechanisms by which the electron density of participating molecules becomes unequally distributed are somewhat different for each of these IMFs, the ultimate consequence is the same. For example, if the temperature of the system is reduced, partial charges on different molecules (or different parts of the same molecule) will attract each other, and when the kinetic energy of the molecules is reduced sufficiently, this will eventually lead to a phase change (e.g., liquid to solid) during which IMFs are formed with the release of energy to the surroundings. Overcoming such IMF interactions requires an input of energy, resulting in a phase change (e.g., solid to liquid). Extending this reasoning to chemical reactions leads to an understanding of how the reacting molecules are attracted to initiate the bond breaking/bond formation steps of a chemical reaction. For this reason, we believe that helping students understand both the causes and consequences of IMFs is crucial for students to have a deep and useful knowledge of chemistry.

There are two main mechanisms by which molecules acquire an unequal distribution of electrons, which then results in attractions between molecules. Understanding how molecules acquire a permanent dipole, resulting in hydrogen bonding and dipole-dipole interactions, requires a long chain of inference involving the relative effective nuclear charge of the participating atoms and the overall shape of the molecule. That is, to understand these IMFs students must have a working knowledge of bonding, VSEPR, molecular shape, and electronegativity.<sup>51</sup> Although students can (and do) take cognitive shortcuts such as identifying electronegative atoms to determine whether a molecule is polar, these shortcuts may lead to erroneous ideas such that carbon dioxide is polar.<sup>52</sup> The language used to refer to hydrogen bonding adds another problematic issue, since hydrogen “bonds” are not covalent bonds, which is also confusing to students.<sup>3,4,39</sup>

In contrast, understanding how LDFs arise requires less supporting knowledge of molecular structure, shape, and electronegativity. While students must be able to use an “electron cloud” atomic model, they need only understand that this electron density can fluctuate producing transient charges which may then induce fluctuations in nearby molecules, causing these molecules to interact. For this reason, LDFs are introduced early in the transformed general chemistry curriculum, “Chemistry, Life, the Universe & Everything”(CLUE).<sup>37,53</sup> LDFs are important noncovalent interactions; they operate between all molecular species, and while they are often discussed as being the “weakest” of all the IMFs, in fact they are dependent on molecular size and provide a significant contribution to the total interaction forces for large molecular species.<sup>54</sup>

### ***Causal Mechanistic Reasoning and LDFs***

Mechanistic reasoning “is a powerful thinking strategy that allows one to explain and make predictions about phenomena.”<sup>5</sup> Russ and co-workers propose that “...mechanisms account for observations by showing that underlying objects cause local changes in the system by acting on one another”.<sup>6</sup> Krist et al. in their proposed framework provide epistemic heuristics to support mechanistic reasoning.<sup>5</sup> They emphasize that an essential piece of a mechanistic explanation is the term “underlying objects”; that is, mechanistic reasoning involves thinking across scalar levels, and the “underlying objects” are at least one scalar level below the phenomenon to be explained. In chemistry, we are often discussing phenomena at the atomic-molecular level (e.g., why do molecules react?); in our own work then we have taken the underlying objects at a scalar level below molecules to be electrons. That is, a mechanistic explanation in chemistry should involve the behavior of electrons. While this may align with the organic chemistry idea of mechanisms, it involves much more than simply describing (or drawing) how electrons move during molecular level events but should also encompass the causal reasoning for this behavior.<sup>6</sup> Designing activities that require such mechanistic reasoning can help students connect

the chain of inferences that are required to use molecular level structure to predict and explain observable macroscopic events.<sup>7,29,51</sup>

While some researchers use the term mechanistic thinking to encompass both the causes and the underlying mechanisms by which students can reason, in our work on chemical phenomena we have seen that students sometimes provide causes without mechanisms and mechanisms without causes.<sup>7,29</sup> For example, in our previous work on the development of student reasoning about LDFs,<sup>7</sup> we saw that some students explained the idea that nonpolar substances can “stick together” as the attraction between oppositely charged parts of the individual molecules or atoms. That is, the students are able to identify the cause of the attraction, but not the mechanism by which the charge separation in the nonpolar molecule arises. For example, “Atoms attract to each other [because] the [London] dispersion forces make the partially negative end of one atom attract to the partially positive end of another atom.”<sup>7</sup> In contrast a causal mechanistic explanation involves the development of an instantaneous dipole as electron density rapidly fluctuates. The electrons are the “underlying entities” that produce the cause of the interaction. For example, “The atoms attract because of the London dispersion forces between the two atoms in which an atom with [instantaneous] dipole causing another atom to become an induced dipole. In this scenario, one atom has its electrons shift to certain region of the electron cloud, giving that region a negative charge. This negative region then induces the movement of electrons in another atom, being attracted to the new present positive region of the second atom.”<sup>7</sup>

Alternatively, in our studies on student understanding of acid-base chemistry, some students are able to articulate a mechanism using electron movement (that is they use a Lewis acid-base model) but do not explain why the electrons move in that particular way as a result of the electrostatic interactions.<sup>29</sup> For example, “The O in the H<sub>2</sub>O gives its electrons to the H in the HCl bond, and simultaneously the HCl bond breaks, placing those electrons onto the Cl. This reaction happens because it is more favorable.”<sup>29</sup> In contrast, a causal mechanistic explanation explains why the electrons move in

that way, “The lone pair on the water molecule attracts the Hydrogen from the HCl. The H-Cl bond is broken and forms a new bond with oxygen. The reaction occurs because the partial negative charge on the oxygen attracts the partial positive charge on the hydrogen.”<sup>29</sup>

Because of these nuances in which causality and mechanism may appear separately in student explanations, we have chosen to separate these ideas in the coding of the explanations. So, while some researchers include causality when they discuss mechanistic explanations, in our work we refer to causal explanations, mechanistic explanations, and causal mechanistic explanations where appropriate, with an implicit assumption that causal mechanistic explanations are the most sophisticated.<sup>7,29</sup> In the case of student reasoning about LDFs, we see only causal and causal mechanistic reasoning, because in our earlier work<sup>7</sup> we did not see any students discussing the mechanism for the separation of charge without also discussing the consequences: That is, the resulting dipoles are attracted to each other. In contrast, in our characterization of acid-base reaction reasoning,<sup>29</sup> we did see students discussing a mechanism, including electron movement, without discussing the cause of that movement.

### ***Designing the Task***

As we develop approaches to helping students’ reason about molecular level phenomena, we have developed tasks that prompt students to think about the causes for, and the mechanisms by which, these phenomena occur. There are several reasons why we have chosen to focus on tasks that involve causal mechanistic reasoning:

1. Structuring the prompt so that students have to think both about how a phenomenon occurs and why that phenomenon occurs sends a signal that these are the thinking skills that are valuable and necessary to understand chemistry. If we consider causal mechanistic reasoning to be a valued form of scientific thinking, then it is important that we assess it, as that is what students will find important.<sup>21</sup>

2. When used as an assessment (either formative or summative), these types of tasks can provide more persuasive evidence about what students know and can do. According to the principles of evidence centered design,<sup>22,55</sup> assessment can be considered as an evidentiary argument. That is, assessments should be designed to elicit evidence about what students know and can do with regard to the target of the assessment. Student responses to these tasks can be characterized with an appropriate coding scheme to identify how students reason about the phenomenon under consideration.

3. These tasks provide polytomous student responses. That is, the reasoning students provide is not right or wrong, but can be scored or coded along a range of sophistication. By using such tasks at different time points in a students' education, we are able to detect how student reasoning changes over time.

However, designing an activity that elicits causal mechanistic reasoning is not a simple task. We know that if we ask students to simply predict or rank without providing an explanation, they default to system I type thinking.<sup>56</sup> That is, they use readily available heuristics rather than thinking about the underlying causes and mechanisms of the phenomena they are being asked about.<sup>52</sup> In order to help students reason causal mechanistically, we look to Hammer's resources framework.<sup>12</sup>

This framework posits that students do not necessarily have coherent conceptions (or misconceptions), but rather a wealth of resources that are activated in particular moments.<sup>12</sup> By focusing on what students know and can do (rather than what they do not know), we can identify the resources that students are using to construct models and explanations. Ideally this might allow us to help students produce more sophisticated and complete explanations if we can help students link their ideas in productive ways.

In this context, if we are asking students to explain LDFs causal mechanistically, we need to design the prompt to activate the appropriate conceptual resources (i.e., subatomic particles and electrostatic interactions). In chemistry, this may provide additional challenges. This framework was originally developed in the context of macroscopic physics ideas, and therefore, the resources required to provide a coherent explanation are ideas they may have experienced in their everyday life.<sup>12</sup> In chemistry, though, the phenomena are often taking place at the unseen atomic-molecular level governed by ideas of forces and energy that are notoriously difficult.<sup>2,42,57,58</sup> Therefore, we are asking students to leverage resources that are not only counterintuitive (forces and energy) but also must have been developed via some kind of instruction in chemistry rather than through everyday experiences (electrons, atoms molecules, electron distributions). That is, providing causal mechanistic reasoning in explanations is particularly difficult for chemical phenomena because students must use entities and ideas that they cannot experience at the macroscopic level and that require a good deal of prior instruction to engage with. To help students reason causal mechanistically about LDFs, we need to scaffold the prompts carefully to activate the appropriate resources.

Scaffolding is a process introduced initially by Wood et al.<sup>17</sup> as the process by which someone more knowledgeable is able to help someone less knowledgeable accomplish a task they would otherwise be unable to. While Wood et al.<sup>17</sup> were describing this process in the context of tutoring, since then many have applied the principles of scaffolding, which also aligns with Vygotsky's Zone of Proximal Development ideas,<sup>18,59</sup> in a variety of related fields.<sup>60</sup> Scaffolding an activity is usually supported by (1) identifying what the learner can do without assistance to determine what exactly the scaffolding should target and (2) eventually fading the scaffolding corresponding to an increase of cognitive responsibility on the student.<sup>24,60</sup>

While many use these principles to improve one on one or small group interactions,<sup>17,60,61</sup> others have used these principles to help design assessments/activities.<sup>23,24,62</sup> For example, Ge and Land<sup>62</sup>

acknowledge that the scaffolding can be informed by research in the area to specifically target areas that have been found to be problematic. Reiser<sup>23</sup> suggests that since we cannot interact with students individually to provide the scaffolding, we can structure the question and take cue's from Wood et al.'s<sup>17</sup> work to reduce degrees of freedom, focusing students in on the areas they need to focus on. McNeil et al.<sup>24</sup> shows how fading can be applied in a similar situation as they reduced the presence of the scaffold over the course of the semester. However, there are few studies on the removal of scaffolding and its effects on the way students respond.

We have previously developed an assessment task and coding scheme that allows us to elicit evidence about how students are thinking about LDFs,<sup>7</sup> which involves both a written explanation and drawn model for how nonpolar species can “stick together”, that is, how LDFs originate and operate in nonpolar species. In this paper we adapt and simplify the previous coding scheme so that we can investigate several research questions that focus on the evidence that we can elicit from such questions and the implications for learning over time. We also adapted the original assessment task to incorporate a scaffolded drawing prompt that is intended to support students as they think about the mechanism of LDF formation.

We used the revised coding scheme and assessment task to perform a longitudinal study (over the course of two semesters) to investigate how students' models and explanations change over time, what happens when scaffolding is withdrawn, and how students' responses change between formative and summative assessment tasks. We also performed a cross-sectional study comparing students from prior semesters who were not exposed to scaffolded tasks.

### **Research questions**

The study is guided by these research questions:

1. How does scaffolding impact student responses to prompts about LDFs (study 1)?

2. How do students' explanations and models of LDFs change over a two-semester sequence (study 2)?
3. How do students written explanations and drawn models correlate with each other (study 3)?

### **Methods**

In this section we describe the methodology for all of the studies in this paper. All of the participants in these studies were undergraduate students at a large research-intensive Midwestern university. All participants were informed of their rights as human subjects and gave their permission to have their responses used for research purposes. All of the participants' responses were deidentified before analysis.

Additionally, following analysis we frequently use statistical tests to help understand the significance of patterns in the data. For all statistical tests we use a significance level of  $p < 0.05$ .

Studies 1-3 involve the analysis of a longitudinal study, following ways in which a cohort of general chemistry students draw and explain LDFs over the two-semester sequence. Specifically, these students were enrolled in Chemistry, Life, the Universe and Everything (CLUE), a transformed general chemistry curriculum that is organized around progressions of core ideas in the context of scientific practices.<sup>37,53</sup> In these studies, we examine the impact of the scaffolding of the activity (study 1), how their responses changed over time (study 2), and how their text and drawing responses were related to one another (study 3).

### ***Participants***

For the major longitudinal study reported here, four assessment items involving understanding of LDFs were administered to general chemistry students over the course of the academic year 2015-2016. Of the 290 students enrolled in the sections in which our activities were administered, 260 completed all four of those assessments and gave consent to have their work used for research



purposes. Using a random number generator, we randomly selected 150 of those students to analyze their responses to the tasks (described below) across all four assessments. We selected only 150 of the 260 possible students for practical reasons as this would give us a group large enough to use the appropriate statistical tests while reducing the number of responses that would need to be coded (by almost 900 text and drawings). This group is referred to as the 15-16 cohort in this paper. This group is majority female (64%), and white (71%) and in their first year of college (84%). The mean ACT score for this group is 26.2, and the mean grade (on a 4-point scale) in the course is 3.43 and 3.22 for the first and second semester, respectively (see Appendix B).

Additionally, we carried out a cross-sectional study in which we compared the responses of this cohort with those from 129 students from our previous study<sup>7</sup> (referred to as the 14-15 cohort in this study, corresponding to the 2014-2015 academic year in which they completed the general chemistry sequence). The 14-15 cohort is a subset of the 250 students whose data were initially analyzed in the previous study<sup>7</sup> who then went on to take General Chemistry 2 the following semester. This cohort of students is similar to the 15-16 cohort based on their grade level, sex, race/ethnicity, and ACT score (see Appendix B). Therefore, we claim that the only meaningful difference between these two groups of students with regards to this study is the assessment items they received.

### ***Data Collection***

For the longitudinal study we administered four activities over the two-semester general chemistry (GC) sequence to the 15-16 cohort to explore how students explain and draw LDFs. In this paper we name the items 1-4 corresponding to the formative (item 1), summative (item 2), homework posttest (item 3), and exam posttest (item 4) assessments (Figure 4.1). These items vary in terms of the timing of their delivery, the stakes of the activity, the form of the assessment, and the structure of the prompt. We discuss the details of these activities in the sections that follow.

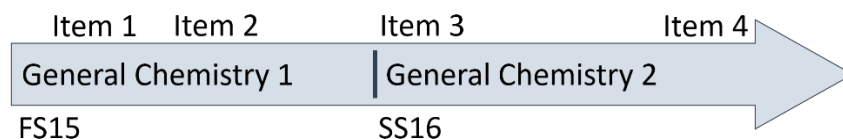


Figure 4.1. Relative timing of the four assessment items relative to the two semesters in the general chemistry sequence in the 2015-2016 academic year. The four assessments consist of a formative homework assessment (item 1), a summative exam assessment (item 2), a homework posttest assessment (item 3), and an exam posttest assessment (item 4).

The items were administered as either a homework or on an examination. Homework activities in this course were administered via beSocratic, a web-based system that allows the collection of both written and drawn responses.<sup>63</sup> This homework is graded for completion rather than correctness to encourage students to explore their ideas rather than look up answers online. Students are not provided with copies of the correct answers, but the questions are discussed in class and often used to anchor the next instructional unit. In contrast, all of the exams in this course were given on paper and credit was awarded to the students on the basis of their correctness. Students are provided with keys to the multiple choice items but not the open response items, but they are free to consult with graduate teaching assistants or undergraduate learning assistants. We consider the homework items “low stakes” assessments while the exam items are “high stakes” assessments.

The text and drawing prompts for all the LDF assessments administered in this study are outlined below (Table 4.1), and the full prompts are shown in Appendix C. In the following section we further describe the context surrounding the individual items.

Table 4.1. Information about the LDF prompts used throughout all four studies.

Item		Prompt		Additional Administration Details			
Name	Type	Text	Drawing	Cohort	Type	Timing	Boxes for Drawing Prompt

Table 4.1 (cont'd)

Previous study <sup>7</sup>	Formative	What do you think causes the helium atoms to move toward each other? Hint: Think about what's happening with the protons and electrons.	Sketch a molecular level picture to show what you think is going on as the atoms move together. Use this space to describe in words what you are trying to represent in your drawing.	14–15	Online homework	Beginning of GC1	1
Item 1	Formative	As the [helium] atoms get closer they attract one another and the potential energy decreases as shown by the circled region. Please explain why the atoms attract and the process by which it occurs. Hint: Think about what's happening with the electrons.	Now, draw a series of molecular level pictures of what's happening within the atoms that causes the atoms to move together. Be sure to include the subatomic particles in your drawing. If there's anything you cannot explain in your drawing, explain it in the box below.	15–16	Online homework	Beginning of GC1	3
Item 2	Summative	Consider the noble gas Kr. In the boxes below, draw a molecular level picture of what happens when two atoms of Kr approach each other. Use the picture to help you explain why the two atoms are attracted to each other.		15–16	Paper exam	Beginning/middle of GC1	3

Table 4.1 (cont'd)

Item 3	Homework posttest	Is there an attraction between two neutral helium atoms? (Yes/No/I do not know) Below, explain in detail why or why not. Be sure to include any evidence you have for your claim and how this supports your reasoning.	Draw a molecular level picture of two helium atoms as they approach one another in the blue box, and explain what is happening in the black box.	15–16	Online homework	Beginning of GC2	1
Item 4	Exam posttest	Consider the noble gas Ar. Draw a molecular level picture of what happens when two atoms of Ar approach each other. Use the picture to help you explain why the two atoms are attracted.		15–16	Paper exam	End of GC2	1

Item 1 (formative) was an instructional homework activity and formative assessment that was administered to students as they were learning about LDFs in class during the first 2 weeks of the first semester of general chemistry (GC1). This formative activity was developed iteratively over the course of several semesters using feedback from analysis of student responses including the work presented in our previous study.<sup>7</sup> In this administration (to the 15-16 cohort) the prompt is scaffolded by (1) asking students to think about the electrons to help them think about the scalar level needed to reason about the attraction, and (2) including three drawing boxes that are intended to prompt students to think about the sequence of events that lead to the attractive interaction (as shown in Figure 4.2).

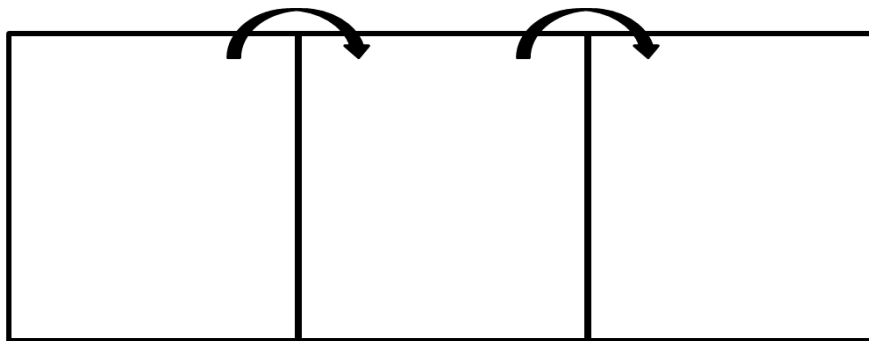
In our previous study,<sup>7</sup> the students in the 14-15 cohort were provided with a prompt that was similar but less specific than the activity described here. In particular, the drawing prompt only had 1

box, rather than the 3 boxes in item 1. In this study, we wanted to provide as much scaffolding as possible to help students learn to construct mechanistic explanations, and it was clear from our previous study<sup>7</sup> that further scaffolding was needed for the drawing prompt.

Item 2 (summative) was included on the first midterm exam administered about 4 weeks into GC1. Mirroring the format of the formative assessment, three boxes were provided to encourage students to draw the mechanism of the attraction. However, the hint reminding the student to think about the electrons was removed.

**Consider the noble gas Kr.**

In the boxes below, draw a **molecular level picture** of what happens when two atoms of Kr approach each other. Use the picture to help you **explain** why the two atoms are attracted to each other.



Explanation:

Figure 4.2. Item 2 drawing prompt including the scaffolded 3 box drawing prompt to encourage students to draw the process by which the interaction formed.

Item 3 was administered on a homework activity given the first week of the second semester. The assessment was not scaffolded in the same way as the previous two so that we could see the effect of removing the scaffolding. While they were still asked to write and draw about what happens when two atoms of a noble gas approach, they were only provided one box in which to draw the atoms approaching.

Item 4 was administered on the last midterm exam in the second semester. Students were informed prior to the exam that there would be first-semester material on the exam, though they were

not told specifically what that material would be. In this case, students were asked about how and why another neutral atom, argon, can attract each other, and again, the scaffolding was removed from the drawing portion of the activity.

### ***Data Analysis***

In our previous study,<sup>7</sup> we developed a holistic coding scheme to characterize how students write and draw about the mechanism by which LDFs form and act. This scheme consisted of six categories (0-6) that emerged from student interview data and analysis of a relatively small number of responses. However, as we continued coding larger numbers of student responses it became clear that the original scheme was too fine-grained to be broadly applicable. Therefore, we simplified our coding scheme by collapsing the codes into three broader categories based on the type of reasoning shown by students: nonelectrostatic, electrostatic causal, or causal mechanistic. Both authors worked to combine these codes through discussion and test coding student responses. Examples and definitions of the three categories of responses are provided in Tables 4.2 and 4.3.

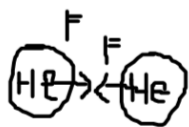
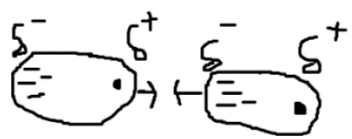
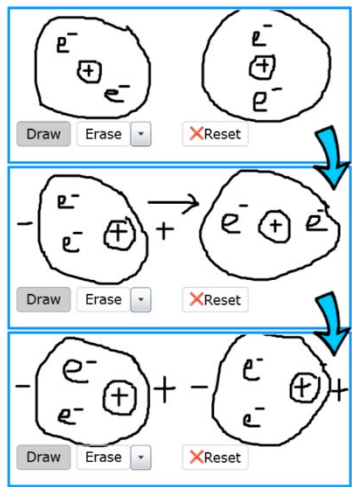
Table 4.2. Coding scheme for text LDF responses.

Type of Response	Previous Level <sup>7</sup>	Salient Text Features	Examples
Nonelectrostatic (NE) text response	0 and 1	The response does not include reasonable electrostatic evidence of the interaction. Instead, the response provides nonelectrostatic evidence or does not address the intermolecular interaction between molecules.	"The two atoms are attracted because of the electromagnetic forces that exist between them"
			"This happens because the two Ar atoms need 8 valence electrons so between the two of them they start pulling at the other to fill its shell. This causes the two to combine, making an Ar <sub>2</sub> molecule"

Table 4.2 (cont'd)

Electrostatic causal (EC) text response	2 and 3	The response indicates that electrostatic charges cause the interaction. Examples of electrostatic causal evidence include subatomic particles, overall charge of the atom, partial charges etc. These responses do not include a mechanism by which a separation of charge occurs.	“The atoms become attracted to [each other] because of the attractions between the protons & electrons. Protons have a (+) charge, while electrons have a (-) charge. Opposite charges attract, and the closer the atoms appear the stronger the attraction becomes”
			“Atoms attract to each other [because] the [London] dispersion forces make the partially negative end of one atom attract to the partially positive end of another atom.”
Causal mechanistic (CM) text response	4 and 5	The response indicates that the interaction occurs due to electrostatic charges and includes the mechanism by which the instantaneous and/or the induced dipole forms.	“The atoms attract because of the London dispersion forces between the two atoms in which an atom with [instantaneous] dipole causing the another atom to become an induced dipole. In this scenario, one atom has its electrons shift to certain region of the electron cloud, giving that region a negative charge. This negative region then induces the movement of electrons in another atom, being attracted to the new present positive region of the second atom.”
			“As the 2 atoms approach each other one of them becomes [instantaneously] dipole due to fluctuation in its electron cloud as most of the electrons are concentrated to one side of the atom making it have a partially positive and negative charges on opposite sides. This partially positive side of the attracts the [electrons] of the other atom making its electron cloud fluctuate and form an instantaneous dipole where both of them attract each other.”

Table 4.3. Coding scheme for LDF drawings. Student drawings used with permission.

Type of Response	Previous Level <sup>7</sup>	Salient Drawing Features	Examples <sup>b</sup>
Nonelectrostatic (NE) drawing	0 and 1	The drawing does not show any reasonable electrostatic interactions or interacting atoms.	
Electrostatic causal (EC) drawing	2 and 3	The drawing includes subatomic particles or charges but does not include pictures showing the process by which partial charges develop.	
Causal mechanistic (CM) drawing	4	The drawing includes the subatomic particles or charges and includes several pictures showing the atoms transitioning into dipoles.	

Level 1 remained the same, capturing responses that do not provide any kind of electrostatic causal reasoning.<sup>7</sup> Typical responses include simply naming an interaction or describing that the atoms attract one another without providing a reason (essentially restating the question). The second category of responses, “electrostatic causal”, combines levels 2 and 3 in which students acknowledge that electrostatic charges are the cause of the interaction but do not discuss how these partial charges arise from neutral atoms.<sup>7</sup> The third category, “causal mechanistic”, includes student responses that discuss how the unequal distribution of charge arises from electron movement to give temporary dipoles, which



then cause the interaction corresponding with levels 4 and 5.<sup>7</sup> This causal mechanistic reasoning aligns with other researcher proposed definitions, in that the mechanism for a phenomenon lies at least one scalar level below the phenomenon being explained.<sup>5,6</sup> That is, mechanistic reasoning must involve a discussion of how the redistribution of electrons leads to the production of transient charges.

While collapsing the codes does reduce the selectivity of the coding scheme (for example, it removes the ability to distinguish between responses that provide the mechanism for forming one vs two dipoles), we believe that the three classifications are a more practical approach to coding large numbers of student responses while still capturing the presence of causal mechanisms. Another advantage of this new characterization scheme is that the classification previously applied to student drawings is now directly comparable to the written explanations. That is, drawings that do not show any kind of charge distribution are nonelectrostatic, those that show charge distribution are electrostatic causal, and those that show the sequence of events that produce the charge distribution are causal mechanistic.

Interrater reliability of the coding scheme was established between author K.N. and an undergraduate research assistant before beginning the analysis (see Appendix B). Analysis of the responses was carried out by Author K.N. using QSR's International Nvivo 10 Software<sup>64</sup> to assist in the coding process. Because of the nature of the prompt and the way data was collected, the authors were aware of the prompt from which the data came.

## **Results and discussion**

### ***Study 1: What effect does the level of scaffolding have on student drawings and explanations***

In this study we compare the student responses from the unscaffolded prompt administered in fall 2014 to those from the new scaffolded prompt, in which students were provided with three boxes to draw the mechanism of LDF formation, which was administered at the same time point in fall 2015 (item 1).

In the previously reported analysis of student responses administered in fall 2014, the drawing portion was not scaffolded; instead, students were provided with a blank drawing space.<sup>7</sup> Using the revised coding scheme on this previously collected data, we see (Figure 4.3) that while 87% of students from the 14-15 cohort provided an electrostatic causal drawing indicating the involvement of charges in attraction between the atoms, the number of students who drew a full causal mechanism was only about 2%.<sup>7</sup> In contrast, the formative prompt (item 1) from fall 2015 which included the scaffolded boxes for drawing the mechanism by which LDFs are generated produced an increase in the full causal mechanistic drawings: 55% of students now produce the full causal mechanism. That is, for these two equivalent cohorts of students who were completing a formative assessment, adding the scaffolded drawing prompt helped many more students to produce causal mechanistic drawings. However, the overall percent of students who provide answers including some electrostatic causal factor (that is, both EC and CM bins) is approximately the same as in fall 2014.<sup>7</sup>

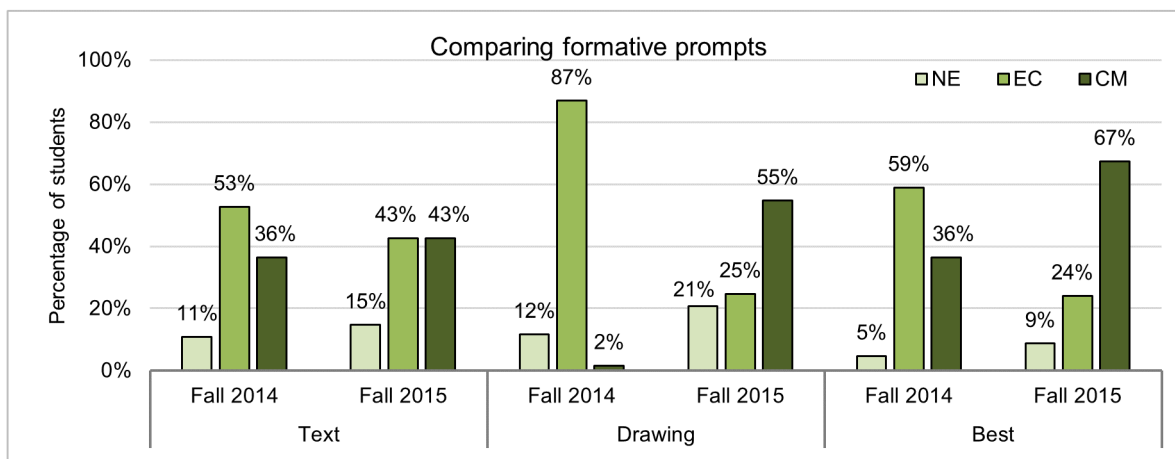


Figure 4.3. Distribution of text, drawing, and best (most sophisticated code for drawing or text) responses for the different formative assessments given during fall 2014 (N = 129) and fall 2015 (item 1, N = 150).

We do not expect that scaffolding the drawing response should affect the written responses, because on the homework, the written prompt appears before the drawing prompt (item 1), and students cannot go back to prior questions. Indeed, there is no significant difference in the students' text responses from the different years ( $\chi^2 = 2.94$ ,  $p = 0.230$ ). In fall 2014, 36% of students provide a CM written response, while that number increases only slightly in fall 2015 (43%).<sup>7</sup>

When we compare their “best” responses (giving each individual student a score corresponding to their most sophisticated code for drawing or text response, CM>EC>NE) we see that the number of students producing a CM drawing or text response almost doubled from fall 2014 to fall 2015. While the text responses look the same for the two groups, by scaffolding the drawing responses we were able to elicit more CM students' responses, including many who did not produce a CM text response. That is, the presence of the scaffolded drawing prompt resulted in the activation of resources associated with the mechanism of the interaction. Since there was little difference in the instructional material that preceded the formative activity between the two years, it is probable that if the drawing prompt in fall 2014 had been similarly scaffolded we would have seen more causal mechanistic drawings than we saw (2%).

***Study 2: How do students' explanations and models of LDFs change over a two-semester sequence***

*Finding 1: A large percentage of students construct a full CM drawing on the first summative exam after being taught the material*

On the summative assessment (item 2), 86% of the students provide a drawing of a full causal mechanism. Including the EC responses, 95% showed separation of charge (Figure 4.4). There are several possible reasons for this improvement:

1. Students may be more likely to provide an effortful response on a summative item that is graded for correctness (item 2) than on the homework for which credit is given for completion (item 1).

2. Item 1 was given when students were learning about LDFs; by the time of the exam, students would have more time to study and assimilate the ideas.

3. Students may have simply memorized the three-pane drawing (although they had not received individual feedback on the homework, it had been reviewed in the next class) due to the close proximity of the assessment to the initial instruction.

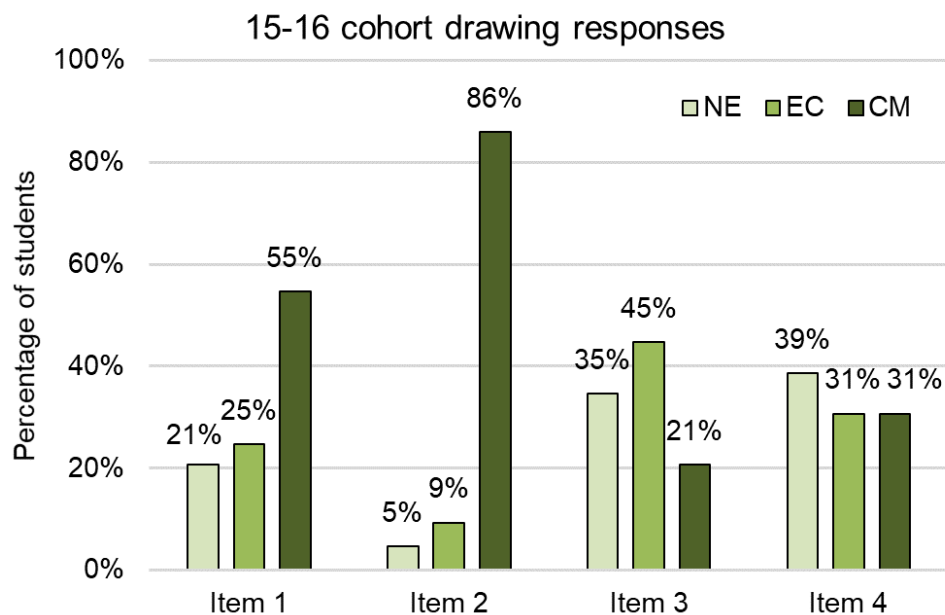


Figure 4.4. Percentage of 15-16 cohort drawing responses (n = 150) in each of the LDF categories for items 1-4 (see Table 4.1 for item descriptions).

*Finding 2: Removal of the scaffold appears to produce a decline in causal mechanistic drawings on unstructured posttest assessments (Items 3 and 4)*

Items 3 and 4 remove the scaffolded 3 box drawing prompt, and there is a corresponding drop in the percentage of CM drawings. At the start of the second semester (item 3), the number of students who provided a drawn causal mechanism is down to 21%, and the percent who show any kind of electrostatic causal drawing is 66% as opposed to 95% on item 2. In addition to the removal of the scaffolding, there are other possible reasons for this decline including the following:

1. These responses were collected several months after the explicit treatment of LDFs in the curriculum, and perhaps students had difficulty in recalling the details of the mechanism.
2. Some students may have simply memorized the drawing for the exam (item 2) rather than deeply understanding the mechanism.
3. The prompt was administered on the low stakes homework system.

Although we cannot be certain of the factors that contribute to the decline in CM drawings, on the final administration of the prompt (item 4) we do see an increase in CM from 21% on item 3 to 31%, though the total percentage of students providing any type of electrostatic causal drawing remains about the same (66% at item 4, 62% at item 3). The slight increase in CM responses on item 4 is probably because of the higher stakes nature of the exam. Despite this improvement, the percentage of CM drawings is still much lower than it was for the summative assessment from semester 1 (item 2). However, the relative stability of the percentage of CM drawings across the two posttests despite an additional 3 months passing may indicate that time from instruction is a less important factor than the lack of scaffolding.

*Finding 3: The initial scaffolded drawing prompt does have an effect on the final number of CM drawings*

Interestingly, the introduction of the scaffolded prompt in fall 2015 did appear to have a long-term impact on the students. Even though the decline in CM drawings after withdrawal of the scaffold is quite marked, in comparison to the responses from fall 2014, the percentage of students who do provide a full drawn causal mechanism on item 4 is significantly higher (31% to 2%) despite the fact that in fall 2014 the material was covered in the same semester that the assessment was given ( $\chi^2 = 41.28$ ,  $p < 0.001$ ,  $\phi = 0.38$ ).

*Finding 4: The text responses do not vary as widely over the four time-points, but there is a general decrease in CM written explanations*

For the first two time points, the percent of EC and CM explanations does not vary much (Figure 4.5). However, in the second semester (items 3 and 4), we see the same “drop off” in performance from the first semester as we did for the drawing responses, perhaps because time passed and some students have forgotten: For item 4 only 20% of students provide causal mechanistic explanations, although in total 77% include an electrostatic causal factor.

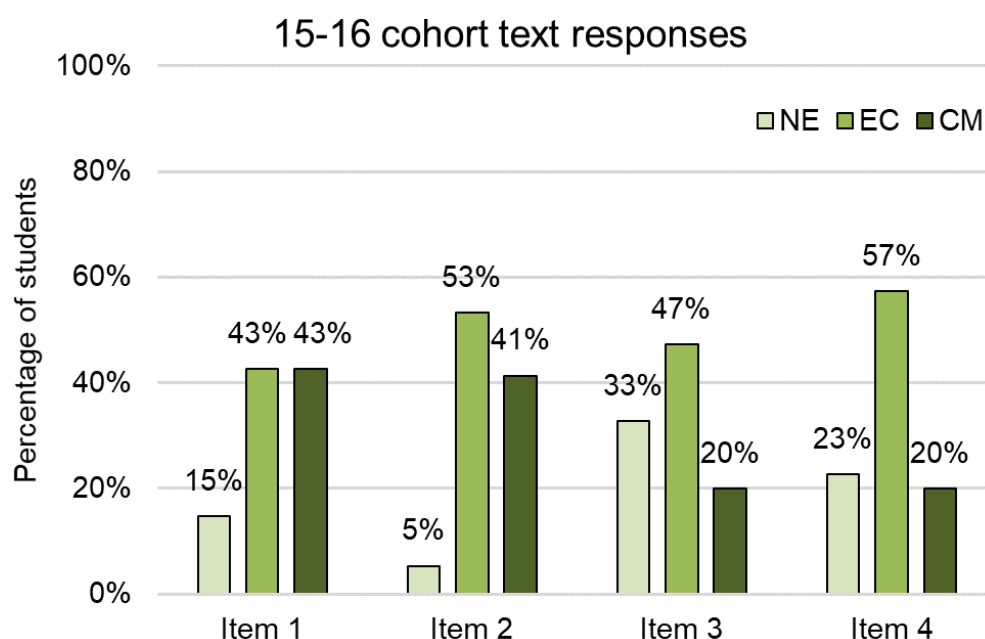


Figure 4.5. Percentage of 15-16 cohort text responses (n = 150) in each of the categories for items 1-4 (see Table 4.1 for item descriptions).

### ***Study 3: How do students written explanations and drawn models relate to one another***

*Finding 1: There is generally an association between text and drawing responses except for EC responses*

While the text and drawing responses were coded separately for, one might imagine that there should be some correspondence between the two types of responses. We used a series of Pearson’s  $\chi^2$  tests<sup>65</sup> to analyze the relationship between students’ text and drawing responses (Table 4). For results that showed a significant association, we determined the strength of the relationship using Cramer’s V,

a modified version of  $\phi$  for contingency tables with more than 2 rows and columns,<sup>65</sup> using Cohen's suggestions for interpreting this value: small - 0.1, medium - 0.3, large - 0.5.<sup>66</sup>

Table 4.4. Pearson's  $\chi^2$  test results exploring the relationship between the text and drawing LDF Responses.

Activity	$\chi^2$ value	P Value	Cramer's V
Item 1	31.44	$P < 0.001$	0.32
Item 3	18.84	$P = 0.001$	0.25
Item 4	50.03	$P < 0.001$	0.41

For items 1, 3, and 4 there is a significant relationship between the text and drawing responses ( $p \leq 0.001$ ) with Cramer's V values ranging from 0.25 to 0.41. For item 2, four of the nine cells (44%) had an expected count less than 5, which violates one of the assumptions of this test. This is because over 90% of the drawing responses were characterized as CM for item 2, and therefore, there were few responses in the remaining cells.

While the  $\chi^2$  test tells us if there is a relationship between the variables, post hoc analysis is necessary to determine what is driving that relationship.<sup>67</sup> To do this, we used the statistical software SPSS<sup>68</sup> to calculate the adjusted standardized residual (from here on called adjusted residual) for each of the cells in the contingency table. This value provides a standardized measure for how different the observed value is from the expected.<sup>67</sup> The sign of the adjusted residual indicates if the observed count is more (positive) or less (negative) than expected, while the magnitude of the value indicates how different the observed count is from the expected.<sup>67</sup> If the adjusted residual is large enough, then we reject the null hypothesis for that particular cell, indicating that particular relationship is driving the significant result of the initial  $\chi^2$  test.<sup>67</sup> To reduce the risk of type 1 error, we used the Bonferroni adjusted critical value (2.78 for a 9-cell table) as the threshold to determine if the cell is a primary driver of the relationship.<sup>69</sup>

For item 1, posthoc analysis revealed that positive associations between the CM text and drawings as well as the NE text and drawings are primary drivers of the significant result (see Appendix D). There is also a negative association between NE text and CM drawing responses. A similar pattern is seen for item 3 with a positive association between NE text and drawings and a negative association between CM text and NE drawing responses driving the significant result (see Appendix D). Item 4 not only had the strongest association from the initial  $\chi^2$  test but was also driven by associations in five of the nine cells: positive associations between CM text/drawings as well as NE text/drawings; negative associations between NE text and EC drawings, NE text and CM drawings, and CM text and NE drawings (Table 4.5).

Table 4.5. Contingency table for the item 4 (exam posttest assessment) text and drawing LDF responses. In each cell the adjusted residual value is reported along with the observed and expected values. Adjusted residuals larger than the Bonferroni adjusted critical value ( $\pm 2.78$  for 9-cell table) are bolded. To visualize the sign and magnitude of the adjusted residuals, the cells are color coded from dark red (most negative) to dark blue (most positive).

Adjusted Residual Values				
Item 4		Drawing		
		NE	EC	CM
Text	NE	<b>6.35</b> Expected: 13.1 Observed: 29	<b>-3.14</b> Expected: 10.4 Observed: 3	<b>-3.56</b> Expected: 10.4 Observed: 2
	EC	-1.78 Expected: 33.3 Observed: 28	1.66 Expected: 26.4 Observed: 31	0.22 Expected: 26.4 Observed: 27
	CM	<b>-4.44</b> Expected: 11.6 Observed: 1	1.24 Expected: 9.2 Observed: 12	<b>3.45</b> Expected: 9.2 Observed: 17

From these posthoc analyses we see that the expected relationship between the students' text and drawing responses is present. There are positive associations between NE text/drawings and between CM text/drawings, and negative associations on mismatched bins, specifically between the NE



and CM bins. We believe these associations between the text and drawing responses indicate that for items 1, 3, and 4 the students' drawings were not memorized without understanding, but instead reflective of their comprehension of the mechanism. We also see that the EC category represents the "messy middle": this may be a reflection of the wide range of responses captured in this category. For example, this category can represent students who have either drawn or written a CM response and do not see the need to provide a full response in their corresponding drawing or explanation. However, it can also capture students who are still trying to figure out the cause of a mechanism, so that a student who describes both atoms as having a single charge and a student who describes the separation of charge in each atom (but not the mechanism of dipole formation) would both be categorized as EC.

*Finding 2: If we use the "best" response - drawing or text - we see a significant increase in CM from item 1 to 2, a decrease from 2 to 3 and then little change from 3 to 4*

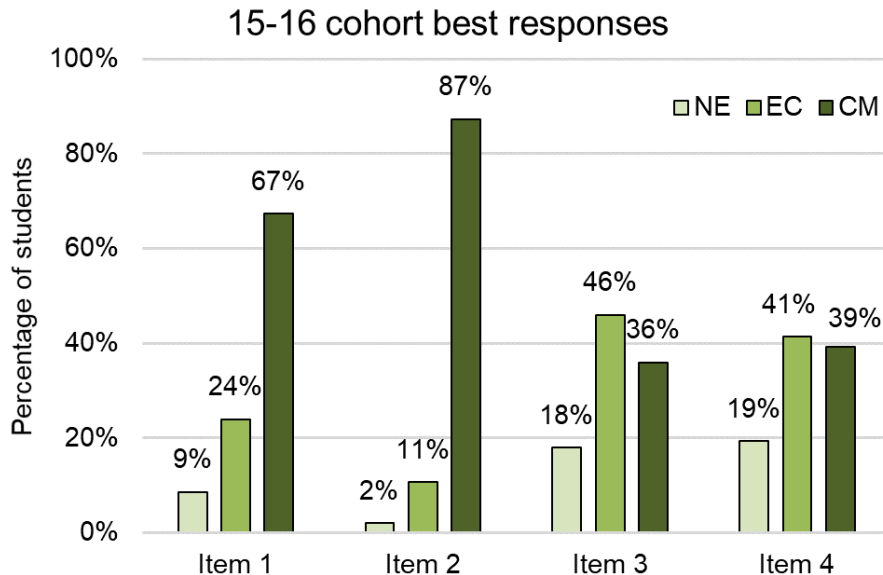


Figure 4.7. Percentage of 15-16 cohort's (n = 150) best responses (based on each student's most highest level text or drawing response) in each of the categories for items 1-4 (see Table 1 for item descriptions).

If we assume that some students will only provide the most detailed response in either the drawn or written response (but not both), we can investigate each student's "best" response over time.

In this case we see an increase in CM from the first formative item (item 1) to the first summative item (item 2, Figure 4.7). As discussed previously this may be because of the increased understanding of students, or the increased effort that students might show on a graded summative exam. When the scaffolded prompt is removed in items 3 and 4, we now see that, while the combined percent of CM responses decreases from first to second semester, there is little difference between the homework or exam posttest response patterns (items 3 and 4). Therefore, we propose that this is evidence for the idea that student responses gathered from homework assessments are most likely representative of their best efforts at the time, despite the fact that they are not being graded for correctness.

### **Summary**

In these studies we have adapted and simplified a previously published coding scheme<sup>7</sup> that characterizes how students explain the origin of LDFs. This has allowed us to investigate the effects of adding and removing scaffolding designed to promote causal mechanistic thinking, the long-term impact of such scaffolding, and how and whether students respond differently to low and high stakes assessments.

We were able to perform these studies with large numbers of students because the instructors allowed us to design and administer four assessment items, two as homework and two on semester exams. Because of the limitations of working with large numbers of students in real world settings, we were not able to design the studies as one might in a laboratory, particularly since half of the assessments were administered on an exam. This means that although we had a number of variables (scaffolding, time from instruction, and formative/summative administrations), we were not able to address each one separately. Nevertheless, we believe that there is sufficient evidence to tease out some more generalizable findings from these studies:

1. Scaffolding of the drawing prompt has an immediate effect on how students represent the origins of LDFs. Comparing responses on item 1 from fall 2014 and 2015, we see that adding the three-box scaffold encouraged students to draw a causal mechanism. The percentage of CM drawings increased from 2% to 55%. Although it is probable that more than 2% of the fall 2014 students could also draw this interaction mechanistically, the appropriate resources were not activated as they answered the unstructured prompt.
2. Scaffolding of the drawing prompt also appears to have a long-term effect on how students represent the origins of LDFs. While the percentage of CM drawings peaks dramatically on item 2, it also falls quite dramatically when the scaffolded prompt is removed. That being said, the percentage of CM drawings on both items 3 and 4 is still much higher than those in fall 2014. Although we do not have long-term data from fall 2014 to spring 2015, it is unlikely that the percent of causal mechanistic reasoning would have increased during this time.
3. Student best responses to versions of the prompt are relatively unaffected by time, or by whether it is a high or low stakes assessment. While there is a large spike on CM drawing on item 2, we do not believe that this is an accurate representation of student understanding but instead is more likely to be a memorized set of representations. We believe that the comparable results from items 3 and 4 are a better representation of what students know about LDFs.

### **Implications For Teaching**

#### ***The use of scaffolding in formative assessments***

What seems clear from our findings is that scaffolding can greatly improve the immediate quality of student responses. When we compare the “best” response from each student from fall 2014 and fall 2015, at item 1 the percentage of CM responses nearly doubled. We also believe that scaffolding did have a long-term effect for some students. Certainly there was a marked drop in CM drawings between items 2 and 3, and it is not possible to determine from this study whether the drop is a

function of the lack of scaffolding or the time from being taught. However, we believe that the relative stability seen in items 3 and 4 indicates that it is the loss of the scaffolding that is the main contributor to the drop in CM drawings. That being said, even by the end of the full year of general chemistry (items 3 and 4) there are many more students providing a CM response than students in the previous study (fall 2014) did just after learning about LDFs. While there is clearly more work to be done on the effects and removal of scaffolding, we do believe that adding scaffolding supports to formative assessments is productive and should be considered whenever possible.

### ***Formative and summative assessments***

Interestingly, there is little difference between the levels of response for formative and summative assessments for items 3 and 4, despite the fact that students believed that item 4 would be graded, and they knew item 3 would not. While some may be surprised that students appear to exert as much effort on the nongraded homework as on the exam, it has been our experience that most (but not all!) students do “try hard”. This may be because the course expectations include the idea that the homework is for their benefit, to help them learn. The instructors attempt to create a culture that does not penalize mistakes on homework, but rather emphasizes that students use these activities to try out ideas. Most online systems are designed to provide immediate feedback, and because of this the nature of the tasks is mostly restricted to calculations, factual recall, or skills. Indeed, there is a danger that instruction and summative assessments will come to align with the restrictions placed upon homework systems rather than help students develop deep, connected, and useful knowledge. We believe that this study provides evidence that formative assessment systems such as beSocratic do not necessarily have to provide immediate feedback. This seems to be evidence for the idea that the very act of constructing the drawings (models) and written explanations actually helps students learn, since they have to reflect on what they know and make connections to answer the question.

We believe this finding provides a strong case for the use of more formative assessments, and for emphasizing their importance in student learning, both explicitly (by spending instructional time on them) and implicitly (by incorporating them into the grading scale for the course).

### ***The importance of long-term post-tests: what do students retain***

In many studies, the effect of interventions is often measured by pre/posttesting, where the posttest is often given immediately after the intervention. In this study our items are administered over a much longer time period. We were able to see whether responses to item 2, which occurred at the end of the instructional unit that covered LDFs (the first summative assessment), were truly indicative of the students having a deep understanding of these interactions. However, as we saw with items 3 and 4, many students did not maintain this level of performance, but these items are likely more indicative of student actual understanding. We believe that this kind of long-term testing is likely to provide more meaningful information about student learning than a posttest that immediately follows an intervention. After all, our goal is to support student learning, not to help them memorize material to pass tests and then forget it.

### **Limitations**


As previously noted, this longitudinal study was conducted with students enrolled in a transformed curriculum that emphasizes scientific practices such as constructing models and explanations, in which students are routinely asked to provide mechanisms for a wide range of phenomena, and where students are not penalized for mistakes on homework. This approach is still quite rare, and therefore, the results here may not be transferrable to more traditional curricula. As we noted earlier, because this study was situated in the actual homework and course examinations, we were not able to vary each variable in a systematic way. Further studies that change, for example, the level of scaffolding in formative (but not summative) assessments may provide further insight into the role of scaffolding and how it affects long-term use of knowledge.


## **APPENDICES**


## APPENDIX A: Permissions


3/21/22, 6:32 PM


Rightslink® by Copyright Clearance Center




 Home

 Help ▾

 Email Support

 Keenan Noyes ▾



**Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study**  
Author: Keenan Noyes, Melanie M. Cooper  
Publication: Journal of Chemical Education  
Publisher: American Chemical Society  
Date: Sep 1, 2019  
*Copyright © 2019, American Chemical Society*

**PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE**

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)  
Comments? We would like to hear from you. E-mail us at [customer@copyright.com](mailto:customer@copyright.com)

Figure 4.8. Permissions to reproduce manuscript in its entirety.

## **APPENDIX B: Studies 1-3 participant demographics and interrater reliability**

Here is a more complete breakdown of the demographics of the 15-16 cohort. Of the 150 students, 54 (36%) are male and 96 (64%) are female. The cohort is majority white (N = 107, 71%), but has 3 American Indian/Alaskan Native students (2%), 12 Asian students (8%), 8 Black or African American students (5%), 5 of Hispanic ethnicity (3%), 5 of two or more races (3%) and 10 international students (7%). Additionally, the students are primarily freshman (N = 126, 84%), but 17 are sophomores (11%) and 7 are Juniors (5%).

For the 129 students in the 14-15 cohort, 61 (47%) are male and 68 (53%) are female. The cohort is also majority white (N = 91, 71%), but has 1 American Indian/Alaskan Native students (1%), 12 Asian students (9%), 7 Black or African American students (5%), 7 of Hispanic ethnicity (5%), 3 of two or more races (2%), 7 international students (5%) and 1 student not reporting their race/ethnicity information. Additionally, the students are primarily freshman (N = 111, 86%), but 11 are sophomores (9%) and 7 are Juniors (5%).

Comparing the 14-15 and 15-16 cohorts using Pearson's chi-square tests,<sup>65</sup> there is no significant difference between the groups in terms of the distribution of grade level ( $\chi^2 = 0.66$ ,  $p = 0.720$ ) or sex ( $\chi^2 = 3.65$ ,  $p = 0.056$ ). There is also a similar distribution of the race/ethnicity of the students with the majority of the students in both cohorts (71% in both instances) being white.

We also compared the standardized test scores of the two groups of students. The majority of the students had ACT scores on file so that was the standardized test we used for comparison. For those that had SAT scores, a concordance table was used to convert it to an ACT score. If they had both ACT and SAT scores on file the higher of the two (when SAT was converted to ACT score) was used to characterize the student. There were some students that did not have either test score on file and were not included in this comparison.



For the 15-16 cohort, the ACT scores range from 18 to 36, are normally distributed and have a mean score of 26.2 (N = 143, standard deviation = 3.296). 7 of the students did not have a standardized exam score on file. For the 14-15 cohort, the ACT scores range from 17 to 34, are normally distributed and have a mean score of 26.5 (N = 124, standard deviation = 2.995). 5 of the students did not have ACT scores on file. Comparing the ACT score of the 14-15 cohort with the 15-16 cohort using an independent t-test<sup>70</sup> we found no-significant difference between the groups,  $t(265) = 0.764$ ,  $p = 0.446$ .

To evaluate the interrater reliability (IRR) of the LDF coding scheme, author KN and an undergraduate research assistant independently coded text and drawing formative assessment responses. The responses were coded in sets of 30 and after each round we calculated the Cohen's kappa value, a measure of agreement which takes into account the probability of agreeing by chance.<sup>71</sup> If the kappa value for that round surpassed 0.7 (which marks the middle of the "substantial" range of kappa values<sup>72</sup>) then we decided that we reached an acceptable level of agreement and could reliably apply the coding scheme. If the threshold of 0.7 was not met, we discussed discrepancies in our coding until 100% agreement was reached, modifying the category definitions if needed and then beginning another round of coding.

We conducted 5 rounds of IRR for the drawing coding scheme before we reached agreement (kappa = 0.80, 93.3% agreement). The text coding scheme proved somewhat more difficult to reach IRR. One reason for this was the initial lack of specificity between EC and CM categories. The process of IRR helped us to elaborate the coding scheme to better differentiate between these bins. Even after more explicitly defining these bins, the presence of a few extraneous responses in each round was consistently keeping the kappa value below 0.7. This is perhaps due to the higher level of sensitivity present with our small test set size of 30. To address this, we increased the test set size to 50 text responses. The first round after increasing the test set size (and the eighth round overall) we passed the threshold, reaching a kappa of 0.78 (88.0% agreement).



### APPENDIX C: Full activity prompts and additional information

What do you think causes the helium atoms to move **towards each other**?

Hint: Think about what's happening with the protons and electrons.

Figure 4.9. Formative LDF text prompt administered to general chemistry 1 students via beSocratic in fall of 2014.

Sketch a molecular level picture to show what you think is going on as the atoms move together.

Draw Erase  

Use this space to describe in words what you are trying to represent in your drawing.

Figure 4.10. Formative LDF drawing prompt administered to general chemistry 1 students via beSocratic in fall of 2014.

As the atoms get closer they **attract** one another and the potential energy **decreases** as shown by the circled region. Please explain **why** the atoms attract and the process by which it occurs.

Hint: Think about what's happening with the electrons.

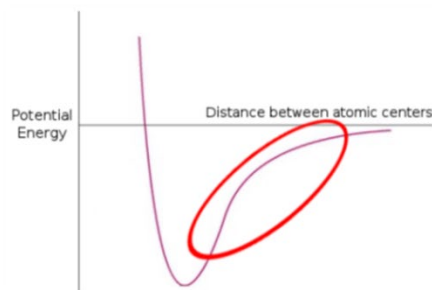


Figure 4.11. Item 1 (formative) LDF text prompt administered to the 15-16 cohort in fall of 2015 during general chemistry 1 via beSocratic.

Now **draw** a molecular level picture of the helium atoms illustrating **the process of attraction** that causes the atoms to move together. Be sure to include the subatomic particles in your drawing.

If there's anything you can't explain in your drawing, explain it in the box below.

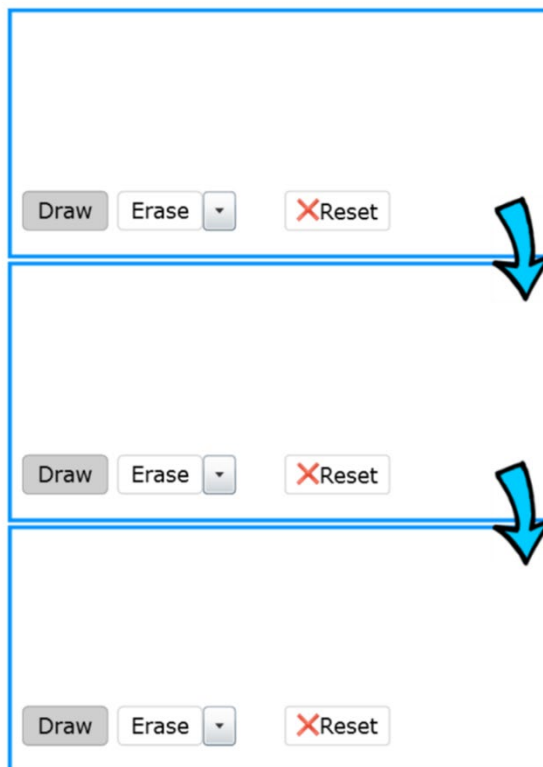
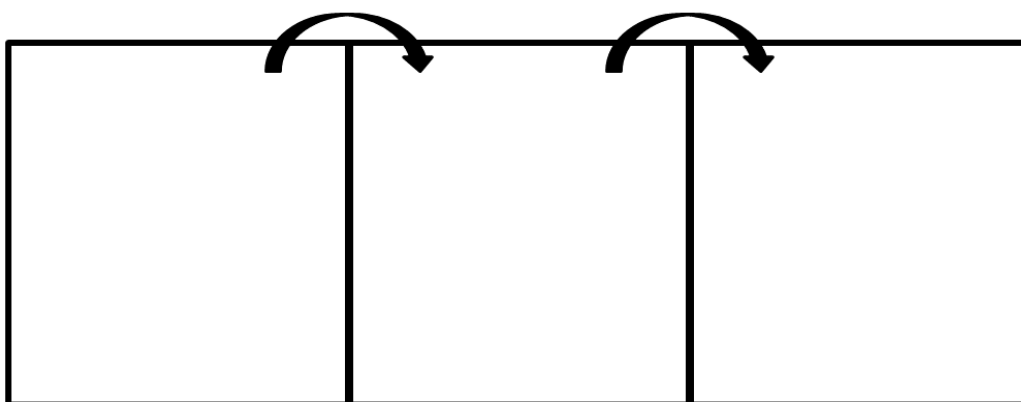


Figure 4.12. Item 1 (formative) LDF drawing prompt administered to the 15-16 cohort in fall of 2015 during general chemistry 1 via beSocratic.

**Consider the noble gas Kr.**

In the boxes below, draw a **molecular level picture** of what happens when two atoms of Kr approach each other. Use the picture to help you **explain** why the two atoms are attracted to each other.



Explanation:

Figure 4.13. Item 2 (summative) LDF text and drawing prompt administered to the 15-16 cohort in fall of 2015 during general chemistry 1 on the first midterm exam.

Is there an attraction between two neutral Helium atoms? ☐ Yes

☐ No

☐ I don't know

Below, explain in detail why or why not. Be sure to include any evidence you have for your claim and how this supports your reasoning.

Figure 4.14. Item 3 (homework posttest) LDF text prompt administered to the 15-16 cohort in spring of 2016 during general chemistry 2 via beSocratic.

Draw a molecular level picture of two helium atoms as they approach one another in the blue box, and explain what is happening in the black box.

Use the erase button to erase part of your drawing. Only use the reset button if you want to erase **everything**.



The interface consists of two adjacent rectangular boxes. The left box has a blue border, and the right box has a black border. Below the blue box, there are three buttons: a grey 'Draw' button, a white 'Erase' button with a small downward arrow, and a white 'Reset' button with a red 'X' icon.

Figure 4.15. Item 3 (homework posttest) LDF drawing prompt administered to the 15-16 cohort in spring of 2016 during general chemistry 2 via beSocratic.

1. Consider the noble gas Ar.

1a. Draw a **molecular level picture** of what happens when two atoms of Ar approach each other. Use the picture to help you ***explain*** why the two atoms are attracted.

Figure 4.16. Item 4 (exam posttest) LDF text and drawing prompt administered to the 15-16 cohort in spring of 2016 during general chemistry 2 on the last midterm exam.

#### APPENDIX D: Additional contingency tables

Table 4.6. The contingency table for the item 1 text and drawing LDF responses. In each cell the adjusted residual value is reported along with the observed value and the value we would expect if there was no relationship between the text and drawing responses (expected value). Adjusted residuals larger than the Bonferroni adjusted critical value ( $\pm 2.78$  for 9-cell table) are bolded. To visualize the sign and magnitude of the adjusted residuals, the cells are color coded from dark red (most negative) to dark blue (most positive).

Adjusted Residual Values				
Item 1		Drawing		
		NE	EC	CM
Text	NE	<b>4.82</b> Expected: 4.6 Observed: 13	-0.76 Expected: 5.4 Observed: 4	<b>-3.26</b> Expected: 12.0 Observed: 5
	EC	-1.72 Expected: 13.2 Observed: 9	2.76 Expected: 15.8 Observed: 23	-0.99 Expected: 35.0 Observed: 32
	CM	-1.72 Expected: 13.2 Observed: 9	-2.22 Expected: 15.8 Observed: 10	<b>3.32</b> Expected: 35.0 Observed: 45

Table 4.7. The contingency table for the item 3 text and drawing LDF responses. In each cell the adjusted residual value is reported along with the observed value and the value we would expect if there was no relationship between the text and drawing responses (expected value). Adjusted residuals larger than the Bonferroni adjusted critical value ( $\pm 2.78$  for 9-cell table) are bolded. To visualize the sign and magnitude of the adjusted residuals, the cells are color coded from dark red (most negative) to dark blue (most positive).

Adjusted Residual Values				
Item 3		Drawing		
		NE	EC	CM
Text	NE	<b>3.66</b> Expected: 17.0 Observed: 27	-2.06 Expected: 21.9 Observed: 16	-1.77 Expected: 10.1 Observed: 6
	EC	-0.90 Expected: 24.6 Observed: 22	-0.23 Expected: 31.7 Observed: 31	1.34 Expected: 14.7 Observed: 18
	CM	<b>-3.17</b> Expected: 10.4 Observed: 3	2.71 Expected: 13.4 Observed: 20	0.40 Expected: 6.2 Observed: 7

## CHAPTER V - DEVELOPING COMPUTER RESOURCES TO AUTOMATE ANALYSIS OF STUDENTS' EXPLANATIONS OF LONDON DISPERSION FORCES

### Preface

In this study, we use a previously developed prompt and coding scheme to characterize students' explanations of the origins of London dispersion forces in order to develop machine learning resources that can carry out such an analysis for large numbers of students. We found that by using large numbers of human coded student responses (N = 1,730) we could subsequently automatically characterize students' responses at a high level of accuracy compared to human coders. This research has been previously published in the Journal of Chemical Education and is reprinted with permission from Noyes, K.; McKay, R. L.; Neumann, M.; Haudek, K. C.; Cooper, M. M. Developing Computer Resources to Automate Analysis of Students' Explanations of London Dispersion Forces. *J. Chem. Educ.* **2020**, *97* (11), 3923–3936. Copyright 2020 American Chemical Society.

A copy of permissions obtained is included in the Appendices. Supporting Information for this manuscript is included in the Appendices.

### Introduction

With our rapid advances in technology, automating assessment of students' responses has become an emerging possibility for the field of education. However, the use of automated text analysis technology is relatively new in the field of chemistry education research. While there are many assessment systems that integrate forced choice or numerical response tasks, the assessment of student constructed written responses to complex chemistry prompts is not as available. Such assessment tasks can provide instructors with important information about what students know and can do; however, there are currently limited resources that allow for meaningful analysis of student explanations of chemical phenomena. Here, we use our work on students' explanations of the origins of London dispersion forces (LDFs) to explore an approach to the development of machine learning resources. We

also investigate whether these resources can code undergraduate students' responses similarly to how humans code those responses, and whether these resources can detect different signals (i.e., different ways in which students explain this phenomenon).

### ***The importance of assessments***

Assessment of student learning can be thought of in terms of evidence-based arguments.<sup>22,25</sup> Using this framework, evidence (in the form of student responses) is gathered from assessment items and subsequently used to support arguments about what students know and can do. Once instructors have such evidence, they can use it to evaluate the depth of student learning and to revise learning materials and instructional methods to support more robust understanding. In this way, assessments become more than just a way to evaluate students, but also an important tool to support students' learning by providing the instructor important feedback about the design and effectiveness of educational materials and teaching strategies.<sup>21</sup>

However, this design cycle is only effective if the assessment tasks elicit appropriate evidence about student learning. While instructors may infer that students are using appropriate reasoning to answer a question, if that question does not explicitly elicit such evidence, there is a strong likelihood that some students are using rules of thumb and learned or taught heuristics to answer questions.<sup>48,52</sup> For example, many students who can answer multiple choice questions about intermolecular forces (IMFs) and their role in physical properties have been shown to construct drawn representations of IMFs acting within a molecule (rather than between molecules).<sup>39</sup> It has been noted that students may use the presence of hydrogen bonding to predict relative boiling points, while at the same time having erroneous ideas about the nature of hydrogen bonding itself.<sup>48,73,74</sup> Indeed, the idea that boiling water produces hydrogen and oxygen becomes less surprising when we recall that students are told that boiling water breaks hydrogen bonds. The fact that there are many students who go through years of chemistry instruction without understanding the nature of IMFs, and the fact that their instructors were



unaware of this problem, may be because many traditional assessments do not elicit explicit evidence of the students' understanding of IMFs.

This process is even more challenging because the design of the task and associated coding scheme is crucial to whether appropriate reasoning can be elicited and captured. In this study, we characterized students' responses on the basis of a previous rubric we developed that captures the degree to which a student's explanation provides a causal mechanism for the phenomenon of LDFs.<sup>7,75</sup> This type of analysis is important because supporting causal mechanistic thinking is an important goal of science education.<sup>1</sup> Developing an understanding of how and why unseen entities behave and give rise to phenomena gives learners the ability to generate robust explanations and make predictions.<sup>5</sup> This approach goes beyond simply identifying what problematic ideas a student might have. Instead, by identifying the resources students use to explain the mechanism, we can begin to understand how best to support student learning.

### ***Designing assessments***

While there is certainly a place for multiple choice assessments, such items typically do not provide the kind of evidence that would allow us to make convincing arguments about what students know and can do, particularly if we want to go beyond recognition of fragmentary information or algorithmic problem solving. Whether a student gets the answer right or wrong, their thinking is not visible, so we cannot know how the student arrived at that answer or the type of reasoning they have engaged in. An ideal way to explore student thinking would be to ask them directly, for example, in an interview setting. However, such a method becomes impractical in classroom settings where an instructor may be responsible for several hundred students. Compromises must be made, so instructors instead often use multiple choice or constructed response assessments (in which the student must generate a text response). However, these two types of assessments do not necessarily elicit the same student understanding. For example, Nehm and Schonfeld<sup>76</sup> used multiple choice, constructed response,

and interview questions to assess their students' understanding of natural selection. They found that the understanding conveyed in the interview responses aligned more closely with the constructed responses compared to the multiple choice responses. Other studies have reported similar findings, that multiple choice questions tend to overestimate what students know compared to constructed response questions.<sup>77,78</sup> Additionally, answering only multiple-choice questions or questions that do not elicit reasoning does not explicitly provide students with the opportunity to reason and reflect on their responses. This process is integral to learning; indeed, asking deep explanatory questions is one of the few pedagogical techniques to promote learning that is supported by strong evidence.<sup>79</sup>

However, asking such questions typically requires multiple rounds of prompt design, hand coding of responses, and refinement of the prompts from the resulting evidence in order to better elicit student ideas. For example, the constructed response questions used by Nehm and Schonfeld<sup>76</sup> were originally developed by Bishop and Anderson<sup>28</sup> and were subjected to multiple rounds of design, coding, and redesign. We have previously reported on our efforts to elicit and characterize stronger evidence about IMFs, acid-base chemistry, simple nucleophilic substitutions, and LDFs.<sup>7,29,39,80,81</sup> The assessments in each of these studies went through a similar iterative design process. Assessment items that are too vague typically do not produce rich responses, while prompts that are too specific tend to signal to students what the desired responses should be. Prompts that do not provide enough direction about what is required, or that do not activate appropriate resources may lead students who might otherwise provide a rich response to give a more simplistic answer. The goal is to find the prompt that is “just right”, allowing students to tell us what they know.

For example, consider the prompt we used in this study: our previously developed LDF prompt.<sup>7,75</sup> We designed this question to explore students' explanations of how and why neutral atoms attract. We wanted to know if students could leverage their knowledge of the electrons and protons within the atoms and unpack their properties (e.g., their charges, how they move) to explain this

phenomenon. Such an explanation that links behaviors of the entities a scalar level below the phenomenon would be evidence of causal mechanistic reasoning, a powerful form of reasoning in science.<sup>5</sup> Our initial efforts to elicit this kind of response resulted only in surface-level descriptions. In order to develop a question that elicited causal mechanistic reasoning and a subsequent coding scheme to characterize those responses we interviewed students, piloted multiple prompts, and analyzed hundreds of students' responses. For more information about this process, see Becker et al.<sup>7</sup> and Noyes and Cooper.<sup>75</sup>

### ***The role of formative, or low-stakes, assessments***

Formative assessment typically refers to assessments that are low-stakes, that provide students with an opportunity to “try out” ideas without penalty, and to receive feedback, ideally leading to improvement over time.<sup>82,83</sup> Generally, and particularly for large enrollment courses, it is not feasible to use such items for assessments that must be responded to on a daily basis. Indeed, this is one reason why many large enrollment courses have defaulted to machine scorable multiple choice and fill in the blank items that can be easily scored and even provided with automated feedback. In our work, we do use items that require explanations, arguments, and drawings. In these assessments, students are awarded credit for completion, not for accuracy; that is, we do not grade the answers but instead provide aggregate feedback and discussion in larger groups. However, if we want to learn how students are meeting the challenges of such items, it becomes necessary to hand score which is time-consuming and expensive both in terms of resources and personnel. Generally, we have found that students' responses on these formative assessments are quite similar to their responses on analogous summative constructed response items.<sup>75</sup>

### ***The case for machine learning***

Machine learning may provide an answer to this conundrum by automating the analysis of the students' responses to such items. There are growing numbers of reports describing such approaches to

the analysis of short, concept-based, text responses in a variety of STEM education areas including acid-base chemistry,<sup>84–86</sup> the definition of randomness in statistics,<sup>87</sup> the role of the stop codon in genetics,<sup>88</sup> and biological mechanisms of weight loss,<sup>89</sup> among others. We have been working with a machine learning group, the Automated Assessment of Constructed Response (AACR) group,<sup>90,91</sup> to develop machine learning resources capable of analyzing the LDF question that we previously developed.<sup>7,75</sup> The AACR group has developed tools available on the web, such as the Constructed Response Classifier (CRC) tool which is discussed in this paper, that use open-source machine learning techniques to create computer models that are capable of mimicking human analyses by identifying key disciplinary concepts or reasoning patterns in student responses. Once these computer models have been developed, they can be used to analyze large numbers of new text responses from students just as an expert would but requiring only a few minutes. A tool like this would then provide instructors a way to incorporate constructed response questions more routinely in their teaching. For reasons we will discuss later, such resources are not intended to give high-stakes feedback for individual students but, rather, useful and timely information about how a set of students understands or reasons with a particular idea.

### ***Overview of the Constructed Response Classifier***

One of the biggest challenges with automating analysis of student responses is developing computer resources that perform well and that are able to recognize nuanced and complex explanations in short text responses. AACR's CRC is a promising tool for machine learning analysis because (1) nearly all of the process is automated and (2) this technology uses several machine learning algorithms in an ensemble.<sup>92</sup> The first point is important because while other software, such as SPSS Modeler, can perform lexical analysis, much more human input is needed. For example, Dood et al. used SPSS Modeler for automated analysis of student responses, and while the software could identify common terms in the responses, humans needed to specify synonyms and create categories and rules within the software in order to develop the predictive models.<sup>84,85</sup> This means that the researchers need to develop

a meaningful way to analyze the responses, code the responses, and then translate their subsequent scheme into the software manually.

In contrast, the CRC contains a set of open-source, supervised machine learning classification algorithms developed by Jurka et al.<sup>93</sup> which can be used to predict scores of responses after “learning” from a set of previously scored student responses. As part of this routine, text in student responses is automatically extracted and parsed into *n-grams*, words and sets of words up to a defined number *n*, which are used as independent variables in the classification algorithms. Such an approach reduces the need for developing lexical resources like defining term dictionaries or creating combinatorial rules (see Nehm et al. and Kaplan et al.).<sup>87,94</sup> That is, the researcher only needs to develop and apply a coding scheme; the machine learning tools handle the rest thereby reducing the time and effort needed from researchers to develop these resources.

Additionally, the classification procedure developed by Jurka et al.<sup>93</sup> combines results from 8 machine learning algorithms to predict a single overall score for each response. By using multiple classification algorithms, the resulting predictions are generally more accurate than using only one algorithm (e.g., Optiz and Maclin).<sup>95</sup> The outcome of this process is that, once trained with human coded responses, the CRC is able to analyze “raw” student responses and predict codes on the new set of data.

### ***Challenges with machine learning and potential solutions***

Even with the machine learning techniques that the CRC uses, the fundamental challenges of automated analysis remain. That is, these computer models are mimicking human analysis of text responses, and therefore, the performance of the model is highly dependent on the rubric and the text responses used to train the model. In developing our computer resources, we tried to address both of those dimensions.

The first dimension is the rubric used to code the responses. Since the computers are mimicking the human analysis, if the humans cannot code reliably, neither will the machine. The rubric we developed to characterize students' LDF explanations was the product of analyzing student interviews, homework assignments, and exam responses, from which we iteratively refined the rubric to capture all the ways students could explain this phenomenon using causal mechanistic reasoning. The end result is a rubric that can capture reasoning and can be used reliably.

The second dimension is that of the text responses. Automated analysis techniques work by identifying patterns in the words and terms used by responses classified into each coding category so that, when presented with a new response, the model can analyze the words in the response to make a prediction. The better the model can identify the key features of each coding category (based on the patterns of words in the response), the more likely the computer is to score that response like the human. This requires a large number of human coded responses (in each coding category) upon which the computer model can be built. Therefore, we collected and took the time to code a very large number of student responses so critical components of each coding category would become salient. In our original analyses,<sup>7</sup> we defined six coding categories for student responses, but in subsequent work, we consolidated these into three coding categories.<sup>75</sup> As we discussed in our previous work,<sup>75</sup> we consolidated the number of coding categories to capture causal mechanistic explanations more concisely and also create a more practical scheme for humans to use to analyze thousands of responses. This also assisted the development of automated resources; by decreasing the overall number of coding categories, we increased the number of responses in each coding category.

It is not just the number of responses that is important but also their content. If we want to develop resources that can be useful across different institutions and courses, then the model must be trained with responses from a variety of learning environments. Otherwise, if these resources encounter a response that explains the phenomenon causal mechanistically, but using words not captured in the

training set, the model would be unable to code the response accurately. For example, Ha et al.<sup>96</sup> explored how accurately a model developed from responses collected at one institution could code responses from another institution. They found that, while the computer could accurately code some key concepts related to evolutionary change for students at both institutions, it did not accurately code all of the key concepts that they intended to capture. Ha et al.<sup>96</sup> note that one of the issues affecting the accuracy of the models across the different institutions is difference in language patterns unique to each institution. Our study is similar to that of Ha et al.<sup>96</sup> in that we also collect responses from multiple undergraduate institutions and different student populations to get a wider range of the ways students explain our phenomenon under study. However, we build upon their findings in this study by developing, and subsequently testing, a model from a combination of responses from all the institutions to capture any differences in the lexical patterns of the responses. Collecting responses across different institutions to develop and test our resources is also important so that the technologies we develop are equitable; that is, these resources are able to characterize the responses from students of all different backgrounds like a human coder would. By addressing these challenges, we hope to develop machine learning resources that can code student reasoning as part of their LDF explanations as well as humans can. Such a resource could be used to give instructors meaningful feedback about how their students understand the mechanism by which this IMF operates to better support student learning.

### **Research questions**

1. How does the machine coding for causal mechanistic explanations compare to humans?
2. How does the machine coding for different groups of students compare?

### **Methods**

#### ***Strategy for developing machine learning resources***

In the methods section, we describe the process by which we collected and analyzed data to develop supervised machine learning models. Our strategy was to first collect and human code many

student responses, from a variety of contexts. Then, these coded responses were used to train the machine learning algorithms in order to develop resources that are capable of coding new responses in the same way as a human coder. The accuracy of the developed models was tested by applying the computer model to new sets of student responses and then comparing the results to the human coding of a subset of these new responses.

### ***Participants***

Responses from four groups of students from three different undergraduate institutions are analyzed here. The data were collected in accordance with each institution's IRB protocols, and at each institution, students consented to having their responses used for research purposes. Students' responses were deidentified before analysis. In the sections below, we provide some additional context for each of the four groups of students. We present more thorough descriptions of the demographics of each group in Appendix B. We note that the demographic information reported was determined by the Registrar's office from each institution and therefore has limitations (e.g., the conflation of a student's racial and ethnic identity, the reporting of only male and female gender identities).

#### ***Group 1A***

Group 1A is composed of students in the first semester of general chemistry from a large midwestern public research institution (which we call institution 1 in this paper). Specifically, these students were taking the first part of a 2 semester transformed general chemistry curriculum CLUE (Chemistry, Life, the Universe and Everything)<sup>37</sup> in the fall semester. While General Chemistry 1 is also offered in the spring, a majority of students at this institution take General Chemistry 1 in the fall. Of the 2,497 students enrolled in this course in fall 2015, 91% responded to our activity and consented for their response to be used for research (N = 2,284). We used 950 of those responses for computer model training. We discuss more about these responses and how they were selected in the later section in this paper about model development. We analyzed more responses from this group than any other because



the initial model development process requires lots of responses and we collected the most responses from this group. This group of 950 students was primarily White (69%), and about half of the group was female (55%) (Appendix B).

#### *Group 1B*

We also collected responses from General Chemistry 1 students at institution 1 in the spring of 2016. In this paper, we describe these “off-sequence” students as group 1B. Some students take such sections because the university has required them to complete additional math courses to prepare for general chemistry. This may be why the difference in mean ACT math scores for group 1A (mean = 25.8, N = 1,925) and group 1B (mean = 24.5, N = 776) is statistically significant (Appendix B). Like group 1A, the students in group 1B also had the CLUE general chemistry curriculum.

Of the 1,070 students enrolled in this course, 86% of the students (N = 915) responded to our prompt and consented for their responses to be used for research. Of those 915, we randomly selected 350 responses (using a random number generator) for this study. We selected only a portion of the total number of responses for machine learning development and testing because we wanted to have a group similar in size to groups 2 and 3. Like group 1A, the 350 randomly selected group 1B students were primarily White, (70%) and about half were female (54%) (Appendix B).

#### *Group 2*

Students in group 2 were enrolled at another large midwestern public research university, which we call institution 2 in this paper. We recruited this group from the first semester of a 2-semester general chemistry course in spring of 2018. Unlike institution 1, this institution used a traditional curriculum. The textbook listed on the course syllabus was Chemistry: The Central Science (14<sup>th</sup> ed.) by Theodore Brown.<sup>97</sup> We note that in this curriculum students received less explicit instruction about the construction of causal mechanistic explanations than those students in the CLUE curriculum, where such reasoning is an explicit focus of the course.

Of the 721 students enrolled, 53.3% of students (N = 384) responded and consented for their responses to be used for research purposes. The majority of those who responded to the activity were female (65%). While we did not collect information about the racial/ethnic identities of the students in this group (we did not get IRB approval to collect this information from this institution), we can approximate the demographic makeup of this group from the information available on the registrar's website about the entire institution. For the entire undergraduate student body (N = 23,856), the students were primarily White (73%) (Appendix B).

### *Group 3*

Group 3 is made up of students from institution 3, a large southeastern public research university, who had the first semester of general chemistry in either fall 2016 or fall 2017. This general chemistry course also used the CLUE curriculum but in a different format from institution 1. At institution 1, the course was taught in large lecture sections (between 360 and 450 students per section) with smaller weekly recitations, while at institution 3 the course was taught in a partially flipped classroom format, where 100-200 students spent the majority of the time during class working in groups, receiving instruction and additional activities as homework outside of the classroom.<sup>98</sup>

Of the 449 students who received the activity over the two years, 77.1% of students (N = 346) submitted responses and consented for their work to be used for research purposes. Of these students, the majority were Hispanic (71%), and 8% were White (Appendix B). This is quite different from the racial/ethnic identities of the students from institutions 1 and 2 where no more than 10% of the students identified as Hispanic. Additionally, just over half of this group was female (52%).

### ***Question prompt***

To probe students' understanding of how LDFs arise, we asked students to explain why two helium atoms attract one another. This prompt was developed as part of our previous work eliciting causal mechanistic explanations of LDFs.<sup>7,75</sup> Originally, we developed this prompt to include a

corresponding drawing component, but for the purposes of this study, we focus only on automatically analyzing the text portion. While groups 1A, 1B, and 3 still received that drawing component, it was administered on a separate slide after the text prompt. When the student reached the drawing prompt, they could not return to the text prompt to change their response. Practically, this means that the student answered the text prompt without ever seeing the drawing prompt; therefore, it appears that we are able to code students' text responses independently from their drawings.

As the atoms get closer they **attract** one another and the potential energy decreases as shown by the circled region. Please explain **why** the atoms attract and the process by which it occurs.  
Hint: Think about what's happening with the electrons.

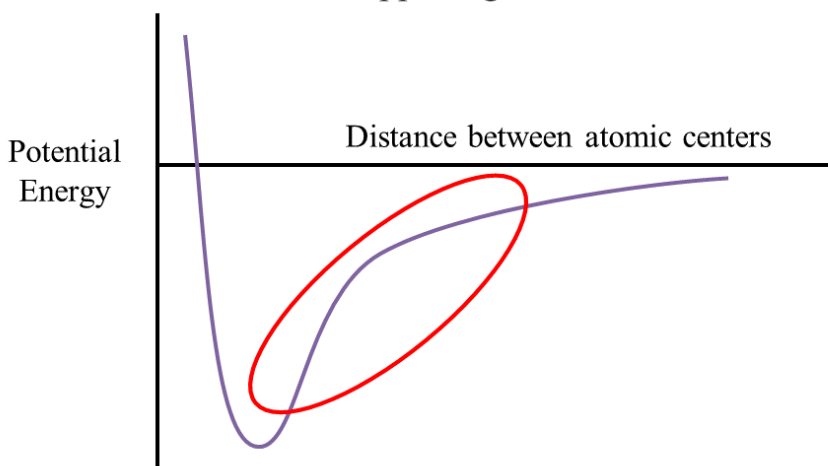


Figure 5.1. LDF prompt.

For groups 1A, 1B, and 3, this prompt (Figure 5.1) was included as part of their homework activities. These formative assessments were a required part of the course, but the students received course credit for completing the activity, not for the correctness of their response. Both institutions 1 and 3 used the online beSocratic system<sup>63</sup> for homework which allowed us to collect responses digitally, a helpful feature when automating analysis of the responses. Additionally, since this question was included in their homework, the students completed this activity shortly after instruction of LDFs. In the

CLUE curriculum, introduction of IMFs including LDFs occurs early as these ideas are continually built upon throughout the course.

For group 2, we administered the activity in a different manner. In this course, the instructor gave the students the opportunity to answer this question for a small amount of extra credit, but it was not a required part of the course. This may explain why the response rate was lower for institution 2 compared to institutions 1 and 3 where the question was part of a required homework activity. Since it was not part of a broader activity, the prompt was asked as a standalone question. Additionally, the instructor did not have access to beSocratic and so instead used the online survey system Qualtrics<sup>99</sup> to collect the students' text responses (we did not administer the drawing component to this group). The timing for this activity also differed as it was given to students near the end of the semester. We acknowledge that these conditions differed substantially from those at institutions 1 and 3, but the purpose of this study is to collect a lexically diverse set of responses to train and test machine learning resources, not to compare across groups. By giving this prompt in contexts where the administration of the prompt, student demographics, and instructional environment differ, we accomplished this goal.

### ***Coding scheme***

To analyze the responses, we used the coding scheme that we had previously developed to characterize the degree to which the students engaged with causal mechanistic reasoning.<sup>75</sup> This holistic, mutually exclusive coding scheme places responses into one of three categories: nonelectrostatic (NE), electrostatic causal (EC), and causal mechanistic (CM) (Table 5.1). NE responses fail to provide any electrostatic evidence for this interaction. Electrostatic causal responses discuss the role of electrostatic attractions in this interaction but do not include the mechanism by which these interactions form. Causal mechanistic responses go further, explaining the happenings at the scalar level below: how the electrons temporarily can localize on one side of atom which results in the separation of charge that causes this interaction.

Table 5.1. Overview of coding scheme from Noyes and Cooper.<sup>75</sup>

Type of Response	Text Features	Student Example
Nonelectrostatic (NE) text response	The response does not include reasonable electrostatic evidence of the interaction. Instead, the response provides nonelectrostatic evidence or does not address the intermolecular interaction between molecules.	“The two atoms are attracted because of the electromagnetic forces that exist between them”
Electrostatic causal (EC) text response	The response indicates that electrostatic charges cause the interaction. Examples of electrostatic causal evidence include subatomic particles, overall charge of the atom, partial charges, etc. These responses do not include a mechanism by which a separation of charge occurs.	“Atoms attract to each other [because] the [London] dispersion forces make the partially negative end of one atom attract to the partially positive end of another atom.”
Causal mechanistic (CM) text response	The response indicates that the interaction occurs due to electrostatic charges and includes the mechanism by which the instantaneous and/or the induced dipole forms.	“As the 2 atoms approach each other one of them becomes [instantaneously] dipole due to fluctuation in its electron cloud as most of the electrons are concentrated to one side of the atom making it have a partially positive and negative charges on opposite sides. This partially positive side of the attracts the [electrons] of the other atom making its electron cloud fluctuate and form an instantaneous dipole where both of them attract each other.”

### ***Human coding of responses***

Supervised machine learning relies on a set of “labeled” data in order to train or develop the computer model. As used here, the human codes assigned to students’ responses were the “labels” necessary for training the computer model. We report descriptive information about the response length and provide some additional example responses in Appendix C.

Before conducting any coding, we deidentified all student responses and established inter-rater reliability (IRR) between the researchers. In this process, two people independently coded small sets of

students' responses and then calculated Cohen's kappa, a measure of agreement that also considers the probability of agreement by chance.<sup>71</sup> Once the Cohen's kappa value surpassed 0.7, a value corresponding to a "substantial" level of agreement,<sup>72</sup> we determined that we had reached IRR. At this point, we could begin coding the responses that would later be used to train and test the automated resources. We provide an overview of the process by which we established IRR and conducted subsequent coding in the following paragraphs and Figure 5.2.

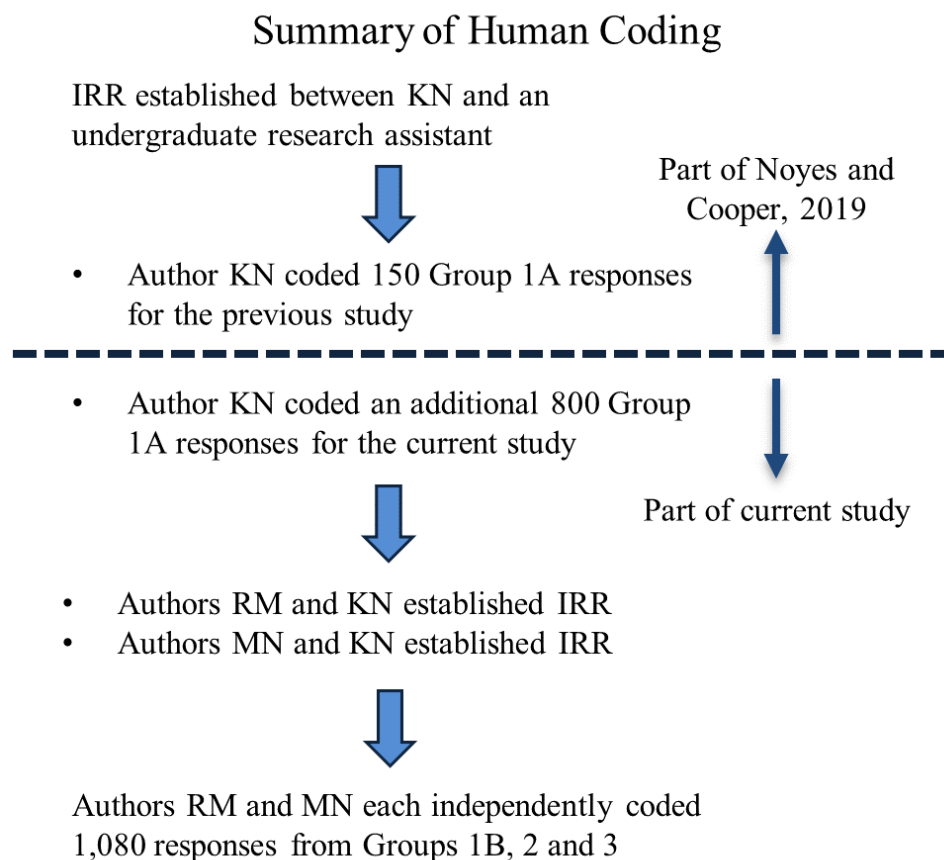


Figure 5.2. Summary of the human coding reported in this study with the distinction as to which data was reported in our previous study, Noyes and Cooper.<sup>75</sup>

Before coding the group 1A responses, author K.N. and an undergraduate researcher conducted IRR. After eight rounds of coding (30 responses in rounds 1-7 and 50 responses in round 8), we established IRR (Cohen's kappa = 0.78, 88% agreement). During this process, the two researchers met

after each round of coding to discuss disagreements and, if needed, refine the coding categories. This process is discussed more in depth in the Supporting Information S1 of Noyes and Cooper.<sup>75</sup> Author K.N. then coded 950 responses from group 1A to train the initial computer model. We note that Author K.N. coded 150 of these responses as part of a previous study.<sup>75</sup>

The coding of responses from groups 1B, 2, and 3 was carried out by authors R.L.M. and M.N. Before coding, both authors separately conducted IRR with author K.N. using sets of 40 previously uncoded responses from group 1A. Authors R.L.M. and K.N. reached IRR after two rounds (Cohen's kappa = 0.88), and authors M.N. and K.N. also reached IRR after two rounds (Cohen's kappa = 0.81), with both Cohen's kappa values corresponding to "Almost Perfect" agreement.<sup>72</sup> While there are other statistical tests to calculate agreement between three coders such as the weighted Cohen's kappa, authors R.L.M. and M.N. began working on this project at different times and therefore were not trained on this coding scheme simultaneously.

With IRR established, both authors R.L.M. and M.N. coded the selected responses from groups 1B, 2, and 3. To minimize bias, the authors were not aware of the institution affiliation of the responses they were coding. These responses were coded individually in batches of approximately 100 responses, and then compared. When comparing, the two authors discussed any discrepancies between their individual codes and assigned a final, mutually agreed upon code to any disputed responses. This continued until the responses of an entire group were coded. This process was then repeated for all three groups. We characterized the initial level of agreement between the coding of authors R.L.M. and M.N. for each group by calculating Cohen's kappa and percent agreement values for their initial codes (Table 5.2).

Table 5.2. The initial level of agreement for each of three cohorts between authors RM and MN.

Group	Number of Students	Cohen's Kappa	Percent Agreement
1B	350	0.71	81%
2	384	0.74	89%
3	346	0.69	81%

### ***General overview of how the CRC works***

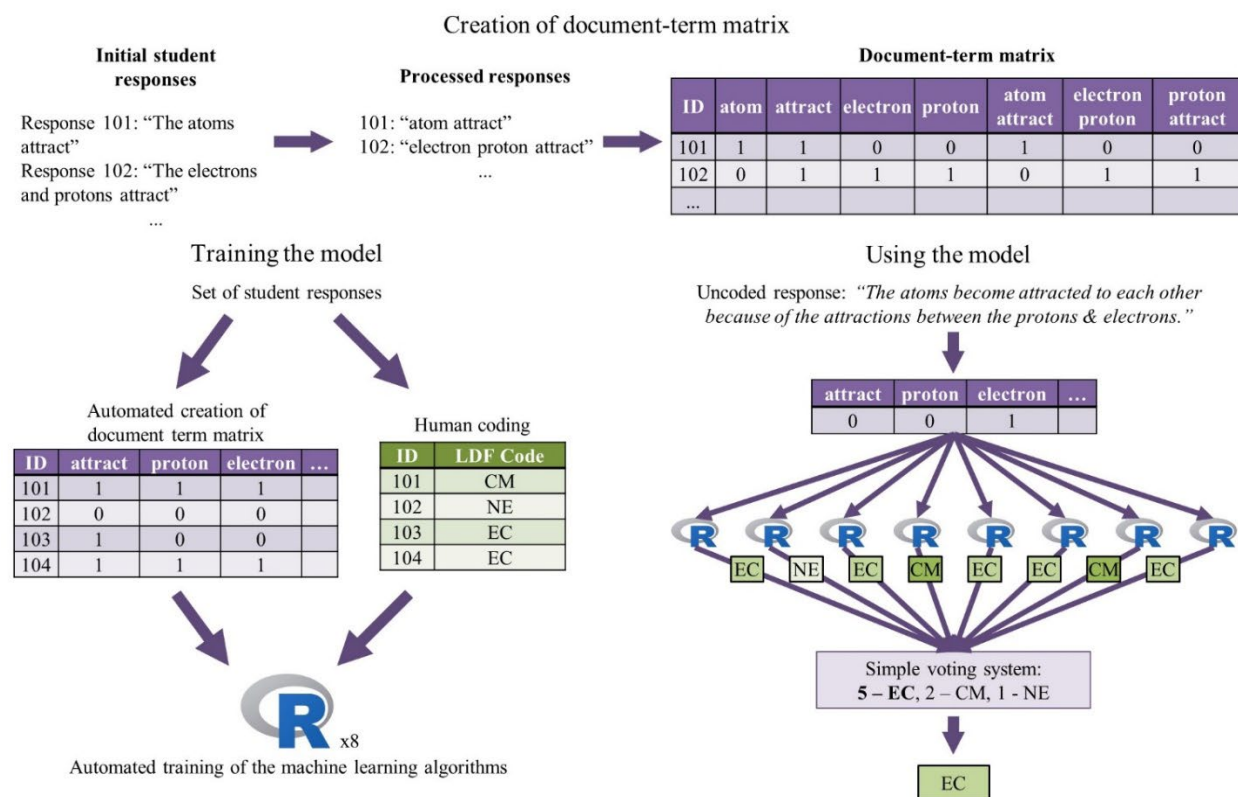
The AACR group has developed machine learning tools to automate the analysis of students' written responses. We used AACR's CRC web application (the developed model is now accessible through the AACR project website: [beyondmultiplechoice.org](http://beyondmultiplechoice.org)) to develop resources to mimic our human coding of the responses to our LDF prompt. This app uses a series of eight machine learning algorithms derived from the open-source statistical package RTextTools developed by Jurka et al.<sup>93</sup> to predict the code for a response based on human coded responses (see also Sieke et al.<sup>92</sup>). Jurka et al.<sup>100</sup> provide a more detailed description of this package and its function, but we briefly describe the inner workings of the CRC tool here.

Before the training responses are input into RTextTools, the CRC "cleans" the responses, based on options selected by the user. In this process, the responses undergo "stemming" so that the suffixes (e.g., "attraction" becomes "attract"), stop words (e.g., "and", "the", "a", "in"), and numerical characters are removed. This leaves the important terms for the lexical analysis. The cleaned responses are then loaded into RTextTools.

Initially, the set of training responses is used to create a document-term matrix.<sup>101</sup> A document-term matrix parses out all of the individual words (unigrams) and pairs of words (bigrams) for each of the responses and captures it in a matrix. In Figure 5.3 we present a hypothetical document-term matrix



for a hypothetical pair of simple student responses. With hundreds of longer student responses, this matrix can get very large, very quickly.



downsides of any one machine learning algorithm, eight different algorithms are used to construct eight different predictive models: support vector machines,<sup>102</sup> supervised latent dirichlet location,<sup>103</sup> logitboost,<sup>104</sup> classification trees,<sup>105</sup> bagging classification trees,<sup>106</sup> random forests,<sup>107</sup> penalized generalized linear models,<sup>108</sup> and maximum entropy models.<sup>109</sup> When presented with a set of responses, the CRC cleans the responses, and then, RTextTools generates a document-term matrix for each particular response. All eight models then use the presence or absence of the unigrams and bigrams in the document-term matrix to classify each of the new responses into one of our three codes. Whichever code is picked by the most models is the machine's final consensus code for the response. In the unlikely event of a tie, the code assigned to the response is the first defined code with the maximum number of votes. For this coding scheme, the codes were defined in the following order: NE, EC, CM.

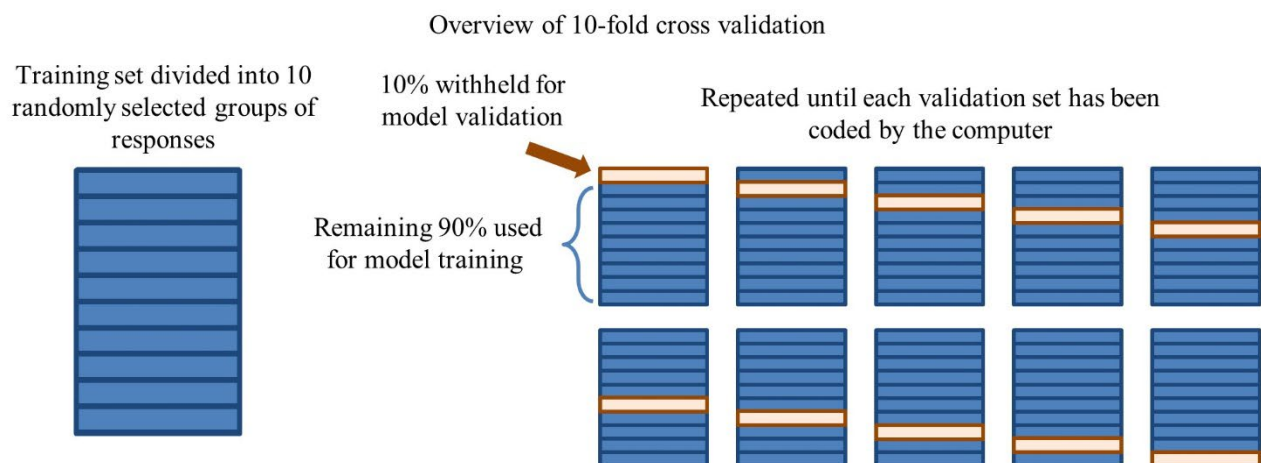


Figure 5.4. Overview of 10-fold cross-validation procedure for assessing the accuracy of the developed computer model.

One way we assessed the accuracy of the computer model is through a 10-fold cross-validation (Figure 5.4). This cross-validation method provides insight into the accuracy of the model without the need for any additional human coded responses; that is, only the human coded training set is needed. In the cross-validation (which is automatically carried out by the CRC), the training set is divided into 10 randomly selected subgroups, each corresponding to 10% of the training set. Each subgroup is then

coded by a new computer model trained with the remaining 90% of responses. This ensures that the same responses are not used simultaneously in the computer model training and testing. This process is then repeated a total of 10 times until the entire training set has received a computer predicted code. The CRC then calculates the agreement (using several statistics like Cohen's kappa) between the human and computer predicted codes generated in the cross-validation. By using the cross-validation built into the CRC, we can get an idea about how accurate the final computer model (trained on *all* the responses in the training set) would be without human coding any additional data. In this study, we also conducted additional tests to ensure that the final computer model is accurate. For this additional testing, we used the final computer model to predict codes for new sets of human coded responses from groups 1B, 2, and 3 which were not included in the training set and calculated the agreement between human and computer coding.

## **Results and discussion**

### ***Developing an initial model to characterize LDF responses***

The first stage in the development of a robust model was to use 150 group 1A responses coded as part of a previous study<sup>75</sup> to train the computer model. Before training the computer model, we used the spellcheck feature in Microsoft Excel to identify and fix errors in the responses (correcting misspelled words, deleting duplicate words) to help the computer evaluate the words present in the responses rather than the spelling. The accuracy of this first model compared to the human coding using the cross-validation procedure was "moderate", reaching a Cohen's kappa of 0.58.<sup>72</sup> This initial stage did not have enough responses to develop an accurate model. More responses were needed to increase the lexical diversity of the training set to help the model better identify the patterns of words associated with each code. Author K.N. coded additional sets of 100 group 1A responses (randomly selected using a random number generator) to include in the training set. After the addition of 100 more responses, we trained a new computer model and found that the Cohen's kappa value (calculated from the cross-

validation) had increased to 0.63. Each set of 100 responses was added iteratively to the training set, to train new models and assess the new model's performance using the cross-validation procedure (Figure 5.5). This process continued until no more meaningful improvement of the cross-validation Cohen's kappa value was observed. After 950 total responses were coded, the Cohen's kappa from cross-validation reached 0.77, signifying "substantial agreement" between the computer and human coding.<sup>72</sup> Recall that the initial human IRR was 0.78 for group 1A responses. Williamson et al. propose that one metric for a successful model is having a quadratic weighted kappa greater than 0.7.<sup>110</sup> In this study, we treated these categories as nominal rather than ordinal consistent with our assumptions from our previous study;<sup>75</sup> therefore, we did not use quadratic weighted kappa for our data. Our Cohen's kappa measure, a more conservative estimate, was above this target for model performance.

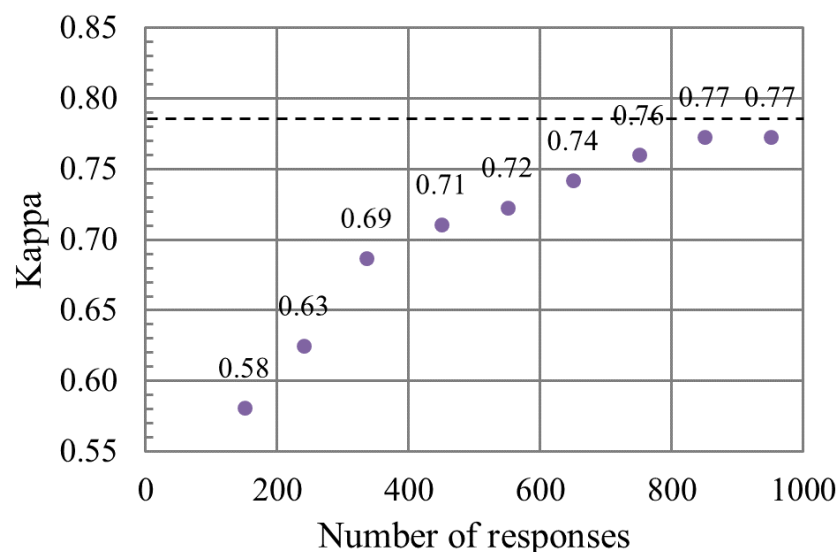


Figure 5.5. Agreement between the human and computer described by the Cohen's kappa value calculated in the 10-fold cross-validation as a function of the size of the number of responses used to train (and also validate) the computer model. The dashed line at 0.78 indicates the Cohen's kappa value for the human-human IRR with the responses.

In this paper, we will refer to the computer model trained on the 950 group 1A responses as the "initial model". A crosstab illustrating the coding agreement between the human and computer codes (determined in the cross-validation) for the initial model is shown in Table 5.3. Of the 950 coded, 813

were coded the same by both the human and computer, corresponding to a proportion of 0.86 (accuracy value). Although the human and computer disagreed on 137 responses, these disagreements occurred in a mostly symmetric manner. For example, while the computer coded 35 EC responses as NE, it also coded 30 NE responses as EC. The result was that these disagreements had a smaller impact when considering the overall distribution of responses for the entire group.

Table 5.3. Crosstab relating the number (and percentage of total) of reference human scores to the predicted computer scores for the training set of the initial model.

Initial model training set		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	178 (18.7%)	35 (3.7%)	4 (0.4%)	217
	EC	30 (3.2%)	379 (39.9%)	46 (4.8%)	455
	CM	1 (0.1%)	21 (2.2%)	256 (26.9%)	278
	Sum	209	435	306	950

Besides Cohen's kappa and accuracy, the sensitivity and specificity values provided important information about the performance of each of the bins in the model. Sensitivity is the proportion of correctly scored positive cases by the computer model, and specificity is the proportion of negative cases correctly scored by the computer model. For example, of the 209 human coded nonelectrostatic responses, the computer correctly coded 178 of those responses corresponding to a proportion of 0.85 (sensitivity value). Additionally, of the 741 responses that the human did not code as NE (i.e., coded as EC or CM instead), the computer coded 702 of those responses as not NE resulting in a specificity value of 0.95. For this model, all bins had both a high sensitivity (ranging from 0.84 to 0.87) and high specificity (ranging from 0.85 to 0.97). All of these factors indicated that the model was sufficiently trained and ready to characterize new responses, so long as they were also from this group of students. To make

sure that our model had captured all the ways a student might explain this LDF causal mechanistically, we needed to include responses in our training set from other groups of students.

### ***Expanding our initial model with responses from 3 other groups***

Now that we had a model that was working well with responses in a single context, we needed to know if the model performed well with different groups of students who may have approached this task differently or had different vocabularies. Using the responses from groups 1B, 2, and 3, we explored how the model performed with groups of students who differed from each other in terms of when they were taking general chemistry, their general chemistry curriculum, the circumstances under which they responded to the task, and the racial/ethnic makeup of the group.

After authors R.L.M. and M.N. coded the responses from these three groups, we set aside 100 randomly selected responses from each group for later testing of the machine learning models. The rest of the responses were added to the initial model to create a new computer model ( $N = 1,730$ ) which we call the “combined model” in this paper. As before, we spellchecked the responses included in the training set to give the algorithms the best chance of identifying the important patterns relevant to each category. With responses from a variety of different groups in the training set, the resulting combined model may be better able to capture other ways students explain this phenomenon that were previously not captured with the initial model.

Table 5.4. Crosstab relating the number (and percentage of total) of reference human scores to the predicted computer scores for the training set of the combined model.

Combined model training set		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	513 (29.7%)	67 (3.9%)	5 (0.3%)	585
	EC	64 (3.7%)	633 (36.6%)	71 (4.1%)	768
	CM	1 (0.06%)	39 (2.3%)	337 (19.5%)	377
	Sum	578	739	413	1730

On the basis of cross-validation, the agreement between the combined model and the human scoring was very similar to that of the initial model (Table 5.4); the Cohen's kappa and accuracy values (0.78 and 0.86 respectively) did not change much, and the ranges of sensitivity and specificity values (sensitivity, 0.82 - 0.89; specificity, 0.86 – 0.97) were very similar to those of the initial model. On the basis of these metrics, it seemed that the combined model performed well but no better than the initial model. This might be because more than half of the responses in this training set came from a single group of students (1A). This meant that metrics evaluating model performance from the cross-validation were primarily reflective of how the model codes responses from group 1A, making it harder to see how the new combined model was able to predict responses from groups 1B, 2, and 3. To get a better sense of how the combined model fared compared to the initial model for those groups, we used both models to score the sets of responses from groups 1B, 2, and 3 withheld from the combined model.

### ***Testing the model performance***

To conduct this test, we used the 100 coded responses from each group (1B, 2, and 3) that had already been human coded but not included in the training set. The responses used to test the models were not spellchecked to simulate how an instructor might apply this tool in practice, where raw

students' responses may be used. We report the agreement between the computer model predictions and the human consensus scores in Table 5.5. Note that the computer scores are compared to the human consensus scores; in other words, the final codes that authors R.L.M. and M.N. agreed upon after discussion. Their initial agreement (before discussing the responses) for each test set is also reported in Table 5.5. We included additional information about the alignment of human codes with both the initial and combined models in Appendix D.

Table 5.5. Cohen's kappa value and percent agreement between human coders and with the computer models for groups 1B, 2 and 3.

Group	Human - Human agreement	Human Consensus – Initial Model Agreement	Human Consensus – Combined Model Agreement
1B N = 100	0.74 (83%)	0.74 (83%)	0.74 (83%)
2 N = 100	0.72 (89%)	0.67 (86%)	0.72 (89%)
3 N = 100	0.64 (78%)	0.77 (86%)	0.80 (88%)

For group 1B, both the combined and initial models showed the same level of agreement with the human consensus scores. This made sense considering that the initial model was trained with group 1A responses, which were from students attending the same institution and taking the same general chemistry curriculum; they differ primarily in the semester that they took the course. For group 2, the combined model had slightly better agreement with human consensus scores than the initial model. The combined model correctly scored an additional three responses, all within the nonelectrostatic category. The majority of group 2 responses were classified as nonelectrostatic, so it could be that the addition of the group 2 responses in the combined model better allowed this model to correctly characterize nonelectrostatic responses.



For group 3, the combined model again showed higher level of agreement with the human consensus scores compared to the initial model, raising the Cohen's kappa value to 0.80. Interestingly, even for the initial model, the Cohen's kappa value (0.77) was quite a bit larger than the value for the agreement between the human coders for this same set (0.64). The lower level of agreement between the human coders was due to disagreements about how to code vague responses that were "edge cases" between the nonelectrostatic and electrostatic causal bins. It is promising that both computer models handled these new differences between the NE and EC bins well and that the addition of the group 3 responses to the combined model continued to improve its performance.

Further investigation of the responses misclassified by the combined model revealed that the computer struggled with the same responses as the human coders. If we consider the 300 test set responses from groups 1B, 2, and 3, the combined model misclassified 40 responses compared to the human consensus codes. Of those 40 responses, the human coders initially disagreed on how to code 19 of those responses (47.5%). Meanwhile, if we consider the remaining 260 responses to be correctly classified by the combined model, the human coders initially disagreed on only 31 of those responses (11.9%), a much lower proportion. Looking further into the responses misclassified by the combined model, we did not find that they were occurring primarily with any one particular code (see Appendix D). We also examined the content of those misclassified responses and found that the bulk were either "edge cases" or a particularly atypical explanation.

Notably, all the computer models tested showed little to no degradation in agreement to human scores when compared to human-human agreement. In other words, the Cohen's kappa value for the human-computer model agreement was not much different than the Cohen's kappa value for the human-human agreement. This was particularly true for the combined model which performed very near or above the human-human agreement level. Even the model showing the largest degradation from human-human agreement measures, the initial model for group 2 students, still performed at an

acceptable level, as the degradation was less than a suggested threshold of 0.1 difference (see Williamson et al.<sup>110</sup>).

All of these results suggest that the combined model, with additional responses from groups 1B, 2, and 3, can characterize student responses like human coders would, even improving upon the accuracy of the initial model coding of groups 2 and 3. While these improvements are modest, it does seem that they are the result of coding the responses from these different groups of students who may explain this phenomenon in ways we had not captured before. This is supported by the fact that the gains in agreement were only seen with groups 2 and 3, whose responses likely differ the most from group 1A. The high level of agreement between the computer and human coding for both of these groups is noteworthy. The students in group 2 are using a different curriculum in which they are not explicitly taught to generate causal mechanistic explanations, but still the computer model works. The high level of accuracy of the computer codes for group 3 is also important because this group of students is primarily Hispanic while the bulk of the other students in our computer model training and testing are White. This aligns with the joint recommendations for educational testing put forth by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, that in developing methods of scoring constructed responses (in particular automated scoring methods) we must be cognizant of the different subgroups of students in our testing populations and work to ensure that these resources are valid for all subgroups.<sup>111</sup> This is some evidence that these resources provide an equitable approach to the automated analysis of explanations; that is, these resources provide meaningful and accurate information that aligns with the human coding for diverse student populations.

### ***Detecting signal at the group level***

While we have achieved good agreement between the humans and the combined model, we have not reached perfect agreement (although we note that the agreement is at least as good as

human-human coding). Although there are bound to be errors in codes of individual student responses, the overall distribution of codes in a group (e.g., a class) may still be informative if the model is accurate overall and errors in misclassifications are symmetrical (see Tables 5.3 and 5.4 and Figure 5.6). In such a case, the predicted distribution of the group may only be minimally affected by errors and may still be valuable for instructors, even if some specific individuals have been misclassified. That is why in Figure 5.6 the human and computer codes look nearly identical despite there being disagreements on individual responses (see Appendix D for statistical analyses). For this reason, these resources should only be used for giving the instructor group-level information and should not be used for high stakes individual student information (i.e., points for providing a causal mechanistic response on a summative assessment).

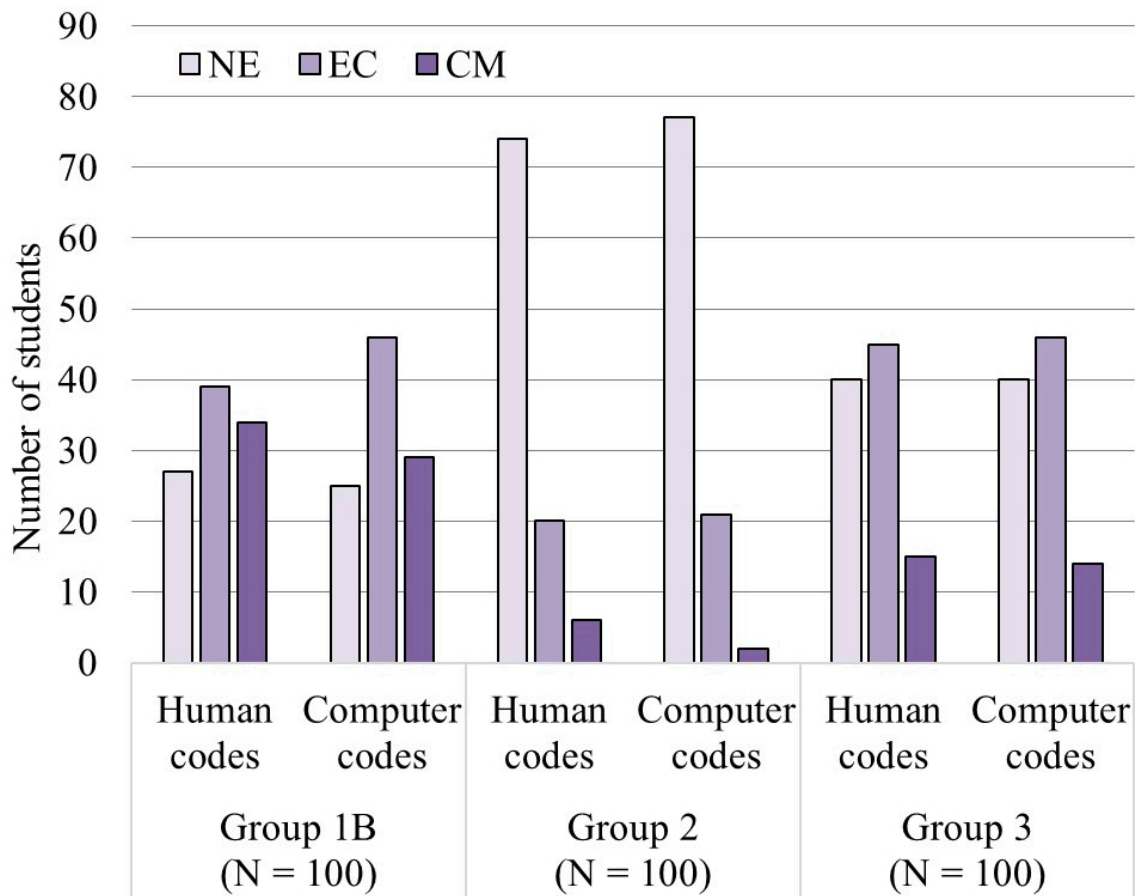


Figure 5.6. Distribution for LDF codes for the subsets of 100 responses from groups 1B, 2, and 3 as coded by humans (consensus score) and the computer (combined model). The human consensus score represents the agreed upon codes of authors R.L.M. and M.N. after discussion. The computer scores were coded by the combined model, which was developed using responses from groups 1A, 1B, 2, and 3.

Characterization of how a group of students responds to the prompt can give us important information about how that group is able to explain this phenomenon. Instructors can then use this information to better understand what their students can do and respond accordingly to better support their learning.<sup>21</sup> Such models should be able to detect different patterns of responses, for example, from different groups of students. Indeed, when we look at the group-level responses for the three test sets (Figure 5.6), we do see different distributions of responses for the different groups. For example, in comparison of groups 1B and 2, there is a markedly different distribution of responses: there are more EC and CM responses in group 1B compared to group 2, which is mainly NE responses. We cannot say

the exact cause for this difference as the prompt administration differed between the two groups. What we can say though is that, regardless of the cause, our model can detect that not many group 2 students are providing EC or CM responses. For the group 3 students, we see more CM and EC responses than are present for group 2. Again, we cannot say the cause of the difference, but our model is able to capture that there is a difference.

It is also worth reiterating that the combined model coding looks almost identical to the human coding. Not only is the model performing accurately, but also by coding responses from these other groups we have ensured that (1) their responses are part of the model we have developed and (2) our final combined model has been tested with responses from these different groups. That is, the resources we have developed are working well for a greater diversity of students in a variety of contexts compared to our initial model.

### **Limitations**

As technology continues to develop and grow more sophisticated, the ability to automate the analysis of student assessment will improve. It is important to remember, however, that these tools cannot replace the human role in analysis outright. While computers can identify patterns in responses, we need expert input to determine if these patterns are meaningful. The importance of collaboration in developing reliable and efficient computer resources that are meaningful to the chemistry education community cannot be understated. This is a growing field where both current advances in the education and machine learning domains are taken into account. Additionally, we must understand that this technology is not perfect. We agree with other automated analysis experts that we should be hesitant about using such resources for making high-stakes decisions, in particular at the level of the individual response.<sup>84,85</sup> Instead, we view the most appropriate use of these technologies for providing feedback to instructors to better support student learning.

We acknowledge that the prompt asked students to provide both a drawing and text response, meaning that students' text explanations are only one part of the story. It is possible that students provided additional thinking in their drawings that was not captured in their text responses. We hope that in the future there could be a corresponding analysis of student drawings. For now, we hope that instructors can use a survey of student drawings along with the results of the automated analysis of students' text responses to inform their teaching.

Unfortunately, because we know that changes in prompt structure change the response it elicits,<sup>29</sup> this computer model should be limited to use with this specific prompt. It remains to be seen if these developed scoring models could still work with different interacting neutral entities (like argon). However, machine learning resources that automate the analysis of prompts exploring other phenomena can certainly be developed using the CRC. On the basis of our experience, developing these resources requires a large number of student responses, prior studies on prompt development and associated coding, and a great deal of time spent on human coding. This may limit the creation of more machine learning models, particularly by busy faculty in charge of teaching these courses. These faculty can still use other models and questions developed by other researchers on the AACR website, but it may serve as a barrier for the addition of more models.

Coding and adding more models from different populations of students improved the model performance for responses from groups 2 and 3. It may be that coding more responses from other populations of students would increase the model performance further. At some point, however, it will be necessary to stop adding more responses, when the time and effort it takes to collect and code more responses does not justify marginal gains in model performance (see Figure 5.5). However, this does not eliminate the need to validate coding carried out by this model, in particular with new populations of students that have not yet been included in the model development.

One further limitation is that such analyses are not currently able to provide the kind of individualized formative feedback that a human reader might (if they had the time). As noted earlier, in our large enrollment classes, feedback from these analyses is provided to the group, rather than individually, and students are encouraged to reflect on and rework their responses. While this is not a substitute for the kind of Socratic dialogue that might be ideal, at the present time we do not have the capability for this kind of interaction.

### **Conclusion and implications for teaching and research**

By using machine learning, we have developed a tool that instructors can use to better understand how their students can explain the origins of LDFs, an intermolecular force that is important throughout chemistry and biology. With knowledge of what their students can do, instructors can then modify their teaching practices to better support their students' learning. Additionally, with the ability to process large amounts of data, departments could use this tool to understand the impact of the instruction across different courses or time.

The use of open-ended explanatory questions is one of the few instructional techniques that is supported by strong evidence<sup>79</sup> (that is, multiple studies across multiple populations), yet for some institutions, especially in lower-level or high-enrollment courses, this approach may be ruled out because of the huge commitment of time and personnel that are needed to grade or evaluate such tasks. Here we show how analysis of one such task can be automated, and by including a range of institutions, student demographics, and curricula we have developed a model that appears to be more robust than simply using student data from one cohort. Additionally, the coding scheme we automated is not just picking out predefined keywords, but also it is capturing the patterns in the text response that correspond to different types of student reasoning. That is, we are characterizing more than just the presence of a student idea, but also how they use that idea as well. This speaks to the power of the

CRC's ability to mimic sophisticated human coding based on the lexical patterns in the students' responses.

While this item and associated scoring model are now available for use ([beyondmultiplechoice.org](https://beyondmultiplechoice.org)), it should be noted that a great deal of time and resources were expended not only in the human coding of the training data that made the model so robust, but also in the design of the prompt that elicited student reasoning. Clearly it is not feasible for individual instructors to design their own assessments and expect such results without similar expenditures. However, if researchers collaborate and pool their data for various tasks, it is feasible to build up a library of items and associated models where instructors can input their own student data.

The increased availability of these kinds of models means that, even for large enrollment courses, assessments need not be limited to forced choice, calculations, or one-word answers that tend to emphasize fragmentary or rote knowledge. The very act of constructing deep explanatory responses is linked to the development of more robust knowledge frameworks, and the more often students are asked to engage in this kind of activity, the more useful their knowledge will become. Answering these kinds of questions, where reasoning is necessary, requires that students understand the material. As has been shown in numerous research studies, the use of vocabulary terms alone does not necessarily correspond with understanding; it is instead the constructed reasoning responses that tend to elicit evidence of understanding. More opportunities to engage in this kind of activity can only improve learning.





## **APPENDICES**


## APPENDIX A: Permissions


3/21/22, 6:31 PM


Rightslink® by Copyright Clearance Center




 Home

 Help ▾

 Email Support

 Keenan Noyes ▾



**Developing Computer Resources to Automate Analysis of Students' Explanations of London Dispersion Forces**

**Author:** Keenan Noyes, Robert L. McKay, Matthew Neumann, et al  
**Publication:** Journal of Chemical Education  
**Publisher:** American Chemical Society  
**Date:** Nov 1, 2020

*Copyright © 2020, American Chemical Society*

**PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE**

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)  
Comments? We would like to hear from you. E-mail us at [customer@copyright.com](mailto:customer@copyright.com)

Figure 5.7. Permissions to reproduce manuscript in its entirety.

## APPENDIX B: Additional demographic information about the participants

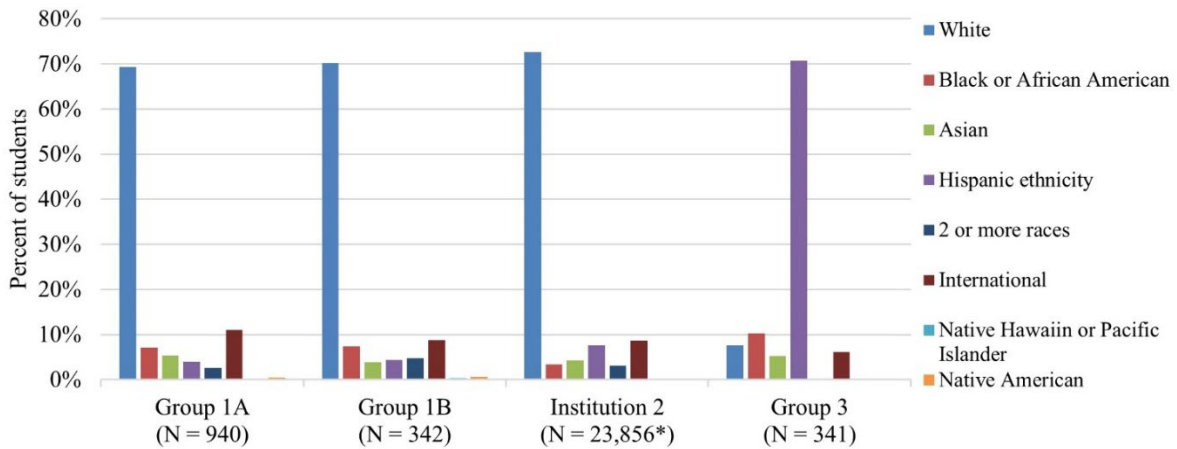


Figure 5.8. The racial and/or ethnic identities of the students from each of the four groups used in this study. We did not collect the racial and/or ethnic identities from the students of Institution 2, so instead we report the information for the entire undergraduate body for that academic year from the registrar's website. For Groups 1A, 1B, and 3 some students did not report their demographic information which is why the number of students described in this figure is slightly less than the number of students coded. Due to the racial/ethnic identity categories used by this institution's registrar, for Group 3 the category for Asian students also includes Pacific Islander students.

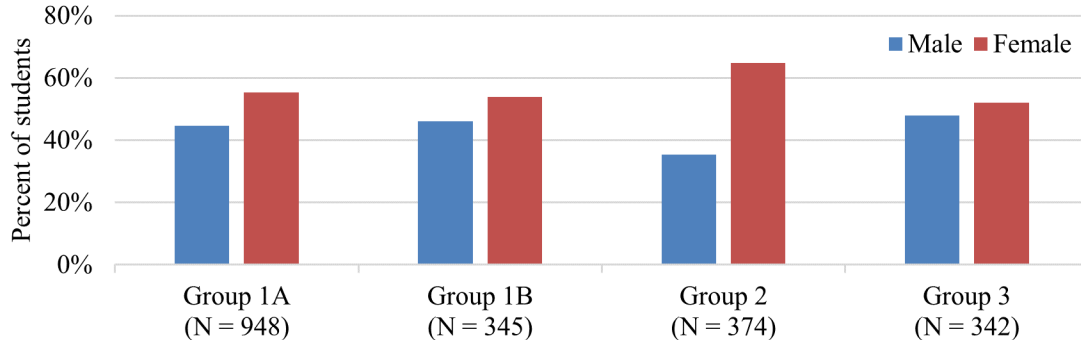


Figure 5.9. The percentage of male and female students in each of the four groups used in this study.

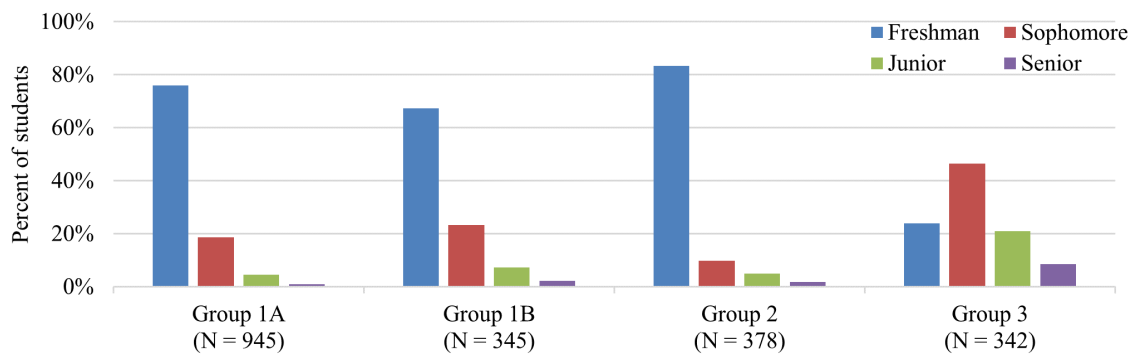


Figure 5.10. The percentage of students in each group broken down by their grade level.

We conducted an independent samples t-test<sup>70</sup> to compare the ACT Math scores of the students in Group 1A and 1B (N = 1,925 and N = 776 respectively). Some students did not have their ACT Math scores on file which is why the number of students in each group for this statistical test is lower than the total number of students in each group as reported in the manuscript (Group 1A N = 2,284 and Group 1B N = 915). Using this statistical test we found that the mean ACT Math score for Group 1A (mean = 25.83, standard deviation = 4.12) and Group 1B (mean = 24.47, standard deviation = 3.97) is significant,  $t(2699) = 7.82$ ,  $p < 0.001$ . To evaluate the effect size, we calculated Cohen's  $d = 0.33$ . The Cohen's  $d$  values of 0.2, 0.5, and 0.8 correspond respectively to small, medium, and large effect sizes.<sup>70</sup>

### APPENDIX C: Additional information about the nature of the responses

To provide some background on the constructed responses we analyzed, here we report some descriptive information for all 2,030 written responses we collected from the students in Groups 1A, 1B, 2, and 3. The following statistics reported are for responses that we had not spellchecked nor processed (e.g., word stemming) using the Constructed Response Classifier (CRC). We found the mean and median character length for these responses to be 235 and 204 characters respectively (Figure 5.4). Also, we found the mean and median word length to be 39 and 34 words respectively (Figure 5.12).

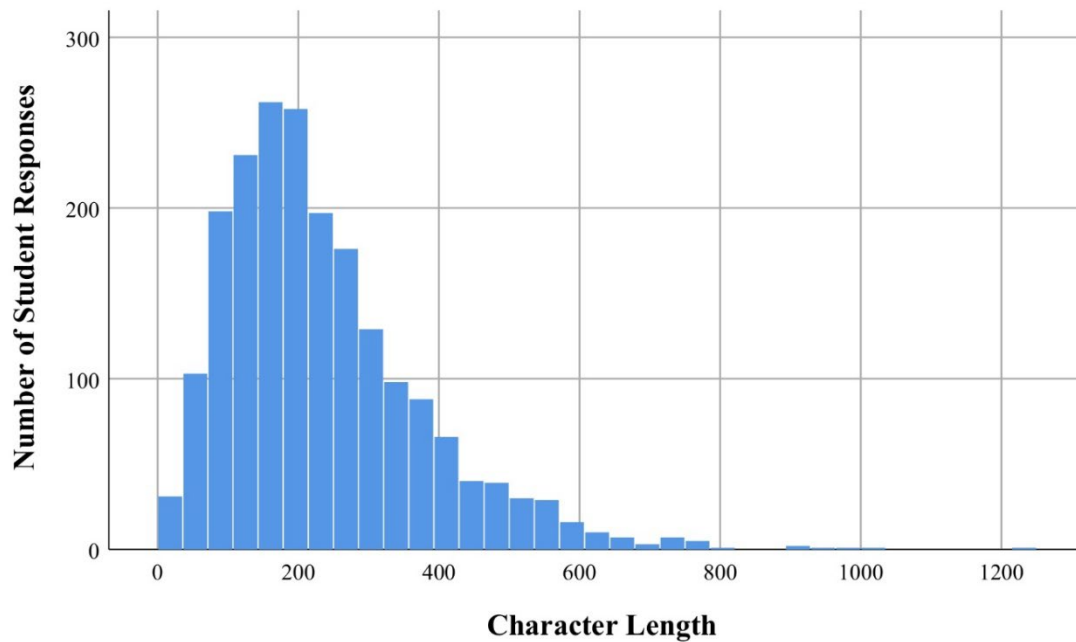


Figure 5.11. Histogram of the character lengths of all the student responses in Groups 1A, 1B, 2, and 3 (N = 2,030).

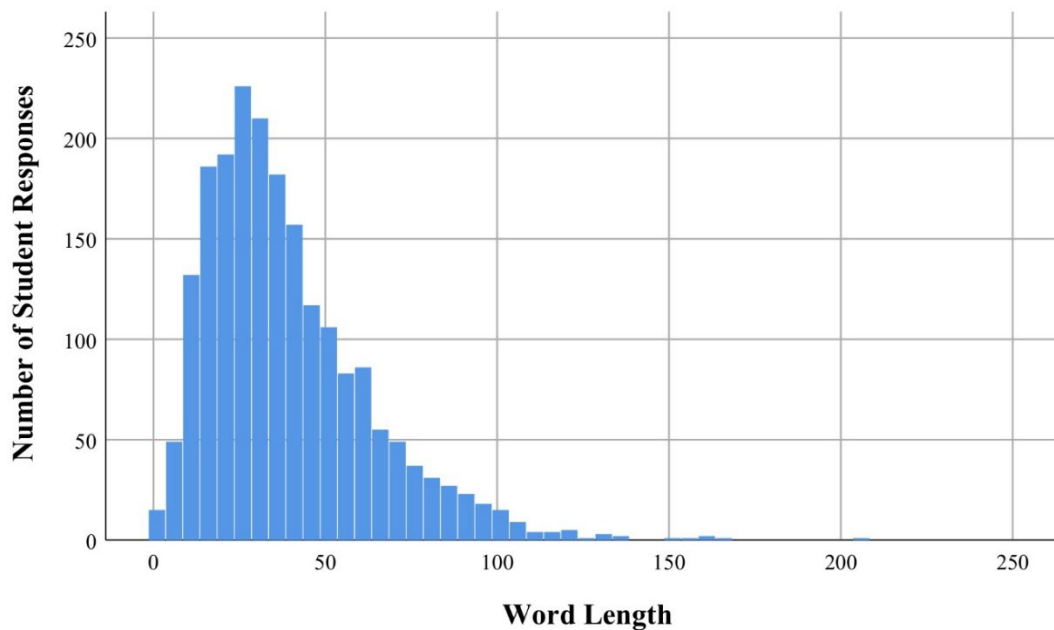


Figure 5.12. Histogram of the word lengths of all the student responses in Groups 1A, 1B, 2, and 3 (N = 2,030).

In Table 5.6 we present a randomly selected set of responses spanning across all of the groups (1A, 1B, 2, and 3) and all of the LDF codes (NE, EC, CM) to provide some additional examples of the student responses. These responses have not been spellchecked nor processed by the CRC. To generate this set, we randomly selected 10 responses from each group and human code. For example, there are 10 NE responses from Group 1A, 10 EC responses from Group 1A, 10 CM responses from Group 1A, 10 NE responses from Group 1B, etc. The LDF codes reported in Table 5.1 are the human LDF codes. As two human coders analyzed each response from Groups 1B, 2, and 3, the scores reported in Table 5.6 for those groups are the are human consensus scores.

Table 5.6. Example LDF responses.

Group	Code	Response
1A	NE	The atoms attract due to London Dispersion Forces. Attractive Coulombic interactions cause atoms to approach. Potential energy then decreases as they approach. They oscillate due to they electron clouds.
1A	NE	The atoms attract because of the london dispersion forces that are so strong between them.
1A	NE	They attract because the potential energy is decreasing. The electrons are also changing phases making them move closer to one another.
1A	NE	It occurs when two atoms can achieve a lower energy state together than they can separately, causing them to attract.
1A	NE	The distance between the two atoms decrease. The force of the two atoms become bigger and the speed become faster. The potential energy has change to kinetic energy.
1A	NE	Atoms attract due to London Dispersion forces. The atoms try to take electrons from each other so that they can both become stable.
1A	NE	The atoms attract because they are the same charge and it is the transfer of kinetic and potential energy
1A	NE	The atoms attract due to London Dispersion attractive forces, in which the electrons are temporarily attracted to one another. When the potential energy is at a minimum, the electromagnetic attraction is equal to the electromagnetic repulsion.
1A	NE	The atoms are attracting because the force of attraction is taking place which is making them come together. In the circle area on the graph the electron clouds are getting closer.
1A	NE	As the atoms move closer because the attractive force. the potential energy decrease. Then, when the potential energy be at the minimum the attractive fore will be the same the repulsive energy so it will decrease.
1A	EC	The atoms attract because of London Dispersion Forces: as an instantaneous dipole forms in one atom, an induced dipole forms in another atom, and the opposite charges are attracted to each other.
1A	EC	Because the electrons are getting closer and closer so they begin to repelling each other.
1A	EC	London Dispersion forces cause the two atoms to attract. The positive and negative dipoles attract each other. The electrons cause the outer cloud to change shape and attract another atom. Potential energy is decreasing at that point because they are going the fastest at that point.

Table 5.6 (cont'd)

1A	EC	The atoms begin to pull towards each other and attract due to London dispersion forces. The opposite charged atoms are attracted to one another causing them to move towards each other, causing the potential energy to decrease because the objects are moving closer together therefore requiring less work.
1A	EC	The atoms attract because they have opposite charges. One charge has to be negative and the other charge has to be positive.
1A	EC	The atoms are attracted to one another because of the London Dispersion Force. This force is allowing them to get closer and closer but it really determines where the electrons are in the electron cloud that surrounds the nucleus. Like forces repel which would have to mean both atoms have opposite charges when they are being pulled together by the London Dispersion Force.
1A	EC	The atoms attract because of London's law, the electrons have electron clouds that if they get too close they repel.
1A	EC	The atoms are initially attracted by London Dispersion Forces with the negative electrons being attracted to the positive side of the other atom.
1A	EC	As the atoms get closer they attract one another because of the London Dispersion Force which is the fundamental force that occurs between anything that is neutral. The process that is happening occurs because the charges within the atom are becoming partially positive on one side and partially negative on the other which is attracting a non polar atom and then creates a dipole.
1A	EC	The atoms attract because the electrons of each atom want to go towards the positively charged nucleus of the other atom.
1A	CM	The atoms attract due to London Dispersion Forces. This causes an unequal amount of electrons on one side of the electron cloud. This makes the other Helium atom to form an instantaneous dipole which is then attracted through electrostatic means.
1A	CM	The atoms attract to do opposite charges attracting, and the electrons move to another side of the atoms (London dispersion model). As they get closer, some of the charges are attracted to the other, and push the rest of the electrons to the other side, causing them to soon touch, which is decreasing the potential energy and increasing the kinetic energy as the distance decreases.
1A	CM	Due to London Dispersion Forces, in one of the atoms there would be an instantaneous uneven distribution of the electrons causing more of a negative charge on one side and a positive side on the other. The other atom would have an induced dipole where the negative charges would go closer to the positive side of the other atom. The opposite charges (positive from one atom and negative from the other atom) would result in an attractive force causing the atoms to come together.
1A	CM	London Dispersion Forces attract the two atoms together because the atoms are polarized when the electrons go to one side.



Table 5.6 (cont'd)

1A	CM	The atoms attract due to london dispersion forces. In other terms, when electrons are more on one side of the atom then the other, the atom is more negatively charged on that side of the atom and more positively charged on the other; therefore, when two atoms approach each other in this manner, they are attracted to each other because opposites attract causing london dispersion forces (induced dipoles).
1A	CM	The atoms attract each other because of the London Dispersion Force. This happens because the electrons become redistributed in proximity to another molecule (instantaneous dipole)
1A	CM	Due to the London Dispersion Forces, the electrons shift to one side of the atom and the nucleus shifts to the other creating a polar atom. The polar sides attract the opposite charge of another atom causing them to stick together.
1A	CM	The atoms attract because of London Dispersion Forces, during london dispersion, the electron dense parts of the atoms polarize a side and cause the atoms to attract each other and move together and stick.
1A	CM	Because of London dispersion forces, the atoms attract. At any given time, the electrons in an electron cloud could be unevenly balanced, causing a polarized atom known as a dipole. This dipole can cause a nearby atom to become an induced dipole and the more positively charged side of one atom will have a stronger attraction to the negatively charged side of the other atom.
1A	CM	At a random time, electrons in one atom are concentrated on one side. Because the electrons in a nearby molecule are repelling each other, so the distribution of electron density in the second molecule is the same as the first one. A dipole is formed and the atoms attract each other.
1B	NE	The london dispersion forces between the cloud of the electrons let them attract each other.
1B	NE	The atoms attract to each other because their is less space between them, thus the decrease of potential energy. The closer the atoms are, the less potential energy that is present.
1B	NE	There is london disperation force between atoms which attract each other together. potential energy decreases because attractive force closed together
1B	NE	The reason why the atoms attract one another is because they have less space to move and more energy is stored in less space.
1B	NE	The potential energy decreases as the atoms get closer together because the electron clouds of each atom interact which can form a chemical bond. When they get really close they reach equilibrium and potential energy increases as they repel each other.
1B	NE	As they approach each other the London Dispersion Forces grow stronger, the atoms are more strongly attracted to each other; the systems potential energy decreases and is converted into kinetic energy, the atoms move faster.

Table 5.6 (cont'd)

1B	NE	As the atoms move closer together, the potential energy increases because it is an attractive force while the atoms are moving closer together.
1B	NE	The atoms move towards each other because the two nucleuses are attracted. They oscillate because they are in an isolated system so they do not lose energy. The potential energy decreases because as the atoms move together the kinetic increases causing the potential to go down.
1B	NE	Because the gravitational forces between them make them to attract to each other
1B	NE	The atoms in the display attract to each other due to the fact that they are opposite signs. Since they are the opposite sign, the interaction between the proton and electron cause them to be pulled towards each other. The electrons in one atom collides/connect's to the electrons in the other atom.
1B	EC	The atoms attract because of the London dispersion force. This process occurs because the left side of the atom in the instantaneous dipole and the right side becomes more positive. Then induced dipole attracts to the more positive side of the instantaneous dipole.
1B	EC	The atoms attract because of the induced polarization in the electron clouds.
1B	EC	The atoms attract because it is an attractive force. The electron cloud on one atom attracts the nucleus on the other atom.
1B	EC	Atoms are attracted to each other by London dispersion forces. When the atoms get too close their electron clouds overlap, meaning two like charges are forced against each other, causing them to repel.
1B	EC	The potential energy is decreasing as well as the distance. Also, there are London dispersion forces that are acting on the atoms. The atoms have a partial negative and partial positive charge that attracts one another.
1B	EC	As the atoms come closer to each other, the London Dispersion forces cause the electrons of the each atom to repel each other. The atoms now have a more positive side and a more negative side. The negative side of one atom is drawn to the positive side of the other atom.
1B	EC	The atoms attract because of the instantaneous dipole where the atom has a charge. The unlike charges attract each other and cause an increase in kinetic energy which in turn would cause a decrease in potential energy. This is all done by dispersion forces.
1B	EC	The atoms are attracted to each other due to London Dispersion Forces. The partial positive charge of one of the atoms is attracted to the partial negative charge of the second atom. They become more attracted to each other as they become closer, up until a certain point.
1B	EC	Atoms are attracted to each other because of London Dispersion Forces. This is where the negatively charged side of one atom is attracted to the positively charged side of the other. Once this happens two atoms are in an induced dipole and are attracted to each other.

Table 5.6 (cont'd)

1B	EC	The two atoms are attracting each other due to London Dispersion forces. One side of the atom has a slightly more positive/negative charge which attracts the other atom. This attraction causes the kinetic energy of the system to increase which in turn, lowers the potential energy.
1B	CM	The atoms attract each other due to London Dispersion forces, specifically by electron density distortions within the atoms. This happens when all or most of the randomly moving electrons within the atom at happen to all be on one side of the atom. This causes the atom to turn into a dipole with a positive and negative "side" to the atom. The positive side then attracts the electrons of a nearby atom to make that atom a dipole. Now these two atoms are attracted to each other.
1B	CM	there is in instantaneous dipole on one of the atoms which causes a instantaneous diploe on the other atom which in turn makes them attract each other because for an instant one side of the atom is partially negative and the other side is partially positive due to all or most of the electrons collecting on one side of the nucleus.
1B	CM	The atoms are attracted to each other due to the differences in their charges which is caused by an instantaneous dipole. This is when the electron cloud fluctuates and there are more electrons on one side of the electron cloud than on the other-this causes one side to be more negatively charged than the other. When this happens in one atom, it happens in nearby atoms as well. This is called an induced dipole.
1B	CM	The atoms are attracted to eachother due to a dipole. One atom attains an instantaneous dipole which makes the electrons in the atom move to one side making it more positive on one side and more negative on the other. Once the other atom gets closer to the instantaneous atom it attains an induced dipole which also makes the elctrons of that atom align on one side making it more positive on one side and more negative on the other. The more positive end of the instantaneous atom attracts the more negative end of the induced atom and this causes the attraction.
1B	CM	A dipole of one atom induces the atom with out dipole. This process is called the London Dispersion Force which is the instantaneous, uneven distribution of the electron cloud. One side of the atom is partially negative which is attracted to the partially postive side of the other electron which makes them stick together
1B	CM	The atoms are attracted due to a london dispersion force. Since electrons are always moving, for one spilt second they are gathered along one sign of the atom. This atom makes one side of the atom partially negative, allowing the nuclues to make the other side partially positive, creating an instataneous dipole.. Since opposites attract, the partially positive end of one atom attracts the partially negative end of the other atom.
1B	CM	The electrons in the electron cloud are attracted to the positive protons in the nucleus in the other atom so they start to move toward each other. This is called the London Dispersion Force which happens when the electrons are all on one side of the atom creating an Instentanous Dipole and then the other atom is the same but it is called an Induced Dipole which is why they are attracted to each other.

Table 5.6 (cont'd)

1B	CM	London Dispersion forces act on the atoms causing induced dipoles which result in attractive forces. This is caused by uneven distribution of electrons.
1B	CM	The atoms attract due to london dispersion forces. Electrons in one atom randomly have an uneven distribution of electrons which creates an instantaneous dipole in that atom. This atom the creates an induced dipole on a neighbouring atom. This inturn creates a field with an attractive interaction between the two atoms. The potential energy in the system decreases as the atoms move closer together.
1B	CM	With all atoms the electrons may fluctuate. When there is distortion this usually results in one side of the atom will being partially negative making the other side partially positive. This atom is known as an instantanous dipole. The neighboring atom will be an induced dipole and the partially positive side will be attracted to the partially negative side on the other atom due to opposites attracting. This allows the atoms to stick together.
2	NE	The molecule is cooling down and the electrons slow down then form bonds together to form liquids and solids.
2	NE	The atoms do not have as much energy.
2	NE	The electrons are slowing down because because the temperature is getting colder. As the temperature gets cooler, the container starts to compress. The electrons are become attract to to fill the the volume that they have.
2	NE	They are attracting each other, because they are undergoing a phase change. Gaseous molecules tend to stay farther apart, but as the atoms undergo a phase change and turn into a liquid, the atoms come closer together, therefore attracting one another.
2	NE	they are attracted to each other because they only need 2 valence electrons to be complete, so they will attract one another in order to complete this. As the temperature lowers the He atoms move slower and the attraction between one another becomes stronger.
2	NE	The atoms attract because the of the energy between the atoms.
2	NE	The process that is happening is condensation because the atoms are condensing to form a liquid. The atoms attract because when the temperature decreases the force between them gets stronger and pulls them together.
2	NE	the electrons are bonding
2	NE	The atoms attract each other because the lowered temperature causes the electrons to attract
2	NE	The atoms and the electron become closer together because temperature is lowering. When the temperature lowers, the electrons cannot move as fast and making them a solid.

Table 5.6 (cont'd)

2	EC	The negative charge on the electrons becomes attracted to the positive charge of the nucleus of the other atom once they're close enough, but if they're too close the negativity of the electrons will repel
2	EC	The positive end is attracted to the negative end of the atom, causing attraction
2	EC	The electrons of He atom one are attracted to the protons of He atom 2 and vice versa
2	EC	the electrons are attracting to the positive care of the nucleus in the other He atom
2	EC	As the atoms become closer together, the attraction between the electrons and protons of the different atoms increases. Additionally, this attraction will continue until the electron-electron repulsion increases enough to a point where it equals the attraction.
2	EC	The electrons of each helium atom are attracted to the nucleus of the other atom. This makes them pull together.
2	EC	This is because of the dipole moments of the He atoms.
2	EC	As the two helium atoms approach each other, their electrons are attracted to the positive nucleus of the other helium atom. The condensation of the He atoms is caused by intermolecular forces.
2	EC	as they approach each other, ones nucleus attracts the others electron causing them to stick together at the right temp and pressure
2	EC	The electrons from one atom are being attracted to the protons in the nucleus of the other
2	CM	During the process, the electrons are moved around so that the two helium atoms are able to attract to each other and this is due to their intermolecular forces moving these electrons to a certain side.
2	CM	The two helium atoms are attracted to each other through an intermolecular force called London Dispersion. The electrons in the adjacent helium atoms occupy positions in which a temporary dipole is formed between the two atomes
2	CM	The two electrons of the helium atom will sometimes appear on the same side simultaneously. When this happens it creates a slightly negative charge on that side and a slightly positive charge on the other side. These faint polarities cause weak dispersion forces between other atoms that are exhibiting similar behavior. One atom's slightly negative portion will attract to another atom's slightly positive portion, and vice versa. This causes the atoms to approach and attract each other without actually bonding in any way.
2	CM	The electrons are moving around nucleus erratically and when the majority of those electrons happen to end up on the same side it creates a temporary dipole making one side slightly more negative than the other and attracted to a slightly positive dipole of another atom

Table 5.6 (cont'd)

2	CM	The atoms begin to attract one another as they approach each other because the random motion of the constantly moving electrons allows them to do so -- when both electrons on each Helium atom end up on the same side of their respective atoms, i.e. if the 2 electrons on each atom are both on the left side of their atoms at the exact same time, then a dipole moment can happen. The atoms attract one another because the placement of their electrons allows them to sit side-by-side since the negative charges are concentrated opposite of the "positive" space (without any electrons).
2	CM	As the temperature is lowered the He atoms lose energy which causes them to move slower. He atoms are attracted to each other through dispersion forces. A dispersion force is created when both of the electrons in an atom happen to be on the same side of the atom; this makes an instantaneous dipole moment that disappears as quickly as it appears. Usually this is a fairly weak attraction. However, as the He atoms move slower and slower this dispersion attraction becomes more and more relevant, pulling the atoms together until it eventually becomes a liquid.
2	CM	The electrons go to the bottom of each atom to form negative charge, so the top of each atom become positive, so the sample can form unit cells and become very stable.
2	CM	Instantaneous dipoles cause an attraction between the atoms. The electrons can be arranged asymmetrically to make the atoms polar and therefore attract one another.
2	CM	The electrons of one helium atom shift to one side allowing a positive charge to be exposed on the other side of the atom. The nucleus of the first helium atom attracts the electrons from the other helium atom, resulting in the attraction between two helium atoms.
2	CM	The electron cloud of a helium atom contains two electrons, which can normally be expected to be equally distributed spatially around the nucleus. However, if at any point the electron distribution is uneven, it results in an instantaneous dipole. This weak and temporary dipole influences other neighboring helium atoms through electrostatic attraction and repulsion and it induces a dipole on nearby helium atoms. The strength of dispersion forces increases as the number of electrons in the atoms or non polar molecules increases.
3	NE	The potential energy decreases because the closer the atoms get and attract each other, the faster it moves which means that the kinetic energy is increasing.
3	NE	they are attracted because of the charge of the atoms and the kinetic energy increases.
3	NE	the atoms are repelling towards each other because they are the same charge, the potential energy decreases and the kinetic energy increases. The distance between the masses are decreasing meaning that the potential energy decreases as well.
3	NE	they attract because of the electrostatic attraction to each other, when they atoms get close to each other the kinetic rises while the potential decreases.
3	NE	Because when the atoms get closer, the attractive force will appear and it will induce the charge of the electron of neighbor atom. And this will lead to decreases in potential energy.

Table 5.6 (cont'd)

3	NE	atoms are attracted to each other through london dispersion forces. As the distance decreases, the amount of energy required to keep them together increases.
3	NE	As the atoms get closer, both atoms attract one another because of the London dispersion forces of attraction. The greater the distance, the greater the attraction of the two atoms.
3	NE	The electrons are attracting each other at the circled point, which is why the potential energy is dropping and the kinetic energy would be increasing. The atoms are attracting because electrostatic energy is an attractive force.
3	NE	As the atoms attract and get closer together, their kinetic energy increases and their potential energy decreases.
3	NE	When two atoms come close to one another, Some of the electrons in the outer shells can interact with electrons in other atoms in quite a complex way. This can cause atoms to take on a temporary charge from a nearby atom. this means that the two atoms polarize each other: one becomes more positive and one becomes more negative. Then these two atoms may be strongly attracted to each other, and form what is called a 'chemical bond'.
3	EC	As one atom forms a dipole it induces a dipole in another and the atoms begin to attract one another and move toward one another with increasing kinetic energy. This interaction between the two atoms as they attract one another causes the potential energy to decrease.
3	EC	The atoms attract because of the charges in the atoms. The electrons and protons.
3	EC	As distance decreases, potential energy decreases and kinetic energy increases. Because of electrostatic forces the two atoms attract one another until their electron clouds begin to overlap and, consequently, repel one another.
3	EC	The two different charges attract each other pulling the two atoms closer but when they get close enough they don't have enough energy to combine and split away
3	EC	The process is called electromagnetic force, and what occurs is that the opposite charges of the atoms attract each other and the closer they get to each other, the faster they attract, causing the kinetic energy to increase and the potential energy to decrease.
3	EC	The Atoms attract because since opposite charges attract, the protons in an atom's nucleus attract neighboring atoms' electrons. When this begins to happen the two atoms will begin to get closer and the attraction increases which increases the speed in which they travel. As speed increases, potential energy begins to turn into kinetic. As the atoms come into contact their surrounding electron clouds which are the same charge will repel the atoms causing kinetic energy to turn back into potential as they get further from each other
3	EC	The atoms attract because they are opposite charges and they attract by the attraction force. The potential energy goes down because the kinetic energy is going up.

Table 5.6 (cont'd)

3	EC	The atoms are attracted to each other to the electrostatic force between them. Electrostatic force is caused due to the different charges the particles have.
3	EC	The atoms attract because you have one atom that has a dipole moment in which one side of the atom is more negatively charged than the other. That is to say one part of the atom has a partial negative charge and the other side is more partially positive. As this atom gets closer to another atom, the partially negative side of the atom repels the electrons in the other atom making it have a partially positive charge on one side that is attracted to the partially negative charge of the other atom.
3	EC	I was unable to get the aplet to work for me this time but, according to the graph above; The Electron cloud of one is inducing a charge in the other causing an attractive force.
3	CM	The reason the atoms attract to one another is because at any given moment, there is a non-zero probability of the electrons of an atom shifting to a particular area of the electron cloud. This causes a dipole, and the atom becomes gains a charge. When this atom gets near enough to another atom, it induces a dipole in that atom, and when both atoms have the dipole, the opposite poles attract each other.
3	CM	The two Helium atoms are coming together due to London Dispersion attraction. The London Dispersion attractions occurs to uneven distribution of electrons. The negative or the side where the electrons are most abundant in one helium atom is attracting to the positive or the side that the electrons are not most abundant in the other helium atom. These opposite sides attract. As they are coming closer together they are losing potential energy due to this force of attraction thats causing movement to one another converting the potential energy to kinetic energy.
3	CM	The electrons are all going to one side and this all happening because of the london dispersion forces. The atoms attract each other due to this.
3	CM	As the atoms attract each other, the electrons are being shifted in a way that the nucleus is vulnerable to another atom where the electrons have moved from the nucleus. This is called London Dispersion Force.
3	CM	The atoms are dipoles, which means that more electrons are on one side of the atom, which creates a charge instead of the atom being neutral. The different charges attract each other. As the distance between the two atoms decreases, the strength of the attraction increases which in turn causes the the atoms to speed up. An increase in speed shows an increase in kinetic energy and a decrease in potential enegy.
3	CM	Electrons develop an instantaneous unequal concentration throughout the atoms which is known as Londons Dispersions Forces. The atom then has a slightly positive and slightly negative concentration, or a dipole. The slightly negative side attracts the slightly positive sides of other dipoles and are therfor the opposite charges are attracted towards each other. According to the potential energy formula, potential energy decreases as the distance between two objects decreases as well like that of two helium dipoles.



Table 5.6 (cont'd)

3	CM	the electron of one atom moves towards the side facing the other atom causing the protons of the atom its facing to move to the side closest to the electron of the first atom. This causes them to attract due to opposite charges, this is called Londond Dispersion Forces.
3	CM	The atoms attract because atoms have electron clouds and at one instance one side of the atom could be partially charged while the other side is negatively charged. This instantaneous dipole will induce another atom because the negatively charged part of the instantaneous dipole will repel the electrons in the new atom. This causes the new atoms's electrons to move to one side. The postively charged side of the new atoms will thenn be attracted to the negatively charged part of the first atom. The atoms will be attracted to each other. London Dispersion Forces are at work here.
3	CM	The eletrons tend to hang around certain areas of the atom, this gives the atom polarity as the negative charges stick one side more than the other. This causes the atoms to attract one another.
3	CM	atoms are attracting each other due to the londond dispersion forces. the position of the atom allows them to form dipoles; their electrons locate at various locations that allow the opposite charges (electrons in the electron cloud, and protons in the nuclei of the other atom) to attract and bring them closer togehter; as the distance between the atom decreases their potential energy decreases, until it is transferred into kinetic energy. Until the electron clouds overlap, which makes the atoms repel because they both are negatively charged, and equals repel.

#### APPENDIX D: Additional results of groups 1B, 2, and 3 test set coding

Table 5.7. Crosstab relating the number of reference consensus human scores to the predicted computer scores using the initial model for the Group 1B test set.

Group 1B Initial model		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	22	5	0	27
	EC	2	35	2	39
	CM	1	7	26	34
	Sum	25	47	28	100

Table 5.8. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the combined model for the Group 1B test set.

Group 1B Combined model		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	21	5	1	27
	EC	3	35	1	39
	CM	1	6	27	34
	Sum	25	46	29	100

Table 5.9. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the initial model for the Group 2 test set.

Group 2 Initial model		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	67	6	1	74
	EC	4	16	0	20
	CM	0	3	3	6
	Sum	71	25	4	100

Table 5.10. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the combined model for the Group 2 test set.

Group 2 Combined model		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	71	3	0	74
	EC	4	16	0	20
	CM	2	2	2	6
	Sum	77	21	2	100

Table 5.11. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the initial model for the Group 3 test set.

Group 3 Initial model		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	36	3	1	40
	EC	5	39	1	45
	CM	0	4	11	15
	Sum	41	46	13	100

Table 5.12. Crosstab relating the number (and percentage of total) of reference consensus human scores to the predicted computer scores using the combined model for the Group 3 test set.

Group 3 Combined model		Human reference scores			
		NE	EC	CM	Sum
Computer predicted scores	NE	36	3	1	40
	EC	4	40	1	45
	CM	0	3	12	15
	Sum	40	46	14	100

To better understand any biases of the combined model function based on particular LDF codes, we analyzed the test set of responses from Groups 1B, 2, and 3 (N = 300). For those responses, the human and combined model assigned the same LDF code to 260 responses, disagreeing on the remaining 40 responses. Of the responses misclassified by the combined model: 14 (35.0%) were NE, 22 (55.0%) were EC, and 4 (10.0%) were CM. Of the 260 responses correctly classified by the combined model: 128 (49.2%) were NE, 91 (35.0%) were EC, and 41 (15.8%) were CM. Using a Pearson  $\chi^2$  test,<sup>65</sup> we found no significant difference in the distribution of LDF codes based on whether the combined model correctly classified the response ( $\chi^2 = 5.93$ ,  $df = 2$ ,  $p = 0.051$ ). This suggests that combined model misclassifications are not more likely to occur with a particular LDF code.

In Figure 5.6 of the main text, we report the distribution of human consensus and computer (combined model) LDF coding of the test sets of Groups 1B, 2, and 3. Here we report the results of a series of Pearson's  $\chi^2$  tests<sup>65</sup> comparing the distribution of LDF codes. We found no significant difference in the distribution of LDF human and computer codes for all three groups (Table 5.13).

Table 5.13. Results of Pearson's  $\chi^2$  tests comparing the human consensus and combined model codes.

Group	$\chi^2$ value	P value
1B N = 100	1.05	0.59
2 N = 100	2.08	0.35
3 N = 100	0.05	0.98

## CHAPTER VI - USING MACHINE LEARNING RESOURCES TO EXPLORE THE LONG-TERM IMPACT OF CLUE ON UNDERGRADUATE STUDENTS' EXPLANATIONS OF LONDON DISPERSION FORCES

### Introduction

#### *CLUE and intermolecular forces*

The idea that atoms and molecules interact with one another is a core idea in chemistry.<sup>37,38</sup> These interactions are important because they give rise to macroscopic phenomenon, like why ethanol has a lower boiling point than water. When these interactions occur between (rather than within) molecules they are called intermolecular forces (IMFs). Despite their importance in chemistry, prior research has shown that students struggle to understand the causes and consequences of IMFs.<sup>4,46</sup> However, the transformed general chemistry curriculum Chemistry, Life, the Universe, and Everything (CLUE) has been shown to help students develop a deeper understanding of IMFs.<sup>46,47,75,112</sup>

CLUE is a rethinking of the general chemistry curriculum, taking a core-idea centered approach to instruction.<sup>37</sup> By anchoring all concepts in the course back to at least one of four core ideas (electrostatic and bonding interactions, energy, atomic/molecular structure and properties, and change and stability), CLUE aims to lead students to develop a well-connected understanding of chemistry. One of those core ideas, electrostatic and bonding interactions, encompasses IMFs. By emphasizing the importance of these interactions and how they are connected to all these different ideas in chemistry, the hope is that students will develop a more productive understanding of IMFs and the phenomenon which these interactions govern.

Encouragingly, there is a growing body of research showing that CLUE can promote a deep understanding of IMFs. For example, Cooper et al. (2012) found that CLUE students, compared to those in a traditional curriculum, were more likely to indicate that they could use the structure of a molecule to predict its properties (by incorporating ideas about the IMFs it would experience).<sup>112</sup> Williams et al.<sup>46</sup> found that CLUE students are more likely than traditional students to identify intermolecular forces as

occurring between (rather than within) molecules. Stowe et al.<sup>47</sup> found that students in the high school adaptation of CLUE were better able to leverage the strength of hydrogen bonds (an IMF) to explain why ethanol has a higher boiling point than dimethyl ether. In addition to these studies which have looked at hydrogen bonds and IMFs more broadly, we have studied in great detail how students use causal mechanistic reasoning think about London dispersion forces (LDFs), a type of IMF which occurs between all atoms and molecules (even neutral species).<sup>7,75,113</sup>

### ***Causal mechanistic reasoning and LDFs***

Causal mechanistic (CM) reasoning is a type of thinking that is greatly valued in science. CM reasoning involves connecting ideas across scalar levels, leveraging the properties and behaviors of the underlying entities to explain how and why a phenomenon occurs.<sup>5,6</sup> It is this connection across scalar levels that makes this type of reasoning so powerful. By understanding how the underlying entities behave, the learner can better predict how the phenomenon might be affected if those entities are modified in any way, or how those entities give rise to similar phenomena. For example, if a baker understands how the flour, salt, water, and yeast all interact in order to produce bread, they can predict what might happen if their yeast has died (they will end up with a rather flat loaf).

In our work, we focused on how students used this type of reasoning to explain what happens when two neutral atoms approach. That is, how do students reason causal mechanistically about the formation of LDFs. Such an explanation would leverage the charged nature of the subatomic particles within the atom (i.e., a scalar level below). Specifically, how the charged subatomic particles within the atom (electrons) can randomly be unequally distributed, leading the atom to temporarily form charged poles which subsequently lead to an attraction between the two neutral species. We previously developed a coding scheme based off student interviews designed to characterize the degree to which students used CM reasoning to explain the formation of this interaction.<sup>7</sup> We then used this coding scheme to characterize how students explain this interaction over the two semesters of general

chemistry.<sup>75</sup> We found that about 40% of the students could provide a causal mechanistic text explanation of the phenomenon on the homework and exam that followed LDF instruction.<sup>75</sup> However, when you include those students who may not have provided a fully CM explanation but at least included electrostatic evidence for the attraction, we found that the vast majority (85%-95%) provided these types of responses.<sup>75</sup> Recall that in prior studies, students struggled to even identify where these interactions occur, so it is very encouraging to see students use a productive core idea to consider this interaction. While these results are exciting and further validate the efficacy of CLUE, this analysis only consisted of 250 students (out of the thousands of general chemistry students served at this institution) and data for only one year at this institution. To make broader claims about the efficacy of CLUE at scale, we need assess many more responses which can be an incredibly resource intensive process. Fortunately, technological advancements may provide a solution.

### ***Machine learning and assessments***

Analyzing student writing, especially for complex explanations and arguments, is not always possible for instructors due to the large numbers of students involved. In these cases, machine learning can provide instructors meaningful feedback on what their students can do. Machine learning provides a way that we can make sense of students' written work without the need for as much human involvement beyond the initial training of the computer models. Already, others have used such technology to automate the analysis of concepts like statistical randomness,<sup>87</sup> acid-base chemistry,<sup>85</sup> and biological matter and energy pathways<sup>89</sup> among others.

Recently, we developed a computer model in collaboration with the Automated Assessment of Constructed Response (AACR) group<sup>91</sup> that can automate the coding of these LDF explanations.<sup>113</sup> This computer model was built using responses from students across three institutions, including some institutions that did not use the CLUE curriculum. By diversifying the types of responses used to train the model, we can better capture the range of responses students might provide. In our previous study, we

found that the computer model could code students' responses with a high level of accuracy compared to human coding.<sup>113</sup> In this study, we explore how such a resource can handle large sets of responses to explore how CLUE students think about LDFs.

### ***Online learning and the global pandemic***

In addition to exploring the long-term impacts of CLUE, examining these large sets of students' responses also provided us the opportunity to see how students adapted to the shift to online learning in the wake of the COVID-19 pandemic. This emergency move to remote teaching has presented numerous challenges to instructors and students, for example adapting teaching techniques to the online environment and overcoming technological issues.<sup>114,115</sup> At the time of writing, we are over two years into this pandemic and what started as a temporary fix has become a sort of "new normal". Interestingly, many of the same strengths and challenges associated with online learning in times without a crisis have been also identified in the shift to online learning that has occurred during this pandemic. For example, Song et al.<sup>116</sup> identified that students found increased flexibility in terms of time and location to be an advantage of online learning, but they also noted that the lack of community and physical interaction created difficulties. Even though our technological capacities have dramatically increased since the early 2000s, researchers studying the impacts of the current pandemic have named these same factors as relevant today.<sup>114,115</sup> Despite these similarities, the scale and impact of the COVID-19 crisis is unprecedented and has impacted students in other ways as well. For example, Wang et al.<sup>117</sup> identified a growing mental health crisis among undergraduate students, specifically citing conditions brought on by the pandemic and the resulting shift to online learning environments. Such research is quite recent still, and we need to continue to understand how our students have been affected so we can better support them. Thanks to these machine learning resources, it is now possible for researchers to get a snapshot of how students may have been affected in a quick and efficient manner—useful characteristics of a tool in a rapidly evolving crisis.



### **Research questions**

1. How has the long-term adoption of CLUE impacted students use of causal mechanistic reasoning to explain the attraction between neutral species (i.e., LDFs)?
2. How do “on” and “off” sequence students compare in their explanations of LDFs?
3. How has the pandemic-induced shift to online learning impacted students’ explanations of LDFs?

### **Methods**

#### ***LDF task***

To explore students’ understanding of LDFs, we administered our previously developed task (hereon referred to as “LDF task”) in which students had to draw and explain the process by which two neutral helium atoms would attract one another as they approach.<sup>7</sup> For the purposes of this study, we only analyzed their text responses. We administered this question via BeSocratic,<sup>63</sup> an online system for collecting students written responses, as a homework activity and credit was given to the students based on completion and not correctness.

#### ***Participants***

We administered this prompt to undergraduate general chemistry students at a public midwestern research intensive institution. We started collecting responses in fall 2015 and continued every subsequent semester through spring 2022 (except for spring 2017—we did not collect student responses then due to a logistical error). We present the full student counts in Table 6.1. The students were enrolled in the first semester of the transformed CLUE general chemistry curriculum.<sup>37</sup> In total, we collected and analyzed 20,954 student responses over 8 years. All students included in this study gave consent for their responses to be used for research purposes, and their responses were deidentified before analysis in accordance with our IRB protocol.

Table 6.1. Number of student responses collected by time and course location.

Year	Semester	Course location	Responses analyzed
2015	Fall (“on” sequence)	In-person	2,284
2016	Fall (“on” sequence)	In-person	2,159
2017	Fall (“on” sequence)	In-person	2,217
2018	Fall (“on” sequence)	In-person	2,241
2019	Fall (“on” sequence)	In-person	2,135
2020	Fall (“on” sequence)	Online	1,777
2021	Fall (“on” sequence)	Online	2,074
2016	Spring (“off” sequence)	In-person	930
2018	Spring (“off” sequence)	In-person	817
2019	Spring (“off” sequence)	In-person	1,165
2020	Spring (“off” sequence)	In-person	1,213
2021	Spring (“off” sequence)	Online	1,005
2022	Spring (“off” sequence)	Online	937

As outlined in Table 6.1, there are three variables of interest over which the responses differ: the year, the semester, and the course location. We sampled responses over a range of years to explore the consistency of the course and its impact on students understanding of LDFs. There were no large efforts to change the course from year to year (prior to March 2020 as we will discuss later). We also examined responses from both the fall and spring semesters, corresponding to “on” and “off” sequence offerings of CLUE. Typically, students take this course in the fall semester (“in-sequence”) to align with required sequence of courses many majors require for which the first semester of general chemistry is a

prerequisite (e.g., molecular biology). This is why more sections of the course are offered in the fall compared to the spring.

There are many reasons students may take the course in the spring (“off-sequence”) instead of during the fall. Students may have deferred enrollment or have scheduling conflicts. Another potential reason students may opt to take this course in the spring is because they are required to complete a required math course in the fall. At this institution, students complete a math placement test upon admission, and students who score below a given threshold on this assessment are required to take an additional math course prior to, or concurrent with, their first semester of general chemistry. Given these potential differences, we compared the standardized test scores of these two groups (prior to the shift to online learning). We used the SAT (1600-point) for this comparison. For those students who instead had a SAT (2400-point) or ACT score on file, we used a concordance table provided by the College Board to convert those scores to a SAT (1600-point) score. Using an independent samples t-test, we found there is a significant difference between the SAT (1600-point) scores of the two groups,  $t(12213) = 21.66$ ,  $p < 0.001$ . Specifically, we found that the “on” sequence students had a higher SAT score ( $M = 1226$ ,  $SD = 128.7$ ) compared to the “off” sequence students ( $M = 1172$ ,  $SD = 129.6$ ), and this difference in SAT scores was of medium effect size (Cohen’s  $d = 0.422$ ).<sup>118</sup> Given these differences, we explored how these two groups of students might differ in terms of their explanations of LDFs.

Finally, in March of 2020, the emergence of COVID-19 in the United States forced this university, like many others, to transition from a primarily in-person class to one that is exclusively online. We note that even when this course occurred in-person, the online platform BeSocratic was still used to give homework. Additionally, while it was during the spring 2020 semester that the course transitioned to an online format, the course was still occurring in-person when the LDF task was administered; for this reason, that timepoint is labeled as “in-person” in Table 6.1. Of course, the “in-person” and “online” designations represent just one aspect in which this course changed after March of 2020, and these

students' lives may have changed in other ways due to the emergence of a global pandemic. We acknowledge that these other changes may have also impacted the students' learning, beyond the change in physical location.

### ***Coding scheme***

We developed a scheme<sup>75</sup> to characterize the degree to which explanations of the origin of the LDFs are causal mechanistic, a type of reasoning that connects phenomenon to the behavior of the entities a scalar level below.<sup>5</sup> Our coding scheme categorizes students into three bins: non-electrostatic (NE) responses, electrostatic causal (EC) responses, and causal mechanistic (CM) responses. Ideally students would provide CM responses describing the process in which the subatomic particles localize to give rise to the dipole. However, students might leave out the behavior of the subatomic particles and only describe the electrostatic nature of the interaction (EC response) or leave out the role of charges entirely (NE response). Further description of these categories can be found in our previous publication.<sup>75</sup>

### ***Automated analysis tools***

To analyze the responses in this data set, we worked in collaboration with the AACR group who specializes in using machine learning techniques to conduct lexical analysis.<sup>91</sup> They developed a tool, referred to as the Constructed Response Classifier (CRC) tool, which generates computer models capable of analyzing written explanations like a human could. Human coded responses are inputted into the CRC tool which then uses a series of machine learning algorithms to identify lexical patterns in the responses associated with the human codes, generating a predictive computer model. Previously, we have used the CRC tool to develop computer models capable of applying our LDF coding scheme with a high level of accuracy (i.e., good agreement with the human codes) to code students' responses.<sup>113</sup> The model used in this current study builds upon the model used in our previous study.<sup>113</sup> In that prior study, we previously withheld some human coded responses (N = 300) from the model training set (N = 1,730)

in order to test the model’s accuracy.<sup>113</sup> In this study, we added these responses into the model training set because they were already coded by humans; generally, more responses in a training set leads to better performing computer models. The final model, trained on 2,030 responses, was used to analyze the responses in this study. For more information about the CRC tool and the responses used to train this model, please see our prior publication.<sup>113</sup>

Table 6.2. Confusion matrix for the computer model used to code the responses in this study.

Final training model		Human reference code		
		NE	EC	CM
Computer predicted code	NE	643	90	10
	EC	74	714	72
	CM	2	39	386

This final model has good agreement with the human coding used to train it. Using a 10-fold cross-validation procedure, we assessed the model’s performance. The resulting confusion matrix of this cross-validation procedure is shown in Table 6.2. Cohen’s kappa (a measure of agreement which takes into account the possibility of agreement by chance) was 0.78 indicating this final model had ‘substantial’ agreement with the human scores.<sup>72</sup> The overall accuracy (fraction of agreed upon coding instances between the human and computer codes) was high (0.86). Sensitivity and specificity, metrics which dig deeper into the accuracy of each of the individual codes, were also high (sensitivity range: 0.825-0.894, specificity: 0.877-0.974). Based on this information, we feel confident in the accuracy of the coding carried out by this computer model. We used this model to code 19,654 students’ responses. The remaining 1,300 (950 from fall 2015 and 350 from spring 2016) were human coded as part of the model development process.<sup>113</sup>

### **Statistical tests**

While using automated analysis tools has afforded us many opportunities, we must be cognizant of the potential issues in using a large data set, specifically the impact of the sample size on our statistical tests. Other researchers have noted that when conducting statistical analyses on large sets of data that statistical significance ( $p < 0.05$ ) is all but guaranteed.<sup>118–120</sup> In these instances, it is particularly important that we examine the effect size (which does not depend on the sample size) to assess the magnitude of the difference. That is, even though a difference might be *statistically* significant, it may not be *practically* significant. Given that we analyze over 20,000 responses for this paper, it is likely that some of the tests may have very small p-values, but also have very small effect sizes.

### **Results and discussion**

#### ***Study 1: How has the long-term adoption of CLUE impacted how students use causal mechanistic reasoning to explain the attraction between neutral species***

To better understand the long-term impacts of the adoption of CLUE on student understanding of LDFs, we examined students' explanations of the formation of this IMF from five fall semesters (2015–2019) of this course. However, given that in a typical fall over 2,000 students will take general chemistry, this required the analysis of a very large data set ( $N = 11,036$ ). While we previously human coded 950 responses from fall 2015 as part of our model development process, we used the CRC tool to analyze the remaining 10,086 responses in just a few minutes. We show the results of our analysis in Figure 6.1.

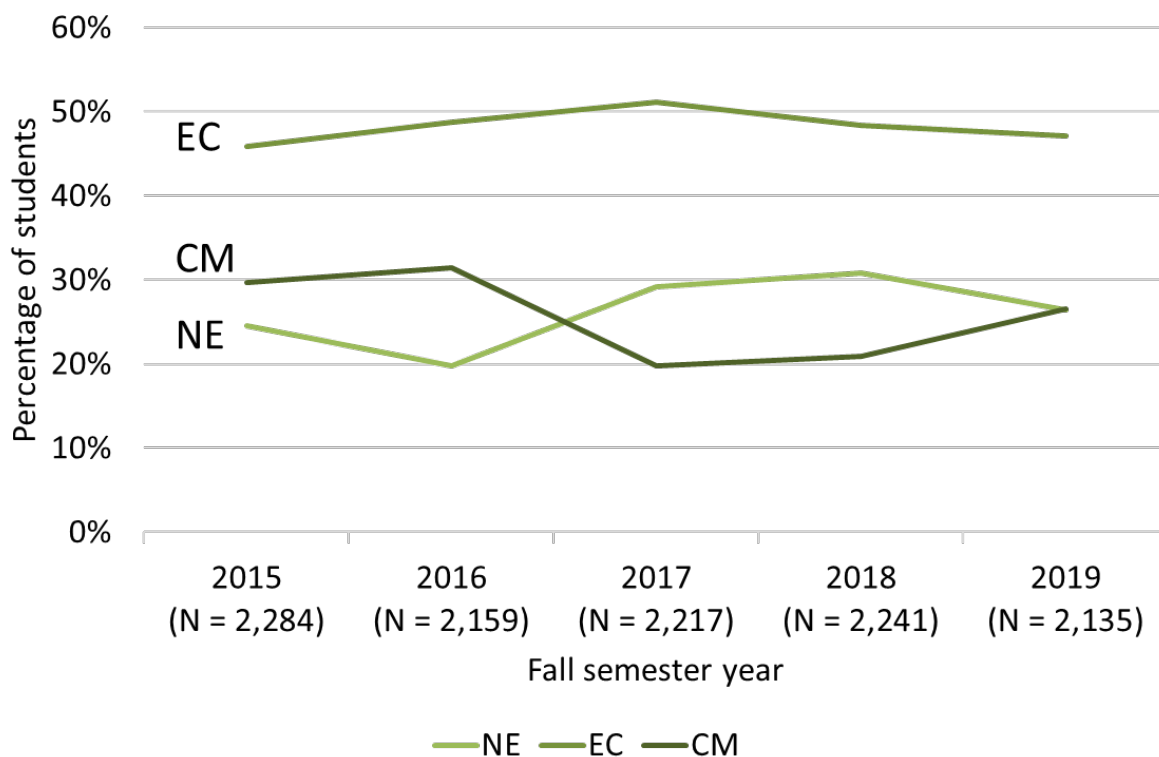


Figure 6.1. The distribution of non-electrostatic (NE), electrostatic causal (EC), and causal mechanistic (CM) explanations to the LDF task from students enrolled in the fall semester of general chemistry (2015-2019).

While there are some variations from year to year, there is a consistent pattern in the responses to the LDF task. About 20-30% of students provide a NE response and another 20-30% provide a CM explanation. Meanwhile, the most popular category remains the EC response with about half of the students providing such an explanation every year. These results are encouraging. They show that this course is consistently helping students to provide productive explanations of this phenomenon with most of the responses including some electrostatic evidence for this interaction while about a quarter of the students go even further to include the mechanism of LDFs. Considering the breadth of literature that has identified intermolecular forces as a difficult idea for students to understand,<sup>3,4,39</sup> this is additional evidence that CLUE can help students to develop a deep understanding of intermolecular forces.

### Study 2: How do “on” and “off” sequence students compare in their explanations of LDFs

To better understand the impact of taking general chemistry “off-sequence” in the spring instead of “on-sequence” in the fall, we used the CRC tool to analyze responses from several spring semesters to compare them with the corresponding prior fall semester. Even without responses from spring of 2017 (which we did not collect due to a logistical error) this required the analysis of another 4,000 responses. We had already coded 350 of the spring 2016 responses as part of the computer model development, but we coded the remaining 3,775 with the CRC tool. The results of this analysis, broken down by academic year and semester, are presented in figure 6.2.

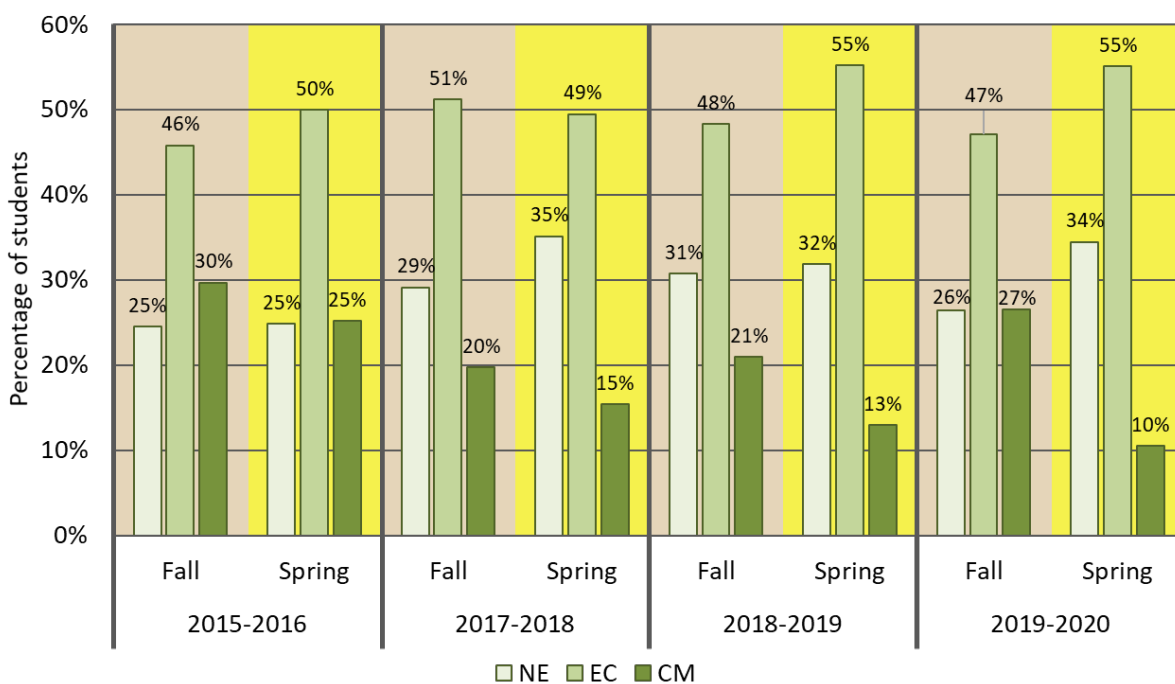


Figure 6.2. The fall and spring response for the in-person academic years 2015-2016, 2017-2018, 2018-2019, and 2019-2020. The fall semesters are color coded with a light brown background while the spring semesters have a yellow background.

To compare the distribution of responses between the fall and spring semesters for each academic year, we conducted a series of pairwise Pearson’s  $\chi^2$  tests of independence (Table 6.3). To account for increased chance of type I error due to repeated tests, we used the Bonferroni corrected alpha value, adjusting the threshold of significance to 0.0125. P-values less than this value are



considered significant. These results show that for three out of the four academic years, there is a significant difference in the distribution of responses based on semester. Across those three years in which significant differences were observed (2017-2018, 2018-2019, 2019-2020), those differences are of small to small-medium effect size based on their Cramer's V value (small – 0.1, medium – 0.3, large – 0.5).<sup>66</sup>

Table 6.3. Results of pairwise Pearson's  $\chi^2$  tests comparing fall and spring semesters each academic year.

Academic year	$\chi^2$ Value	P-value	Cramer's V (if applicable)
2015-2016	7.277	0.026	N/A
2017-2018	13.225	0.001	0.066
2018-2019	34.156	<0.001	0.100
2019-2020	123.116	<0.001	0.192

Across those three years, the most pronounced difference between the two semesters is the consistent drop in CM responses from the fall to spring semester of 5-17%. While this decrease in CM responses is certainly concerning, we should put this difference in context. First, there was not a significant difference in all the semesters, and even in those semesters where there was a significant difference, that difference is small to small-medium effect size. Additionally, even in the “worst” instances, only a third of the students provided NE responses. That is, two-thirds of the students are still providing some electrostatic evidence for this interaction. However, given the importance of CM reasoning, it is of concern that we see just 10% of the students providing a CM response in spring of 2020 (note that this is not because of the pandemic, since as discussed earlier the activity was completed before the courses went online). While the percentage of CM responses in the fall semesters varied between 20-30% over these years, the percentage of CM responses never reached 20% for spring semesters after that first academic year we analyzed. Additional research may be needed to better

understand how meaningful these differences are and how we might best support these “off-sequence” students.

***Study 3: How has the emergence of the pandemic impacted the students’ responses***

In March of 2020, general chemistry instruction shifted to an online format in the interest of safety as the world entered a global pandemic. This shift represented a major change in how this course was normally taught. Now, thousands of students would only be able to interface with this class virtually, requiring drastic modifications to be made to the learning environment. With all these novel changes taking place, it was not clear how students’ learning would be affected. Fortunately, one of the few things in the course that remained relatively unaffected was the homework, which was being administered online even while the class was in-person. Two years into this pandemic, we have responses from 5,793 students collected across four semesters of this course, providing a snapshot of how this shift to online learning during this difficult time has impacted students.

Using the CRC tool, we analyzed all 5,793 responses. These responses came from two fall semesters (2020 and 2021) and two spring semesters (2021 and 2022). Due to the differences between fall and spring semesters, we compared the students’ responses before and after the start of online learning compared against students from those same semesters (i.e., comparing fall with fall responses and spring with spring ones). We present the distribution of responses in Figure 6.3.

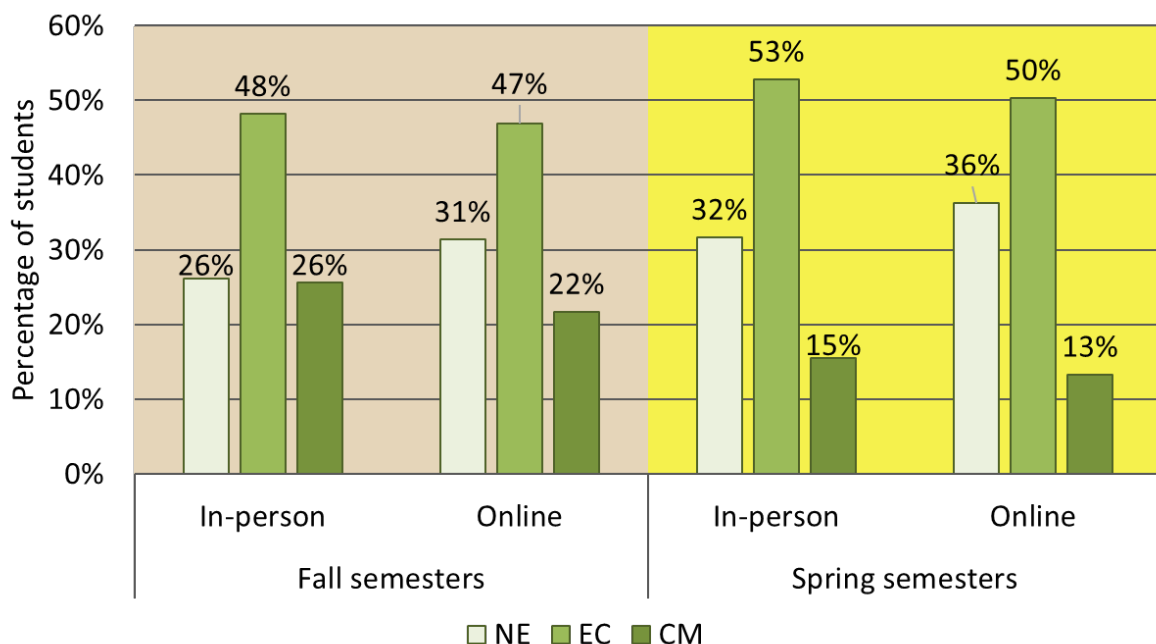


Figure 6.3. The distribution of LDF responses for the in-person and online learning environments (before and after March 2020).

We conducted a pair of  $\chi^2$  tests to see if there was a difference between the learning environments. We found that there was a statistically significant difference based on the learning environment for both the fall ( $\chi^2 = 47.244$ ,  $p < 0.001$ , Cramer's  $V = 0.056$ ) and spring semesters ( $\chi^2 = 14.112$ ,  $p = 0.001$ , Cramer's  $V = 0.048$ ). Given the large sample size, this statistically significant result does not provide us much information about how large this difference is. When we consider the effect size, we see that both Cramer's  $V$  values are below the "small" benchmark (0.1).<sup>66</sup> Even the largest change between any of the time points was just 5% (NE responses for fall semesters). Therefore, we argue that there is no practical difference in the distribution of LDF explanations. Considering just how life upending this pandemic has been, the fact that the distribution of responses to this prompt has remained relatively the same is really encouraging. This highlights how CLUE can still be effective in helping students develop a deep understanding of intermolecular forces even when occurring online during a global pandemic.

### **Limitations**

While this tool provides us important feedback about how students explain LDFs, this is just one question, and one idea out of many in a general chemistry course. To explore how students understand other ideas and other intermolecular forces, more assessments (and machine learning resources) are needed.

### **Conclusions**

If it took just one minute for a human to analyze each response, analyzing this data set would require around-the-clock work for about two weeks. Meanwhile, thanks to the CRC tool, we analyzed this data in about the time it takes to make a cup of coffee. This highlights the major advantage to using an automated resource: faster analysis time and the ability to process large amounts of written responses. Additionally, the responses did not need to be spellchecked or processed before the analysis, which makes this resource easy to use. It is important, however, that the students' responses are collected electronically. Without access to BeSocratic or another electronic platform, we would have had to transcribe the responses first which would make it much more difficult to process large numbers of responses.

From this analysis, we learned more about the long-term impact of CLUE. It appears that students in this course are consistently able to leverage electrostatic evidence to explain LDFs. Furthermore, in the "on-sequence" course between a quarter and a third of students engaged in causal mechanistic reasoning to explain the formation of this intermolecular force. This is important because in addition to the utility of causal mechanistic reasoning, other researchers have identified IMFs as a difficult topic for students to understand. This means that in our data set alone we had over 4,500 students engage in a sophisticated form of reasoning to provide a fully causal mechanistic explanation of this phenomenon.

We also used the CRC tool to explore how the students taking the course at different times (“in-sequence” or “off-sequence”) differed in terms of their LDF explanations. We found a statistically significant difference between those student responses of small effect size as a greater portion of the “off-sequence” students provided NE and EC responses instead of CM responses. While the effect size of this difference was small, we observed this same trend across multiple academic years. This may be an area to study further and see if these observed differences are part of any larger differences between the semesters in how these students engage in CM reasoning or understand other IMFs.

Finally, we used the CRC tool to explore how a change in learning environment (and the emerging pandemic) impacted students’ LDF explanations. Again, we found a statistically significant difference in the distribution of LDF responses, but the Cramer’s V values were very low and below the “small” effect size benchmark. This suggests that while this difference may be statistically significant, it is not practically significant, especially considering very large size of our samples (which makes a statistically significant result more likely). This suggests that even after the emergency shift to online learning, CLUE was still effective in helping students develop a rich understanding of LDFs.

Overall, this study provides important evidence about the utility of CLUE and the utility of machine learning technologies like the CRC tool. In this analysis, we examined nearly 20,000 responses, a feat that would not have been possible without the use of AACR’s CRC tool. Such resources can allow for instructors to move beyond multiple choice questions, which can over-estimate students’ knowledge,<sup>77,78</sup> and use constructed response questions to get feedback on their students’ thinking. Understanding what our students know plays an important role in determining how best to support students learning. Going forward, as more computer models are developed for use with the CRC tool and other machine learning resources, we can expand our capabilities as we become able to analyze large sets of responses like those discussed in this study. The question prompt and machine learning resources are available at [beyondmultiplechoice.org](https://beyondmultiplechoice.org).<sup>90</sup>

## CHAPTER VII - EXPLORING CONNECTIONS BETWEEN STUDENTS' EXPLANATIONS OF LONDON DISPERSION FORCES AND POTENTIAL ENERGY

### Introduction

In chemistry, energy and forces are not just two important core ideas, they are inseparably linked. We often say that changes in the forces and interactions in a given system give rise to changes in the potential energy (PE) of a system. Alternatively, we also say that kinetic energy transferred through collisions, or energy carried in photons, results in forces being overcome. This close connection ideally provides an important learning opportunity—we can leverage students' understanding of one to support their understanding of the other.

Unfortunately, there is plenty of evidence that students often struggle to understand these ideas.<sup>2-4,39,58</sup> However, we have found that students enrolled in the transformed general chemistry curriculum Chemistry, Life, the Universe, and Everything (CLUE) have a richer understanding of intermolecular forces, both from a mechanistic standpoint (how they arise)<sup>7,75</sup> and how IMFs help predict chemical and physical properties.<sup>47,112,121</sup> This potentially affords us the opportunity to study how students understand the connections (and misconceptions) between energy and forces.

### *Prior research on understanding of potential energy*

It is difficult to overstate the importance of energy in science. According to *A Framework for K-12 Science Education* from the National Research Council,<sup>1</sup> energy is a core idea in the physical and life sciences. In addition to its importance within the individual disciplines, energy is also a crosscutting concept. That is, energy is a productive lens that experts use to identify key aspects that allow them to better understand phenomena.<sup>122</sup> That being said, understanding energy is demonstrably difficult.<sup>2</sup> This is partially due to its abstract nature; energy is not a tangible entity, but rather a theoretical construct which helps us to predict how a system behaves, or an accounting scheme that allows us to track energy across systems.<sup>57</sup> Time and time again, researchers have found that the idea of energy creates issues for

students and instructors alike.<sup>2,73,123,124</sup> It is clear that traditional approaches have not been successful thus far—what then should we expect our students to know and be able to do with energy?

In the *Framework*, the authors propose a simplified approach to energy in which there are two forms of energy: kinetic and potential.<sup>1</sup> At the atomic and molecular level, kinetic energy describes the movement of atoms while PE describes the forces experienced by entities within an electric field. In this paper, we focus on how students understand PE, a concept which has proven to be difficult for students.<sup>125–128</sup> In the context of undergraduate chemistry, little research has been done exploring how students think about PE. One study by Nagel and Lindsey<sup>125</sup> found that chemistry students struggled to understand the relationship between PE and distance. Another study by Becker and Cooper<sup>126</sup> found that students saw PE as the ability (potential) to undergo some change or as something that could be stored but were unable to connect these ideas to the electrostatic forces which underpin PE.

#### ***Prior research on understanding of intermolecular forces***

Like energy, forces play an important role in science, and this is especially true in chemistry. Intermolecular forces, which occur between atoms and molecules, are frequently discussed in general chemistry and explain a wide range of macroscopic phenomenon, like why geckos can walk up walls, ice cream melts on a hot summer day, and how DNA can replicate and maintain the genetic code. These forces are electrostatic in nature, meaning that they are caused by the attractions and repulsions of unlike and like charged entities.

Intermolecular forces, like energy, have traditionally been a difficult idea for students to understand. For example, Henderleiter et al.<sup>3</sup> found that organic chemistry students often rely on rote memorization to identify where hydrogen bonding (another intermolecular force) would occur. This strategy led many students awry when they tried to predict the relative boiling points, misremembering

which functional groups would experience stronger forces.<sup>3</sup> In our earlier work, we have shown that many students struggle to identify that these intermolecular forces occur between molecules.<sup>39</sup>

The transformed CLUE curriculum has been shown to improve students' understanding of intermolecular forces.<sup>46,47,121</sup> CLUE is centered around four core ideas that form the basis of most chemical phenomena, two of which are energy and (electrostatic) forces. Earlier studies have shown that CLUE students are able to both identify where intermolecular forces occur, link these forces to macroscopic phenomenon, and produce rich causal mechanistic explanations of these forces.<sup>46,47,75,112,121</sup>

Because of the close relationship between forces and energy, it may be that CLUE students are able to connect their understanding of forces to energy. We have some evidence of CLUE students making these connections already. Stowe et al.<sup>47</sup> found that a group of students in a high-school adaptation of CLUE were asked to compare the boiling points of dimethyl ether and ethanol, over a third of the students explained the connection between the strength of the hydrogen bonding forces and the increased energy needed to overcome those forces. In a similar prompt, Kararo et al.<sup>121</sup> surveyed undergraduate CLUE students and found that over half of these students leveraged energy and interactions to explain the relative boiling points. While these studies provide valuable and exciting information about the impacts of CLUE, they did not explicitly explore why the hydrogen bonding interactions were stronger (i.e., the leveraging the role of electrostatics). In this study, we try to further explore this relationship between forces and energy using the lens of causal mechanistic reasoning.

### ***Causal mechanistic reasoning***

Causal mechanistic reasoning is a powerful way of thinking that is highly valued in science. Approaching a phenomenon through this lens requires unpacking the properties of the underlying entities to explain how their spatial and temporal behavior causes the phenomenon (at the scalar level above) to occur.<sup>5,6</sup> This type of reasoning is powerful because understanding what occurs a scalar level



below enables the learner to make better predictions about what may happen if those entities are changed, or how those same underlying entities give rise to related phenomenon.<sup>5,6</sup> Developing this type of understanding of phenomenon is a defining characteristic of expertise and not just in science. Consider an expert car mechanic. Through years of experience and training, a mechanic understands how the machinery under the hood of a car allows it to function. If the car begins to behave abnormally, they can identify which of the underlying components may be creating this issue and fix the problem.

We have extensively studied how students use causal mechanistic reasoning to think about the formation of London dispersion forces.<sup>7,75,113</sup> That is, how the temporary distortions of the electron distribution in atom can lead charged ends to form on an otherwise neutral species, and that the oppositely charged ends can then attract one another. Such an understanding connects the properties of the entities a scalar level below (the charged nature of the electrons and protons) to the interactions experienced by the atoms (the formation of LDFs). In our prior research, we have found that vast majority of CLUE students can identify the electrostatic origin of LDFs, and that this understanding persists through the two-semester general chemistry sequence.<sup>75</sup> Furthermore, we have found that many of these students have a causal mechanistic understanding of LDFs, linking the role of the subatomic particles to the formation of these charged ends.<sup>75</sup> We believe that developing such an understanding should give students the ability to make sense of related phenomenon, like changes in PE, that call upon these same ideas (e.g., the charged nature of the subatomic particles).

### ***Resources perspective***

To make sense of what students know, it is important that we ground our work in a theory of learning. Constructivism posits that students are not blank slates, and that new knowledge is built upon existing knowledge in the mind of the learner.<sup>8</sup> The resources perspective elaborates on these ideas, providing a mechanism for how that knowledge is built.<sup>12,14</sup> According to this theory, the mind is filled with small fragments of knowledge, called resources, which are contextually activated and used to

reason through different situations.<sup>12,14</sup> These resources may or may not be linked to other resources, which can then be activated as an entire cluster or coordination class.<sup>10,12,14</sup> Through this lens, learning is the addition of new resources and connections between resources. The more these resources are called upon in different contexts, the stronger the connections become, and the more readily they can be used by the learner. Therefore, we strive to create as many opportunities as possible for students to activate and connect their resources in productive ways.

The resources perspective aligns well with causal mechanistic reasoning, which emphasizes the connections between entities and their properties, as well as linking across scalar levels. This highlights and further supports the utility of causal mechanistic reasoning. By calling upon entities a scalar level below, unpacking their properties, and linking their behavior to the overall phenomenon students can strengthen the connection between those resources. Given the strong connections between energy and forces in the natural world, perhaps we can leverage causal mechanistic reasoning and electrostatics to help students to build connections between these resources in their minds.

In this study, we aim to explore this connection investigating both how students use causal mechanistic reasoning to reason through the PE changes which arise from the formation of LDFs as well as LDFs themselves. This will provide insight into how best we can support students' learning of these important core ideas which may be useful to them in chemistry and beyond.

### **Research questions**

1. How do students use causal mechanistic reasoning to explain differences in PE minima of interacting neutral species?
2. How do students' causal mechanistic explanations of the formation of LDFs compare to their explanations of differences in PE minima?

3. How do students' causal mechanistic explanations about LDFs and the depth of PE wells impact students' responses about associated macroscopic phenomena?

## **Methods**

### ***Participants***

We administered our tasks to two groups of undergraduate general chemistry students at a large midwestern public research university in fall 2018 and fall 2020. Students from both semesters were enrolled in the first semester of CLUE, the transformed core-idea centered general chemistry curriculum.<sup>37</sup> In fall 2018, the course was taught in person as usual; however, due to the COVID-19 pandemic, in fall of 2020 the instructors taught the course online via the video conferencing software Zoom. In fall 2018 we collected responses from 2,445 students (97.2% response rate) and in fall 2020 we collected 2,059 responses (87.7% response rate). All students included in this study consented for their work to be used for research purposes and we deidentified their responses before analyzing the data in accordance with our IRB. All names included in this manuscript are pseudonyms.

For our analysis, we randomly selected 250 students using a random number generator from both the fall 2018 and fall 2020 semesters. We selected sets of this size so that we would have a sample size large enough to conduct meaningful statistical tests without having to analyze responses from over 2000 students. While we analyzed responses from two different semesters, our goal was not to uncover how these two groups differ. Rather we used the responses from both groups to better understand how students in this course understand the relationship between forces and energy. While these two groups of students are similar in some ways, for example in terms of their standardized SAT test scores (independent samples t-test FS18 mean: 1204, FS20 mean: 1224,  $p = 0.087$ ), the outbreak of COVID-19 drastically changed the college experience of those enrolled in fall of 2020. At the very least, the global pandemic shifted this class, which had been taught in-person during all prior semesters, to be taught online. Researchers have just begun to reveal how difficult the global pandemic has been for students,

both inside and outside the classroom, which may have affected students' performance on our research activities.<sup>114,115,117</sup> Despite these circumstances outside of our control, we feel that we can still learn much from these students.

### ***Data collection***

For this study, we administered two tasks, which we refer to as the "LDF task" and "PE task", in sequence on an exam given several weeks after these concepts were first introduced. However, moving the course online in fall 2020 also changed the format of the exam. In fall 2018, this exam was administered as an in-person exam, but in fall 2020, given the virtual nature of the course, students completed the exam on their own outside of the class. In addition to the change in physical location, students in fall 2020 had extended time to complete the exam (over the course of a weekend) compared to the students in fall of 2018 who completed the exam in one hour and twenty minutes.

### ***LDF text and drawing tasks***

The tasks we used to probe students' understanding of the origin of LDFs originated from our previous work.<sup>7,75</sup> In these tasks, we asked the students to draw two neutral atoms approaching each other (LDF drawing task) and to use their picture to explain the attraction between the neutral atoms (LDF text task) (Figure 7.1). In fall 2018 and 2020 we situated the phenomenon in the context of pairs of interacting neon and argon atoms respectively. Besides the change in atoms, there were no other meaningful changes in the LDF tasks between the two semesters.

**Consider the noble gas Ne.** In the boxes below, draw a series of atomic level pictures to show what happens when two atoms of Ne approach each other to form a stable interaction. Use pictures to help you explain why the two Ne atoms are attracted to each other.

Drawing:                      Explanation:

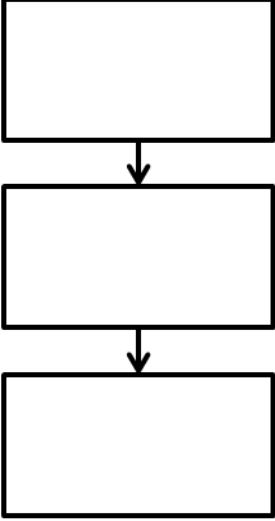


Figure 7.1. The LDF text and drawing tasks given in fall 2018. We modified the format of these questions to reduce the size for this manuscript.

#### *PE task*

To probe students understanding of PE, we asked students a series of questions related to the relationship between PE and strength of electrostatic interactions. In PE task version 1 (Figure 7.2) given to the fall 2018 students, we asked students to draw the PE curves of two interacting neon atoms and two interacting helium atoms as they approach each other. We then asked the students to identify the differences between the two PE curves and explain why those differences exist. We hoped that in their explanations they would connect the differing number of charged subatomic particles to the strength of the attraction and, therefore, a lower PE minimum. In the second version of the PE task (Figure 7.2) given to the fall 2020 students, we still asked the students to draw the PE curves of two sets of interacting atoms (argon and neon), but this time we included additional language to encourage students to think back to their LDF responses (in which they were prompted to unpack the electrostatic nature of the attraction). We also added two new questions to this version asking students to identify

which species (argon or neon) had a higher boiling point and to explain this difference using the PE curves they had drawn. We describe the rationale for the changes between PE tasks version 1 and 2 in the results and discussion section.

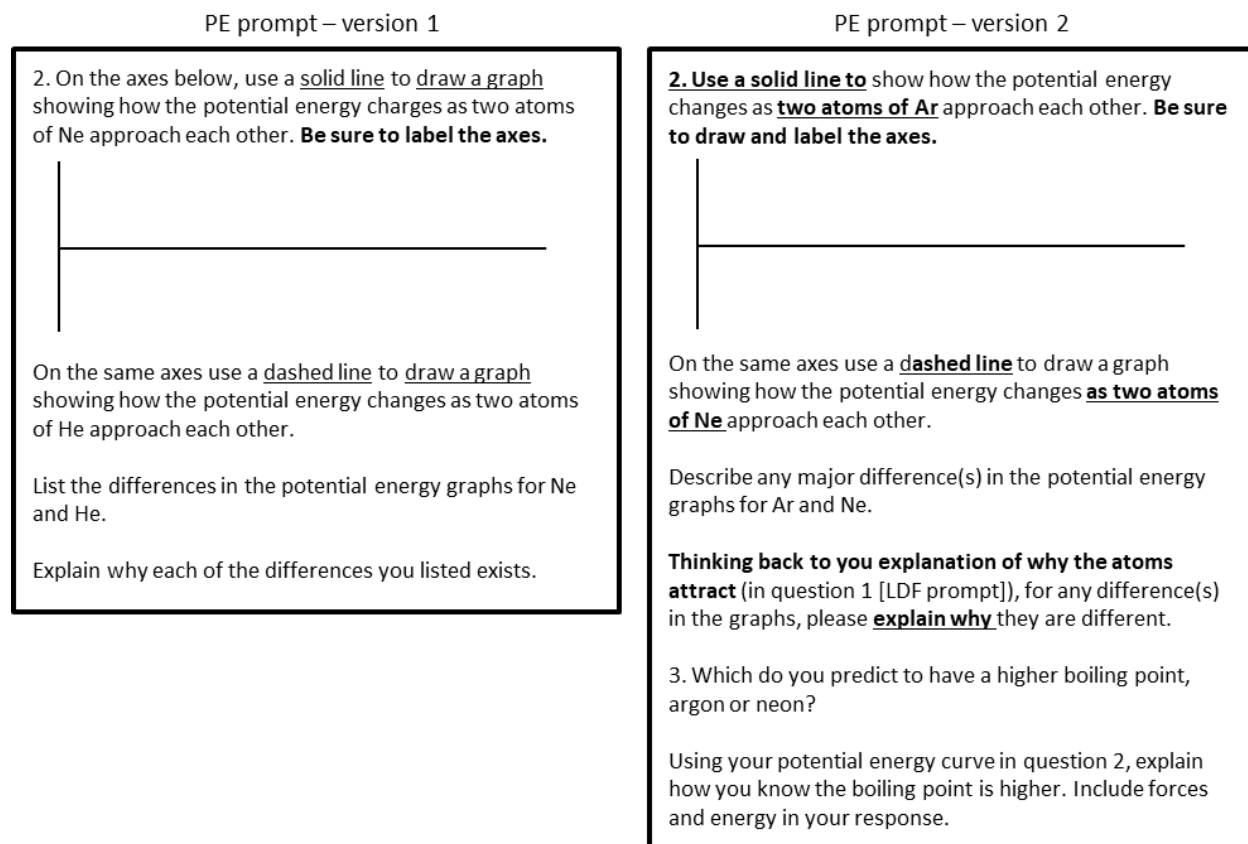


Figure 7.2. Both versions of the PE task. The LDF tasks (Figure 7.1) are question 1 in this figure. We modified the format of this question to reduce the size for this manuscript.

### ***Coding schemes***

To make sense of the students' responses, we used three different coding schemes to capture how the students used causal mechanistic reasoning to respond to: the LDF text task, the LDF drawing task, and the PE task.

#### ***LDF coding scheme***

For the LDF text and drawing responses we used our previously published coding schemes.<sup>75</sup>

Both the LDF text and LDF drawing coding schemes bin the responses into one of three mutually exclusive categories: non-electrostatic (NE), electrostatic causal (EC), or causal mechanistic (CM). NE

responses only describe the interaction or attribute the interaction to a factor that is not explicitly electrostatic, like simply naming the London dispersion force or drawing two uncharged atoms approaching one another. EC responses include the role of electrostatics in the interaction by discussing or drawing the charges in each atom. CM responses expand upon the EC responses by explaining or drawing how these neutral atoms develop charged ends, due to the localization of electrons on one side of the atom, which then allow the oppositely charged ends of each atom to attract. We highlight the critical features of each bin and provide example responses in Tables 2 and 3 of our prior publication.<sup>75</sup> Based on emergent patterns in this data set, we modified one part of the requirements for CM drawing category to require the drawing to explicitly show a change in electron cloud shape or the movement of electron.

#### *PE coding scheme*

For the PE task, we set out to develop a coding scheme analogous to the LDF coding scheme, capable of assigning a single mutually exclusive code characterizing engagement in causal mechanistic reasoning. To develop this scheme, authors K.N. and R.L.B. open coded 180 responses collected in fall 2018 to identify emergent themes and patterns in the students' explanations. To do this, the two authors independently analyzed sets of thirty responses, capturing the ideas students used to compare the PE minima. The two researchers met after each set to discuss their findings and reflect upon the themes identified in previous sets. After open coding 180 responses, we felt that we had generated an exhaustive list of the themes present in their explanations and therefore could continue with the analysis process.

Next, we condensed those themes into meaningful mutually exclusive coding categories to characterize the degree to which they engaged in causal mechanistic reasoning. As we hoped, students identified two differences in the PE minima: the depth of the PE minimum (position on the y-axis) and the internuclear distance at which the PE minimum occurs (position on the x-axis). During the open

coding process, we found that students provided richer explanations of the role of the subatomic particles when comparing the depth of the PE minima. Therefore, we focused on the themes associated with explanations of the PE well depth when developing the coding scheme. Additionally, we were interested in capturing causal mechanistic reasoning, so we focused on students' ideas related to the entities a scalar level below (i.e., the subatomic particles) and their electrostatic properties. Finally, we created three coding categories: non-electrostatic (NE), partially causal mechanistic (PCM), and causal mechanistic (CM) (Table 7.1).

Table 7.1. Descriptions and examples of the coding categories for characterizing responses to the PE Prompt.

Type of Response	Description	Student Examples
Non-electrostatic (NE)	The student solely described the phenomenon or attributed the difference in the PE minima to a non-electrostatic feature at the scalar level of the atom like "size".	"The Ne graph has a deeper PE well because the atom is larger, so it needs more energy to break interactions"
Partially causal mechanistic (PCM)	The student discussed the role of the subatomic particles (entities a scalar level below) to explain the difference in PE well depth but did not unpack their charged nature. Alternatively, they leveraged the role of charge to explain the different depths of the PE well but did not connect electrostatics to the subatomic particles.	"Ne has more electrons and therefore a larger electron cloud. This makes the attractive forces of the 2 atoms much stronger and therefore the [sic] need much more energy to overcome the attractive forces" "The well is further down on Ne than He because Ne is bigger in size thus being stronger in charges and forces."
Causal mechanistic (CM)	The student explained the role of charge in causing the difference in the PE well depth and linked those charges to the subatomic particles.	"Ne has a deeper well because it has more electrons and protons therefore the slightly negative side of the dipole and the slightly positive side of the dipole are stronger, causing a stronger interaction which needs more energy in order to be overcome."



NE responses either solely described the different depths of the PE minima or attribute the depth of the well to a non-electrostatic property at the scalar level of the atom, such as the atom's mass or size. PCM responses provided evidence of some component of causal mechanistic reasoning but did not include all the components. Either the response attributed the difference in PE minimum to the number of subatomic particles within each atom without unpacking their electrostatic properties or they leveraged the charged nature of the atom without discussing the subatomic particles. Finally, those students who received the CM code have put all the pieces together, explaining that the difference in the depth of the PE minima is due to the larger partial charges present in one set of atoms which occurs due to those atoms having more electrons and protons. We provide further description and examples of each bin in Table 7.1.

### ***Data analysis***

#### ***Fall 2018 data***

Before analyzing the responses, we first established interrater reliability (IRR) for both the LDF and PE coding schemes. In this process, a pair of researchers independently coded sets of forty randomly selected responses and checked their initial agreement by calculating Cohen's kappa, a quantitative measure of agreement which accounts for the probability of agreement by chance.<sup>71</sup> If this initial Cohen's kappa value surpassed 0.7, corresponding to "substantial" agreement,<sup>72</sup> then we determined that we had reached IRR and could begin the coding process. We calculated Cohen's kappa value using SPSS.<sup>129</sup> If we did not reach this threshold Cohen's kappa value, we came to consensus on our disagreed upon responses and then repeated the process with a new set of forty randomly selected responses until we did pass this threshold. For the fall 2018 LDF responses, authors R.B. and R.L.M. reached IRR for the text responses after two rounds (initial Cohen's kappa value = 0.80) and three rounds of drawing responses (initial Cohen's kappa value = 0.92). For the 2018 PE responses, authors K.N. and R.B. reached IRR after coding two rounds of responses (initial Cohen's kappa value = 0.82). Two

coders (LDF text and drawing: authors R.L.M. and R.B.; PE: authors R.B. and K.N.) then independently analyzed the 250 fall 2018 students' responses and met afterwards to reconcile any disagreements. The coding results presented in this paper are the final human consensus codes. By analyzing the data in this way, we could further ensure the reliability of the final codes.

#### *Fall 2020 data*

We repeated the IRR process before we coded the fall 2020 responses because some time had passed since our analysis of the fall 2018 data. We also changed the coders due to our researchers' availability. For the fall 2020 responses, authors R.B. and K.N. again coded the PE responses, but authors R.L.M and K.N. coded the LDF text and drawing responses. For the fall 2020 LDF responses, authors R.L.M and K.N reached IRR after one round of the LDF text responses (Cohen's kappa value = 0.88) and one round of LDF drawings (Cohen's kappa value = 0.96). For the fall 2020 PE responses, authors K.N. and R.B. reached IRR after two rounds (Cohen's kappa value = 0.96). We analyzed their LDF and PE responses in the same manner as the fall 2018 data, with both authors independently coding the students' responses, meeting afterwards to reconcile any disagreements.

#### *Statistical tests*

After coding the responses, we used SPSS<sup>129</sup> to conduct Pearson's  $\chi^2$  tests of independence to make sense of the relationships between the groups of responses. This statistical test evaluates how different the distribution of observed responses is from the distribution we would expect if there was no relationship between the two variables.<sup>65</sup> If the results of this analysis revealed that  $p < 0.05$ , we determined there to be a significant difference in the distribution of responses and therefore an association between the variables. To evaluate the effect size of any significant results, we calculated Cramer's V, an extension of Phi for cross tables with more than two rows or columns. According to Cohen, Cramer's V values of 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes respectively.<sup>66</sup>

Where appropriate, we also conducted post-hoc analyses by calculating the adjusted standardized residuals (ASRs) for each cell in the cross table (representing the relationship between a specific pair of categories), quantifying the difference between the observed and expected value (assuming no relationship between the variables).<sup>67</sup> The greater the difference between the two values, the larger the magnitude of the ASR. If the ASR surpassed the critical value, we determined that specific pair of categories was driving the initial  $\chi^2$  result. We used the Bonferroni adjusted critical value for this analysis to appropriately account for increased risk of type 1 error (i.e., false positives).<sup>69</sup>

## **Results and discussion**

### ***Study 1: How do students use causal mechanistic reasoning to explain differences in PE minima of interacting neutral species***

#### *Results of PE task version 1*

In response to the first version of the PE task, we found that few students provided causal mechanistic explanations (CM: 8%) (Figure 7.3). Instead, most of the students (NE: 60%) did not discuss electrostatics explicitly in their response. For example, as Joe explains: *“The Ne graph has a deeper PE well because the atom is large, so it needs more energy to break interactions.”* While Joe implies that the stronger interactions require more energy to overcome, they only leverage the size of the atom to answer the question without any mention of charged subatomic particles that led to the stronger interaction.

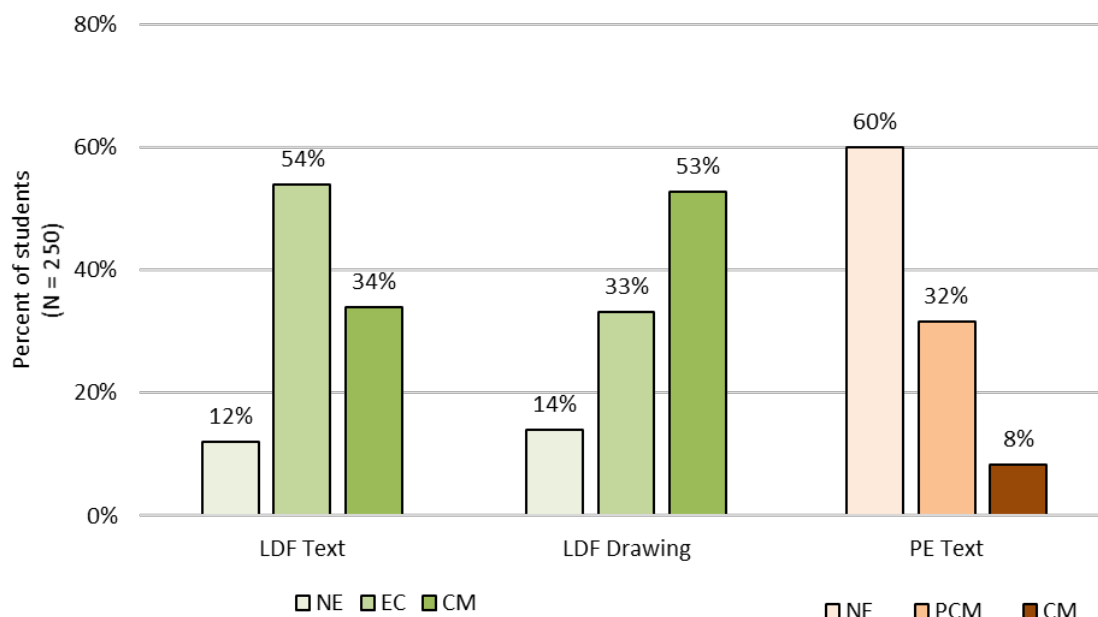


Figure 7.3. Distribution of fall 2018 students' responses to the LDF text task, LDF drawing task, and PE task version 1.

Another 32% provided some part of a causal mechanistic response (PCM), the bulk of which (24% of the total) compared the number of subatomic particles (e.g., electrons) to explain the difference in PE well depth. Consider this response from Sarah who explains that the PE minima for interacting Ne atoms is deeper “...because Neon has more electrons and therefore a larger electron cloud. The larger the electron cloud, the stronger the interactions keeping the particles together. Since the bonds (sic) are stronger it would take more energy to separate them.” While Sarah steps down a scalar level and considers the number of electrons in neon compared to helium, they do not unpack the charged nature of the subatomic particles and instead reference the size of the atom. The prevalence of this heuristic (larger atoms experience stronger interactions) may explain why the majority (87%) of students correctly indicated that the neon PE minimum is deeper even though so many of students provided a NE explanation.

Based on the prevalence of the NE PE explanations, one might assume that these students do not have the necessary electrostatic resources to call upon. This could be the case, however, that would

contradict the results of the LDF task. In response to those tasks, most students provided either an EC or CM LDF text or drawing response (88% and 86% respectively) (Figure 7.3). Both types of responses require the student to leverage ideas related to electrostatics, suggesting they do indeed have electrostatic resources to call upon—the students are just not including these ideas in response to the PE prompt.

Why then are students not using those ideas to respond to the PE task? The contextually dependent nature of resources may provide a clue. It could be that the PE task did not activate students' electrostatic resources. Another possibility is that these resources were activated, but that the PE task did not cue the students to include those ideas in their response. That is, for one reason or another they felt it unnecessary to elaborate upon the role of the charged nature of the subatomic particles in explaining the difference in PE minima.

Considering these results, we modified the PE task to better activate the appropriate resources needed to reason causal mechanistically through this phenomenon, and to cue to students to include those ideas in their explanation. To do this, we included more explicit reference to the LDF prompt, as that prompt was successful in activating the students' electrostatic resources. We also added two questions, extending the change in PE to a macroscopic phenomenon: boiling point (Figure 7.2). By asking students to compare the boiling points of the two neutral species and explain this difference, we hoped students could have an opportunity to expand upon their earlier ideas. To further emphasize this, we specifically asked the students to reference their PE graph when explaining the difference in relative boiling points.

#### *Results of PE task version 2*

In fall of 2020, nearly all the students (92%) correctly answered the PE minima question, indicating that the depth of the PE well for argon below that of neon. Additionally, 97% of the students

indicated that the boiling point of argon was higher than that of neon. Not only did the students again all get the overall phenomenon correct, but there was also a significant change in the causal mechanistic reasoning used by students in their explanations of the PE minima compared to fall 2018 ( $\chi^2=82.79$ ;  $p<0.001$ ; Cramer's  $V=0.41$ ). Far fewer students provided NE responses (25%) and more students provided CM responses (37%) (Figure 7.4). Consider Gene who explains “Argon has more electrons to be shifted at any given time, leading to a stronger polarity & therefore stronger LDFs compared to Neon, thus making the well for Argon’s potential energy deeper.” Gene not only steps down a scalar level identifying the electrons, but also connects the electrons to charges (polarity) and uses that idea to support the stronger interactions experienced by argons. The proportion of PCM responses remained relatively similar, increasing slightly from 32% to 38%. Overall, these results are encouraging as more than a third of the students in fall 2020 provided a causal mechanistic explanation of the difference in PE minima.

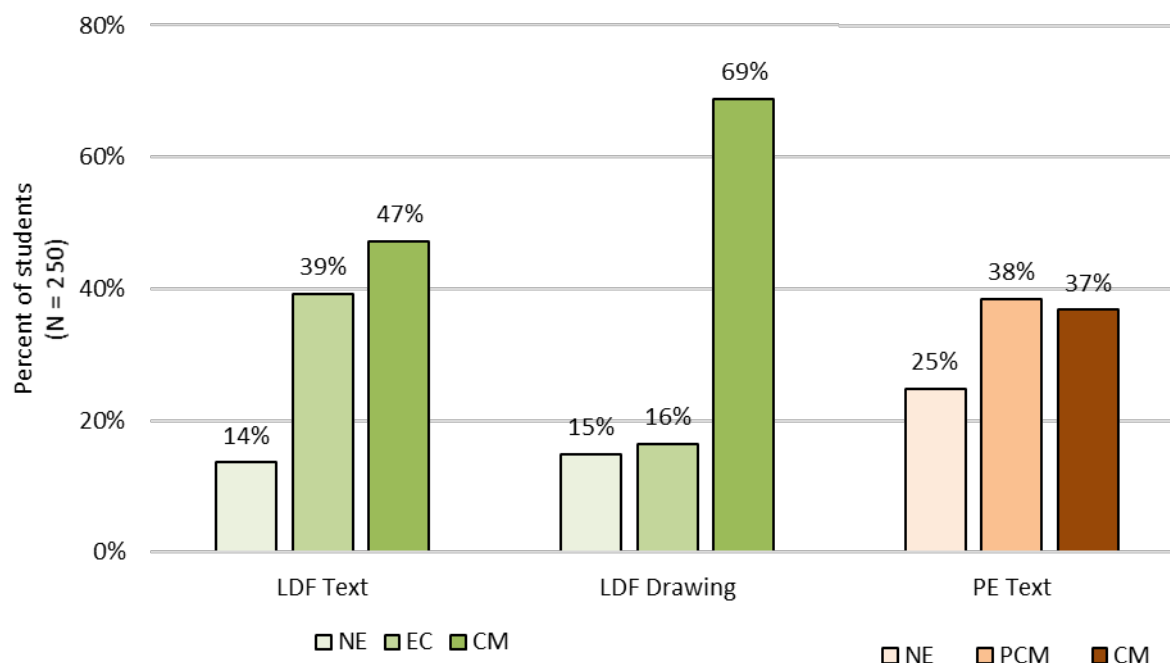


Figure 7.4: Distribution of fall 2020 students’ responses to the LDF text task, LDF drawing task, and PE task version 2.

The PE responses were not the only change we observed as there was also a significant difference in the distribution of both the LDF text ( $\chi^2=11.49$ ;  $p=0.003$ ; Cramer's  $V=0.15$ ) and drawing responses ( $\chi^2=19.55$ ;  $p<0.001$ ; Cramer's  $V=0.20$ ) compared to the fall 2018 responses. Despite no change in the LDF prompt, more students were now providing CM text and drawing responses (47% and 69% respectively). It is important to note though that while there was an increase in CM LDF responses, not *all* of the students provided CM LDF responses. That would suggest the at-home exam environment did not result in widespread cheating. Therefore, we can still be relatively confident that these responses are providing useful insight into what students know and can do. Because there were multiple changes occurring in fall 2020 (e.g., modification of the PE task, move to online learning, at-home assessment environment) we cannot identify which of these changes impacted student responses. However, determining the cause of these differences was never our goal. What matters is that we now have LDF and PE tasks capable of eliciting causal mechanistic explanations allowing us to explore the connections (and misconceptions) between the students' understanding of forces and interactions.

***Study 2: How does students' causal mechanistic explanations of the formation of LDFs compare to their explanations of differences in PE minima***

To characterize the relationship between students' responses to the LDF and PE tasks, we conducted a pair of  $\chi^2$  tests using the fall 2020 responses. We used only the responses from this semester as many more students provided causal mechanistic explanations to the PE prompt compared to fall 2018. With a substantial number of NE *and* CM PE responses, we could identify what types of LDF text and drawing responses were associated with both categories.

This analysis revealed that there is a statistically significant relationship between the PE and LDF text responses ( $\chi^2=13.79$ ;  $p=0.008$ ; Cramer's  $V=0.17$ ) and the PE and LDF drawing responses ( $\chi^2=15.96$ ;  $p=0.003$ ; Cramer's  $V=0.18$ ). Both significant results are of small to medium effect size.<sup>66</sup> This evidence supports the notion that how students reason through the formation of the LDFs is associated with how

they reason through the subsequent changes in PE. To better understand these relationships, we conducted a post-hoc analysis and calculated the ASRs for each pair of categories. By digging deeper into the frequency that students provided certain types of responses (e.g., both a CM LDF text and PE response) we could see which categories were positively or negatively associated with one another. For a 9-cell table, ASRs surpassing the Bonferroni adjusted critical value of  $\pm 2.78$  are significant drivers of the initial significant  $\chi^2$  result. The results of this post-hoc analysis is shown in Figure 7.5.



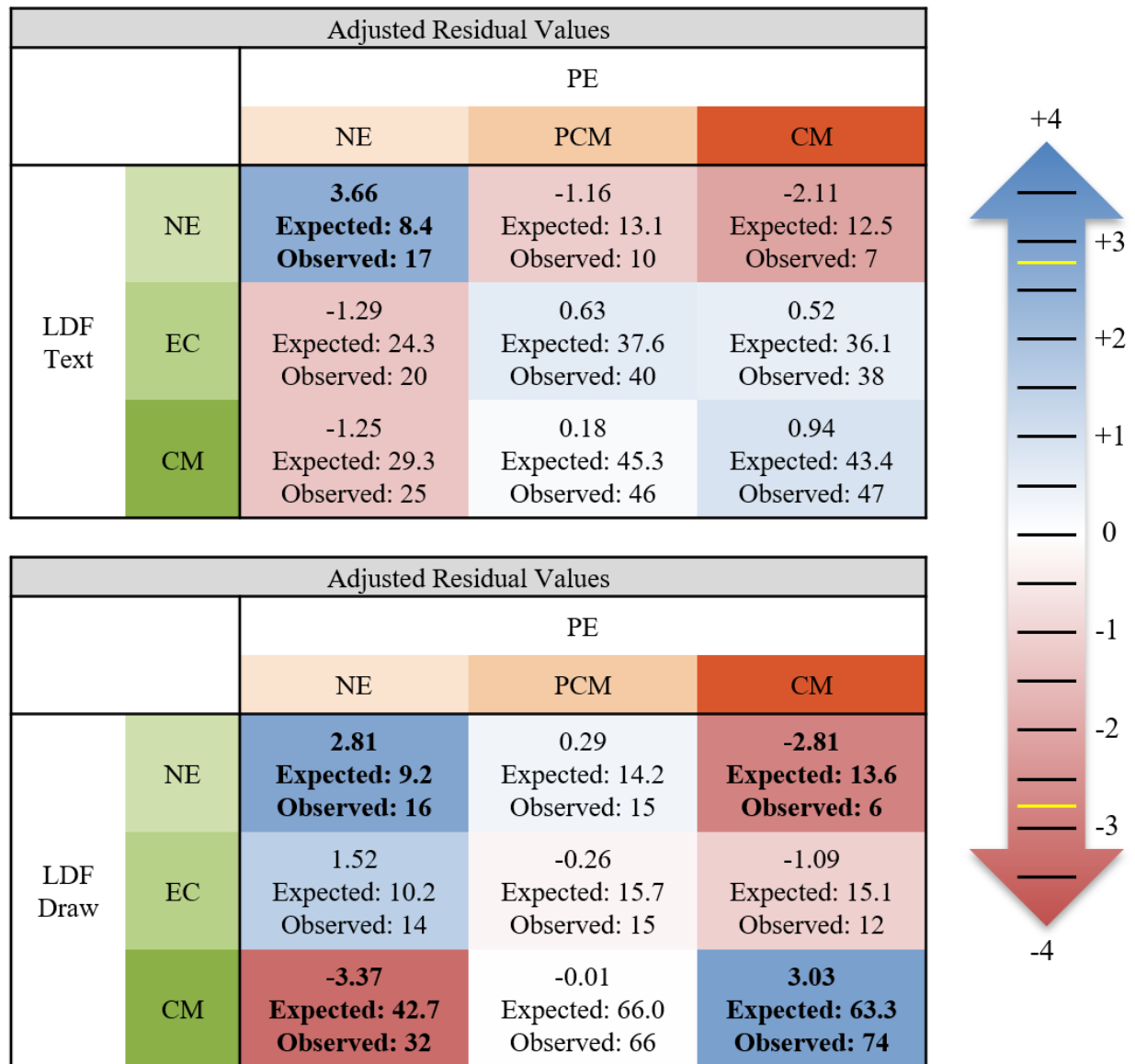


Figure 7.5. The adjusted standardized residuals of the cross table comparing the PE and LDF text responses as well as the PE and LDF drawing responses. The cells are color coded according to the scale on the right. Cells with more positive adjusted residuals (blue) have more observed instances than would be expected if there was no relationship between the variables. Cells with more negative adjusted residuals (red) have less observed instances than expected. Bolded adjusted standardized residuals have a magnitude greater than 2.78 (yellow line on the scale) and are considered significant drivers of the initial  $\chi^2$  result.

When comparing PE and LDF text responses, we found only one ASR surpassed the critical value, indicating that the positive association between the NE LDF text and PE responses was a primary driver of the initial significant  $\chi^2$  result. The other ASRs indicate that none of the other categories are positively associated with a NE response, while the EC, PCM, and CM responses are all positively associated with

one another. However, none of these other ASRs surpass the critical value. This suggests that students who call upon non-electrostatic resources to reason through one task use those same non-electrostatic ideas to reason through the other. The same is not true however for partially or fully causal mechanistic reasoning. If a student provides a fully CM response to one prompt, they may leverage some aspect of causal mechanistic reasoning in the other prompt, but the smaller ASRs suggest that this relationship is weaker. This highlights how difficult these concepts are to understand and articulate.

The ASRs calculated for the PE and LDF drawing responses show a similar relationship for the NE and CM categories. Again, there is a significant positive association between the NE PE and LDF drawing categories, but now there is also a significant positive association between the CM responses as well. The corresponding negative associations in other corners of the cross table are also significant (Figure 7.5). That is, for the LDF drawing and PE responses we see the relationship occurring at both ends. If the student provides an NE or CM response to one task they are more likely to provide the same type of response to the other task.

Interestingly, there are no significant ASRs in either the PE: PCM or LDF drawing: EC categories. Additionally, the ASRs for these categories suggest the exact opposite associations compared to the LDF text and PE responses. This may be an indication of a “messy middle” in which students are figuring out how to piece these ideas together in a productive manner. This would align well with the fragmented nature of knowledge outlined in the resources perspective and highlights the importance of providing students with continued opportunities to engage in causal mechanistic reasoning activating electrostatic resources to explain phenomenon related to both forces and energy.

***Study 3: How do students’ causal mechanistic explanations about LDFs and the depth of PE wells impact students’ responses about associated macroscopic phenomena***

In addition to analyzing students’ explanations for evidence of causal mechanistic reasoning, we also examined students’ predictions about the macroscopic phenomenon. That is, what is the

relationship between students' understanding of the underlying entities and their ability to predict what occurs at the scalar level above. In this study, we discuss how students responded to the prediction questions in the PE task. Specifically, their predictions of which species would have a lower PE minimum (fall 2018 and fall 2020) and which species would have a higher boiling point (fall 2020 only).

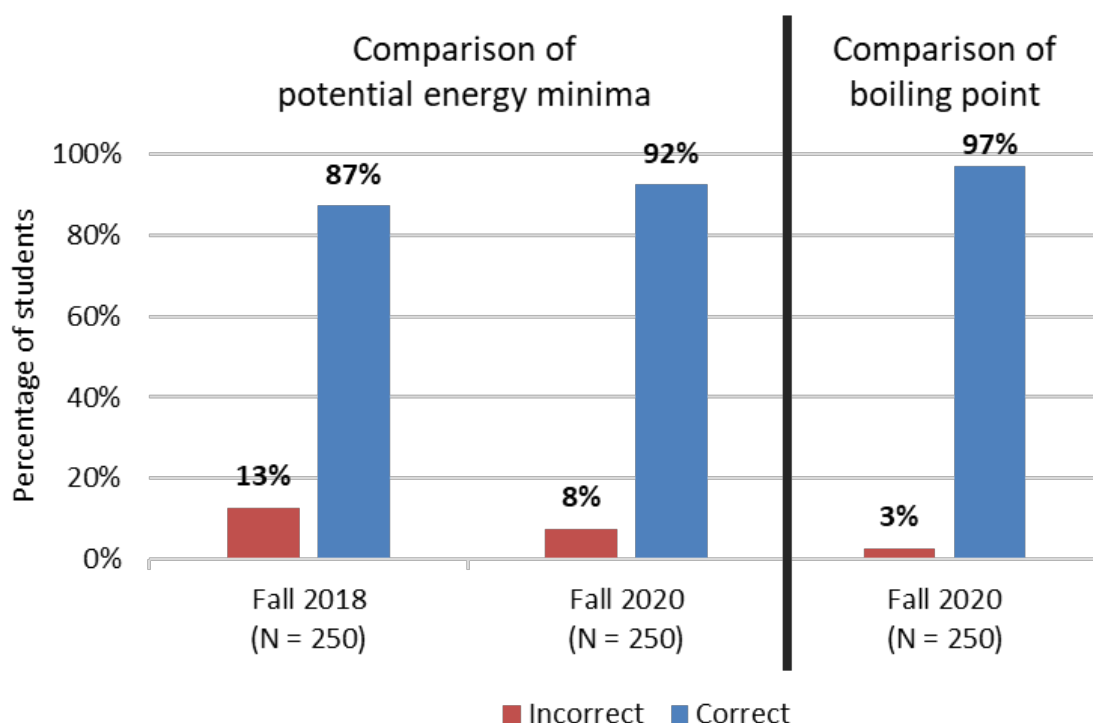


Figure 7.6. The distribution of correct and incorrect students' predictions about the relative depth of the PE minima (left of black divider) and boiling point (right of black divider). Note that in fall 2018 we did not ask the students to compare the boiling points.

In fall 2018, the overwhelming majority of the student (87%) correctly predicted that the interacting neon atoms would have a deeper PE well than the helium atoms (Figure 7.6). It is notable that these students were able to correctly predict the relative depths of the PE wells even though 60% of those students in fall 2018 provided a non-electrostatic explanation for this phenomenon (Figure 7.3). However, as we discussed in study 1, the prevalence of non-electrostatic PE explanations could be the result of the question design as the majority of those students included productive ideas in response to the LDF task (Figure 7.3).

In fall 2020, many more students provided a causal mechanistic explanation to the PE task (37%), but the percentage of students correctly predicting the relative depth of the PE minima remained relatively unchanged (92%) (Figure 7.6). Granted, that is partially due to the fact the students in fall 2018 did very well on this question. Additionally, nearly all the fall 2020 students (97%) correctly predicted the other phenomenon, that argon would become a gas at a higher temperature.

These results indicate two important things. First, that all these students, across both semesters, were very successful in correctly predicting the macroscopic phenomenon. This is certainly encouraging, but even though about 90% of the students in both fall 2018 and 2020 correctly predicted the phenomenon, the distribution of causal mechanistic explanations of the PE well depth look very different for the two semesters. That is, even though these two questions both explored PE minima, they provide very different pictures of students' understanding. Finally, even though in fall 2020 over 90% of the students correctly predicted that argon would have a deeper PE well and a higher boiling point, only about a third of the students provided a causal mechanistic explanation of the depth of the PE minimum. This could indicate that students were relying on heuristics, like size, instead of ideas of forces and interactions to predict the boiling point. While this p-prim "more is more" (bigger atom-larger boiling point, larger PE well), provides an appropriate answer in this instance, this is not always the case.<sup>48</sup> This highlights the importance of asking students to explain. In order to see if students know *how* and *why* phenomenon occur, we need to ask them those questions.

### **Limitations**

The presence of the global pandemic during the administration of the FS20 activity impacted students in ways we are only just beginning to understand. Despite these circumstances, we feel we can still learn valuable lessons from these students that might help us to provide better instruction and support for learning in the future.

While we elicited rich, causal mechanistic explanations from these students, it is important to note that the CLUE curriculum (in which these students were enrolled) places explicit emphasis on the core ideas of energy and forces and interactions. Students from other curricula, which place less emphasis on these ideas, may not be as likely to provide these types of causal mechanistic explanations.<sup>113</sup>

### **Conclusion**

In our prior work, we explored how students think about LDFs both because intermolecular forces are important in chemistry and because of the close relationship between forces and energy, another critical idea in science. In this study, we investigate this connection, between understanding forces and energy in more depth. Specifically, we examined how students leveraged causal mechanistic reasoning to compare the PE minima, how their responses related to their explanations of the formation of LDFs, and whether these explanations were connected to predictions about associated macroscopic phenomena.

While we previously developed materials to explore students' understanding of LDFs,<sup>7,75</sup> we needed to repeat this process for a phenomenon relevant to PE. We built upon our LDF task, asking students to graph how PE changes with distance for two pairs of interacting neutral species (which would experience LDFs). Our first attempt primarily elicited non-electrostatic responses. That is, students leveraged ideas like size rather than the electrostatically charged subatomic particles most students called upon when explaining the formation of the LDFs in the question before. In response, we revised the task to better activate the relevant electrostatic resources and cue the students to include those ideas in their response. However, we cannot know if these modifications to the task led to resulting changes in the responses as the course and assessment environment changed simultaneously. It is reasonable to assume that these other factors influenced the students' responses in some way

considering that in fall of 2020 we observed more causal mechanistic LDF text and drawing responses despite only modifying the PE task.

Regardless of the specific cause, we were encouraged to see that over a third of these students now provided fully causal mechanistic accounts of the difference in PE minima with another third leveraging some aspect of causal mechanistic reasoning. The fact that so many students are engaging in causal mechanistic reasoning, unpacking the properties and behaviors of the entities a scalar level below, is important. This type of reasoning is highly valued in science given its powerful predictive and explanatory nature.<sup>5,6</sup> However, these students may not be using causal mechanistic reasoning to predict the overall phenomenon. For example, despite most students providing a non-electrostatic response to the PE task, nearly all the students correctly predicted which pair of neutral atoms would have the lower PE minima. These students may instead be relying on associated ideas such as size, potentially leveraging p-prims “more is more”,<sup>12,48</sup> to correctly answer the question without using core ideas like electrostatic interactions to reason through the phenomenon. Furthermore, in fall 2020 even though only 37% of students provided a fully causal mechanistic explanation of the PE minima difference, 97% correctly predicted the relative boiling points. This highlights the importance of asking how and why phenomenon occur. This not only allows us to collect evidence of causal mechanistic reasoning, information which we can use to better support students learning, but also to send a message to students that this type of reasoning is valued.

Exploring the relationship between students PE responses and LDF text and drawing responses revealed that there is indeed a relationship between how students engage in causal mechanistic reasoning to explain both phenomena. This supports the idea that forces and energy are linked, and that developing a deep understanding of forces may help students to better understand the abstract concept of energy. However, our analysis also revealed that this relationship is modest, and that the partially causal mechanistic PE responses were not strongly associated with any of the LDF responses. This

highlights how difficult an idea energy is to grasp, and how messy this learning process can be. We need to give students the opportunity to think about the relationship between forces and energy to connect these resources in productive ways. These are both core ideas in chemistry which must be woven into the fabric of the course so that students may have a foundation they can call upon in their future learning.

## CHAPTER VIII - A DEEP LOOK INTO DESIGNING A TASK AND CODING SCHEME THROUGH THE LENS OF CAUSAL MECHANISTIC REASONING

### Preface

The purpose of this paper is to share the iterative process we used to design a task that elicits causal mechanistic reasoning and how the subsequent student responses can be analyzed. The task development was approached using (1) a resources perspective of learning, (2) principles of scaffolding, and (3) evidence-centered design, for which we specified evidence that would be considered a fully causal mechanistic explanation. To characterize students' responses to this task, we developed a coding scheme which can be used to code explanations based on the presence or absence of three key ideas relevant to this phenomenon. This research has been previously published in the *Journal of Chemical Education* and is reprinted with permission from Noyes, K.; Carlson, C. G.; Stoltzfus, J. R.; Schwarz, C. V.; Long, T. M.; Cooper, M. M. A Deep Look into Designing a Task and Coding Scheme through the Lens of Causal Mechanistic Reasoning. *J. Chem. Educ.* **2022**. Copyright 2022 American Chemical Society.

A copy of permissions obtained is included in the Appendices. Supporting Information for this manuscript is included in the Appendices.

### Introduction

Our interdisciplinary team of researchers has been studying how students connect and use ideas across the disciplines of chemistry and biology at different scales. In particular, we have chosen the lens of causal mechanistic reasoning (CMR), which connects phenomena to the behaviors and interactions of entities at lower scalar levels and provides a powerful predictive and explanatory strategy.<sup>130</sup> By engaging students in tasks that promote CMR, our goal is to provide approaches that are appropriate across a range of courses and contexts, and through analysis of students' explanations, can inform how we might better help students make interdisciplinary connections. However, the development and evaluation of such tasks is by no means trivial. This paper presents the process by which we designed



one such task and the resulting coding scheme that will allow us to characterize student mechanistic explanations both within and across disciplines. While the task and coding scheme will eventually be used in this way, it is the development process and qualitative analysis guiding the revisions which we describe in this paper—such analysis is a finding in and of itself.<sup>131</sup>

### ***Causal mechanistic reasoning***

Our work is based on the causal mechanistic reasoning framework (hereon referred to as the CMR framework) outlined by Krist et al.<sup>130</sup> This is a simplified framework based on that described by Russ et al.,<sup>132</sup> designed to work across content areas and for written assessments, which involves three steps. The first is to consider the level below the target phenomenon. The second is to identify and unpack the behaviors and interactions of entities at that lower level. And finally, the third is to connect how the lower-level behaviors give rise to the target phenomenon.<sup>130</sup> These steps, both individually and taken together, form a powerful thinking strategy that is central to all science disciplines. It is possible that causal relationships, and the mechanisms that underly them, could provide an important connection between chemistry and biology, as phenomena at the atomic, molecular, and cellular levels are the result of behaviors and interactions of entities at smaller scalar levels. Thus, being able to reason causal mechanistically might better help students navigate between these courses and make predictions about biological phenomena using chemistry ideas. For these reasons, we used this CMR framework to guide the development of a task and subsequent evaluation of responses to that task.

### ***Resources perspective of student learning***

In the development of tasks, we approach learning through a *resources* perspective, which theorizes that students call upon context-dependent conceptual and epistemological resources to make sense of phenomena.<sup>133,134</sup> This is a departure from previous research identifying misconceptions, which is based on the underlying assumption that students hold coherent and intact conceptions, which should be challenged and replaced with more expert conceptions.<sup>134,135</sup> While research into misconceptions has

revealed the range of problematic ideas students hold, this approach tends to emphasize a deficit view of learning and does not adequately account for how these novice ideas are replaced with, or evolve toward, more expert ones.<sup>135</sup> In contrast, the resources perspective takes a constructive approach to student thinking by focusing on the knowledge pieces that students *do* have in contexts and how students use those ideas in productive ways.<sup>133,134</sup> Through this lens, the aim of teaching is not to replace the students' ideas, but to design instruction that activates appropriate and productive resources for the specific context, allowing students to use and advance those ideas.

This perspective on student thinking has important implications for how knowledge is transferred to new contexts, including across disciplinary boundaries. The resources perspective suggests that transfer is not the movement of intact ideas, but rather the activation of similar resources in different contexts.<sup>134</sup> For example, when considering why atoms interact in a chemical context, conceptual resources related to electrostatic interactions and forces may be activated. When then asked to consider a biological phenomenon, such as why a ligand binds to an enzymatic binding site, the goal would be to activate these same cognitive resources in the new situation. If the student can repeatedly use and coordinate such resources in productive ways across multiple contexts, these connections may strengthen. However, developing expertise requires many years of experience working across a range of contexts to develop a connected framework of knowledge. It is therefore important that we give our students ample opportunity to use and develop their resources productively, and one such way we aim to do this is through formative assessment tasks.

### ***Assessment design***

Assessment of student learning can be thought of as a process by which evidence is elicited to make an argument about what students know and can do.<sup>136</sup> The general approach to the design of assessments involves specifying the type of cognition (or theory of learning that the assessment is designed to assess), and, in this case, we are using the resources perspective. Then, we should define

what types of observations would produce data that can be ultimately interpreted as evidence of learning. To do this, we used a modified evidence-centered design (ECD) approach,<sup>137</sup> which requires we first identify the resources (or evidence) we wish to elicit from students. That is, what would it look like for a student to reason causal mechanistically through the phenomenon of protein-ligand binding. We decided that explanations which appropriately leverage the electrostatic properties of the atoms and/or amino acids to explain the attraction between the protein and ligand would be evidence of CMR. While other ideas such as the size, shape, or orientation of the ligand also influence its ability to bind with a protein, the impact of all these characteristics on binding are inherently electrostatic in nature—maximizing attractive interactions and minimizing repulsive ones. Further, while we know that protein-ligand binding occurs in an aqueous environment, which brings with it the added complexity of entropic changes, we chose not to probe such ideas in this activity. It is our experience that designing tasks that probe for all these ideas at the same time leads to confusion and problems with coding the student responses.

The next step in ECD is to develop a task capable of eliciting such evidence. The task design is key as it provides the context that determines which of the students' resources are activated. One way to attempt to elicit particular resources is to provide scaffolding that supports students' reasoning and guides them to an appropriate response. Wood et al. coined the term scaffolding while exploring the methods by which an expert may help a beginner accomplish a task that they would be unable to do on their own.<sup>138</sup> This approach may be considered as a way to help students traverse Vygotsky's Zone of Proximal Development (ZPD) - the gap between what a learner can do on their own and what they can do with assistance from an expert.<sup>139,140</sup> While scaffolding was initially studied in one-on-one or in-person environments, this approach has also been applied to assessment design.<sup>141</sup>

Wood et al. identified several productive scaffolding techniques in their initial paper, two of which have been key for the development of our task: reduction in degrees of freedom and marking

critical features.<sup>138</sup> Both focus the learners' attention on specific ideas that we, the experts, have deemed important to the phenomenon. Additionally, these techniques align well with Hammer's resources perspective: by attempting to focus the learner's attention to relevant productive ideas, we can activate those resources so they may then use those and any other closely linked resources to reason through the task at hand.<sup>133</sup>

A good task though, must do more than support the activation of relevant resources in students' minds; it must also communicate what information we expect them to include in their response. However, like Goldilocks and her porridge, a task must be carefully balanced, providing enough information for the student to understand what is being asked, but not so much information that it can be answered without thoughtful effort. To design such a task, we drew on the work done by Graulich et al. to incorporate scaffolding through contrasting cases.<sup>142–144</sup> In their work, they asked students to compare two phenomena using a series of questions helping students to consider (1) what explicit structural features are different between the two cases, (2) how the phenomenon at hand differs for each case, and finally (3) how the resulting implicit properties of the structures lead to this change. This scaffolding design was illuminating, both because of its ability to foreground aspects of the phenomenon that might otherwise be missed, and the subtlety of the scaffold. To answer the question, the student must go down to the scalar level where the two cases differ, thereby providing a more "natural" way of framing the scalar level at which we wanted them to explain the phenomenon.

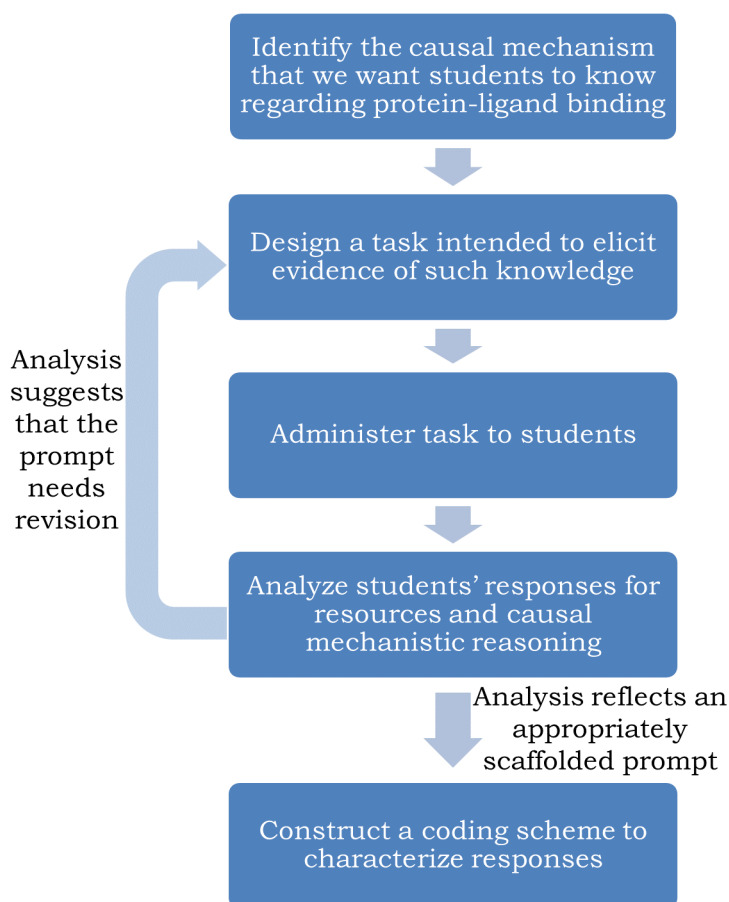


Figure 8.1. Our iterative process of designing a well-scaffolded assessment task.

Of course, designing an assessment that provides the right amount of scaffolding often takes multiple rounds of task design and evaluation of the students' responses (Figure 8.1). The iterative nature of this process is important to designing high-quality assessments.<sup>136</sup> While some researchers have shared portions of this process when discussing their tasks,<sup>145–148</sup> in this paper we provide an in-depth look at how we designed a task which reveals how students reason through the mechanism of protein-ligand binding, detailing the task revisions and underlying decisions which drew from principles of ECD, scaffolding, and the resources perspective. Although this project is part of a larger interdisciplinary research endeavor, in this paper, we only discuss the development of this assessment

(hereon referred to as the “PL task”, for Protein-Ligand) and the subsequent coding scheme used to characterize students’ explanations in relation to CMR.

### **Research questions**

The following questions will be addressed in this paper.

1. What is the impact of different types of scaffolding on the resources students use to respond to the task?
2. In what ways can we characterize the degree to which students are engaging in CMR to explain this phenomenon?

### **Methods**

#### ***Overview of the rationale for the design of the PL task and associated coding scheme***

We chose to situate our task in the phenomenon of protein-ligand (PL) binding, because, while it is positioned in a molecular biology context, it requires that students use core ideas from chemistry to construct a mechanistic explanation. To construct such an explanation, the student must think about the scalar level below the phenomenon, unpack the properties and behaviors of the entities at that scalar level, and link those properties and behaviors to the phenomenon. In this context, an ideal response identifies the charges or partial charges of the atoms or functional groups in the protein and ligand (that is where the electron density is distributed) and explains that the oppositely charged entities experience an attractive noncovalent interaction. Furthermore, we designed the task to incorporate two contrasting cases so that students could compare two potential binding sites and explain, using the strength of the charges, which site would bind most strongly and why their selected site preferentially binds the ligand. Then, using responses from the final version, we aimed to characterize the ideas that students used as well as their engagement with CMR to develop the coding scheme.

The development of this task and coding scheme required several iterations, and, in the methods section, we describe the different groups of students whose responses we analyzed during this

process. As our analysis of the students' responses drove the iterative redesign of the PL task, we describe key features of each task version and the rationale for our design decisions in the results section. The coding scheme has both analytic and holistic components to it and, while the development of the analytic rubric required several rounds of refinement, in this paper we present only the final coding approach. However, in the Appendix B, we include a detailed discussion of our decisions about several earlier bins that were refined, removed, or combined with others to ultimately reach our final analytic rubric. We have included this discussion in accordance with calls for greater transparency surrounding coding scheme development and application.<sup>131</sup>

### ***Participants***

For the development and testing of different versions of this task and coding scheme, we collected and analyzed the responses of students enrolled at a large midwestern public research institution, including participants from Molecular Biology (MB), General Chemistry 2 (GC2), and Organic Chemistry 1 (OC1) and Organic Chemistry 2 (OC2). The majority of these students are familiar with being asked to explain phenomena as the general and organic chemistry courses from fall 2018 and spring 2019 have been transformed, using the three-dimensional learning (3DL) approach,<sup>149,150</sup> and the molecular biology course is also undergoing transformation using the 3DL approach. The fall 2019 OC2 course had not been transformed using the 3DL approach. Unfortunately, due to logistical constraints (e.g., not all courses were offered every semester, existing agreements with instructors for data collection), we could not administer every version of our task to students in all these courses. Students were offered a small amount of extra credit in their course for completing our activity. The number of student responses collected and analyzed is shown in Table 8.1. All the students in this study consented for their work to be used for research purposes, and their responses were collected and deidentified in accordance with our IRB protocol. The names included in this manuscript are all pseudonyms.

Table 8.1. Overview of administration of each version of the PL task.

Semester	PL Task Version	Course	Responses collected (response rate %)	Responses analyzed (task development)	Responses analyzed (coding scheme development)
Fall 2018	1	Organic chemistry 1	44 (75%)	20	-
Fall 2018	1	Molecular biology	94 (68%)	20	-
Spring 2019	2	Organic chemistry 2	74 (74%)	20	-
Spring 2019	2	Molecular biology	313 (74%)	20	-
Summer 2019	3	General chemistry 2	61 (73%)	29	61
Fall 2019	Final	Molecular biology	121 (55%)	-	60
Fall 2019	Final	Organic chemistry 2	300 (85%)	-	60

Most student responses were collected using the online assessment system beSocratic, which allows students to draw and write free-form responses.<sup>151</sup> Early versions (1 and 2) were administered to MB students on a hard-copy worksheet. For the final version, all students responded using beSocratic. These activities consist of a series of “slides”, on which students can write or draw. In versions 2 and 3, we administered alternate versions (described in the results section) of the activity to test small, specific aspects of the task. Students in these semesters randomly received either the original or alternate version (full student counts presented in Appendix C).

### ***Analysis guiding task development***

Our iterative analyses of the students’ drawings and explanations guided the development of the PL task. The first author qualitatively analyzed small sets of responses (randomly selected via random number generator) from each course and each version of the task to identify themes and



patterns in the students' responses (Table 8.1). We selected more responses to analyze from version 3 to ensure that we could stop making substantive changes to the task after this version. Due to differences in the assessment medium, the first author was aware of the course background of each student (i.e., if the response came from a biology or chemistry student) during the analysis.

In this analysis, the first author examined the students' entire response (instead of their responses to each individual question) to explore all the resources they used to approach the overall phenomenon. Then, we compared the ideas that students used with our ideal causal mechanistic explanations of the phenomenon. That is, did the students identify and unpack the properties and behaviors of entities at a scalar level below to explain the phenomenon. After analyzing the responses from one version of the task, we shared our findings with our larger interdisciplinary team to discuss changes we could make to the scaffolding of the following version of the task to better activate the appropriate resources and elicit CMR. Examining the students' responses also provided us the opportunity to reflect on our evidence statements and determine what would be reasonable for a student to include in their explanation. We repeated this process three times until the task appropriately cued the students to provide a causal mechanistic response. In an effort to be as transparent as possible about our analysis that drove the task refinement, we have provided the full responses from all the students in the Appendix D.

### ***Analysis guiding coding scheme development***

Following version 3, we stopped making substantive changes to the PL task, which is why we used responses from both this version and the final version to guide the coding scheme development. Using these more effective tasks, the first and second authors qualitatively analyzed larger sets of responses from version 3 (N=61) and the final version (N=120) of the PL task (Table 8.1). By analyzing more responses at this stage, we hoped to compile a more exhaustive list of the ways students responded to our question and use that information to build an analytic rubric. To capture the presence

or absence of the ideas in a student's response, we initially used an analytic approach, rather than a holistic one, which would require assigning a single code to each response as a whole. However, once the final analytic rubric was determined, it was used to develop holistic codes based on different combinations of the presence or absence of the key ideas included in the explanations. During development of the analytic rubric, the authors were aware of the course background of the students they were coding to provide context for the ways in which the students answered the task. In this process, the first and second authors independently analyzed the responses in sets of 30, meeting afterward to discuss the ideas students included in their explanations and how those ideas relate to the CMR framework. Drafts of the analytic rubric were continually modified based on discussions between the first and second authors, as well as the input of our larger interdisciplinary team, and the iterative process continued until we felt that we had accounted for all the key ideas relevant to constructing a causal mechanistic explanation, of which there were three. At this point, we developed the holistic scheme characterizing if the student included some, all, or none of the key ideas in the analytic rubric. We recognize the importance of the researchers' roles in decision-making about codes, so we have included a more detailed discussion of the analytic rubric development in Appendix B.

Once the analytic rubric and holistic scheme had been finalized, the first and second authors used this approach to code the 181 responses used for coding scheme development. To determine inter-rater reliability we calculated Cohen's kappa for our initial coding.<sup>152</sup> For the analytic bins, Cohen's kappa ranged from 0.859 to 0.945, and for the holistic codes Cohen's kappa was 0.873. All of the Cohen's kappa values were greater than 0.8 suggesting high (almost perfect) agreement.<sup>153</sup>

## **Results and discussion**

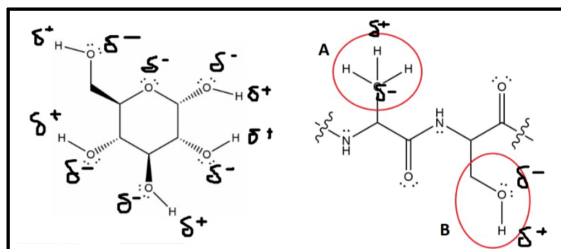
### ***PL task development***

In this section, we discuss the key features of each version of the PL task and our analysis of the responses which guided the task revisions. Figures 8.2, 8.3, and 8.4 show abbreviated versions of each task in the body of the manuscript, and images of each full task version can be found in Appendix E.

#### *PL task version 1*

In the initial version of the PL task, we chose to use a glucose molecule as the ligand and asked students a series of questions addressing how and why such a molecule interacts with a hypothetical peptide chain that had both polar and nonpolar regions (Figure 8.2). This protein-ligand pair was chosen because glucose can form multiple hydrogen bonds and is introduced in most introductory MB courses, so students are familiar with its structure. To activate students' resources related to the electrostatic nature of the interaction, we first asked students to draw the partial charges present on a glucose molecule and explain why they drew the charges in those locations. We also included a hint asking students to consider the role of the subatomic particles to lead them to think about the entities a scalar level below. This process was then repeated for a hypothetical peptide chain featuring an alanine and serine amino acid. Then, students were asked to predict which amino acid would interact most strongly with glucose and explain why the interaction was stronger (Figure 8.2).

[Below], a glucose molecule is shown. Please draw any appropriate partial charges in this molecule.



**Why did you draw the charges where you did in the glucose molecule? Hint: think about the role of the subatomic particles.**

*I placed partial negative charges around the oxygen molecules because they are highly electronegative, and partial positives on the H's because they are bonded to the oxtgen [sic] atoms.*

**Here is part of a peptide showing the two amino acids A and B. Please draw any appropriate partial charges in the side chains (circled) of each amino acid.**

*[Drawing shown in black rectangle above]*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The top partial positive charge is spread out amongst the hydrogens, because they are attached to a carbon which is bigger and therefore more electronegative. On B, the partial negative is on the oxygen and the partial positive is on the Hydrogen. because the oxygen is highly electronegative.*

**Which amino acid would glucose more strongly interact with?**

*B*

**Why does the glucose interact more strongly with that amino acid?**

*B because glucose can form hydrogen bonds with the oxygen.*

Figure 8.2. The response from Isabella to PL task version 1.

As we directly asked the students about the location of the charges, it was not surprising that all the students mentioned charge somewhere in their response. However, only about half of the students (OC1,  $N=10$ , 50%; MB,  $N=9$ , 45%) explicitly leveraged the charges to explain the attraction between their selected amino acid and glucose. One potential issue is that some students (OC1,  $N=6$ , 30%; MB,  $N=8$ , 40%) only named the attractive force (i.e., hydrogen bonding) without explicitly linking the interaction to the charges present. For example, Isabella (Figure 8.2) assigned the appropriate partial charges to the alcohol functional group in serine and glucose but explained that glucose would interact more strongly with side chain B “because glucose can form hydrogen bonds with the oxygen”. Without explicitly connecting these ideas, it is unclear if this student understood the role of charges in hydrogen bonding, especially since they also drew partial positive and negative charges on the alanine side chain. This made

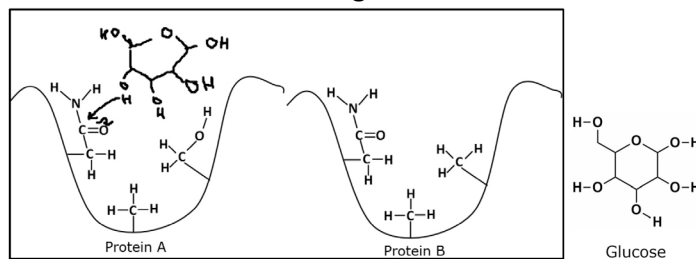
it difficult to determine if the students truly understood the electrostatic nature of this interaction or if they had simply memorized that hydrogen bonds occur between hydroxyl groups—a strong possibility as many studies have identified noncovalent interactions as a difficult idea for students to understand.<sup>154–</sup>

<sup>158</sup> These findings led us to believe that the task overemphasized the role of charges, which may have led students to discuss charges because we asked them to and not because they viewed those ideas as relevant in this context.

#### *PL task version 2*

To more appropriately cue ideas related to electrostatics, in PL task version 2 we removed the questions which asked about the locations of the partial charges in each molecule. Instead, we asked the students to draw glucose binding to one of two potential binding sites and explain (1) why their selected site had the better binding site and (2) what caused glucose to bind to the site (Figure 8.3). By reducing the scaffolding, we could see if binding alone was enough to activate students' resources related to electrostatics. We also tested the order of the two explanation questions, creating an alternate PL task version 2 (see Appendix E) in which we asked for the cause of the protein-glucose binding before we asked why glucose preferentially binds to one site.

**Pick the binding site you think is most likely to bind glucose (shown [below]) and draw a possible way glucose could bind in the binding site.**



**Explain why the protein you chose has the better glucose binding site and how the differences in the site cause this difference in binding. Include specific references to the figures.**

*Protein A has a better glucose binding site because it has an additional OH group. This additional OH group can be added onto the glucose ring so that the OH group can help the biological system because O is necessary in these systems.*

**Use your drawing to help explain what causes glucose to bind to the protein.**

*Glucose binds to the protein because the OH groups are able to attack at the carbonyl and break the double bonded O. This creates a single bonded O and an OH group to the C. This helps for the ring to break and glucose to attach to the protein.*

Figure 8.3. The response from Katrina to PL task version 2.

We also changed the hypothetical binding sites in the task to include 3 amino acids, with both sites containing at least one polar amino acid (asparagine). As both sites could now interact with glucose, students would have to compare the relative strengths of the interactions with each site to determine which site would preferentially bind glucose. We hoped this would lead more students to discuss the relationship between the magnitude of the charge and the strength of the interaction rather than solely stating that one site could interact with glucose and the other could not.

We found that the ordering of the two explanation questions did not impact the students' responses; instead, the broader changes to the task had a more notable impact. Specifically, the reduction in scaffolding related to electrostatics caused fewer students (OC2,  $N=10$ , 50%; MB,  $N=1$ , 5%) to include charge in their explanations. Without appropriate activation from the task, the students' course enrollment (OC2 or MB) appeared to determine which resources were activated. For example, the majority of OC2 students ( $N=14$ , 70%) approached this phenomenon as a reaction (instead of an interaction) in which the oxygen in glucose acted as a nucleophile to attack the carbonyl carbon in

asparagine (see Katrina's response in Figure 8.3). While not appropriate in this context, trying to predict how different species react with one another is a reasonable strategy for an organic chemistry course, in which reactivity and reactions are strongly emphasized.

Unlike the OC2 students, the MB students addressed the noncovalent binding of glucose to the protein. However, only one MB student discussed the role of charges in the binding. A few more MB students (N=5, 25%) used polarity instead of charge to describe the properties of certain groups. The majority, however, (N=13, 65%) named the noncovalent interaction, typically hydrogen bonding, without leveraging charge or polarity. Even though students identified that more hydrogen bonds formed with protein A, they did not explain *why* the atoms formed these interactions, so it is unclear if the students understood the electrostatic nature of this interaction or if they used a memorized heuristic (e.g., hydroxyl groups form hydrogen bonds). We highlight two students who discussed hydrogen bonding in Appendix F. It is possible that some MB students were using resources related to electrostatics when reasoning through noncovalent interactions; however, without the appropriate framing for the task, students may have felt that simply identifying the presence of hydrogen bonding was a sufficient answer. These themes in the OC2 and MB student responses suggest that the task (1) did not activate the appropriate resources related to electrostatic ideas necessary to explain this phenomenon, or (2) did not cue the students to include those ideas in their explanations.

#### *PL task version 3*

In the third version of this task, we tried to activate relevant electrostatic resources in a different way by replacing glucose with a positively charged magnesium ion ( $\text{Mg}^{2+}$ ), a biologically relevant metal ion (Figure 8.4). By changing the ligand to  $\text{Mg}^{2+}$  students could no longer use the ligand as a nucleophile and, also, it would no longer be sufficient for students simply to identify "hydrogen bonding" as the reason for binding. As we found no impact from the order of the explanation questions, in version 3 we used the order of the questions in the original PL task version 2 and arranged the

questions so the students could see both questions at the same time. In this iteration, we tested an alternate version of the activity where we replaced the asparagine, instead of serine, with alanine in protein B to identify which combination of amino acids best elicited causal mechanistic responses (see both versions in Appendix E).

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

**Explain why the protein you chose has the better magnesium binding site and how the structural differences in the site cause this difference in binding. Include specific references to the figures.**

*The magnesium ion has a +2 charge and will be attracted to the 2 slightly negative oxygen molecules that are present in protein A. The negative charge provided by a single oxygen molecule in protein B is not as strong as the charge in protein A.*

**Explain what causes the magnesium ion to bind to the protein making specific references to your drawing of the magnesium ion in the binding site.**

*The magnesium ion will form an ionic-dipole interaction with the slightly negative oxygen molecules in protein A. Electrostatic forces will keep the magnesium ion at the binding site until there in [sic] an introduction of energy into the system.*

Figure 8.4. The response from Conor to PL task version 3 alternate.

As with version 2, we found no difference between the original and alternate versions of the activity (which varied the amino acids in protein B); instead, the change of ligand appeared to have the largest impact to the students' responses. With this version, about half ( $N=16$ , 55%) of the students explicitly identified the role of charge in the binding of  $\text{Mg}^{2+}$  to the protein, and 34% ( $N=10$ ) of the students both identified the role of charge *and* used the strength of the charge to explain the preferential binding (example in Figure 8.4). Additionally, only a few students ( $N=5$ , 17%) named the noncovalent interaction without including evidence of the role of electrostatics. Other students



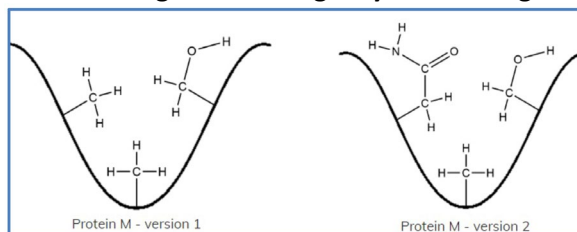
provided evidence that was not explicitly electrostatic in nature, such as the size of the groups in each protein or the number of bonds present in each site, to explain the binding of the ligand.

On the basis of these responses, we felt we had reached an appropriate amount of scaffolding. By switching from glucose to  $\text{Mg}^{2+}$ , we appeared to activate the resources related to electrostatics for those students who saw electrostatics as relevant to protein-ligand binding and encouraged them to include those ideas in their explanation. This change also avoided the unresolved issue of whether students were using the idea of hydrogen bonding appropriately, and at the same time, reduced the scaffolding so that not all students felt forced to discuss charge.

#### *PL task final version*

Having struck (we believe) a good balance of scaffolding, the final version of the PL task (Figure 8.5) has the same major features as version 3. In addition, we included two lessons learned from the original and alternate versions of the second and third iterations of this activity. First, we used the ordering of the explanation questions from the alternate version 2, in which the students explained what causes the protein-ligand binding before answering why their selected site preferentially binds  $\text{Mg}^{2+}$ . Although the ordering did not result in major differences in student responses, we used this order because it aligns better with the CMR framework since the student identifies and unpacks the properties and behaviors of the entities first (in explaining what causes the protein-ligand binding) before then using those properties and behaviors to explain why one site better binds the ligand. Second, we decided to use the distractor protein amino acids from the alternate PL task version 3 featuring two alanine residues and a serine residue. We chose this version because both binding sites contain the amino acid serine, which has a hydroxyl functional group. By having this functional group in both sites, we hoped to further reduce the number of students focusing on the presence or absence of hydrogen bonding as the sole reason for the binding of  $\text{Mg}^{2+}$ .

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



In the space below, explain what causes the magnesium ion to bind to the protein making specific references to your drawing.

*[space for student response]*

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

*[space for student response]*

Figure 8.5. The final version of the PL task.

### ***PL coding scheme development***

In developing the coding scheme, we leaned on work done by Jescovitch et al. in which a combination of analytic and holistic approaches were used to characterize explanations of a phenomenon.<sup>159</sup> Our goal was to characterize students' responses holistically, but, in order to do so, we first developed an analytic rubric to capture the presence or absence of specific ideas in the students' explanations. From the analytic rubric, we identified three key ideas which, when taken together, represent a fully causal mechanistic explanation. These ideas, while related, are not dependent on each other, and thus, we can use the analytic rubric as a means of identifying the combinations of conceptual pieces in students' reasoning. Thus, rather than attempting to make sense of the entire response all at once to assign a single holistic code, we could use the more structured analytic rubric to characterize the response by capturing the presence or absence of each of these three ideas. The resulting combinations of ideas in the analytic rubric can then be used to form a holistic scheme consisting of three codes: non-causal mechanistic (CM), partially CM, or fully CM. When a response provides evidence of an understanding of all three ideas in the analytic rubric, it is given the holistic code fully CM. In previous studies, some of the authors have characterized explanations about chemical phenomena as causal,

mechanistic, or causal mechanistic;<sup>7,146,160,161</sup> however, the scheme we developed for this task emerged from this set of student responses and discussions with our interdisciplinary team, resulting in an approach that is distinct from (and should not be compared to) previous publications.

#### *Developing the analytic rubric*

We developed the analytic rubric based on emergent themes from our analysis of the students' responses and deductive themes based on the CMR framework.<sup>130</sup> That is, we leaned on the CMR framework and the principles of ECD to define what ideas are required for a fully CM explanation of  $\text{Mg}^{2+}$  binding preferentially to a protein site, but also tailored our rubric to be reflective of the knowledge pieces that students provided in their responses.

In the initial stages of development, we tried to capture as many ideas as possible with a range of analytic "bins". For example, does the student mention oxygen? Do they say oxygen is negative? Do they mention electrons? Do they identify a noncovalent interaction? Do they identify polar groups? During discussions between the first and second authors, both the bins themselves and the options for each bin were frequently refined, eventually resulting in our final analytic rubric which, after its use, can then be used to code responses holistically through the lens of CMR. Several decisions, such as whether to combine, refine, or remove conceptual bins, were made in this process (see Appendix B); however, we want to be clear that while the resulting rubric is one way that we feel appropriately captures the ideas relevant to CMR in the context of protein-ligand binding, one might envisage others that are also effective.

Here, we present our analytic rubric and how it can be used to code responses holistically in terms of CMR, followed by a description of, and example responses for, each of these codes. As discussed earlier, a fully CM explanation for this phenomenon involves identifying the (partially) negative atoms in the binding sites, their attraction to  $\text{Mg}^{2+}$ , and the idea that one site is more negative

and therefore more strongly attracts the  $Mg^{2+}$ , causing the preferential binding. This fully CM explanation can be deconstructed into three key ideas which make up our analytic rubric: an understanding of (1) *the attraction between oppositely charged species*, (2) *the negative or polar nature of atoms and amino acids in the binding sites*, and (3) *the larger number of negative entities in one version, causing a stronger attraction to  $Mg^{2+}$* . These three ideas make up the analytic rubric specific for this task, but they also encompass the more general ideas laid out in the CMR framework.<sup>130</sup> We use the analytic rubric as the first step of characterizing a response by designating “yes” or “no” to each category depending on the presence or absence of that particular idea in the response. For example, a response may explain the attraction of oppositely charged species (“yes” for the first idea), but not identify a lower-level charged entity (“no” for the second idea), nor compare the magnitude of the charge between sites (“no” for the third idea). In another response, the combination may be “no” “yes” “no”, for each respective bin. Once we have identified the combination of relevant ideas the response includes (based on this analytic rubric), it is used to assign a holistic CM code. During the process of finalizing this rubric, we encountered several “edge cases” or responses that fell right on the line between receiving a “yes” or a “no”. We discuss these edge cases in Appendix B.

### ***The holistic scheme***

As noted, a response could show the presence or absence of any of the three key ideas based on the evidence provided. To this end, leaving the categories independent of each other both (1) allows us to identify where students may be struggling to construct a complete CM response and (2) provides a more structured method of coding. To characterize each student’s explanation, we used the analytic rubric to keep track of how many ideas (of the three that we defined) they included, ultimately resulting in three holistic codes: non-CM, partially CM, and fully CM (Figure 8.6). Figure 8.7 and the following subsections provide examples and more detailed descriptions of the different codes.

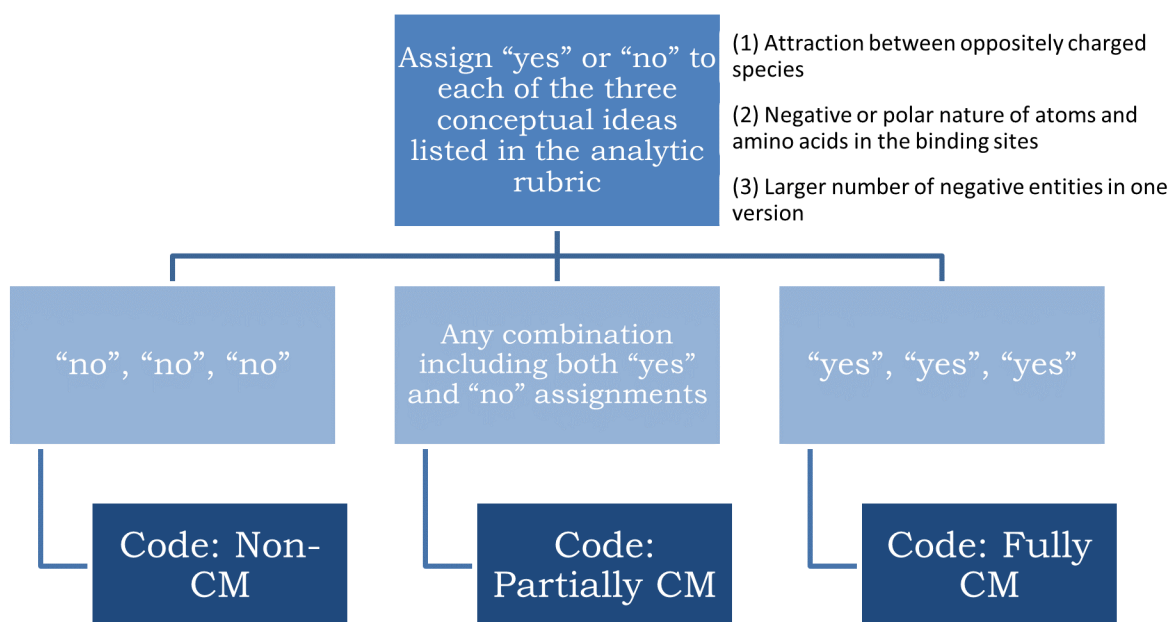


Figure 8.6. The process of using the analytic rubric to assign a holistic code for each response.

*Non-CM.* A non-CM explanation is one that does not provide evidence of any of the key ideas in the analytic rubric. For example, Claudia wrote, *"I believe proten [sic] B has the better binding site. Protein A already has a lot binded to it unlike the left structure on protein B. So i believe it will bind to that structure to sort of mimic protein A's structure. Im [sic] not totally sure about this. However I believe it will choose the arealeast [sic] with the least amount of bonds already in order to create more and equal it out."* Claudia focused on a physical aspect of the binding site, noting that the site they chose had "the least amount of bonds". Although it is an interesting heuristic, identifying the number of bonds in each site is not productive in explaining this phenomenon and does not provide evidence of an understanding of electrostatics, so this receives "no" for each idea captured in the analytic rubric and, therefore, is coded as non-CM.

Student Text Response	Student Drawing	Type of explanation
<p>"This protein has the better magnesium binding site because its easier to bind and the structural differences in the site causes a difference because it makes the structure longer. The extra carbon helps the magnesium ion to bind to the protein" -Teddy</p>	<p>Protein M - version 1</p> <p>Protein M - version 2</p>	Non-causal mechanistic
<p>"I picked version one to bind <math>Mg^{2+}</math> to O. This is because O is usually negatively charged, so there is a high chance that it will bind to Mg which is positively charged. Oxygen is negatively charged. According to Coulomb's law, opposite charges attract. This is why I think <math>Mg^{2+}</math> will bind with O" - Simone</p>	<p>Protein M - version 1</p> <p>Protein M - version 2</p>	Partially causal mechanistic
<p>"Protein A has a better binding site because the there are two polar molecules with partial negative charges compared to only one polar molecule in Protein B. Intermolecular forces bind the magnesium ion to Protein A. The positively charged magnesium ion interacts with the partial negative oxygens in the two polar molecules within the binding site through ion-dipole interactions." -Lois</p>	<p>Protein A</p> <p>Protein B</p>	Fully causal mechanistic

Figure 8.7. Examples of student engagement in causal mechanistic reasoning.

*Partially CM.* A partially CM response includes any combination of “yes” and “no” for the ideas in the analytic rubric (i.e., it has one or two of those ideas, but not all three). For example, consider

Wayne's response in Figure 8.8, which was given the combination "yes" "yes" "no" from the analytic rubric. Wayne provided an appropriate explanation of the role of electrostatics in causing  $\text{Mg}^{2+}$  to bind; however, rather than using electrostatics to explain the preferential binding, they invoked an additional resource saying, "Version 1 has more space around the primary alcohol on the rightmost amino acid, allowing for the Magnesium to have an easier time binding to the site." Again, this other resource (shape/accessibility) may be a useful heuristic; however, it is not what causes the preferential binding and leads Wayne to the incorrect answer.

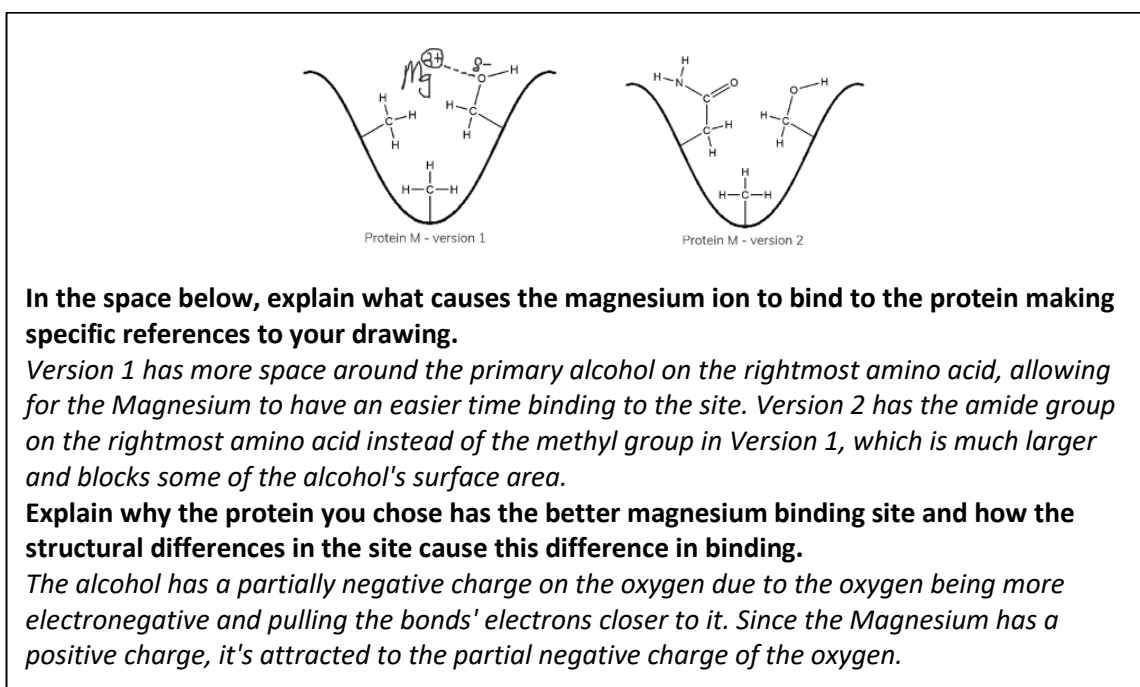


Figure 8.8. Wayne's drawing and text responses.

*Fully CM.* Responses are coded as fully CM explanations for this phenomenon by providing evidence of all three key ideas in the rubric. For example, Triston wrote, "Protein 2 has a better magnesium binding cite [sic] because there is a larger partial negative charge due to having two oxygens present. Protein 1 only has one oxygen to give a partial negative charge which will not attract nor stabilize the Mg ion as well. The protein 2 that I chose has the better binding cite [sic] because it has more partial negative (the two oxygen containing molecules) that attract the positive charge on the Mg

2+ ion.” Triston leveraged electrostatics to not only explain the cause of  $\text{Mg}^{2+}$  binding, but also to explain why the version they chose is the better binding site.

Some students discuss the lack of polar groups in one version or the presence of hydrophobic/nonpolar groups, making it a poor binding site for  $\text{Mg}^{2+}$  (see Appendix B for our decisions regarding responses that discuss polarity instead of charge). For example, Jordan said “*Version 2 has the better magnesium binding site because there are uneven distributions of charges (C=O, O-H) in some of the side chains that allow for the positively-charged magnesium ion to bind with the negative charges of the oxygen atoms. Version 1 has very hydrophobic side chains (C-H) so they would shy away from the charged ion.*” Jordan did not explicitly say that one site is more negative than the other, but they noted that version 2 has negative charges to bind  $\text{Mg}^{2+}$ , while version 1 has hydrophobic groups which “*would shy away from the charged ion.*” While this response does contain an extra idea that is not strictly correct (i.e., the idea that hydrophobic groups are repelled by the ion), we believe Jordan did enough in comparing the charged nature of the two sites. Ultimately, the students constructing fully CM responses link their ideas about electrostatics to the phenomenon of  $\text{Mg}^{2+}$  binding preferentially to one of the sites because it is more negative and, therefore, more strongly attracts the ion.

### **Limitations**

This study was designed and carried out primarily in the context of chemistry and biology courses which emphasize the importance of constructing causal mechanistic explanations for phenomena and give the students the opportunity to do so on assessments. It is probable that in another setting where the curriculum does not focus on these types of explanations that student responses to the task would be different. Additionally, we recognize that the number of responses analyzed for task versions 1 and 2 are somewhat low ( $N=40$ ), and it may be that analyzing more responses would have resulted in a different final task. However, this report is intended to show how we



compromised to find a process that was feasible in a reasonable amount of time. That being said, the overall task and coding scheme design still took almost two years.

### **Conclusions and future work**

This paper describes the process by which we developed (1) a task to elicit causal mechanistic explanations of how species bind in a simplified biological system and (2) an associated coding scheme to make sense of those explanations. To do this, we drew upon literature related to evidence-centered design, the resources perspective, scaffolding, and causal mechanistic reasoning. Designing tasks to elicit rich explanations about how and why phenomena occur is difficult; we hope that the lessons we learned during this process can help other instructors and researchers to design effective assessments to collect evidence about what students know and can do.

In accordance with ECD, we first determined what we want students to know about this phenomenon and what work products we would accept as evidence of such an understanding.<sup>137</sup> These evidence statements allowed us to identify which resources the task must activate (e.g., electrostatics), which then influenced the design of the initial task. We iteratively refined the task three times, guided by our analysis of the students' responses, to better activate productive resources in the students' minds by modifying the scaffolding. This required us to reflect, carefully considering: (1) what resources students used to construct an explanation of the phenomenon, (2) what aspects of the current task activated those resources, and (3) how we might modify the scaffolding of the task to activate the appropriate resources. This process was not trivial, and the iterative design of the task was crucial to the development process, ensuring that the responses elicited most accurately reflected the students' understanding.

In comparing iterations, we found that seemingly small changes to the task (i.e., switching the ligand from a glucose molecule to a  $\text{Mg}^{2+}$  ion) dramatically changed how the students responded to the

task. This highlights how aspects of the task can activate a particular set of resources, either productive or unproductive, that the student then uses to reason through the phenomenon. In our case, we found that the association between alcohol functional groups and hydrogen bonding is so strong that, in order to elicit explicit explanations of electrostatics, we had to change the ligand. Species such as glucose are still important molecules however, and cannot be avoided forever, so additional research is needed to explore how students understand hydrogen-bonding interactions.

In addition to activating the appropriate resources, we found that the framing of the task impacted what students chose to include in their written explanations. For example, we wanted students to explain *how* and *why* the entities interacted with one another instead of just describing *that* they interacted with one another, so we modified the task to cue the students to provide such information. However, we needed to include just enough scaffolding to (1) activate the productive resources necessary for CMR and (2) provide enough information about which ideas we expected students to include in their response, without overspecifying.

To design this task, we had to balance multiple factors. For example, instead of probing all the ideas relevant to protein-ligand binding (such as shape or entropy), we chose to ask students about a simplified version of this phenomenon centered around the role of electrostatic attraction. We hope in the future we can ask students, particularly in upper-level courses, about more complex phenomena, but we know students struggle to understand the electrostatic basis of these noncovalent interactions.<sup>146,154–156</sup> These ideas are hard, and it would not be productive to overwhelm the students for the sake of scientific accuracy. At the same time, we were also cognizant of the potential to overcue students, which may encourage more rote learning behaviors.<sup>162</sup> How then can we provide just the right amount of support to students? It is not easy to strike this balance, but we argue that iterative design and being responsive to what our students say and do should play an important role in this process.

Once the final task version was determined, we began developing a coding scheme including both analytic and holistic components which can be used to characterize responses. We found that beginning with an analytic rubric allowed for more structure and detail in the initial stages of characterizing the data. The final analytic rubric, which was the result of several iterations, consists of three key ideas which were distilled both from our ideal causal mechanistic explanation as well as the ideas students included in their responses. From these three key ideas, holistic codes emerged based on evidence of the number of key ideas in the response. Thus, we could characterize responses as non-CM, partially CM, or fully CM in this context of protein-ligand binding. This process of designing a coding approach with both analytic and holistic aspects was insightful, and the approach will be used to characterize chemistry and biology students' responses to the PL task in forthcoming publications.

Constructing and evaluating tasks in this way is philosophically different than many approaches to designing assessments, in which students may not be provided with scaffolded cues. Our goal is to determine what it is that students know and can do, rather than letting them rely on memorization or heuristics which may not be backed by a robust understanding of the underlying concepts. Designing and using such tasks is crucial if we are to help students to go beyond rote memorization of isolated facts; we need to both understand how students engage in sophisticated forms of reasoning (like CMR) and provide opportunities on assessments for students to use those types of reasoning. Furthermore, by including questions which invoke CMR on assessments, we send a message to the students that understanding *how* and *why* phenomena occur is important and valued.<sup>163</sup>

This work is a step toward our larger goal of supporting students' interdisciplinary understanding of science and their ability to use CMR. While in this paper we share a developed task and coding scheme to assess explanations of protein-ligand binding, our team has also developed materials in the contexts of protein structure-function and phenotypic variation which we report out in future publications.<sup>164</sup> In forthcoming studies, we plan to use these tasks and coding schemes to explore

how undergraduate students respond at different time points in their science degree programs and in the context of different disciplines. By using these tasks, we hope to gain a better understanding of how students may or may not be engaging in CMR and to develop more effective ways to support CMR both within and across disciplines, with the ultimate goal of providing opportunities for students to make meaningful connections between chemistry and biology and develop more powerful reasoning strategies in science.

## **APPENDICES**

## APPENDIX A: Permissions

3/21/22, 6:24 PM

Rightslink® by Copyright Clearance Center



Home



Help ▾



Email Support



Keenan Noyes ▾



### A Deep Look into Designing a Task and Coding Scheme through the Lens of Causal Mechanistic Reasoning

Author: Keenan Noyes, Clare G. Carlson, Jon R. Stoltzfus, et al

Publication: Journal of Chemical Education

Publisher: American Chemical Society

Date: Feb 1, 2022

Copyright © 2022, American Chemical Society

#### PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)

[CLOSE WINDOW](#)

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)  
Comments? We would like to hear from you. E-mail us at [customer@copyright.com](mailto:customer@copyright.com)

Figure 8.9. Permissions to reproduce manuscript in its entirety.

## **APPENDIX B: Analytic rubric development and application**

As noted in the body of the manuscript, the coding scheme development was an iterative process requiring several rounds of application, modification, and refinement. It is considered best practice to be as transparent as possible regarding development and application of coding schemes.<sup>131</sup> Qualitative data is indeed messy, and analysis of such data requires the researchers to make decisions which greatly impact the results; therefore, we discuss several of our decisions here, in the hopes of convincing our readers that our resulting coding scheme is valid, reliable, and reflective of the data. However, we recognize that had others been involved using a different theoretical framework, the resulting coding scheme may have been different.

Although our goal was to characterize explanations holistically, assigning a single code to the response, we drew on work from Jescovitch et al. and used an analytic approach to support the assignment of holistic codes.<sup>159</sup> By using an analytic approach initially, we could capture a wider range of ideas in the students' responses. After the development of our final PL task was complete, the first author created an initial analytic rubric based on the themes and patterns of the students' responses he uncovered during his analysis of the responses guiding the earlier task development. These "bins" served as a starting point for the rubric development:

1. Is the metal drawn binding to the correct protein version?
2. Is the metal drawn interacting with the oxygens?
3. Do they mention an intermolecular force?
4. Do they mention charges?
5. Do they describe the attraction between positive and negative entities?
6. Do they indicate that oxygen is electronegative?
7. Do they mention electrons?
8. Do they relate binding back to the number of covalent bonds?

9. Do they describe any of the groups as polar/hydrophilic?

10. Do they discuss the energy changes associated with binding?

At this point, author C.G.C. joined author K.N. in the analysis process to refine the analytic rubric. The two authors independently analyzed 181 students' responses in sets of approximately 30 responses (Table 8.1 in the manuscript body), meeting to discuss the responses and revise the bins to better characterize the data. Initially, the options for each bin were "yes", "no", or "unclear". Typically, the presence of many "unclear" responses suggested that the criteria for receiving "yes" or "no" needed to be revised. As the result of this analysis, all the bins were either modified, condensed, or removed.

For example, we eventually removed the two questions pertaining specifically to the drawing. Instead, we revised the criteria for all the bins so that receiving "yes" or "no" was based on an overall understanding of both the students' text and drawing responses. Further analysis led to the removal of bins 3, 4, 8, and 10. We removed the bin "Do they mention charges", because it overlapped with the bin "Do they describe the attraction between positive and negative entities". Bins 8 and 10 were removed because few students discussed energy changes or the number of covalent bonds, and because these ideas were not necessary for what we deemed a causal mechanistic explanation. While they include interesting ideas and resources that we may investigate in future publications, their use is not the focus of this paper.

One significant point in the development of the analytic rubric was identifying what entities would qualify as going a "scalar level below", as defined by the literature on causal mechanistic reasoning.<sup>130,132</sup> To explore this, we added a bin to capture "what is negative?" We expanded the possible options for this bin to capture the scalar level the student referenced: "electrons", "atom", "functional group", "amino acid", or "entire binding site". After reflecting on the task, we defined the phenomenon under consideration (protein-ligand binding) as occurring at the "protein level" and, thus,



the immediate level below that would include atoms, functional groups and amino acids – since these are the entities which make up the protein and give the protein its properties and behaviors. However, prior to this decision, we considered electrons as the appropriate level. This is likely because authors K.N and C.G.C are chemists, but after discussions with our interdisciplinary team, as well as our observations in the data (e.g. few students went to this lower level of electrons), we decided that atoms and amino acids would suffice as the “scalar level below”. Thus, our final analytic rubric does not include the bin “do they mention electrons”.

As we refined our rubric, the bins increasingly focused on the presence of ideas related to charge and polarity in the students’ responses. The six bins characterizing these ideas included:

- Do opposite charges attract? (yes, no)
- What is negative? (atom, functional group, amino acid, entire binding site)
- Do they explain why that entity is negative? (yes, no)
- Does a polar entity attract  $Mg^{2+}$ ? (yes, no)
- What is polar? (atom, functional group, amino acid, entire binding site)
- Do they explain why that entity is polar? (yes, no)

We added the polarity bins because these ideas were showing up more frequently than we originally expected in students’ responses. Initially, we were not convinced that only discussing polarity (in the absence of explicit mention of charge) would be sufficient for a CM explanation; however, both the data and our discussions with our interdisciplinary team led us to consider polarity as an appropriate resource in explaining this phenomenon. First, many students who discussed polarity also discussed charge. For example, Andy wrote *“This protein version 2 has the better magnesium binding site because the molecule I drew it interacting with is polar and will have partial charges on each side, which allows magnesium (a charged ion) to be able to interact with it...”*. Similarly, Jessica wrote, *“I chose version 2*

*because it has more polar molecules with more electronegative atoms with partial negative charges which would attractive the positively charged Mg more."* Second, as our biology colleagues reminded us, polarity is a frequently used and productive resource when reasoning through biological interactions, due to the scale at which these interactions occur; it is a more concise way to identify uneven charge distributions within macromolecules and does not require explicit attention to the relative locations of electrons. It is true that a positive ion would not bind just anywhere on a polar substance, because it would only be attracted to the partially negative atoms, however, we decided that identifying an amino acid (or functional group) as polar or nonpolar and recognizing the attraction of ions to polar entities was a sufficient and reasonable explanation for this phenomenon.

In the final analytic rubric, we combined the bins involving charge and polarity because both were acceptable resources in the explanation of  $Mg^{2+}$  binding. The analytic bins which ask if the students explain *why* that entity is polar or negative aimed to capture explanations which discussed electrons or electronegativity. This bin was eventually removed, since this subatomic level is not required to construct what we have deemed a fully CM explanation. We also changed the "what is negative/polar?" bin, focusing instead on if the student identified a negative or polar entity that is a scalar level below (either an atom, functional group, or amino acid). In the end, we distilled the analytic rubric down to three bins related to charge/polarity which capture the key ideas corresponding to a causal mechanistic explanation of protein-ligand binding: (1) *the attraction between oppositely charged species*, (2) *the negative or polar nature of atoms and amino acids in the binding sites*, and (3) *the larger number of negative entities in one version, causing a stronger attraction to  $Mg^{2+}$* . With this final analytic rubric, there were a handful of "edge cases", responses falling on the line between "yes" and "no", we encountered which we discuss below.

The first edge case we discuss is "the attraction between oppositely charged species". Sometimes, students discussed electrons (typically on the oxygen atom) attracting the  $Mg^{2+}$  ion without

explicitly stating that electrons are negative. This made it difficult to determine if the student understood that there was an attraction between oppositely charged species. In these cases, we assumed that students implied that the electrons were negative given the strong association between electrons and being negatively charged as well as their discussion of the attraction with  $\text{Mg}^{2+}$ . For example, Henry wrote, *“The lone pairs of electrons on the oxygen molecule will be attracted to the 2+ charge on the magnesium...”*. Thus, even though the response does not explicitly state that opposite charges attract, we decided to assume they know the negative nature of electrons (and would receive “yes” for this bin). Based on this assumption, the student could also receive “yes” for an understanding of *the negative or polar nature of atoms and amino acids in the binding sites*, depending on what entity they describe as having these electrons. Henry (example above), for example, would also receive “yes” for this bin since they said the electrons were on the oxygen atom and attracted the positive ion.

We also encountered cases in which students drew interactions between positive and negative entities but did not explicitly write about their attraction. In these cases, we considered the drawing of an interaction between oppositely charged species as evidence of understanding their attraction. Consider the drawing in Figure 8.10 from Penny, who wrote, *“Protein 2 cause the ketone has a partial negative charge. Ionic bond between Mg and the O in the ketone group.”* Although Penny does not provide a detailed enough explanation to decisively receive “yes” for this idea, their drawing shows a clear interaction between a partially negative oxygen and the positive  $\text{Mg}^{2+}$  and, thus, they do receive “yes” for *the attraction between oppositely charged species*.

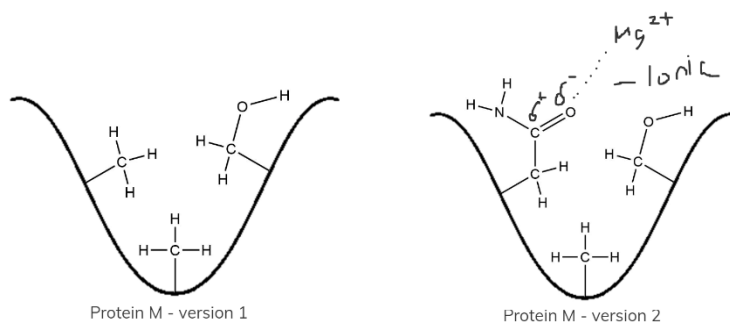


Figure 8.10 Drawing response from Penny.

The second edge cases we discuss is “the negative or polar nature of atoms or amino acids in the binding sites”. We used both the students’ drawings and text responses to determine if they considered a charged entity that is a scalar level below (atom or amino acid). For example, some students did not mention a lower-level entity in their text, but explicitly drew a negative sign on the oxygen atoms, so they still received “yes” for this analytic bin. However, in whatever way they chose to identify the negative entity, either through drawings or their text response, they needed to be explicit. For example, if the student solely stated that *“the difference in charges caused Mg to bind”* they would receive “no” for this bin. Even if they drew magnesium near the oxygen atom they would still receive “no”, because they did not explicitly indicate that the oxygen was negatively charged. Consider David’s drawing (Figure 8.11) and paired text response which said, *“...The difference in charges causes magnesium ions to bind to the protein as it did in my drawing previously.”* To receive “yes” they would have to explicitly draw or discuss the negative charge on that oxygen.

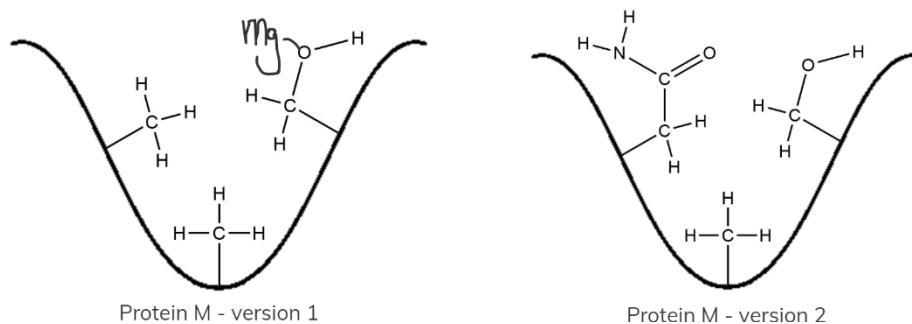


Figure 8.11 Drawing response from David.

The final edge case we discuss is “the larger number of negative entities in one version, causing a stronger attraction to  $\text{Mg}^{2+}$ ”. There were a few different means by which students could receive “yes” for this bin. The most common is students identifying the larger number of oxygens, which are negative, in one site compared to the other. However, there were also students who discussed the lack of polar groups in one site or the presence of hydrophobic groups, which would not attract the  $\text{Mg}^{2+}$ , in one site. As long as the explanation made a clear comparison of charge or polarity between the two sites to explain  $\text{Mg}^{2+}$  binding, then they received “yes” for this bin.

### APPENDIX C: Alternate task administration details

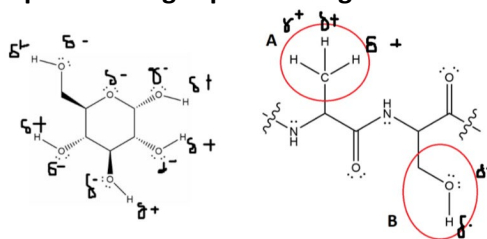
Table 8.2. Details of the administration of the original and alternate versions of the PL task.

PL Task Version	Course	Responses collected	Responses analyzed (task development)	Responses analyzed (coding scheme development)
2 - original	Organic chemistry 2	38	11	-
2 - alternate	Organic chemistry 2	36	9	-
2 - original	Molecular biology	160	13	-
2 - alternate	Molecular biology	153	7	-
3 – original	General chemistry 2	31	15	31
3 - alternate	General chemistry 2	30	14	30

#### **APPENDIX D: Deidentified student responses used in task development**

In this section we provide the full response from every student whose response was used to develop the PL task. To deidentify the students' responses, we used a random number generator to place the students in a random order. We then assigned each student a number starting with "101" and increasing sequentially for each group of students who completed the PL task. To help give further context, we added additional information to the student IDs in the manuscript. This follows the format "task version\_course\_student random id". The possible courses include: organic chemistry 1 (OC1), organic chemistry 2 (OC2), general chemistry 2 (GC2), and molecular biology (MB). For example, the student "V2Alt\_OC2\_105" would be an organic chemistry 2 student who received the alternate second version of the PL task. If the student response was featured in the body manuscript, the assigned pseudonym is given in parenthesis next to the student ID. Responses are copied in their original form, including any grammatical errors. Note that the question stems included have been condensed from their original form (see Appendix E) to conserve space in this section.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Glucose is polar and soluble in water which means that it is attracted to the charged bonds in water through its partial charges on the oxygens.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*As a peptide, teh side chains are attracted to polar molecules and eventually allow the peptide to dissolve. Because of this teh side chains need to be partially charged in order to attract to other molecules and solvents.*

**Which amino acid would glucose more strongly interact with?**

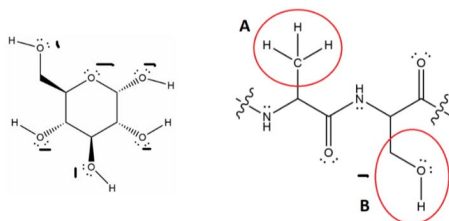
*B*

**Why does the glucose interact more strongly with that amino acid?**

*Glucose interacts more strongly with the OH side chain because of teh side chains polar qualties and oppositely charged molecules. The hydrogen in the peptide side chain can attract to the oxygen atom in the glucose.*

Figure 8.12. Response from student V1\_OC1\_101.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Where ever the elctrons were on the oxygens is where i put the partial negative*

**Why did you draw the partial charges where you did [on the amino acids]?**

*because the carbon is satisfied by forming 4 bonds and a negative on the oxygen because of its electrons*

**Which amino acid would glucose more strongly interact with?**

*A*

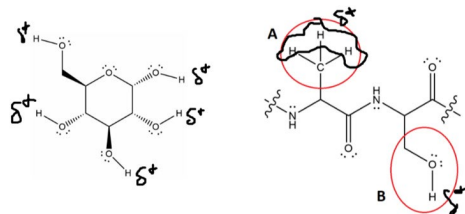
**Why does the glucose interact more strongly with that amino acid?**

*a because it has more oh groups that can form hydrogen bonds so it would have stronger interactions*

Figure 8.13. Response from student V1\_OC1\_102.



[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*The oxygen is an electronegative atom which means it has a strong attraction for electrons. In this picture the oxygen is pulling the hydrogens electrons closer to itself and the h is left with a partially positive charge*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The CH3 has a dispersed partial positive charge while the OH has a strong central partial positive*

**Which amino acid would glucose more strongly interact with?**

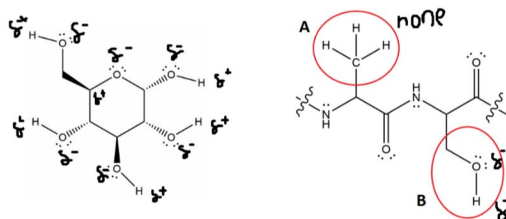
*B*

**Why does the glucose interact more strongly with that amino acid?**

*B because it has more localized partial positives and negatives*

Figure 8.14. Response from student V1\_OC1\_103.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*the effective nuclear charge gives oxygen a high electronegativity, so it will pull electrons towards itself. When the oxygen does this, a partial positive charge is placed on the hydrogens it is connected to, because the electrons (have a negative charge) are being pulled away from the hydrogen.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The carbon is not very electronegative, so it will share the electrons pretty evenly with the hydrogens it is attached to. This results in no partial charges on either the H atoms or the C atom. Since the oxygen is highly electronegative, the electrons in the O-H bond will be pulled toward the oxygen, giving the oxygen a partial negative charge and the hydrogen a partial positive charge.*

**Which amino acid would glucose more strongly interact with?**

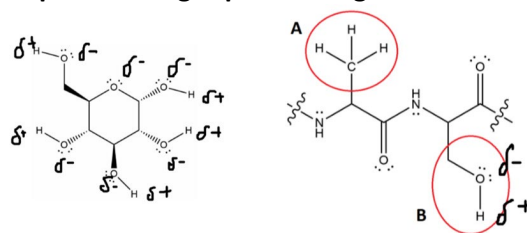
*B*

**Why does the glucose interact more strongly with that amino acid?**

*The glucose has many partial positives (on H atoms) and partial negatives (on O atoms), so they will be attracted to opposite charges (found in the partial positives and negatives on amino acid B) more than neutral charges (found on amino acid A).*

Figure 8.15. Response from student V1\_OC1\_104.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*The partial negative charges on the more electronegative atoms O and the partial positive on the less electronegative atoms H*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The OH group has a partial - and partial + end but the methyl does not due to the lack of separation of electronegativity*

**Which amino acid would glucose more strongly interact with?**

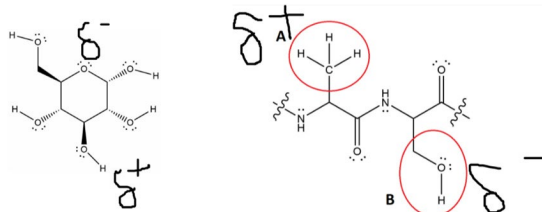
A

**Why does the glucose interact more strongly with that amino acid?**

*The partial negative charge in glucose can react better to the partial positive charge on the H of the OH amino acid because there is a stronger dipole and therefore LDF that can be formed*

Figure 8.16. Response from student V1\_OC1\_105.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*The cyclohexane has equal amounts of OH placement on the right and left sides, then when comparing the top and bottom of the cyclohexane, you notice that the top has the oxygen inside the ring which means that part is partially negative, then the opposite side must be partially positive.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*I drew a partial positive charge on the methyl group because there is a group of protons (hydrogens); however, the OH on the opposite side of the molecule only has one hydrogen then an oxygen with a lot of lone pairs of electrons (negative charges).*

**Which amino acid would glucose more strongly interact with?**

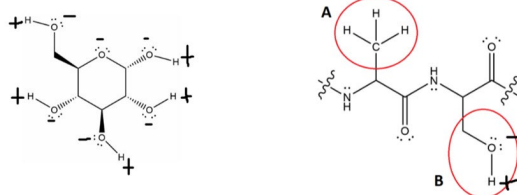
B

**Why does the glucose interact more strongly with that amino acid?**

*Carbons are less likely to have their hydrogens leave when compared to an oxygen because of the electronegativity. The oxygen's high electronegativity allows it to hold most of the electrons (because of the higher electron attraction) which makes the hydrogen that's attached easily available to leave.*

Figure 8.17. Response from student V1\_OC1\_106.

**[Student's drawings of the partial charges present in glucose and the amino acids]**



**Why did you draw the charges where you did in the glucose molecule?**

*the charge depends on the electrons and protons on or near each atom. Oxygen is a more electronegative atom and therefore has more electrons around it than the H. This causes a partially negative charge.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*the methyl group is a neutral group that typically has no charge. the alcohol has a partially negative and postive area. this is due to the osygen being electronegative*

**Which amino acid would glucose more strongly interact with?**

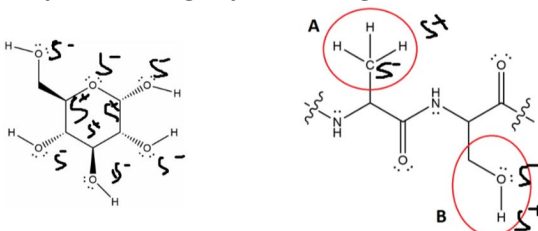
*B*

**Why does the glucose interact more strongly with that amino acid?**

*The glucose will interact more strongly with the alcohol group because it is able to hydrogen bond with it. Hydrngen bonding is a type of interaction between two molecules containing an H bonded to an electronegative atom (O,N,F). This is a strong interaction that takes a substantial amount of energy to overcome. The methyl group is neutral and can only interact via London Dispersion Forces which is a weak interaction, overcame by little energy.*

Figure 8.18. Response from student V1\_OC1\_107.

**[Student's drawings of the partial charges present in glucose and the amino acids]**



**Why did you draw the charges where you did in the glucose molecule?**

*The Oxygen is more electronegative than Carbon so it is polar.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Carbon is more electronegative than Hydrogen. Oxygen is more electronegative than hydrogen.*

**Which amino acid would glucose more strongly interact with?**

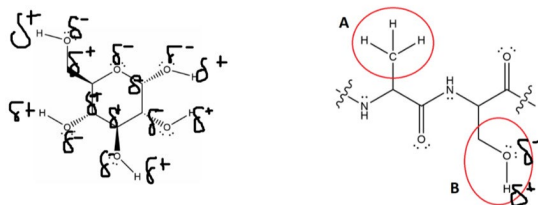
*A*

**Why does the glucose interact more strongly with that amino acid?**

*Because Oxygen is attracted to Carbon because it has a positive partial charge when Oxygen has a negative partial charge*

Figure 8.19. Response from student V1\_OC1\_108.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*i drew partial negatives on the O because they are more electronegative than C and H so they will pull the electrons towards them making H and C more partial positive*

**Why did you draw the partial charges where you did [on the amino acids]?**

*H and C have similar electronegativity so they will share the electrons evenly whereas O is more electronegative than H so it will pull the shared electrons more*

**Which amino acid would glucose more strongly interact with?**

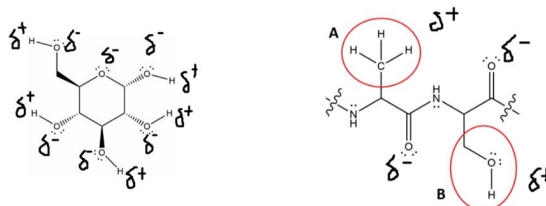
*B*

**Why does the glucose interact more strongly with that amino acid?**

*partial positive from the H will react with the partial negative from the O in glucose and vice versa the O will react with Hs on glucose*

Figure 8.20. Response from student V1\_OC1\_109.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Oxygen is a more electronegative atom, therefore, it is more likely to attract electrons than hydrogen.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The double bonds have highest density of electrons, and they also attach to a electronegative atom, oxygen. Therefore, those are the partial negative charge. For the methyl part, they do not have that many electrons around, therefore, positive charge.*

**Which amino acid would glucose more strongly interact with?**

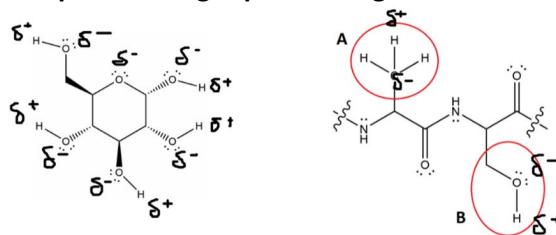
*A*

**Why does the glucose interact more strongly with that amino acid?**

*The partial negative charge tend to attract postive charge. Since glucose is surrounded by negative charges, it is more likely to connect to the positive charge of the amino acid, which is the methyl part.*

Figure 8.21. Response from student V1\_OC1\_110.

**[Student's drawings of the partial charges present in glucose and the amino acids]**



**Why did you draw the charges where you did in the glucose molecule?**

*I placed partial negative charges around the oxygen molecules because they are highly electronegative, and partial positives on the H's because they are bonded to the oxygen atoms.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The top partial positive charge is spread out amongst the hydrogens, because they are attached to a carbon which is bigger and therefore more electronegative. On B, the partial negative is on the oxygen and the partial positive is on the Hydrogen. because the oxygen is highly electronegative.*

**Which amino acid would glucose more strongly interact with?**

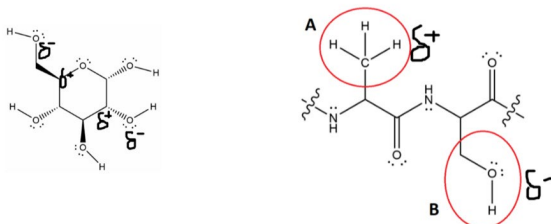
*B*

**Why does the glucose interact more strongly with that amino acid?**

*B because glucose can form hydrogen bonds with the oxygen.*

Figure 8.22. Response from student V1\_OC1\_111 (Isabella).

**[Student's drawings of the partial charges present in glucose and the amino acids]**



**Why did you draw the charges where you did in the glucose molecule?**

*I drew the charges where i did was because the oxygen atom is more electronegative then the carbon and will draw the electrons from the carbon to the oxygen, meaning it will become more negative.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*i drew the oxygen to hydrogen as a partial negative because the oxygen is more negative then the hydrogen which draw electrons to it and away from the rest of the mechanism. Carbon is positive because it is drawing electrons toward the mechanism*

**Which amino acid would glucose more strongly interact with?**

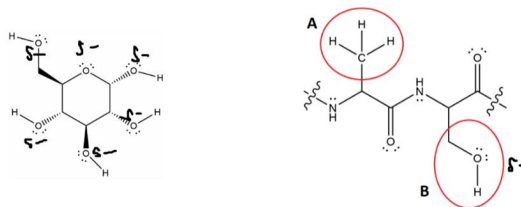
*A*

**Why does the glucose interact more strongly with that amino acid?**

*Glucose would react with the carbon side chain because the carbon would be able to form a bond with oxygen since carbon could be removed and form four bonds with the oxygen with only two lone pairs*

Figure 8.23. Response from student V1\_OC1\_112.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*I drew the partially negative charges around the oxygens because they're the more electronegative atom, meaning it will be able to pull other atom's electrons closer to itself*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The OH is a negative alcohol group. the CH<sub>3</sub> is a non-polar methyl group*

**Which amino acid would glucose more strongly interact with?**

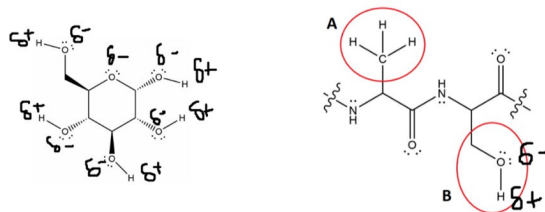
*A*

**Why does the glucose interact more strongly with that amino acid?**

*Because the OH groups in the glucose would be able to interact with the OH groups on the amino acid through hydrogen bonding interactions*

Figure 8.24. Response from student V1\_OC1\_113.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*The oxygen atoms are highly electronegative, and attract the electrons in their respective covalent bonds more strongly than the other atoms in question (hydrogens and carbons).*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The methyl group is fairly equal in how the electrons are distributed between the bonds. The hydroxide group, however, contains an oxygen atom, which has a stronger attraction to the electrons around it than other atoms.*

**Which amino acid would glucose more strongly interact with?**

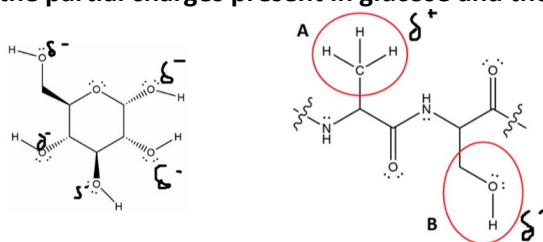
*B*

**Why does the glucose interact more strongly with that amino acid?**

*The hydroxide groups on the glucose would be attracted to the hydroxide group on the amino acid, due to both of them having partial positive and negative charges present. Of course, the oxygens would be attracted to the hydrogens and vice versa. The methyl group is relatively stable, and would not have any partial charges that would interact with the hydroxides on glucose.*

Figure 8.25. Response from student V1\_OC1\_114.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*This is because the oxygen did not have the full amount of bonds making it partially negative.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Because carbon has 3 hydrogen which have a positive charge compared to oxygen with one.*

**Which amino acid would glucose more strongly interact with?**

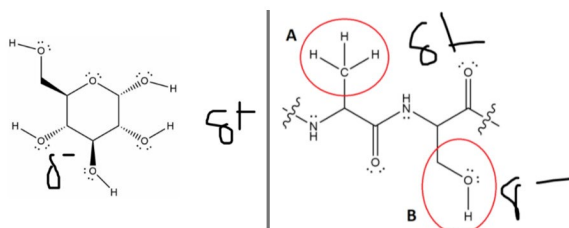
A

**Why does the glucose interact more strongly with that amino acid?**

*Because there would be a stroger attrachion because of the LDF.*

Figure 8.26. Response from student V1\_OC1\_115.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Oxygen is partially negative and hydrogen is partially positive.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The hydrogen molecules in the ch3 molecule is partially positive while the oxygen in the other molecule is partially negative.*

**Which amino acid would glucose more strongly interact with?**

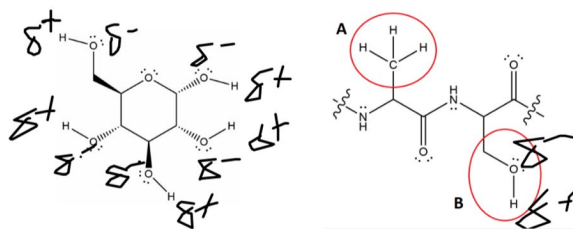
B

**Why does the glucose interact more strongly with that amino acid?**

*This is bvecause the glucose molecule has the partially positive hydrogens on the outside, which would attract the partially negative part of the amino acid (B).*

Figure 8.27. Response from student V1\_OC1\_116.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*oxygen is more electronegative than hydrogen, so th electrons spend more time at the oxygen, giving oxygen a partial negative charge and leaving hydrogen partially positive*

**Why did you draw the partial charges where you did [on the amino acids]?**

*carbon is not much more electronegative than hydrogen, so it will not have a partial charge. however, the oxygen and hydrogen will*

**Which amino acid would glucose more strongly interact with?**

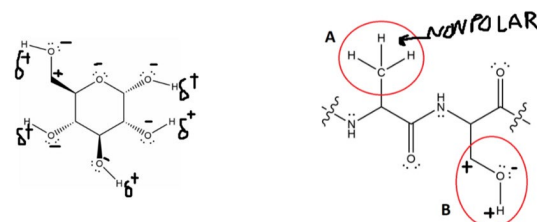
*B*

**Why does the glucose interact more strongly with that amino acid?**

*because of thr strong partial charges. the oxgen will be more attracted to the hydrogen and the hydrogen to the oxygen*

Figure 8.28. Response from student V1\_OC1\_117.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*You have to look at differences in electronegativity and see which element will pull the electrons strongly.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The COH had partial charges because the oxygen has a higher electronegativity. The CH3 is nonpolar because C and H are similar in electronegativity so there is non partial charge. In some instance there is a induced dipole but that would require another CH3 molecule.*

**Which amino acid would glucose more strongly interact with?**

*A*

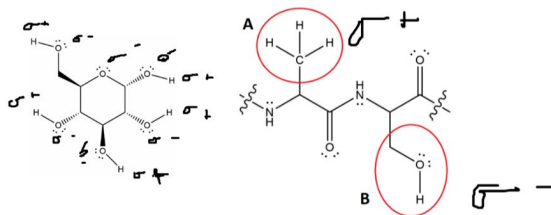
**Why does the glucose interact more strongly with that amino acid?**

*The attraction of opposite charges will allow for interaction.*

Figure 8.29. Response from student V1\_OC1\_118.



[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*O are super electronegative and it's going to pull electrons close to itself.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*CH<sub>3</sub> is non polar and therefore has no charge while OH is electronegative and would have a negative charge.*

**Which amino acid would glucose more strongly interact with?**

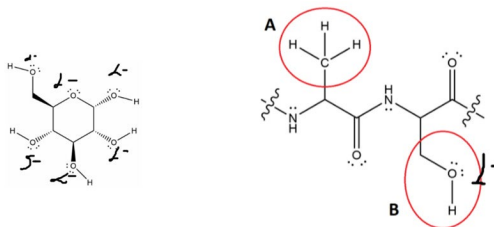
*B*

**Why does the glucose interact more strongly with that amino acid?**

*It would react more strongly because they're both electronegative and bonded with OH which can bond Hydrogen interactions vs A which can only interact through LDFs*

Figure 8.30. Response from student V1\_OC1\_119.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*I drew partial negatives on the oxygen because of how electronegative it and so it should have a partial negative on it*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The OH is still partial negative because of the oxygen however the A part of the molecule is neutral because carbon has four bonds and no free electrons*

**Which amino acid would glucose more strongly interact with?**

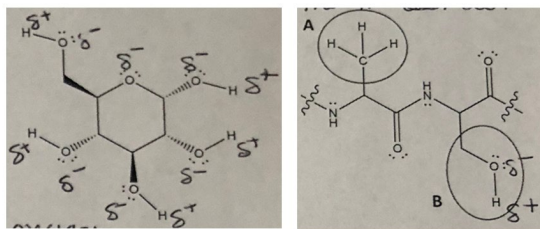
*A*

**Why does the glucose interact more strongly with that amino acid?**

*I think glucose would react more strongly with this part of the molecule because the negative charge would repel part B of the molecule*

Figure 8.31. Response from student V1\_OC1\_120.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*The oxygen molecules have a partial negative charge because they are more electronegative than hydrogen so they pull the electrons closer to the nucleus of the oxygen than the H do.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*There are only partial charges on the O-H bond because C is not electronegative enough to pull the electrons in the bond*

**Which amino acid would glucose more strongly interact with?**

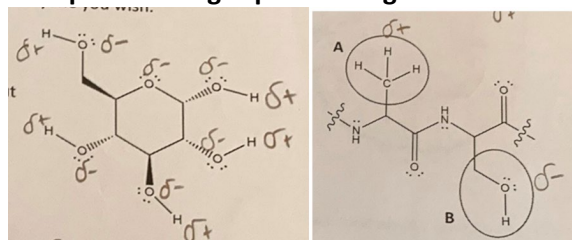
A

**Why does the glucose interact more strongly with that amino acid?**

*Glucose will interact more strongly with the A amino acid because it will be able to hydrogen bond*

Figure 8.32. Response from student V1\_MB\_101.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Oxygen has a negative charge & hydrogen has a positive*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The hydrogens cause a positive charge & the oxygen causes a negative*

**Which amino acid would glucose more strongly interact with?**

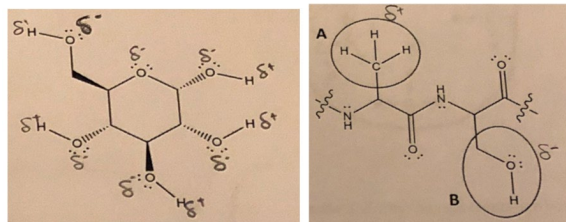
B

**Why does the glucose interact more strongly with that amino acid?**

*The more positive charge of glucose will be attracted to the negative charge of B*

Figure 8.33. Response from student V1\_MB\_102.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*I drew the charges where I did because the oxygen molecules have more electrons not being used in bonds (more negative) where the hydrogen's electrons are being used in the bonds of the molecule*

**Why did you draw the partial charges where you did [on the amino acids]?**

*A doesn't have any valence electrons available for interactions where as B does*

**Which amino acid would glucose more strongly interact with?**

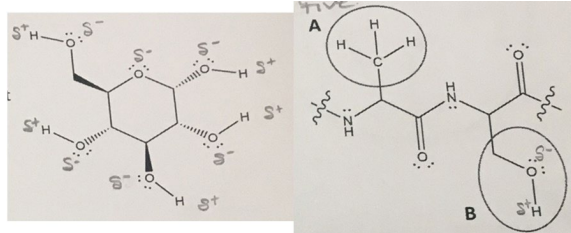
*B*

**Why does the glucose interact more strongly with that amino acid?**

*There are valence electrons able to form interactions with*

Figure 8.34. Response from student V1\_MB\_103.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Oxygen is more electronegative than hydrogen so the electrons will be around the oxygen more. This will give the oxygen a partial negative charge and hydrogens partial positive*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Oxygen is more electronegative the hydrogen, therefore the electrons will be pulled more towards the oxygen. Carbon and hydrogen share electrons almost equally*

**Which amino acid would glucose more strongly interact with?**

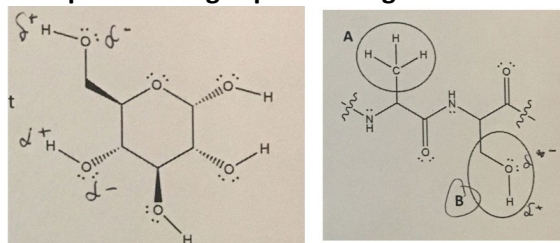
*B*

**Why does the glucose interact more strongly with that amino acid?**

*The partial positive charge on the hydrogen will interact with the partial negative charge on the oxygen in glucose*

Figure 8.35. Response from student V1\_MB\_104.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Because oxygen is more electronegative compared to hydrogen. This is based off electron distribution. O2 has more electrons*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Oxygen has more electrons and is more electronegative*

**Which amino acid would glucose more strongly interact with?**

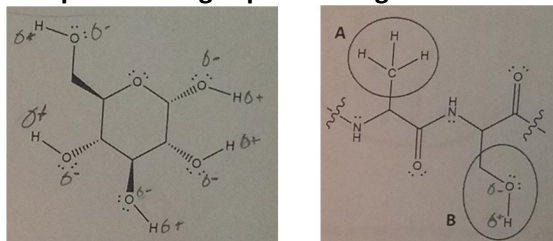
*B*

**Why does the glucose interact more strongly with that amino acid?**

*B is polar and has a stronger partial positive and partial negative charge*

Figure 8.36. Response from student V1\_MB\_105.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*I drew the charges where I did because oxygen is very electronegative while hydrogen is not. So oxygen has a stronger pull on the shared e- between it and hydrogen giving oxygen a partially negative charge and hydrogen a partially positive*

**Why did you draw the partial charges where you did [on the amino acids]?**

*C-H bonds have no disparity (?) in how the electrons are shared so there are no partial charges formed, but both glucose and amino acid B have partial charges leading them to interact more favorably*

**Which amino acid would glucose more strongly interact with?**

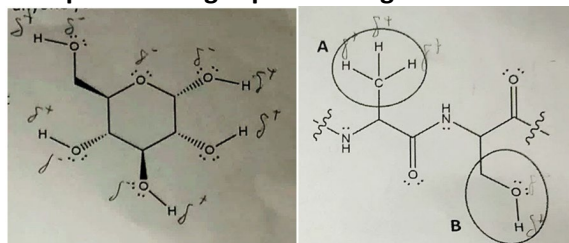
*B*

**Why does the glucose interact more strongly with that amino acid?**

*Because both glucose and B have charges while A does not.*

Figure 8.37. Response from student V1\_MB\_106.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Oxygen is negative. Hydrogen is a proton so positive*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Hydrogen: proton (positive)*

*Oxygen: negative*

**Which amino acid would glucose more strongly interact with?**

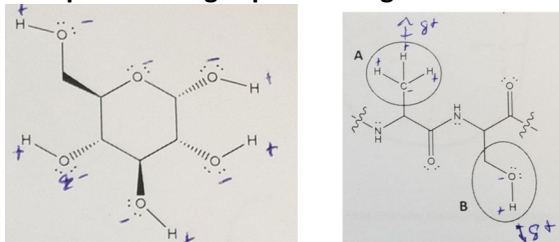
*A and B*

**Why does the glucose interact more strongly with that amino acid?**

*Both b/c each one has 3 bonds made w/ glucose*

Figure 8.38. Response from student V1\_MB\_107.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Hydrogens have a partial pos. charge and oxygen has a partial neg. charge*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Hydrogen has a partial pos. charge and carbon and oxygen have a partial neg. charge.*

**Which amino acid would glucose more strongly interact with?**

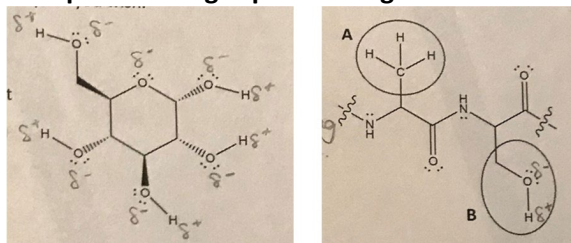
*B*

**Why does the glucose interact more strongly with that amino acid?**

*The oxygen can hydrogen bond with the glucose*

Figure 8.39. Response from student V1\_MB\_108.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Oxygen is more electronegative than Carbon and Hydrogen, so it pulls more electrons towards it giving it that  $\delta^-$  charge.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Oxygen is more electronegative than Hydrogen resulting in electrons being closer to it, making it slightly negative*

**Which amino acid would glucose more strongly interact with?**

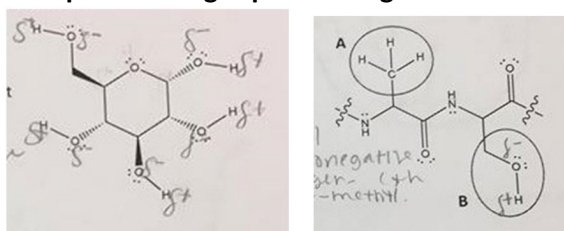
*B*

**Why does the glucose interact more strongly with that amino acid?**

*B, because of the charges that would attract to each other*

Figure 8.40. Response from student V1\_MB\_109.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*To form H bonds, a more electronegative atom (oxygen) must bind with a hydrogen atom*

**Why did you draw the partial charges where you did [on the amino acids]?**

*There are partial charges on amino acid B because it is a polar hydroxyl group. H-bond forms between an electronegative atom (O) and a less e- atom - hydrogen. C and H have similar electronegativities - non polar - methyl.*

**Which amino acid would glucose more strongly interact with?**

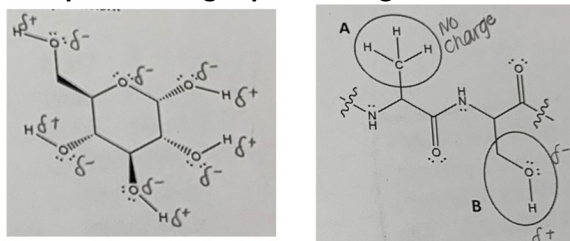
*B*

**Why does the glucose interact more strongly with that amino acid?**

*Both have polar, hydrophilic H-bonds and side chains*

Figure 8.41. Response from student V1\_MB\_110.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*Oxygen is more electronegative because it has more electrons than hydrogen*

**Why did you draw the partial charges where you did [on the amino acids]?**

*OH is  $\delta^-$  and  $\delta^+$  (same reason [as previous]). Methyl group is nonpolar so it doesn't have a charge*

**Which amino acid would glucose more strongly interact with?**

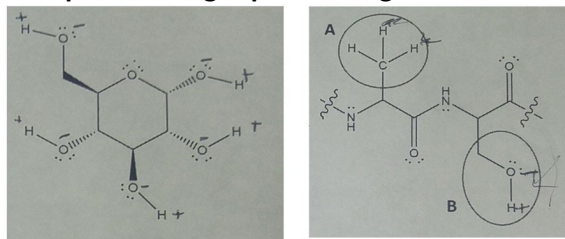
*B*

**Why does the glucose interact more strongly with that amino acid?**

*There's a charge on the molecule and it favorably interacts with water (methyl doesn't)*

Figure 8.42. Response from student V1\_MB\_111.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*They are Polar bonds, the  $-O$  is attracted to  $H^+$  in a polar hydrogen bond*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Hydroxyl groups have polar bonds w/ charges that would attract negatively charged glucose*

**Which amino acid would glucose more strongly interact with?**

*B*

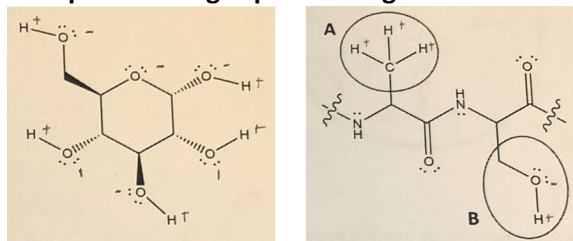
**Why does the glucose interact more strongly with that amino acid?**

*The methyl group is non polar to glucose's polar interaction, hydroxyl + glucose wouldn't repel each other*

Figure 8.43. Response from student V1\_MB\_112.



[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*-OH groups are electronegative*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Hydrogen atoms are positively charged and the -OH group is electronegative*

**Which amino acid would glucose more strongly interact with?**

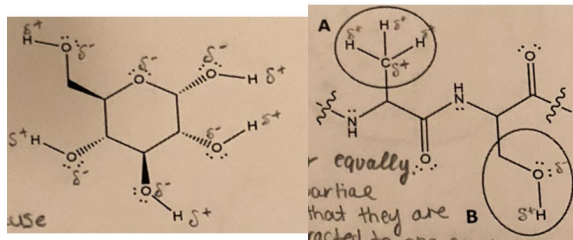
*A*

**Why does the glucose interact more strongly with that amino acid?**

*The hydrogens could connect through a hydrogen bond. They both have the same structure. Both are non-polar.*

Figure 8.44. Response from student V1\_MB\_113.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*I draw the Hydrogen atoms having a partial positive charge because Hydrogen has a positive charge and I drew the oxygen atoms with a partial negative charge because oxygen is negatively charged. This has to do with the electronegativity of the atoms*

**Why did you draw the partial charges where you did [on the amino acids]?**

*I drew amino acid A with all partial positive charges because Carbon and Hydrogen have about the same electronegativity, meaning they will "pull" on each other equally. I drew a partial negative on the oxygen atom and a partial positive on the Hydrogen in amino acid B to indicate that they are attracted to one another*

**Which amino acid would glucose more strongly interact with?**

*B*

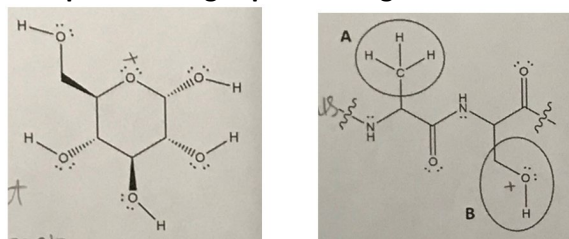
**Why does the glucose interact more strongly with that amino acid?**

*Glucose will interact more strongly with the OH amino acid (B) because it is able to hydrogen bond to glucose, whereas  $\text{CH}_3$  (A) won't be able to interact with glucose b/c it is hydrophobic.*

Figure 8.45. Response from student V1\_MB\_114.



[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*There is one charged particle in glucose, & it is located on the oxygen. I put it there because it was the one particle that wasn't "extended" (?) with other hydrogen or oxygen bonds.*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Only in B, because I, A is a methyl, [illegible] is hydrophobic, uncharged*

**Which amino acid would glucose more strongly interact with?**

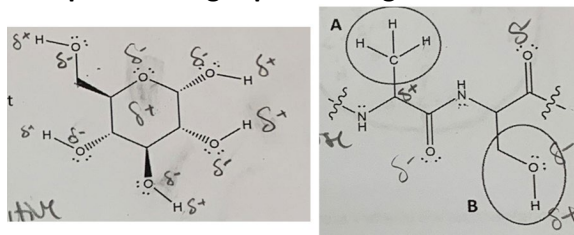
*B*

**Why does the glucose interact more strongly with that amino acid?**

*It can form hydrogen bonds with the glucose. While A, methyl, will not be able to interact favorably*

Figure 8.46. Response from student V1\_MB\_115.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*I drew the charges in those locations because the amount of electrons that surround the oxygen they are relatively negative*

**Why did you draw the partial charges where you did [on the amino acids]?**

*The partial charges are in those locations based on how the elements in those areas share charges*

**Which amino acid would glucose more strongly interact with?**

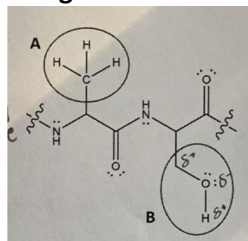
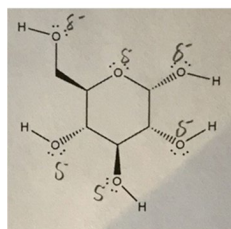
*B*

**Why does the glucose interact more strongly with that amino acid?**

*Glucose would interact more with B because it can form hydrogen bonds with the glucose*

Figure 8.47. Response from student V1\_MB\_116.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*I drew the partial negatives on the oxygen because oxygen is more electronegative than hydrogen and carbon*

**Why did you draw the partial charges where you did [on the amino acids]?**

*Carbon and hydrogen have the same electronegativity so they do not have partial charges. Oxygen is more electronegative than H and CH<sub>2</sub>*

**Which amino acid would glucose more strongly interact with?**

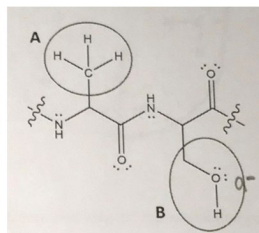
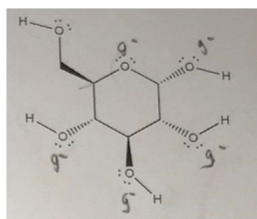
*B*

**Why does the glucose interact more strongly with that amino acid?**

*Because of the H and Oxygens it can form hydrogen bonds*

Figure 8.48. Response from student V1\_MB\_117.

[Student's drawings of the partial charges present in glucose and the amino acids]



**Why did you draw the charges where you did in the glucose molecule?**

*-OH is more electronegative*

**Why did you draw the partial charges where you did [on the amino acids]?**

*-OH is more electronegative*

**Which amino acid would glucose more strongly interact with?**

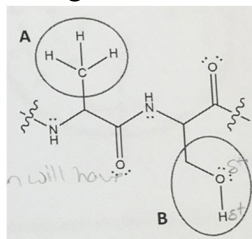
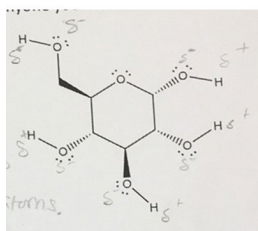
*B*

**Why does the glucose interact more strongly with that amino acid?**

*It is higher in electronegativity to attract more bonds.*

Figure 8.49. Response from student V1\_MB\_118.

**[Student's drawings of the partial charges present in glucose and the amino acids]**



**Why did you draw the charges where you did in the glucose molecule?**

*When hydrogen & oxygen bind the oxygen has a slight negative charge as it attracts electrons a little stronger than hydrogen atoms. Hydrogen has a slight negative charge*

**Why did you draw the partial charges where you did [on the amino acids]?**

*C-H are very similar w/ charges and the electrons do not pull/push one way or another. Same as the above answer Oxygen will have a slight negative charge.*

**Which amino acid would glucose more strongly interact with?**

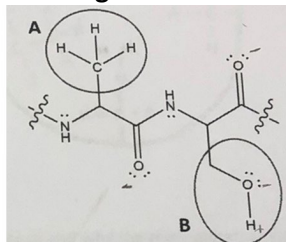
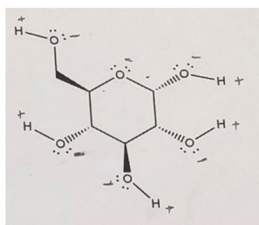
*B*

**Why does the glucose interact more strongly with that amino acid?**

*A is a methyl group and hydrophobic. Because the slight charge there will be greater bonding hydrogen bonding. More favorable interaction.*

Figure 8.50. Response from student V1\_MB\_119.

**[Student's drawings of the partial charges present in glucose and the amino acids]**



**Why did you draw the charges where you did in the glucose molecule?**

*Because the amount of bonds creates a negative charged portion*

**Why did you draw the partial charges where you did [on the amino acids]?**

*CH3 creates a non polar portion so the only charged portions is the single O molecules*

**Which amino acid would glucose more strongly interact with?**

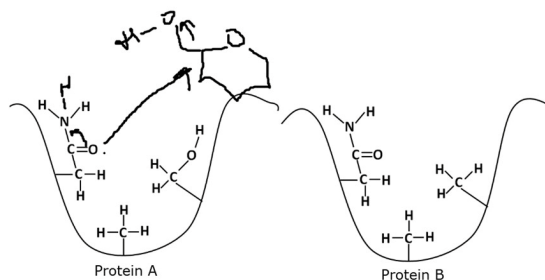
*B*

**Why does the glucose interact more strongly with that amino acid?**

*Because of the partial charges the amino acids are more likely to bond with the opposite charge to become more stable.*

Figure 8.51. Response from student V1\_MB\_120.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

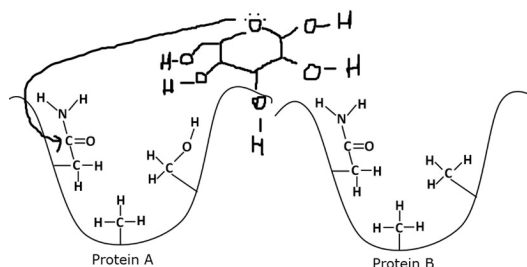
*Separation of charge causes the proteins to bind to the glucose. Oxygen is electronegative and carbons have a partial positive charge on them.*

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

*It has a better glucose binding site because it has a double bond to oxygen making it more reactive than the rest of the elements.*

Figure 8.52. Response from student V2Alt\_OC2\_101.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

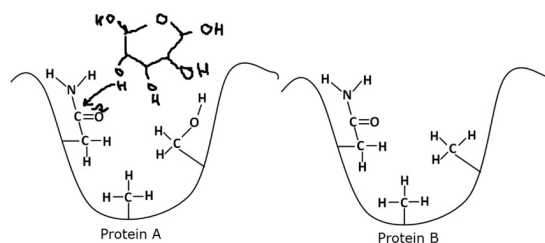
*The glucose has electron rich oxygen that is drawn to a partially positive carbon atom on the binding site.*

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

*I chose protein A because it has an alcohol instead of a ch3. This could make a difference in the reactivity and affect whether or not glucose would bind.*

Figure 8.53. Response from student V2Alt\_OC2\_102.

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



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

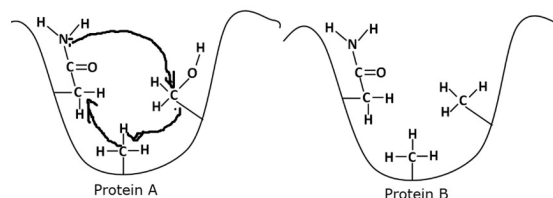
*Protein A has a better glucose binding site because it has an additional OH group. This additional OH group can be added onto the glucose ring so that the OH group can help the biological system because O is necessary in these systems.*

Explain what causes glucose to bind to the protein.

*Glucose binds to the protein because the OH groups are able to attack at the carbonyl and break the double bonded O. This creates a single bonded O and an OH group to the C. This helps for the ring to break and glucose to attach to the protein.*

Figure 8.54. Response from student V2\_OC2\_103 (Katrina).

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



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

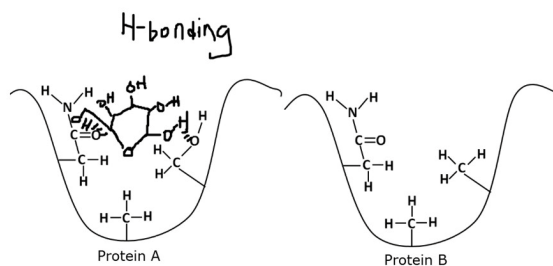
*I pick protein A because it had the OH group that connected to the carbon that shown in glucose and it had NH2 to attack.*

Explain what causes glucose to bind to the protein.

*[blank]*

Figure 8.55. Response from student V2\_OC2\_104.

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



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

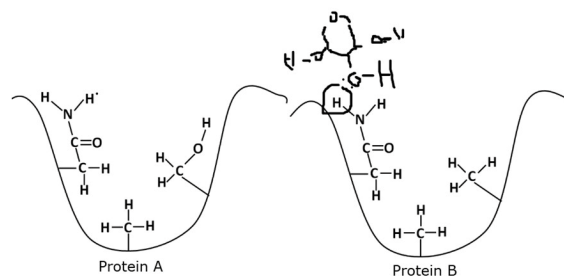
*I believe protein A has the better protein binding site because it has two oxygens that can interact with the glucose through hydrogen bonding of polar groups. In protein B there is only the carbonyl oxygen which could have interaction with hydrogen, but Protein A has this and another oxygen on the opposite side with a hydrogen. That functional group could either hydrogen bond to glucose using the oxygen or the hydrogen.*

**Explain what causes glucose to bind to the protein.**

*The cause of glucose binding to the protein is the electrostatic force of attraction between polar groups. The electronegative oxygen is attracted to the partial positive charge distribution of hydrogen (due to random movement of electron clouds, and the fact carbon is slightly more electronegative), when they hydrogen bond.*

Figure 8.56. Response from student V2\_OC2\_105.

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



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

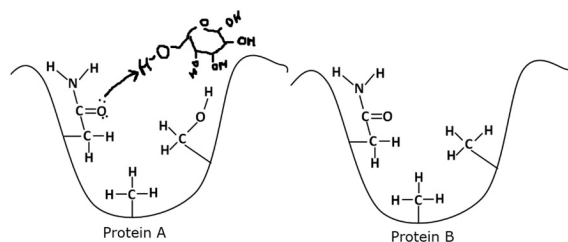
*The protein I chose, protein B has the better Glucose binding site than Protein A because of the carbonyl and the methyl group which can act as a leaving group.*

**Explain what causes glucose to bind to the protein.**

*The electronegative oxygen in the carbonyl is what causes the glucose to bin to protein B. With this, the methyl group is then kicked off and elminated as a leaving group.*

Figure 8.57. Response from student V2\_OC2\_106.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

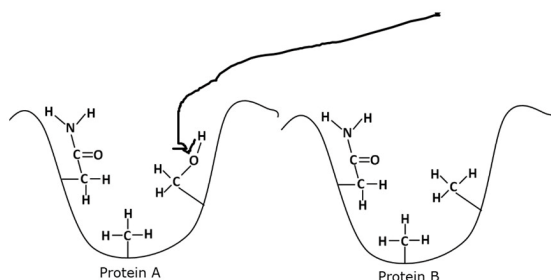
*The lone pair on the oxygen of the carbonyl side chain in protein A would attack a hydrogen on glucose, causing the two to bond.*

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

*Protein A has more oxygens than protein B so therefore, the glucose will be attracted to protein A as more lone pairs can attack it, causing it to bond together.*

Figure 8.58. Response from student V2Alt\_OC2\_107.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

*Protein A has an OH side chain that can hydrogen bond with the OH side chains of glucose.*

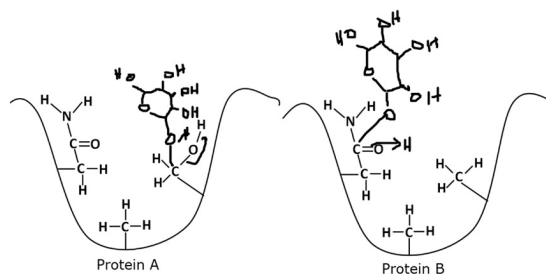
*The H of glucose will hydrogen bond with the O of protein A.*

**Explain what causes glucose to bind to the protein.**

*The OH in protein A is very electronegative and attracts the protons on the glucose.*

Figure 8.59. Response from student V2\_OC2\_108.

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



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

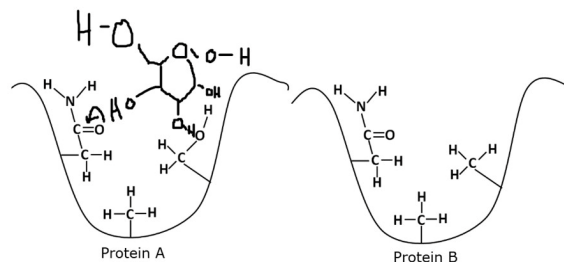
*The OH in protein a is a better leaving group and protein B has no leaving group. So glucose attacks at the carbonyl in the protein B and the OH in the Protein A.*

Explain what causes glucose to bind to the protein.

*OH group off the methyl forms a carbocation and that attacks the carbonyl in protein b and the C from the OH group in protein A. This force to restabilize the carbocation is what drives the reaction.*

Figure 8.60. Response from student V2\_OC2\_109.

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



Explain what causes glucose to bind to the protein.

*I'm not entirely sure, but I think the glucose could bind to protein because it is a very unstable molecule and it is able to oxidize the oxygen on the left side chain on protein A in order to link up for both of them to become more stable.*

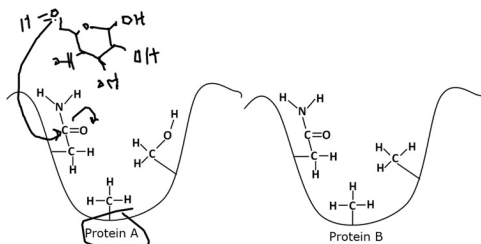
Explain why the protein you chose has the better glucose binding site and how the differences in the site cause this difference in binding.

*I think protein A has a better bonding site because the right side chain on protein A has two pairs of lone pairs on the oxygen that are able to move around in order to link up with glucose, as well as protein A's left most side chain also being able to link.*

Figure 8.61. Response from student V2Alt\_OC2\_110.



Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



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

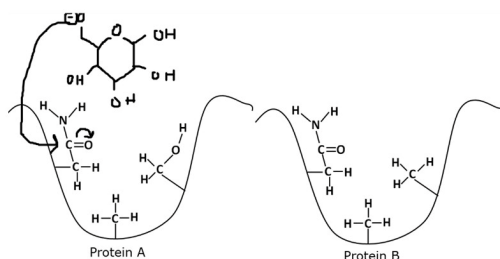
*I chose protien A as the better binding site because when bindng a glucose, i know you need an OH group to bond with a carbonyl carbon and breka the double bond of the carbonyl. After the OH in the glucose ring bings with the carbonyl in protein A, the CH3OH can react with glucose again and make the bond to protein A even stronger.*

**Explain what causes glucose to bind to the protein.**

*Glucose binds to protein because of the long pair on one of the OH groups on the Glucose. The negatively charged lone pair attracted to the partial positive carbonyl carbon, and breaks the double bond on the oxygen, leaving you with a negatively charged oxygen. A hydrogen can protinate the NH3 and the negative charge on the O could potentially push back down and push off the NH3.*

Figure 8.62. Response from student V2\_OC2\_111.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

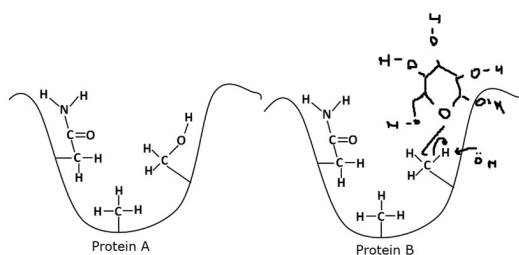
*I think protein A has a better binding sight at the carbonyl OR the oh connected to CH3. If you deproatanate the O on the carbon you have anegative charge that could attract a hydrogen from the glucose molecule. If you use the carronyl and deproatanate the glucose then it can attach to the carbon near the oxygen.*

**Explain what causes glucose to bind to the protein.**

*The carbon near the Nitrogen has a higher partially positive charge because nitrogen is electronegative and electron withdrawing therefore pulling the electrons away from the carbon. this stronger positive charge will attract the negative charge on the glucose more readily*

Figure 8.63. Response from student V2\_OC2\_112.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

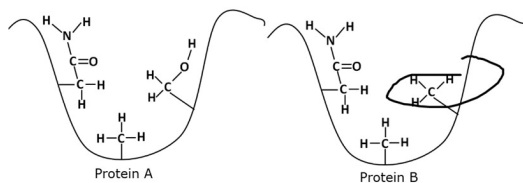
*protein b has on open carbon that is available to be attached to the glucose after the removal of the hydrogen*

**Explain what causes glucose to bind to the protein.**

*the free electrons from the carbon from protein b make a bond with the oxygen in the ring*

Figure 8.64. Response from student V2\_OC2\_113.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

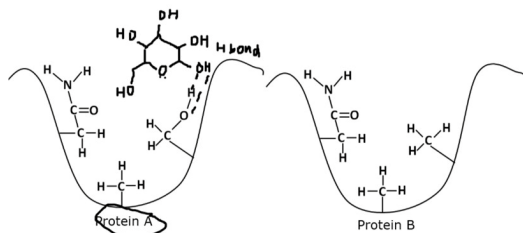
*The protein i chose to have a better binding site to glucose becausse the CH3 on prtein b has a positive charge and will eaily react with a negatively charged substance*

**Explain what causes glucose to bind to the protein.**

*The OH group on the protein will interact with the Oh groups on glucose and will hydrogen bond*

Figure 8.65. Response from student V2\_OC2\_114.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

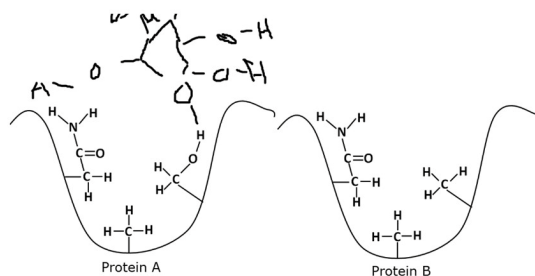
*The binding is due to thhe electrostatic attraction of the acetic hydrogen and the hydroxyl group*

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

*Proteins A and B are the same except for the addition of the hydroxyl group in Protein A. This introduces an acetic hydrogen that doesn't exist in protein B, increasing its reactivity*

Figure 8.66. Response from student V2Alt\_OC2\_115.

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



Explain what causes glucose to bind to the protein.

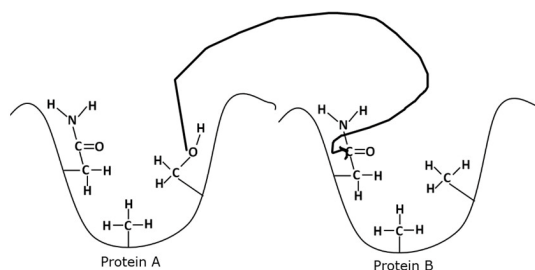
*The lone pair on the O of the glucose molecule donates to the protein, allowing it to bind together*

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

*Protein A has the O with a lone pair, which allows it to bind to the lone pair on the O in the glucose molecule*

Figure 8.67. Response from student V2Alt\_OC2\_116.

Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.



Explain what causes glucose to bind to the protein.

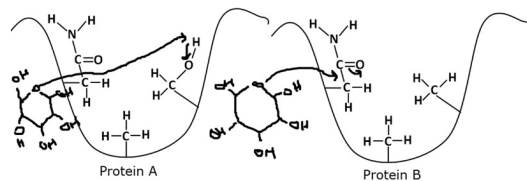
*Tautomerixation I think*

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

*the carbonyl carbon is a good binding site because of its charge distribution*

Figure 8.68. Response from student V2Alt\_OC2\_117.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

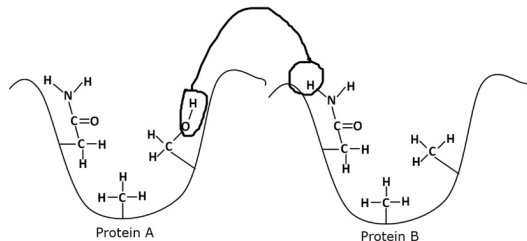
*Glucose needs to attach to a protein. It need to bind because it will activate the protein and this allows the protein to bind to DNA.*

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

*I chose the glucose binding sites because if you bind them there it will activate the protein. If you were to bind them anywhere else it wouldt cause the protien to full activate.*

Figure 8.69. Response from student V2Alt\_OC2\_118.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

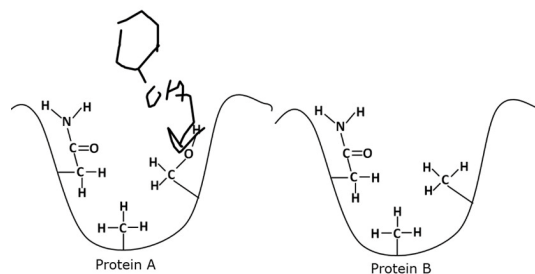
*The OH on protein A attacks the H on protein B and water is formed connecting the Nitrogen and Carbon creating a peptide bond.*

**Explain what causes glucose to bind to the protein.**

*Since water is lost, the nitrogen attaches to the carbon because the nitrogen is nucleophilic and attaches to the positively charged carbon.*

Figure 8.70. Response from student V2\_OC2\_119.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

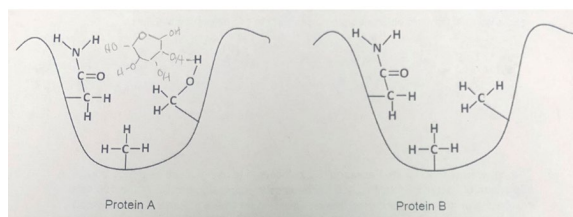
*glucose binds to protein A because of how the molecule has OH bonds which are polar covalent which is a bond that has a charge and can bind to the CH4O of the protein A*

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

*protein B has nonpolar charges on the site which does not allow for the glucose to bind easily to it.*

Figure 8.71. Response from student V2Alt\_OC2\_120.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

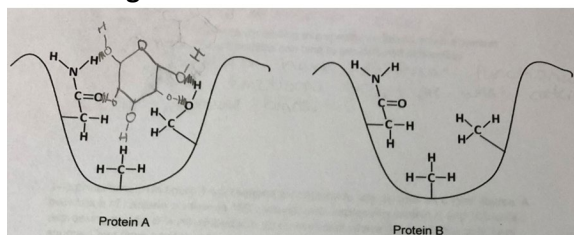
*In protein A, it has one less methyl group. Glucose can bind to this protein by hydrogen bonding to the O-H on the right side of the binding site. Glucose isn't favorable to hydrophobic clustering with methyl groups.*

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

*Protein B has an amino side chain and two methyl side chains. Protein A has an amino side chain, a methyl side chain, and a side chain containing an O-methyl groups form interactions through hydrophobic clustering, which glucose is not favorable to. Therefore protein A, having one less methyl group and the O-H side chain can form a favorable hydrogen bond with the glucose molecule*

Figure 8.72. Response from student V2Alt\_MB\_101.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

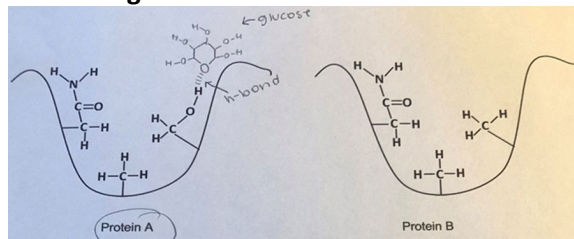
*Protein A allows hydrogen bonding to occur. Whereas Protein B does not replace its O-H group with a CH<sub>3</sub> group which does not allow hydrogen bonding to occur. CH<sub>3</sub> only allows hydrophobic clustering to occur.*

**Explain what causes glucose to bind to the protein.**

*The oxygens in the glucose molecule bind to the hydrogen in the hydroxyl group. Since more bonds can be formed in protein A, Glucose will bind to protein A*

Figure 8.73. Response from student V2\_MB\_102 (Liam).

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

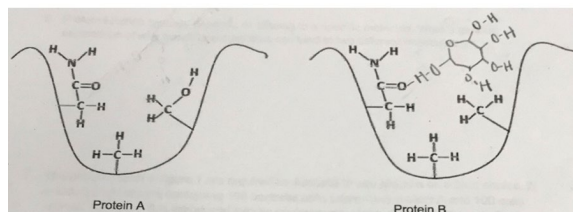
*Protein A is a better binding site for glucose because CH<sub>2</sub>OH is able to form a hydrogen bond with glucose. The nonpolar H is able to hydrogen bond to the lone pair on the highly electronegative oxygen in glucose*

**Explain what causes glucose to bind to the protein.**

*The hydrogen bond between the low electronegative hydrogen and the high electronegative oxygen allows glucose to bind to the protein*

Figure 8.74. Response from student V2\_MB\_103.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

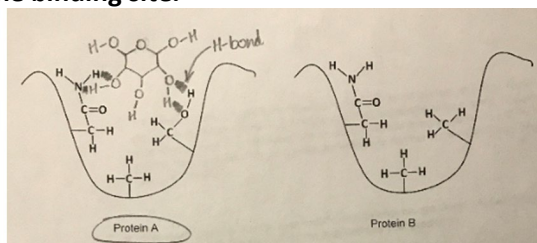
*I think that Protein B has a better glucose binding site. The differences in the sites are that Protein A only has one CH3 while protein B has two CH3. This gives more places for hydrogen bonding and where the glucose can attack on.*

**Explain what causes glucose to bind to the protein.**

*Glucose binds to the protein. This is due to hydrogen bonding that occurs with the CH3*

Figure 8.75. Response from student V2\_MB\_104 (Paul).

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

*Protein A has a hydroxyl group while Protein B has a methyl. Since glucose can undergo H-bonding. Protein A will bind more favorably due to the hydroxyl group also being able to undergo H-bonding while the methyl undergoes hydrophobic clustering*

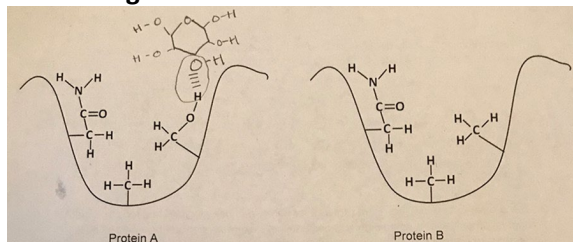
**Explain what causes glucose to bind to the protein.**

*The hydroxyl groups on glucose and on Protein A along with the amino group on Protein A all undergo H-bonding, where the H in the groups are attracted to the O or N of another group and vice versa*

Figure 8.76. Response from student V2\_MB\_105.



**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

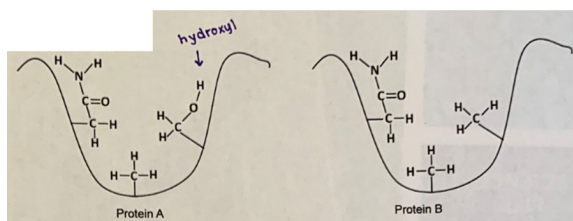
*The glucose molecule has lots of O-H bonds that have the ability to hydrogen bond with other functional groups. Protein A has a hydroxyl group on the right so it can positively interact with the hydroxyl groups in glucose. Meanwhile I did not choose protein B because on the right side it only has methyl groups which cannot H bond with glucose*

**Explain what causes glucose to bind to the protein.**

*The glucose binds to the protein because of H bonding of the hydroxyl groups. The partially positive H atom in the protein interacts with the lone pair of electrons on O in glucose to form a hydrogen bond*

Figure 8.77. Response from student V2\_MB\_106.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

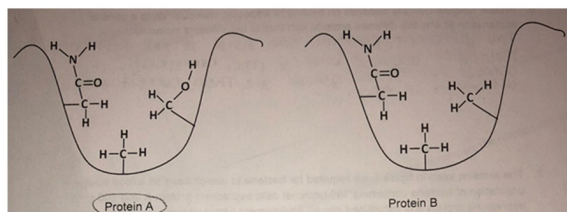
*Protein A has a hydroxyl (O-H) functional group which is polar, allowing this site to interact with the polar glucose molecule and bind*

**Explain what causes glucose to bind to the protein.**

*The highly electronegative glucose molecule pulls in electrons and will interact favorable with the polar side chain within protein A's binding pocket*

Figure 8.78. Response from student V2\_MB\_107.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

*I think that glucose is more liking to bind to protein A's binding site because glucose is polar & so is the binding site of protein A*

**Explain what causes glucose to bind to the protein.**

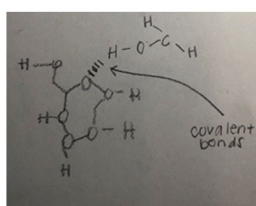
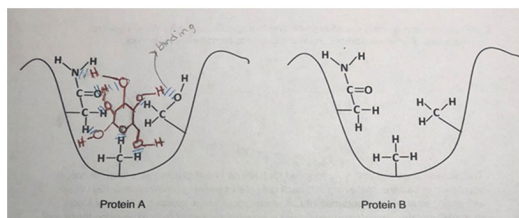


Figure 8.79. Response from student V2\_MB\_108. Note that the student provided a drawing rather than an explanation in response to "Explain what causes glucose to bind to the protein."

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

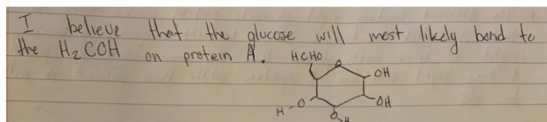
*I picked protein A because it had more spaces for the hydrogens to bind to. Protein B doesn't have enough oxygens and therefore would be more hydrophobic than would be preferred.*

**Explain what causes glucose to bind to the protein.**

*Glucose binds to the protein through hydrophobic clustering and hydrogen bonds*

Figure 8.80. Response from student V2\_MB\_109.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

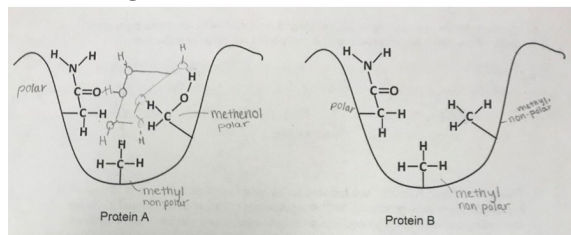
*Glucose will bind to the protein by replacing one of OH on the glucose. The H<sub>2</sub>COH will bind with the glucose forming HCHO & leaving H<sub>2</sub>O. Protein A has more potential to bind with glucose properly*

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

*Protein A & B are very similar, their only difference was the last section of the strand. The glucose molecule will not properly bond with protein B, because it lacks the OH protein A is the specific binding right to glucose changing H<sub>2</sub>COH into HCHO.*

Figure 8.81. Response from student V2Alt\_MB\_110.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

*Protein A provides a better binding site because of the polar side groups which allow for hydrogen bonds to form*

**Explain what causes glucose to bind to the protein.**

*Glucose has several -OH's (hydroxyl) which can bind well with the hydrogens in protein A*

Figure 8.82. Response from student V2\_MB\_111.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**

*[blank]*

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

*Protein A has a better binding site, due to the presence on another oxygen on the binding site. Due to the structure of the glucose, it is most likely to form hydrogen bonds between the H on the glucose and the So present in the first protein. These weak interactions are what bind the subrostrate (glucose) to the binding site.*

**Explain what causes glucose to bind to the protein.**

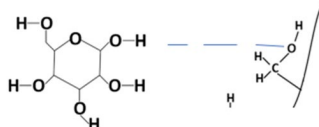
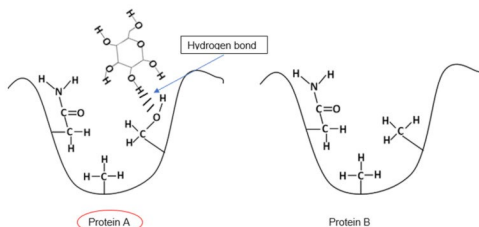


Figure 8.83. Response from student V2\_MB\_112. Note that the student provided a drawing rather than an explanation in response to “Explain what causes glucose to bind to the protein.”

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

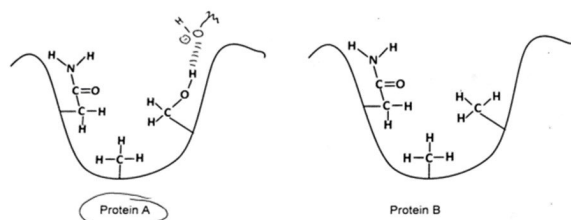
*In Protein A, there is an oxygen bonded to a hydrogen. A hydrogen bond can be formed between a hydrogen attached to an electronegative atom (O) in glucose and the oxygen attached to a hydrogen in Protein A. In Protein B, a hydrogen bond can be formed between N in Protein B and a H from glucose. Since oxygen is more electronegative, it forms better hydrogen bond, thus Protein A is the better binding site.*

**Explain what causes glucose to bind to the protein.**

*The hydrogen bond interaction between an H bonded to O in glucose and the O bonded to H in Protein A causes the glucose to bind to the protein.*

Figure 8.84. Response from student V2\_MB\_113.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

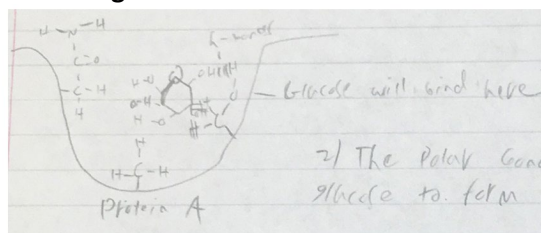
*I believe protein A would be a better bonding site for glucose because it is able to make stronger interactions, more hydrogen bonds. Since, there's more oxygen + nitrogen for the glucose to bond to strongly.*

**Explain what causes glucose to bind to the protein.**

*The dipole-dipole interactions bring the hydrogen + oxygens of each respective chemical together forming hydrogen bonds which are strong interactions between molecules.*

Figure 8.85. Response from student V2\_MB\_114.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

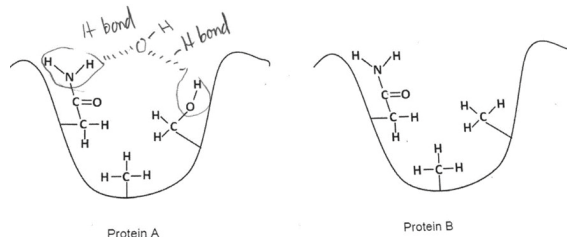
*The polar bonding in protein A will allow for glucose to form more hydrogen bonds to the CH<sub>2</sub>OH*

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

*The polarity in the CH<sub>2</sub>OH will allow for the formation of H-bonds to the O-H of the glucose.*

Figure 8.86. Response from student V2Alt\_MB\_115.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

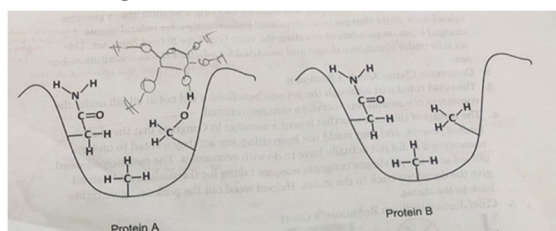
*Glucose could bind to the amino group, or the hydroxyl group via H-bonding. The methyl group does not have the ability to Hbond w/ glucose*

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

*The protein I choose has 2 potential binding sites while protein B only has 1 potential binding site. The 2 methyl groups can't H-bond.*

Figure 8.87. Response from student V2Alt\_MB\_116.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

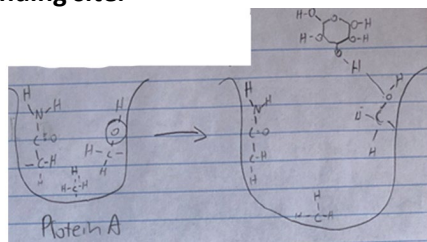
*I think it would bind here because there is hydrogen bonding.*

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

*I had this because the oxygen makes it possible to hydrogen bonding.*

Figure 8.88. Response from student V2Alt\_MB\_117.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

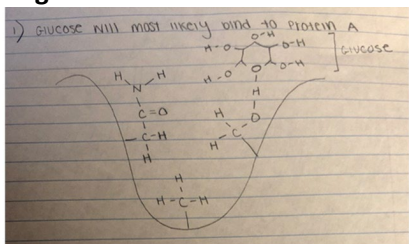
*What causes glucose to bind to the protein is the oxygen with only 2 bonds. The glucose is comprised of oxygen + hydrogen molecules resulting in a perfect fit for hydrogen bonding. One of the hydrogens of the glucose can H-bond with the oxygen of protein A.*

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

*Protein A has a better glucose binding site than protein B because of the only difference between them. The oxygen with 2 single [illegible] is the best spot, that is most open to hydrogen bonding.*

Figure 8.89. Response from student V2Alt\_MB\_118.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



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

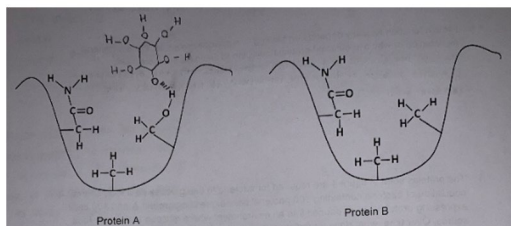
*Protein A has a better binding site because glucose is a polar structure b/c of the multiple –OH groups in the structure. In protein A, it has a an –OH instead of just a –H which causes H to be polar. Protein A + protein B are the same besides that one difference protein A = CH<sub>3</sub>O + protein B = CH<sub>4</sub>*

**Explain what causes glucose to bind to the protein.**

*What causes glucose to bind to a protein is the –H off of the CH<sub>3</sub>O. The –H binds w/ the –O of the glucose molecule*

Figure 8.90. Response from student V2\_MB\_119.

**Pick the binding site you think is most likely to bind glucose and draw a possible way glucose could bind in the binding site.**



**Explain what causes glucose to bind to the protein.**

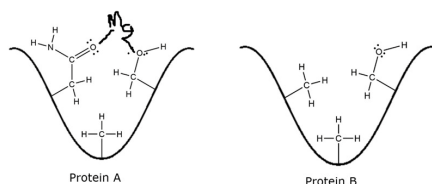
*Glucose would much more likely bind to protein A because it can form hydrogen bonds easier and much more of them than protein B because protein B's hydrogens are bonded to C except for two. The ones bonded to carbon can not form hydrogen bonds.*

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

*Because the H is bonded to an O which means it can form a hydrogen bond. H could also form H-bonds with two other H than are bound to a N.*

Figure 8.91. Response from student V2Alt\_MB\_120.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

*Protein A has more oxygen with lone pairs that are willing to be shared*

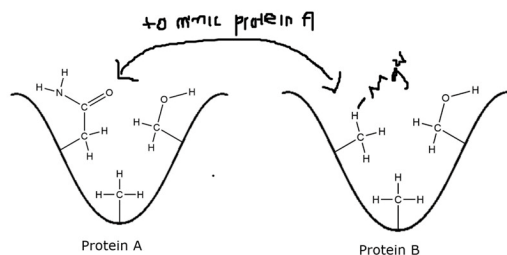
**Explain what causes the magnesium ion to bind to the protein.**

*The 2+ charge of Mg is attracted to the negative polarity of the oxygen*

Figure 8.92. Response from student V3Alt\_GC2\_101.



**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

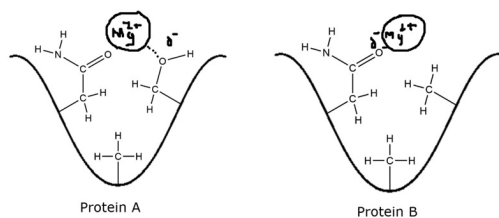
*I believe proten B has the better binding site. Protein A already has a lot binded to it unlike the left structure on protein B. So i believe it will bind to that structure to sort of mimic protein A's structure*

**Explain what causes the magnesium ion to bind to the protein.**

*Im not totally sure about this. However I believe it will choose the arealeast with the least amount of bonds already in order to create more and equal it out*

Figure 8.93. Response from student V3Alt\_GC2\_102 (Claudia).

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

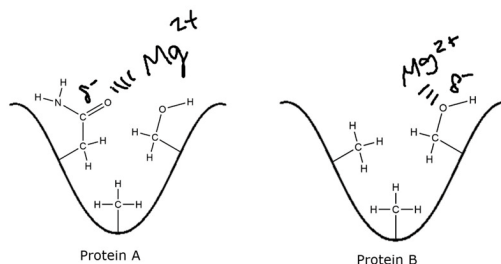
*protien A will offer a better binding site for the magnesium ion because there are two highly electronegative oxygens with lone pairs of electrons that are very attracted to the magnesium cation*

**Explain what causes the magnesium ion to bind to the protein.**

*the partially negatively charged oxygen contains lone pairs of electrons that are highly attractive to the magnesium cation. an ion-dipole will therefore form between the oxygen and the magnesium cation.*

Figure 8.94. Response from student V3\_GC2\_103.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

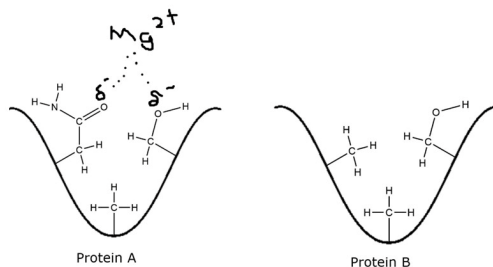
*I drew the magnesium binding to both protein a and b; however, I believe that the magnesium ion would bond to the protein a over protein b because protein a is more polar and has more atoms that have high electronegativity like O.*

Explain what causes the magnesium ion to bind to the protein.

*The partial negative charge on the electronegative atoms allow for a attraction to develop between the magnesium ion and the protein.*

Figure 8.95. Response from student V3Alt\_GC2\_104.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

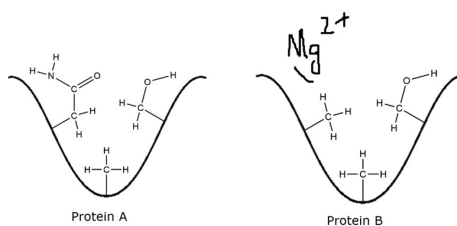
*The magnesium ion has a +2 charge and will be attracted to the 2 slightly negative oxygen molecules that are present in protein A. The negative charge provided by a single oxygen molecule in protein B is not as strong as the charge in protein A.*

Explain what causes the magnesium ion to bind to the protein.

*The magnesium ion will form an ionic-dipole interaction with the slightly negative oxygen molecules in protein A. Electrostatic forces will keep the magnesium ion at the binding site until there in an introduction of energy into the system.*

Figure 8.96. Response from student V3Alt\_GC2\_105 (Conor).

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

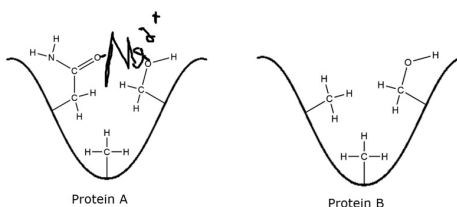
*I'm choosing the protein B. Because the protein A has more bonds and it is more stable, the protein B has less bonds so it is easier to combine with other metal.*

Explain what causes the magnesium ion to bind to the protein.

*It will bond with hydrogen ion because the hydrogen ion is easier to be combined with metal ion*

Figure 8.97. Response from student V3Alt\_GC2\_106.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

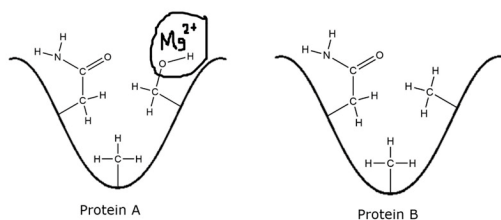
*protein a was picked because it contained stronger intermolecular forces that could bind  $Mg^{2+}$  to the protein such as hydrogen bonding and dipole dipole the oxygens bind to the  $Mg$*

Explain what causes the magnesium ion to bind to the protein.

*intermolecular forces causes the attraction between magnesium and the protein*

Figure 8.98. Response from student V3Alt\_GC2\_107.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

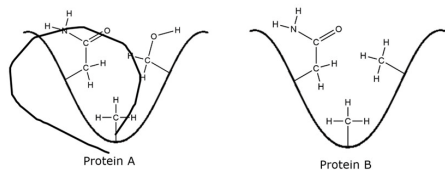
*I chose the  $CH_3OH$  binding site because it has hydrogen bonding interaction which could form the strongest bond with the magnesium through ion dipole interaction*

**Explain what causes the magnesium ion to bind to the protein.**

*IMF's allow  $Mg$  to bind to the protein, I chose  $CH_3OH$  because it is the only binding site that allows for hydrogen bonding interaction which is the strongest IMF*

Figure 8.99. Response from student V3\_GC2\_108.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

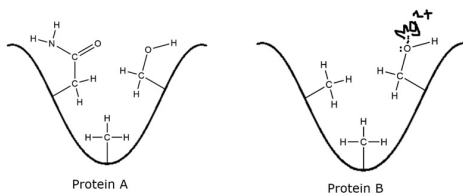
*because there is a binding side for magnesium to bind*

**Explain what causes the magnesium ion to bind to the protein.**

*oh group*

Figure 8.100. Response from student V3\_GC2\_109.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

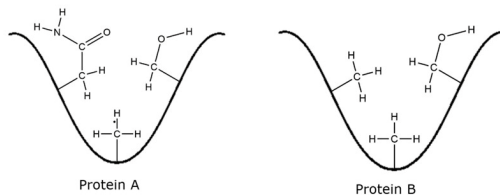
*Protein B has an oxygen available to bond to the magnesium. Oxygen is more electronegative than nitrogen and so would attract the magnesium ion better.*

**Explain what causes the magnesium ion to bind to the protein.**

*In protein B there is an electronegative oxygen that attracts the highly positively charged magnesium ion.*

Figure 8.101. Response from student V3Alt\_GC2\_110.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

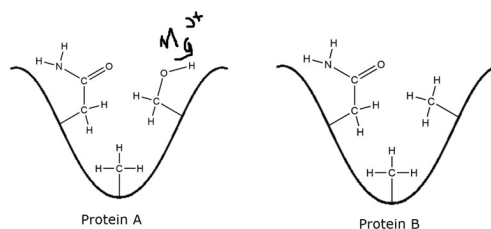
*It has the better binding site because there are more electronegative oxygens willing to share their electrons with the positive Mg.*

**Explain what causes the magnesium ion to bind to the protein.**

*The attractive forces between the negative of the oxygen will be attracted to the positive of the magnesium.*

Figure 8.102. Response from student V3Alt\_GC2\_111.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

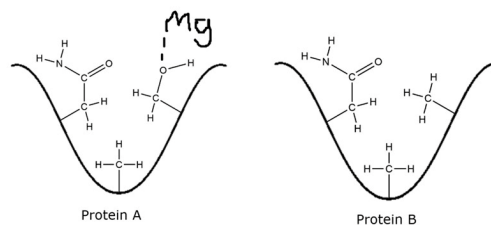
*I think protein A has the binding site because there is an oxygen on the carbon with a hydrogen. The oxygen is partially negative due to its bond with the hydrogen and the Mg has a positive charge making it more likely to be near the oxygen in protein A.*

Explain what causes the magnesium ion to bind to the protein.

*I don't think it would be near protein B because protein B is just carbons and one nitrogen and the nitrogen will have a partially negative charge BUT the partially negative charge on the oxygen will be stronger due to its larger effective nuclear charge and more electronegativity.*

Figure 8.103. Response from student V3\_GC2\_112.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

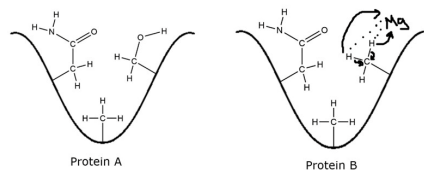
*The only partially negative atom in Protein B is the nitrogen but in Protein A, there is another partially negative atom (Oxygen) that has great electronegativity. This means magnesium would most likely be attracted to Protein A.*

Explain what causes the magnesium ion to bind to the protein.

*Oxygen's high electronegativity (large partial negative charge) attracted the Magnesium cation.*

Figure 8.104. Response from student V3\_GC2\_113.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

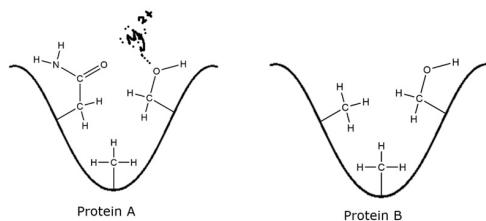
Since Protein B has CH<sub>3</sub>, C is able to give Mg 2 protons, whereas protein A has CH<sub>3</sub>O.

Explain what causes the magnesium ion to bind to the protein.

CH<sub>3</sub> is able to give two protons (H) while Mg gives electrons to C

Figure 8.105. Response from student V3\_GC2\_114.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

I chose protein A to have the better magnesium binding site due to the O-H group on CH<sub>2</sub>OH.

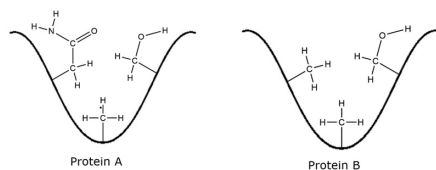
The O has two lone pairs that will bind well with the positively charged magnesium ion.

Explain what causes the magnesium ion to bind to the protein.

The O has two lone pairs that will bind well with the positively charged magnesium ion.

Figure 8.106. Response from student V3Alt\_GC2\_115.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

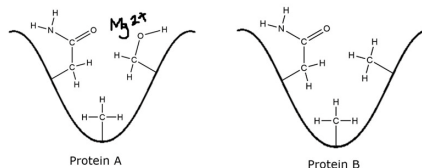
[blank]

**Explain what causes the magnesium ion to bind to the protein.**

[blank]

Figure 8.107. Response from student V3Alt\_GC2\_116.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

*In protein A There are two oxygen with lone pairs that are able to form interactions with the magnesium*

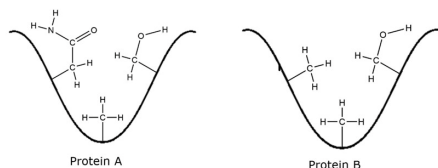
**Explain what causes the magnesium ion to bind to the protein.**

*The protein has one oxygen that can form a bond to the Magnesium and it can form LDFs with other molecules*

Figure 8.108. Response from student V3\_GC2\_117.



Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

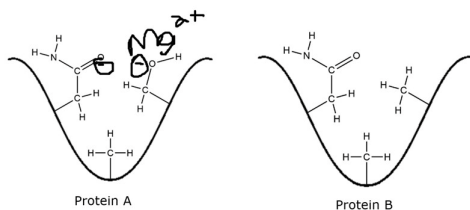
C

Explain what causes the magnesium ion to bind to the protein.

X

Figure 8.109. Response from student V3Alt\_GC2\_118.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

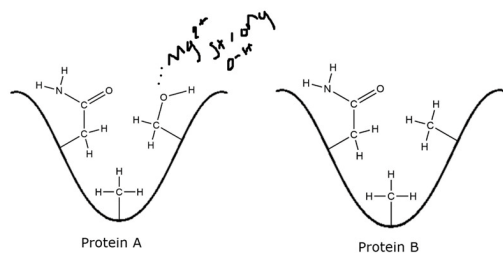
*The protein I chose is better because it has more electronegative atoms available. Therefore the partial negative charges on it are able to attract the positive.*

Explain what causes the magnesium ion to bind to the protein.

*The partial positives on magnesium are attracted to the partial negatives of the oxygens.*

Figure 8.110. Response from student V3\_GC2\_119.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

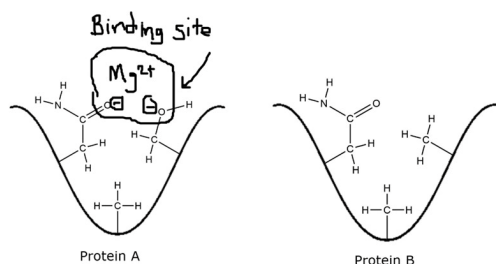
*This protein is the better option because it has more electronegative elements in the protein opposed to the other one which has a lot of nonpolar bonds in it*

Explain what causes the magnesium ion to bind to the protein.

*I think the OH and other electronegative bonds cause the magnesium to bond to it and have a stonger connection*

Figure 8.111. Response from student V3\_GC2\_120.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

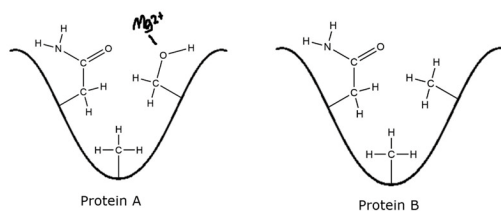
*The protein I chose has the better magnesium binding site because for protein A the negative ends of the amino acid chains were facing towards a way the positive  $Mg^{2+}$  ion can bind to.*

Explain what causes the magnesium ion to bind to the protein.

*What cause the magnesium ion to bind to the protein are the attractive forces (IMFs) present. The negative end of the amino acid is being attracted to the postivie magnesium ion.*

Figure 8.112. Response from student V3\_GC2\_121.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

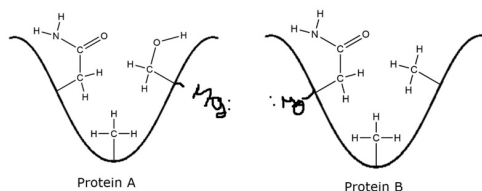
*Protein A has a more suitable oxygen binding site that is not connected to either a double bonded oxygen-carbon interaction, nor would the same space on protein B be applicable due to there being a CH<sub>3</sub> group instead of a CH<sub>2</sub>OH group.*

**Explain what causes the magnesium ion to bind to the protein.**

*Protein A is more electronegative than protein B, and the difference between the two is enough to cause the magnesium 2+ ion to attach to protein A to lower the deficit between the two proteins. It is more structurally sound when bound to A than it would be to B.*

Figure 8.113. Response from student V3\_GC2\_122.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

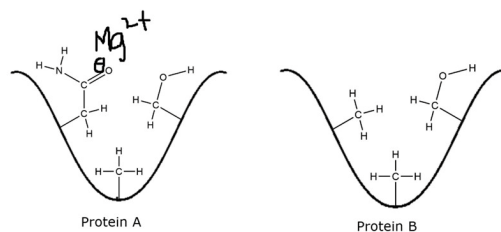
*I thought about the electronegativity of the molecule, since Mg is positively charged, it will be connected to the more negatively charged molecule*

**Explain what causes the magnesium ion to bind to the protein.**

*It would be the Dipole-dipole interaction*

Figure 8.114. Response from student V3\_GC2\_123.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

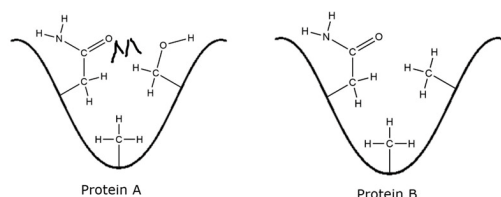
*I think that protein A is a better magnesium binding site because of the double bond on from the C to the O. The O has more of a partial negative with means that is would be more attracted to the  $Mg^{2+}$ . With the double bond present, there are more possible arrangements for the  $Mg^{2+}$  to bind to the cite.*

**Explain what causes the magnesium ion to bind to the protein.**

*The magnesium has a positive charge so it is attracted to the oxygen which has a negative charge.*

Figure 8.115. Response from student V3Alt\_GC2\_124.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

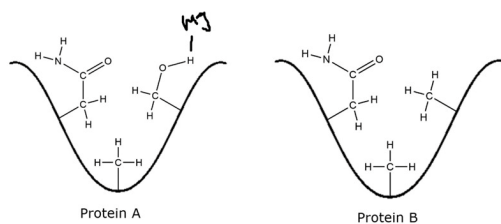
*Since Magnesium is positively charged, it will be drawn to a more negatively charged binding site. This is why I chose Protein A. Besides  $CH_3$ , the amino acid side chains of Protein A contain negative charges.*

**Explain what causes the magnesium ion to bind to the protein.**

*The magnesium ion will bind to the protein, because  $NH_2$  as well as  $OH$  both have negative charges which the positive Magnesium is attracted to*

Figure 8.116. Response from student V3\_GC2\_125.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

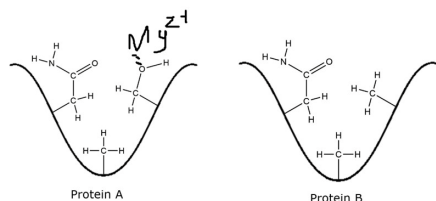
*I think it binds to the hydrogen bonded to the oxygen because it is directly bonded with oxygen and oxygen has strongest electronegativity.*

**Explain what causes the magnesium ion to bind to the protein.**

*I think the transfer of enzymes.*

Figure 8.117. Response from student V3\_GC2\_126.

**Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.**



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

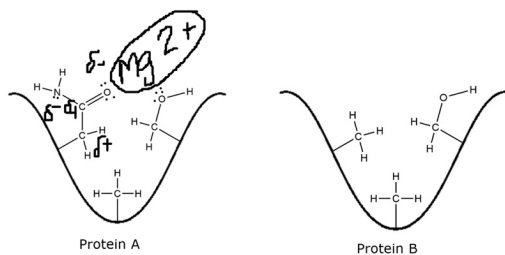
*protein A- O carries -2 charge Mg carries 2 charge they will form a ion-ion attraction*

**Explain what causes the magnesium ion to bind to the protein.**

*protein A has a better set of lone pair on oxygen and can bind much easier*

Figure 8.118. Response from student V3\_GC2\_127.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

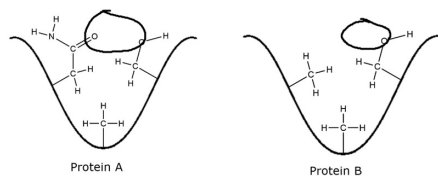
*Protein a can make ion-ion interactions and make ion-dipole interactions bc of h2o which are stronger and attract Mg2+ more.*

Explain what causes the magnesium ion to bind to the protein.

*Protein b can only make ion-ion interactions they are not as strong as the other protein so Mg2+ would not attract as much /:*

Figure 8.119. Response from student V3Alt\_GC2\_128.

Pick the binding site you think is most likely to bind the magnesium ion and draw the ion in the binding site showing why it is binding in that site.



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

*Because the O has lone pairs such that can have interaction with Mg+.*

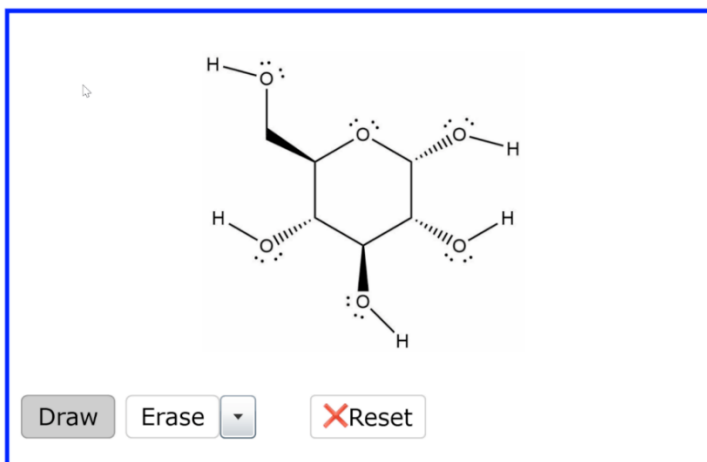
Explain what causes the magnesium ion to bind to the protein.

*Because the O has lone pairs such that can have interaction with Mg+.*

Figure 8.120. Response from student V3Alt\_GC2\_129.

## APPENDIX E: Screenshots of each activity version

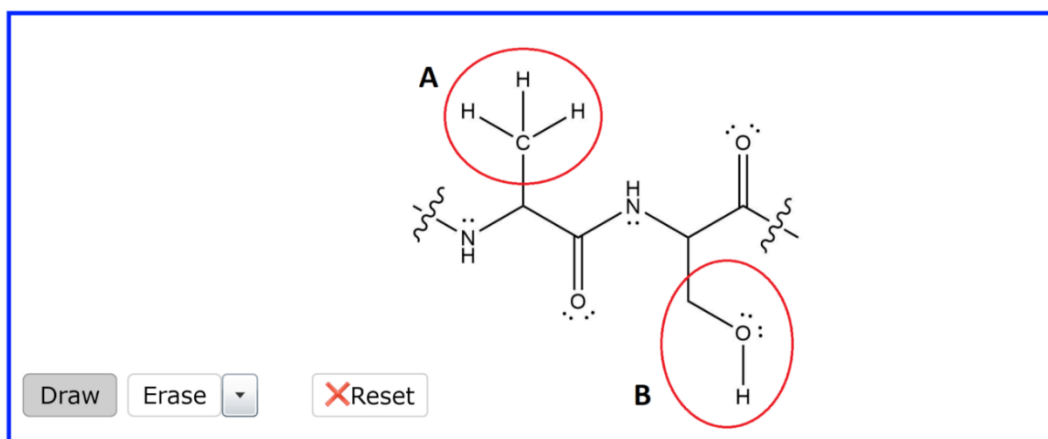
To the right, a glucose molecule is shown. Please draw any appropriate partial charges in this molecule.



Why did you draw the charges where you did in the glucose molecule?  
Hint: think about the role of the subatomic particles.

Figure 8.121. PL task version 1 slide one of three.

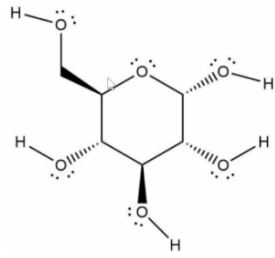
Here is part of a peptide showing two amino acids A and B. Please draw any appropriate partial charges in the side chains (circled) of each amino acid.



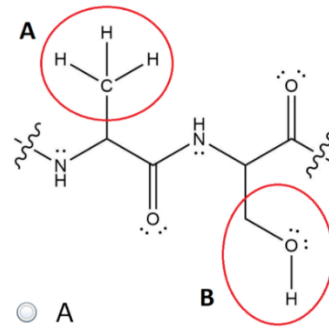
Why did you draw the partial charges where you did?

Figure 8.122. PL task version 1 slide two of three.





Which amino acid would glucose more strongly interact with?



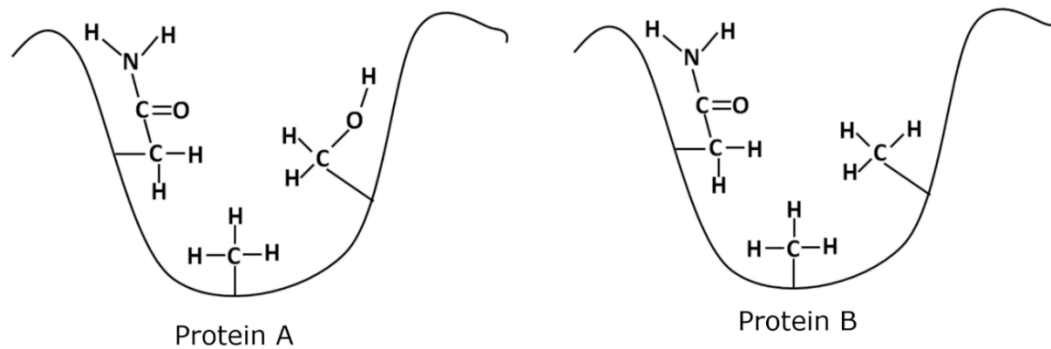
☐ A

☐ B

Why does the glucose interact more strongly with that amino acid?

Figure 8.123. PL task version 1 slide three of three.

The drawings below represent binding sites in two different versions of a protein showing only the atoms in relevant amino acid side chains.



Think about how these binding sites might bind glucose (shown on the right)

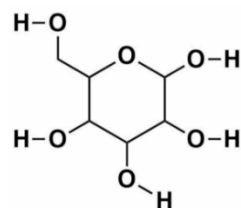


Figure 8.124. PL task version 2 (original and alternate) slide one of four.

Pick the binding site you think is most likely to bind glucose (shown on the right) and draw a possible way glucose could bind in the binding site.

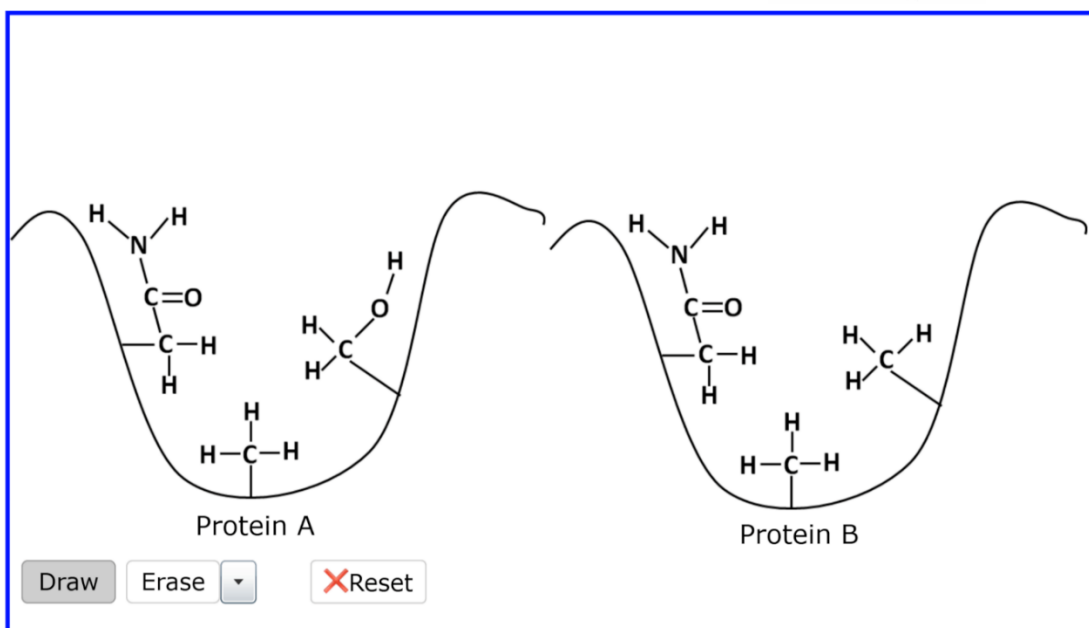
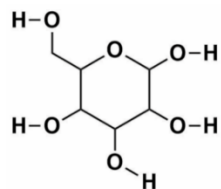


Figure 8.125. PL task version 2 (original and alternate) slide two of four.

In the space below, explain **why the protein you chose has the better glucose binding site** and **how the differences in the site cause this difference** in binding. Include specific references to the figures. Feel free to click the "back" button if you would like to look again at your drawing

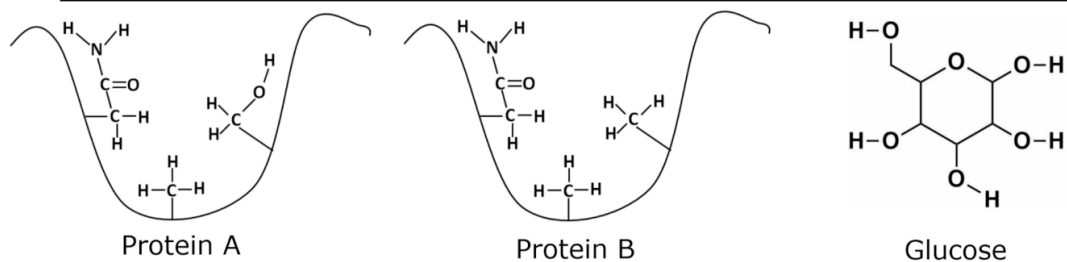
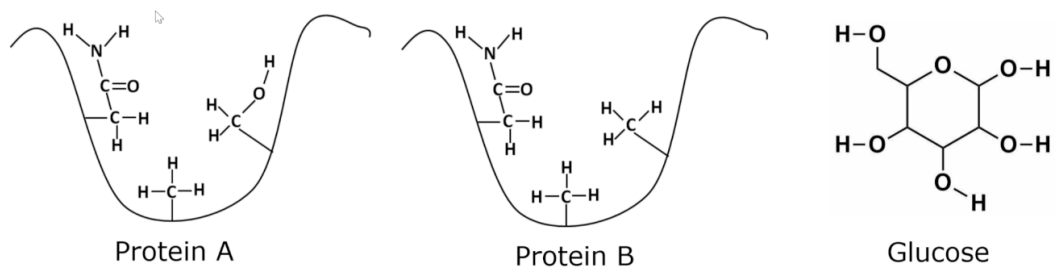


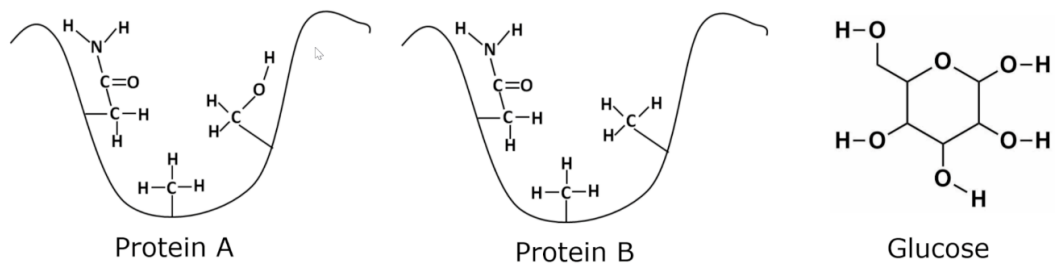
Figure 8.126. PL task version 2 (original) slide three of four.



In the space below, use your drawing to help explain **what causes glucose to bind to the protein.**



Figure 8.127. PL task version 2 (original) slide four of four.



In the space below, use your drawing to help explain **what causes glucose to bind to the protein**. Feel free to click the "back" button if you would like to look again at your drawing

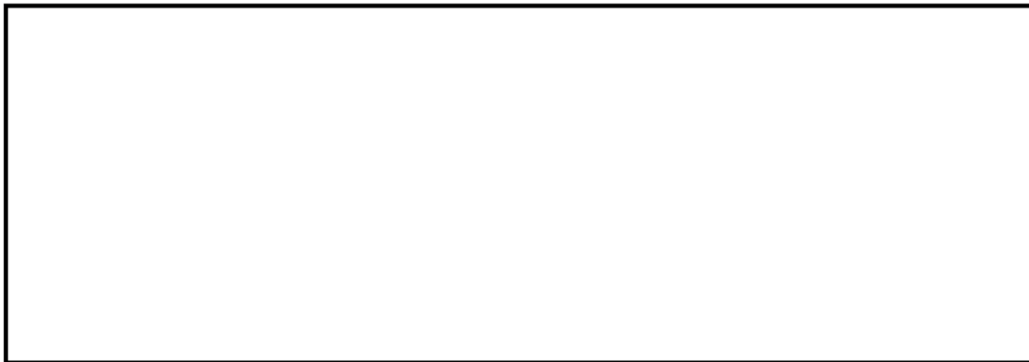


Figure 8.128. PL task version 2 (alternate) slide three of four.

In the space below, explain **why the protein you chose has the better glucose binding site** and **how the differences in the site cause this difference** in binding. Include specific references to the figures.

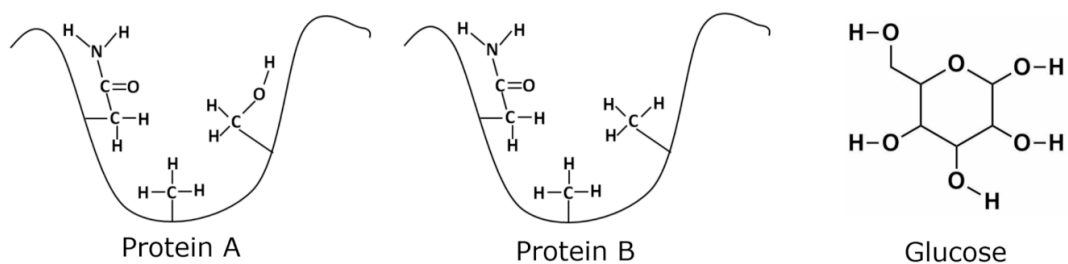


Figure 8.129. PL task version 2 (alternate) slide four of four.

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

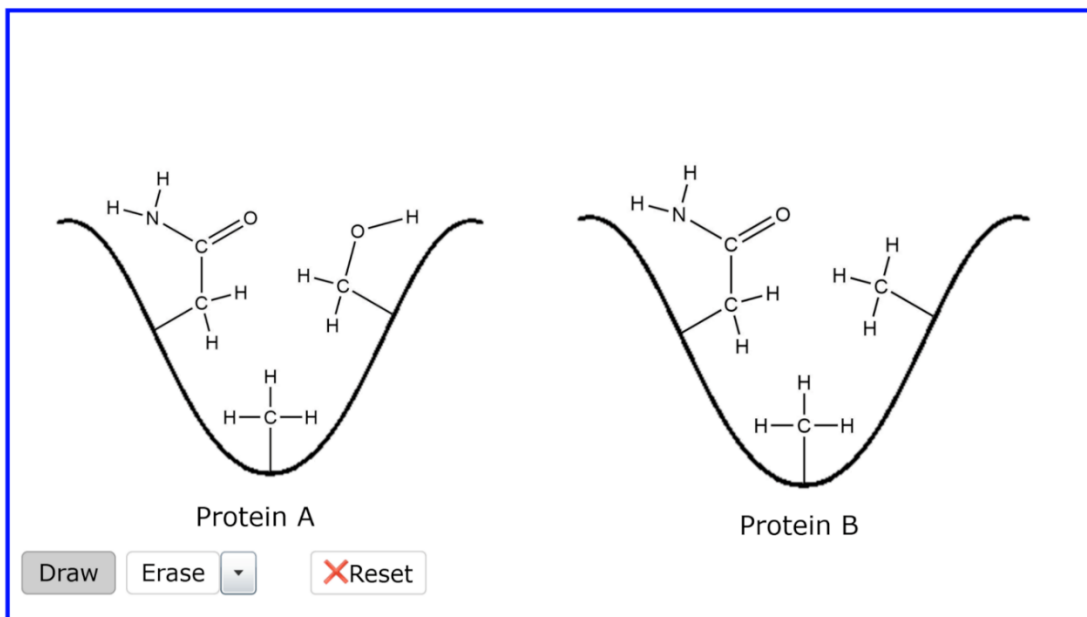
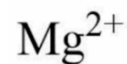


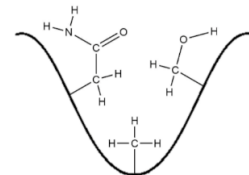
Figure 8.130. PL task version 3 (original) slide one of two.



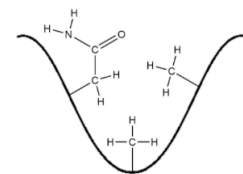
**Explain why the protein you chose has the better magnesium binding site** and how the structural differences in the site cause this difference in binding. **Include specific references to the figures.**

**Explain what causes the magnesium ion to bind to the protein** making specific references to your drawing of the magnesium ion in the binding site.

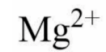
Click the "back" button if you would like to look at your drawing again.



Protein A



Protein B



Magnesium

Figure 8.131. PL task version 3 (original) slide two of two.

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

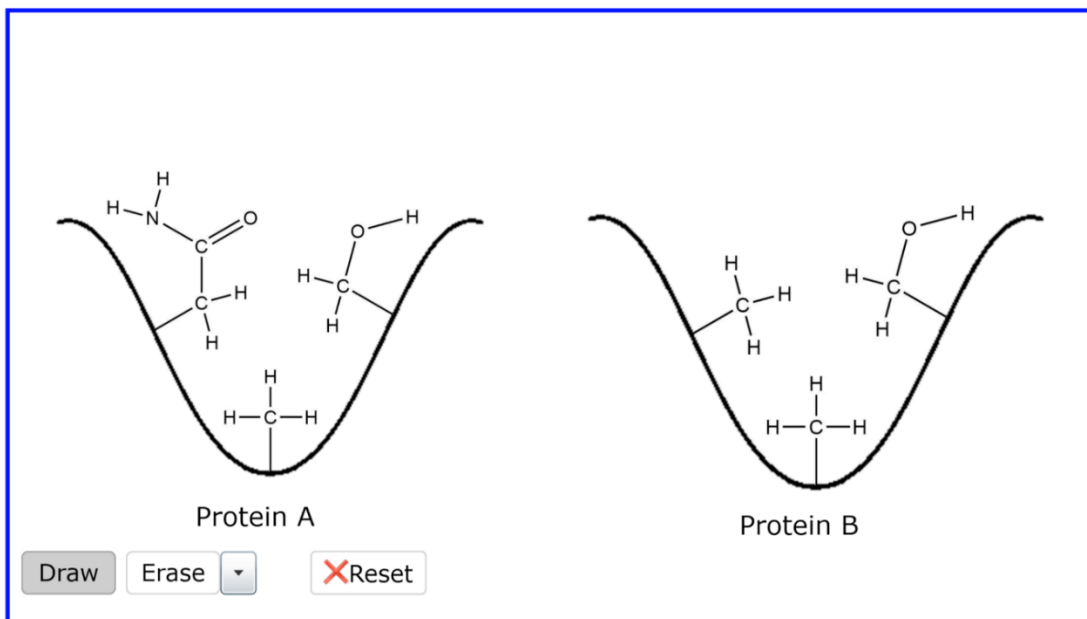
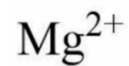
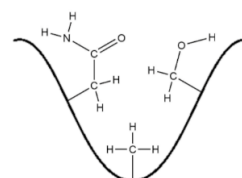


Figure 8.132. PL task version 3 (alternate) slide one of two.

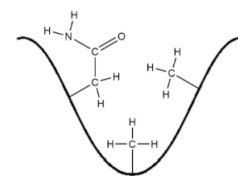
**Explain why the protein you chose has the better magnesium binding site** and how the structural differences in the site cause this difference in binding. **Include specific references to the figures.**

Click the "back" button if you would like to look at your drawing again.

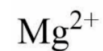


Protein A

**Explain what causes the magnesium ion to bind to the protein** making specific references to your drawing of the magnesium ion in the binding site.



Protein B



Magnesium

Figure 8.133. PL task version 3 (alternate) slide two of two. Note: we accidentally included protein B from the original version of this slide.

The drawings below represent binding sites in two different versions of Protein M showing only the atoms in relevant amino acid side chains. Consider a positively charged magnesium ion (shown on the right). Pick the binding site you think is most likely to bind the magnesium ion and **draw** the ion in the binding site **showing why it is binding** in that site.

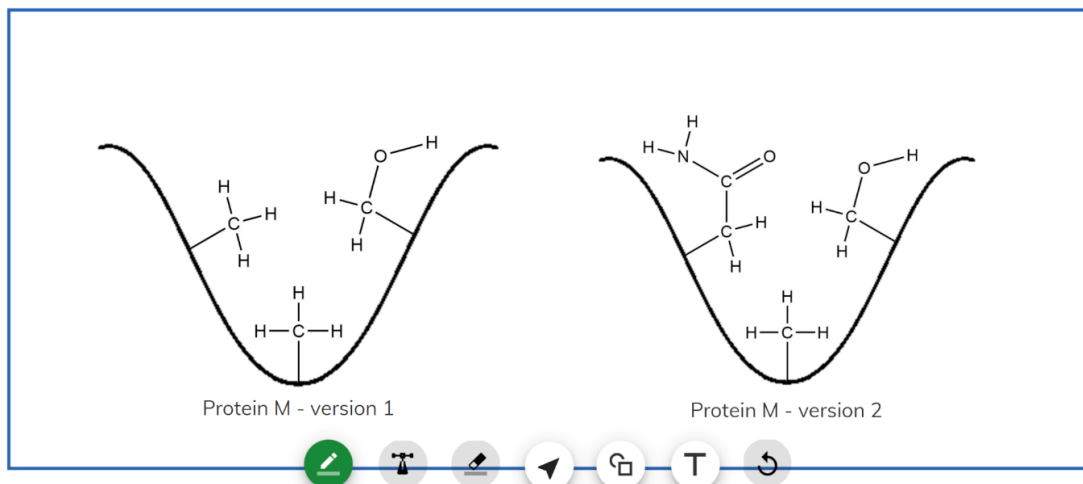
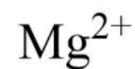


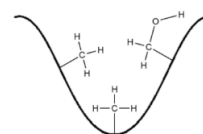
Figure 8.134. PL task final version slide one of two.

In the space below, **explain what causes** the magnesium ion to bind to the protein **making specific references to your drawing**.

Explain here...

Click the "previous" button in the bottom left corner if you would like to look at your drawing again.

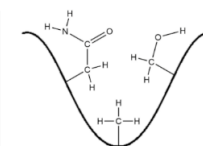
Protein M - version 1



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

Explain here...

Protein M - version 2



Magnesium

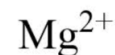
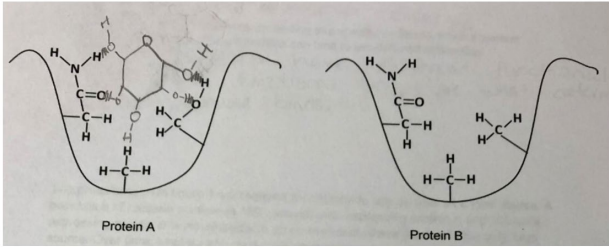


Figure 8.135. PL task final version slide two of two.

## APPENDIX F: Example responses mentioning hydrogen bonding

In Figure 8.136, two students both explained that the glucose would bind with whichever site could form more hydrogen bonding interactions with glucose. However, the students disagreed on which function group forms these interactions (hydroxyl vs methyl groups), which led them to select two different sites. This highlights the importance of students developing a deeper, causal mechanistic understanding of this interaction, one that is well anchored to the core ideas of the discipline.

Liam

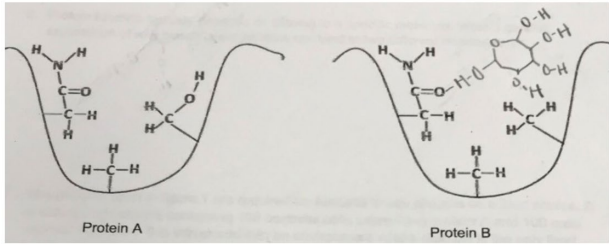


Protein A                      Protein B

**Explain why the protein you chose has the better glucose binding site and how the differences in the site cause this difference in binding.**  
*Protein A allows hydrogen bonding to occur. Whereas Protein B does not replaces its O-H group with a CH<sub>3</sub> group which does not allow hydrogen bonding to occur. CH<sub>3</sub> only allows hydrophobic clustering to occur.*

**Explain what causes glucose to bind to the protein.**  
*The oxygens in the glucose molecule bind to the hydrogen in the hydroxyl group. Since more bonds can be formed in protein A, Glucose will bind to protein A*

Paul



Protein A                      Protein B

**Explain why the protein you chose has the better glucose binding site and how the differences in the site cause this difference in binding.**  
*I think that Protein B has a better glucose binding site. The differences in the sites are that Protein A only has one CH<sub>3</sub> while protein B has two CH<sub>3</sub>. This gives more places for hydrogen bonding and where the glucose can attack on.*

**Explain what causes glucose to bind to the protein.**  
*Glucose binds to the protein. This is due to hydrogen bonding that occurs with the CH<sub>3</sub>*

Figure 8.136. Two MB students' responses to PL task version 2 who explain hydrogen bonding as the cause for the protein-ligand binding but attribute the interaction to two different groups and do not explicitly connect to charge or polarity.

## CHAPTER IX - CONCLUSIONS, IMPLICATIONS, AND FUTURE RESEARCH

### Conclusions

#### *CLUE helps students develop a deeper understanding of LDFs*

In this thesis, I explored how CLUE students understand London dispersion forces (LDFs). IMFs, such as LDFs, are important in chemistry but are often misunderstood.<sup>3,4,39</sup> Recent research has identified that CLUE students can identify where these interactions occur and predict the consequences of these interactions,<sup>46,47,112,121</sup> but in this thesis I dug deeper into how students explain *how* and *why* LDFs occur.

In study 1, we followed a cohort of students across the two-semester CLUE sequence. In the exam following initial instruction of LDFs, almost 90% of the students provided a causal mechanistic response (either text or drawing) in the exam. In the second semester, this number dropped to about 40% of the students, but over 80% still provided at least some electrostatic basis for this interaction. This is important because prior studies have found that students struggle to identify where such interactions occur,<sup>3,4,39</sup> let alone explain how or why they occur. The fact that the overwhelming proportion of the students in our sample left this course able to tie LDFs back to a core idea in chemistry (electrostatics) is very encouraging.

Furthermore, we investigated, in studies 2 and 3, how tens of thousands of CLUE students think about LDFs by using automated technologies. We found that in any given fall semester, about a quarter of those students provided a causal mechanistic explanation for the formation of LDFs on their homework activity, while about 75% provide some electrostatic evidence for the attraction (the percentage drops for those in the “off”-sequence offering of CLUE, but only slightly). Even if just a quarter of the students provide a causal mechanism, this course typically serves over 2,000 students each fall which would mean that 500 students are engaging in this difficult form of reasoning. This is

especially encouraging when we consider that, in study 2, we observed many students (over 70%) from a traditional curriculum providing non-electrostatic responses to this task. Overall, this data shows that CLUE is having a productive impact on students' understanding of this IMF.

***Scaffolding and iterative design are useful in eliciting causal mechanistic reasoning***

Across these studies, I designed assessments to elicit causal mechanistic explanations of IMFs. Eliciting these types of explanations can be difficult;<sup>23</sup> the assessment must be nuanced enough to convey to students what reasoning I want them to include in their explanation, but not so heavy-handed that the activity becomes rote. I found that scaffolding techniques, specifically the reduction of degrees of freedom and marking of critical features, were useful in developing assessments that struck the appropriate balance.<sup>17</sup>

In study 1, we found that it was helpful to include multiple boxes when asking students to “draw what happens” as the two neutral atoms approached one another. By modifying the response environment in this way, we encouraged students to show a process and incorporate the temporal component necessary to address the causal mechanism by which these LDFs occur. In study 5, we found that by using contrasting cases we could focus students' attention on the scalar level at which the structure of the two cases differed. This helped nudge students to step down a scalar level and unpack the properties of those entities, a key feature of causal mechanistic reasoning.

Of course, even the most well-intentioned assessment designer will often find that their initial activity did not perform as expected. In this thesis, I have found iterative design, the process of using students' responses to guide subsequent revisions, to be a critical component of assessment design. It is only once we have seen how the assessment performs “in the wild” that we learn if the intended resources were activated by the assessment. For example, we noticed that students were not interacting with our potential energy question (study 4) in a causal mechanistic manner, so we modified

the activity wording and added an additional phenomenon to the task. When examining protein-ligand binding in study 5, we found the presence of hydrogen bonding stymied causal mechanistic reasoning. Once we modified the phenomenon to incorporate a ligand which did not experience this type of interaction, we were much more successful in eliciting causal mechanistic responses.

### ***The “messy middle” of causal mechanistic reasoning***

One key feature of causal mechanistic reasoning is the unpacking of properties and behaviors of the underlying entities.<sup>5,6</sup> This component is part of what makes causal mechanistic reasoning such a powerful form of thinking. By building strong connections between the entities, their properties, and the resulting behaviors, students may be better able to predict what happens when any one of those parts is changed, or how those entities produce similar phenomenon.

In study 4, we explored the relationship between how students construct causal mechanistic explanations of a phenomenon related to forces (formation of LDFs) and potential energy (relative depth of potential energy wells). After conducting statistical analyses, we identified a significant relationship between how students responded to both tasks. While the effect size of this test is modest, we found that this relationship was driven by positive associations between categories you would expect. For example, those who provided a causal mechanistic explanation to the LDF task were more likely to provide one to the PE task. While there was a similar positive association between the non-electrostatic bins, there was no such relationship for the intermediate bins however (electrostatic causal and partially causal mechanistic).

This was also true when comparing the text and drawing LDF responses in study 1. Those who provided a causal mechanistic text response were more likely to provide a causal mechanistic drawing response, and the same was true for the non-electrostatic text and drawing responses. However, there was once again no positive association between the intermediate electrostatic causal responses. Taken



together, these results suggest that there is this sort of “messy middle” where students are beginning to connect their ideas, but they have not yet fully organized their knowledge. This aligns with resources perspective which posits that knowledge is fragmented and loosely organized.

### **Implications**

#### ***The importance of long-term educational initiatives***

At this point, the evidence that CLUE can help students to develop a deeper understanding of forces and interactions is quite impressive. It begs the question then: what about CLUE has led it to be this successful? While there are a number of features of CLUE that one could certainly point to, I believe one of the most powerful aspects of CLUE is its long-term nature. That is, there is not one activity which highlights the importance of intermolecular forces, but the whole course emphasizes these connections over two entire semesters. While smaller interventions are often logistically easier to implement, learning chemistry is difficult and we need to give students the opportunity to take the time needed to build deep and meaningful connections between the concepts. Going forward, the chemistry education research community at large should continue to explore long-term and structural ways to support our students.

#### ***Automated resources can facilitate the use of explanation questions***

Explanation questions provide us with powerful evidence about what students know and can do. For example, consider the responses to the “identify” and “explain” questions asked in study 4. While nearly all the students were able to correctly identify which pair of interacting atoms had the deeper potential energy well, much fewer could provide a fully causal mechanistic explanation of why this difference exists. That is, depending on what we ask, we get a very different picture of students’ knowledge. However, administering explanation questions is not always practical in large enrollment courses due to the resource intensive nature of the required analysis. Fortunately, automated resources provide a way for instructors and researchers to characterize students’ written responses in a quick and

accurate manner, even for very large groups of students. Additionally, by using constructed response questions which ask students to explain how and why phenomena occur, we send students the message that complex forms of thinking, like causal mechanistic reasoning, is valued.

### ***Supporting students in the “messy middle”***

While these studies have gathered encouraging evidence of students engaged in causal mechanistic reasoning, it is important to acknowledge that not all these students provided fully causal mechanistic accounts of the phenomena. Many are stuck in an intermediate zone, leveraging some of the right ideas, but not all of them. Therefore, we need to continue to give students opportunities to connect these resources in productive ways across multiple contexts. We have made great progress in helping students to use ideas like electrostatics to think about forces and interactions, but we must look for further ways to support students learning.

### **Future directions**

The studies outlined in this thesis primarily explored how students think about role of LDFs in chemical phenomena. Future research in this area should extend beyond just LDFs and explore how students use causal mechanistic reasoning to think about other interactions, like hydrogen bonding. Additionally, the connection between how students think about interactions and energy requires further investigation. While I also explored this relationship, the findings (from study 4) indicated that the link between how students thought about LDFs and potential energy was quite modest. If we can better understand how students think about these two ideas, we can figure out how to better support their learning. Finally, forces and interactions play an important role in biological phenomenon as well. While only the development of the protein-ligand binding activity was discussed, we plan to use this task in the future to explore how chemistry and biology students explain this phenomenon. Beyond this single activity, there is significantly more work to be done in this space to support students' interdisciplinary

understanding of science. For example, future work could investigate how students think about the role of adenosine triphosphate or ATP, a molecule frequently associated with energy in biology contexts.

## REFERENCES

## REFERENCES

- (1) National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, D.C., 2012.  
<https://doi.org/10.17226/13165>.
- (2) Cooper, M. M.; Klymkowsky, M. W. The Trouble with Chemical Energy: Why Understanding Bond Energies Requires an Interdisciplinary Systems Approach. *CBE—Life Sciences Education* **2013**, 12 (2), 306–312. <https://doi.org/10.1187/cbe.12-10-0170>.
- (3) Henderleiter, J.; Smart, R.; Anderson, J.; Elian, O. How Do Organic Chemistry Students Understand and Apply Hydrogen Bonding? *J. Chem. Educ.* **2001**, 78 (8), 1126–1130.  
<https://doi.org/10.1021/ed078p1126>.
- (4) Peterson, R. F.; Treagust, D. F.; Garnett, P. Development and Application of a Diagnostic Instrument to Evaluate Grade-11 and -12 Students' Concepts of Covalent Bonding and Structure Following a Course of Instruction. *Journal of Research in Science Teaching* **1989**, 26 (4), 301–314.  
<https://doi.org/10.1002/tea.3660260404>.
- (5) Krist, C.; Schwarz, C. V.; Reiser, B. J. Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *J. Learn. Sci.* **2019**, 28 (2), 160–205.  
<https://doi.org/10.1080/10508406.2018.1510404>.
- (6) Russ, R. S.; Scherr, R. E.; Hammer, D.; Mikeska, J. Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Science Education* **2008**, 92 (3), 499–525. <https://doi.org/10.1002/sce.20264>.
- (7) Becker, N.; Noyes, K.; Cooper, M. M. Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *J. Chem. Educ.* **2016**, 93 (10), 1713–1724.  
<https://doi.org/10.1021/acs.jchemed.6b00298>.
- (8) Bodner, G. M. Constructivism: A Theory of Knowledge. *J. Chem. Educ.* **1986**, 63 (10), 873.  
<https://doi.org/10.1021/ed063p873>.
- (9) Piaget, J. Part I: Cognitive Development in Children: Piaget Development and Learning. *J Res Sci Teach* **1964**, 2, 176–186. <https://doi.org/10.1002/tea.3660020306>.
- (10) diSessa, A. A. A History of Conceptual Change Research. In *The Cambridge Handbook of the Learning Sciences*; Sawyer, R. K., Ed.; Cambridge University Press: Cambridge, 2014; pp 88–108.  
<https://doi.org/10.1017/CBO9781139519526.007>.
- (11) Posner, G. J.; Strike, K. A.; Hewson, P. W.; Gertzog, W. A. Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Sci. Ed.* **1982**, 66 (2), 211–227.  
<https://doi.org/10.1002/sce.3730660207>.
- (12) Hammer, D. Student Resources for Learning Introductory Physics. *Am. J. Phys.* **2000**, 68 (S1), S52–S59. <https://doi.org/10.1119/1.19520>.

- (13) Minstrell, J. Facets of Students' Knowledge and Relevant Instruction. In *Research in physics learning: theoretical issues and empirical studies*; Duit, R., Goldberg, F., Niedderer, H., Eds.; Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel: Kiel, 1992; pp 110–128.
- (14) Hammer, D.; Elby, A.; Scherr, R. E.; Redish, E. F. Resources, Framing, and Transfer. In *Transfer of Learning from a Modern Multidisciplinary Perspective*; Information Age Publishing Inc.: Greenwich, CT, 2005; pp 89–119.
- (15) *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*; National Academies Press: Washington, D.C., 2000. <https://doi.org/10.17226/9853>.
- (16) Mislevy, R. J.; Haertel, G. D. Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice* **2007**, 25 (4), 6–20. <https://doi.org/10.1111/j.1745-3992.2006.00075.x>.
- (17) Wood, D.; Bruner, J. S.; Ross, G. The Role of Tutoring in Problem Solving. *J. Child. Psychol. Psys.* **1976**, 17 (2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>.
- (18) Vygotsky, L. S. Interaction between Learning and Development. In *Mind in Society: The Development of Higher Psychological Processes*; Cole, M., John-Steiner, V., Scribner, S., Souberman, E., Eds.; Harvard University Press: Cambridge, MA, 1978; pp 79–91.
- (19) Bruner, J. S. The Inspiration of Vygotsky. In *Actual minds, possible worlds*; Harvard University Press: Cambridge, MA, 1986; pp 70–78.
- (20) Shepard, L. A. Linking Formative Assessment to Scaffolding. *Educational Leadership* **2005**, 63 (3), 66–70.
- (21) Shepard, L. A. The Role of Assessment in a Learning Culture. *Educ. Researcher* **2000**, 29 (7), 4–14. <https://doi.org/10.3102/0013189X029007004>.
- (22) Mislevy, R. J.; Almond, R. G.; Lukas, J. F. *A Brief Introduction to Evidence-Centered Design*; Research Report RR-03-16; Educational Testing Service, 2003.
- (23) Reiser, B. J. Scaffolding Complex Learning: The Mechanisms of Structuring and Problematising Student Work. *Journal of the Learning Sciences* **2004**, 13 (3), 273–304. [https://doi.org/10.1207/s15327809jls1303\\_2](https://doi.org/10.1207/s15327809jls1303_2).
- (24) McNeill, K. L.; Lizotte, D. J.; Krajcik, J.; Marx, R. W. Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *Journal of the Learning Sciences* **2006**, 15 (2), 153–191. [https://doi.org/10.1207/s15327809jls1502\\_1](https://doi.org/10.1207/s15327809jls1502_1).
- (25) *Knowing What Students Know: The Science and Design of Educational Assessment*; Pellegrino, J. W., Chudowsky, N., Glaser, R., Eds.; National Academies Press: Washington, D.C., 2001.
- (26) Brandriet, A.; Rupp, C. A.; Lazenby, K.; Becker, N. M. Evaluating Students' Abilities to Construct Mathematical Models from Data Using Latent Class Analysis. *Chem. Educ. Res. Pract.* **2018**, 19 (1), 375–391. <https://doi.org/10.1039/C7RP00126F>.

- (27) Becker, N. M.; Rupp, C. A.; Brandriet, A. Engaging Students in Analyzing and Interpreting Data to Construct Mathematical Models: An Analysis of Students' Reasoning in a Method of Initial Rates Task. *Chem. Educ. Res. Pract.* **2017**, *18* (4), 798–810. <https://doi.org/10.1039/C6RP00205F>.
- (28) Bishop, B. A.; Anderson, C. W. Student Conceptions of Natural Selection and Its Role in Evolution. *J. Res. Sci. Teach.* **1990**, *27* (5), 415–427.
- (29) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid–Base Reactions. *Journal of Chemical Education* **2016**, *93* (10), 1703–1712. <https://doi.org/10.1021/acs.jchemed.6b00417>.
- (30) Jin, H.; Anderson, C. W. Developing Assessments For A Learning Progression on Carbon-Transforming Processes in Socio-Ecological Systems. In *Learning Progressions in Science*; Alonzo, A. C., Gotwals, A. W., Eds.; SensePublishers: Rotterdam, 2012; pp 151–181. [https://doi.org/10.1007/978-94-6091-824-7\\_8](https://doi.org/10.1007/978-94-6091-824-7_8).
- (31) Bodé, N. E.; Deng, J. M.; Flynn, A. B. Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *J. Chem. Educ.* **2019**, *96* (6), 1068–1082. <https://doi.org/10.1021/acs.jchemed.8b00719>.
- (32) Macrie-Shuck, M.; Talanquer, V. Exploring Students' Explanations of Energy Transfer and Transformation. *J. Chem. Educ.* **2020**, *97* (12), 4225–4234. <https://doi.org/10.1021/acs.jchemed.0c00984>.
- (33) Caspari, I.; Kranz, D.; Graulich, N. Resolving the Complexity of Organic Chemistry Students' Reasoning through the Lens of a Mechanistic Framework. *Chemistry Education Research and Practice* **2018**. <https://doi.org/10.1039/C8RP00131F>.
- (34) Graulich, N.; Caspari, I. Designing a Scaffold for Mechanistic Reasoning in Organic Chemistry. *Chem. Teach. Int.* **2020**, *3* (1), 19–30. <https://doi.org/10.1515/cti-2020-0001>.
- (35) Graulich, N.; Schween, M. Concept-Oriented Task Design: Making Purposeful Case Comparisons in Organic Chemistry. *J. Chem. Educ.* **2018**, *95* (3), 376–383. <https://doi.org/10.1021/acs.jchemed.7b00672>.
- (36) Moreira, P.; Marzabal, A.; Talanquer, V. Using a Mechanistic Framework to Characterise Chemistry Students' Reasoning in Written Explanations. *Chemistry Education Research and Practice* **2018**. <https://doi.org/10.1039/C8RP00159F>.
- (37) Cooper, M. M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116–1122. <https://doi.org/10.1021/ed300456y>.
- (38) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLOS ONE* **2016**, *11* (9). <https://doi.org/10.1371/journal.pone.0162333>.

- (39) Cooper, M. M.; Williams, L. C.; Underwood, S. M. Student Understanding of Intermolecular Forces: A Multimodal Study. *J. Chem. Educ.* **2015**, *92* (8), 1288–1298. <https://doi.org/10.1021/acs.jchemed.5b00169>.
- (40) Gayford, C. ATP: A Coherent View for School Advanced Level Studies in Biology. *Journal of Biological Education* **1986**, *20* (1), 27–32. <https://doi.org/10.1080/00219266.1986.9654772>.
- (41) Johnstone, A. H.; Mahmoud, N. A. Isolating Topics of High Perceived Difficulty in School Biology. *Journal of Biological Education* **1980**, *14* (2), 163–166. <https://doi.org/10.1080/00219266.1980.10668983>.
- (42) Novick, S. No Energy Storage in Chemical Bonds. *Journal of Biological Education* **1976**, *10* (3), 116–118. <https://doi.org/10.1080/00219266.1976.9654072>.
- (43) Kohn, K. P.; Underwood, S. M.; Cooper, M. M. Energy Connections and Misconnections across Chemistry and Biology. *Cell Biology Education* **2018**, *17* (1). <https://doi.org/10.1187/cbe.17-08-0169>.
- (44) Cooper, M. M.; Posey, L. A.; Underwood, S. M. Core Ideas and Topics: Building Up or Drilling Down? *J. Chem. Educ.* **2017**, *94* (5), 541–548. <https://doi.org/10.1021/acs.jchemed.6b00900>.
- (45) Cooper, M. M.; Stowe, R. L. Chemistry Education Research—From Personal Empiricism to Evidence, Theory, and Informed Practice. *Chem. Rev.* **2018**, *118* (12), 6053–6087. <https://doi.org/10.1021/acs.chemrev.8b00020>.
- (46) Williams, L. C.; Underwood, S. M.; Klymkowsky, M. W.; Cooper, M. M. Are Noncovalent Interactions an Achilles Heel in Chemistry Education? A Comparison of Instructional Approaches. *J. Chem. Educ.* **2015**, *92* (12), 1979–1987. <https://doi.org/10.1021/acs.jchemed.5b00619>.
- (47) Stowe, R. L.; Herrington, D. G.; McKay, R. L.; Cooper, M. M. The Impact of Core-Idea Centered Instruction on High School Students' Understanding of Structure–Property Relationships. *J. Chem. Educ.* **2019**, *96* (7), 1327–1340. <https://doi.org/10.1021/acs.jchemed.9b00111>.
- (48) Cooper, M. M.; Corley, L. M.; Underwood, S. M. An Investigation of College Chemistry Students' Understanding of Structure–Property Relationships. *J. Res. Sci. Teach.* **2013**, *50* (6), 699–721. <https://doi.org/10.1002/tea.21093>.
- (49) Othman, J.; Treagust, D. F.; Chandrasegaran, A. L. An Investigation into the Relationship between Students' Conceptions of the Particulate Nature of Matter and Their Understanding of Chemical Bonding. *International Journal of Science Education* **2008**, *30* (11), 1531–1550. <https://doi.org/10.1080/09500690701459897>.
- (50) Pierri, E.; Karatrantou, A.; Panagiotakopoulos, C. Exploring the Phenomenon of “change of Phase” of Pure Substances Using the Microcomputer-Based-Laboratory (MBL) System. *Chem. Educ. Res. Pract.* **2008**, *9* (3), 234–239. <https://doi.org/10.1039/B812412B>.
- (51) Cooper, M. M.; Underwood, S. M.; Hilley, C. Z. Development and Validation of the Implicit Information from Lewis Structures Instrument (IILSI): Do Students Connect Structures with Properties? *Chem. Educ. Res. Pract.* **2012**, *13* (3), 195–200. <https://doi.org/10.1039/C2RP00010E>.



- (52) Maeyer, J.; Talanquer, V. The Role of Intuitive Heuristics in Students' Thinking: Ranking Chemical Substances. *Science Education* **2010**, 94 (6), 963–984. <https://doi.org/10.1002/sce.20397>.
- (53) Cooper, M. M.; Klymkowsky, M. W. CLUE: Chemistry, Life, the Universe and Everything <https://clue.chemistry.msu.edu/> (accessed 2018 -12 -02).
- (54) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, 1996.
- (55) Underwood, S. M.; Posey, L. A.; Herrington, D. G.; Carmel, J. H.; Cooper, M. M. Adapting Assessment Tasks To Support Three-Dimensional Learning. *Journal of Chemical Education* **2018**, 95 (2), 207–217. <https://doi.org/10.1021/acs.jchemed.7b00645>.
- (56) Kahneman, D. *Thinking, Fast and Slow*; Farrar, Straus and Giroux: New York, 2011.
- (57) Feynman, R. P.; Leighton, R. B.; Sands, M. *The Feynman Lectures on Physics*; Addison-Wesley: New York, 1963.
- (58) Boo, H. K. Students' Understandings of Chemical Bonds and the Energetics of Chemical Reactions. *Journal of Research in Science Teaching* **1998**, 35 (5), 569–581. [https://doi.org/10.1002/\(SICI\)1098-2736\(199805\)35:5<569::AID-TEA6>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1098-2736(199805)35:5<569::AID-TEA6>3.0.CO;2-N).
- (59) Stone, C. A. What Is Missing in the Metaphor of Scaffolding? In *Contexts for Learning: Sociocultural Dynamics in Children's Development*; Oxford University Press: New York, NY, 1993; p 15.
- (60) van de Pol, J.; Volman, M.; Beishuizen, J. Scaffolding in Teacher–Student Interaction: A Decade of Research. *Educational Psychology Review* **2010**, 22 (3), 271–296. <https://doi.org/10.1007/s10648-010-9127-6>.
- (61) Hogan, K.; Pressley, M. Scaffolding Scientific Competencies within Classroom Communities of Inquiry. In *Scaffolding Student Learning: Instructional Approaches and Issues*; Brookline Books: Cambridge, MA, 1997; pp 74–107.
- (62) Ge, X.; Land, S. M. A Conceptual Framework for Scaffolding Ill-Structured Problem-Solving Processes Using Question Prompts and Peer Interactions. *Educational Technology Research and Development* **2004**, 52 (2), 5–22.
- (63) Bryfczynski, S. BeSocratic: An Intelligent Tutoring System for the Recognition, Evaluation, and Analysis of Free-Form Student Input. Ph.D. Dissertation, Clemson University, Clemson, SC, 2012.
- (64) *NVivo Qualitative Data Analysis Software*; QSR International Pty Ltd., 2012.
- (65) Green, S. B.; Salkind, N. J. Two-Way Contingency Table Analysis Using Crosstabs. In *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*; Pearson Education Inc.: Upper Saddle River, NJ, 2011; pp 366–376.
- (66) Cohen, J. A Power Primer. *Psychol. Bull.* **1992**, 112 (1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>.

- (67) Agresti, A. Inference for Two-Way Contingency Tables. In *Categorical Data Analysis*; Wiley series in probability and statistics; Wiley: Hoboken, NJ, USA, 2013; pp 69–112.
- (68) IBM Corp. *SPSS Statistics for Windows*; IBM Corp: Armonk, NY, 2017.
- (69) MacDonald, P. L.; Gardner, R. C. Type I Error Rate Comparisons of Post Hoc Procedures for I j Chi-Square Tables. *Educ. Psychol. Meas.* **2000**, *60* (5), 735–754.  
<https://doi.org/10.1177/00131640021970871>.
- (70) Green, S. B.; Salkind, N. J. Independent-Samples t Test. In *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*; Pearson Education Inc.: Upper Saddle River, NJ, 2011; pp 175–182.
- (71) Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20* (1), 37–46. <https://doi.org/10.1177/001316446002000104>.
- (72) Landis, J. R.; Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33* (1), 159. <https://doi.org/10.2307/2529310>.
- (73) Barker, V.; Millar, R. Students' Reasoning about Basic Chemical Thermodynamics and Chemical Bonding: What Changes Occur during a Context-Based Post-16 Chemistry Course? *International Journal of Science Education* **2000**, *22* (11), 1171–1200.  
<https://doi.org/10.1080/09500690050166742>.
- (74) Taber, K. S. Building the Structural Concepts of Chemistry: Some Considerations from Educational Research. *Chem. Educ. Res. Pract.* **2001**, *2* (2), 123–158.  
<https://doi.org/10.1039/B1RP90014E>.
- (75) Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, *96* (9), 1821–1832.  
<https://doi.org/10.1021/acs.jchemed.9b00455>.
- (76) Nehm, R. H.; Schonfeld, I. S. Measuring Knowledge of Natural Selection: A Comparison of the CINS, an Open-Response Instrument, and an Oral Interview. *Journal of Research in Science Teaching* **2008**, *45* (10), 1131–1160. <https://doi.org/10.1002/tea.20251>.
- (77) Hubbard, J. K.; Potts, M. A.; Couch, B. A. How Question Types Reveal Student Thinking: An Experimental Comparison of Multiple-True-False and Free-Response Formats. *CBE-Life Sci Educ* **2017**, *16* (2), ar26. <https://doi.org/10.1187/cbe.16-12-0339>.
- (78) Lee, H.-S.; Liu, O. L.; Linn, M. C. Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. *Appl. Meas. Educ.* **2011**, *24* (2), 115–136.  
<https://doi.org/10.1080/08957347.2011.554604>.
- (79) Pashler, H.; Bain, P. M.; Bottge, B. A.; Graesser, A.; Koedinger, K.; McDaniel, M.; Metcalfe, J. *Organizing Instruction and Study to Improve Student Learning*, 2007.  
<https://doi.org/10.1037/e607972011-001>.

- (80) Crandell, O. M.; Kouyoumdjian, H.; Underwood, S. M.; Cooper, M. M. Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. *J. Chem. Educ.* **2019**, *96* (2), 213–226. <https://doi.org/10.1021/acs.jchemed.8b00784>.
- (81) Crandell, O. M.; Lockhart, M. A.; Cooper, M. M. Arrows on the Page Are Not a Good Gauge: Evidence for the Importance of Causal Mechanistic Explanations about Nucleophilic Substitution in Organic Chemistry. *J. Chem. Educ.* **2020**, *97* (2), 313–327. <https://doi.org/10.1021/acs.jchemed.9b00815>.
- (82) Black, P.; Wiliam, D. Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice* **1998**, *5* (1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- (83) Sadler, D. R. Formative Assessment: Revisiting the Territory. *Assessment in Education: Principles, Policy & Practice* **1998**, *5* (1), 77–84. <https://doi.org/10.1080/0969595980050104>.
- (84) Dood, A. J.; Fields, K. B.; Raker, J. R. Using Lexical Analysis To Predict Lewis Acid–Base Model Use in Responses to an Acid–Base Proton-Transfer Reaction. *J. Chem. Educ.* **2018**, *95* (8), 1267–1275. <https://doi.org/10.1021/acs.jchemed.8b00177>.
- (85) Dood, A. J.; Dood, J. C.; Cruz-Ramírez de Arellano, D.; Fields, K. B.; Raker, J. R. Analyzing Explanations of Substitution Reactions Using Lexical Analysis and Logistic Regression Techniques. *Chem. Educ. Res. Pract.* **2020**, *21* (1), 267–286. <https://doi.org/10.1039/C9RP00148D>.
- (86) Haudek, K. C.; Prevost, L. B.; Moscarella, R. A.; Merrill, J.; Urban-Lurain, M. What Are They Thinking? Automated Analysis of Student Writing about Acid-Base Chemistry in Introductory Biology. *Cell Biology Education* **2012**, *11* (3), 283–293. <https://doi.org/10.1187/cbe.11-08-0084>.
- (87) Kaplan, Jennifer J.; Haudek, Kevin C.; Ha, Minsu. Using Lexical Analysis Software to Assess Student Writing in Statistics. *Technology Innovations in Statistics Education* **2014**, *8* (1).
- (88) Prevost, L. B.; Smith, M. K.; Knight, J. K. Using Student Writing and Lexical Analysis to Reveal Student Thinking about the Role of Stop Codons in the Central Dogma. *CBE—Life Sciences Education* **2016**, *15* (4), ar65. <https://doi.org/10.1187/cbe.15-12-0267>.
- (89) Sripathi, K. N.; Moscarella, R. A.; Yoho, R.; You, H. S.; Urban-Lurain, M.; Merrill, J.; Haudek, K. Mixed Student Ideas about Mechanisms of Human Weight Loss. *LSE* **2019**, *18* (3), ar37. <https://doi.org/10.1187/cbe.18-11-0227>.
- (90) Automated Analysis of Constructed Response <https://beyondmultiplechoice.org/> (accessed 2022 -04 -06).
- (91) Haudek, K. C.; Kaplan, J. J.; Knight, J.; Long, T.; Merrill, J.; Munn, A.; Nehm, R.; Smith, M.; Urban-Lurain, M. Harnessing Technology to Improve Formative Assessment of Student Conceptions in STEM: Forging a National Network. *CBE-Life. Sci. Educ.* **2011**, *10* (2), 149–155. <https://doi.org/10.1187/cbe.11-03-0019>.
- (92) Sieke, S. A.; McIntosh, B. B.; Steele, M. M.; Knight, J. K. Characterizing Students’ Ideas about the Effects of a Mutation in a Noncoding Region of DNA. *LSE* **2019**, *18* (2), ar18. <https://doi.org/10.1187/cbe.18-09-0173>.

- (93) Jurka, T. P.; Collingwood, L.; Boydston, A. E.; Grossman, E.; Atteveldt, W. V. *RTextTools: Automatic Text Classification via Supervised Learning*; 2012.
- (94) Nehm, R. H.; Ha, M.; Mayfield, E. Transforming Biology Assessment with Machine Learning: Automated Scoring of Written Evolutionary Explanations. *J Sci Educ Technol* **2012**, *21* (1), 183–196. <https://doi.org/10.1007/s10956-011-9300-9>.
- (95) David Opitz; Richard Maclin. Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research* **1999**, *11*, 169–198.
- (96) Ha, M.; Nehm, R. H.; Urban-Lurain, M.; Merrill, J. E. Applying Computerized-Scoring Models of Written Biological Explanations across Courses and Colleges: Prospects and Limitations. *LSE* **2011**, *10* (4), 379–393. <https://doi.org/10.1187/cbe.11-08-0081>.
- (97) Theodore L. Brown; H. Eugene LeMay, Jr; Bruce E. Bursten; Catherine J. Murphy; Patrick M. Woodward; Matthew W. Stoltzfus. *Chemistry: The Central Science*, 14th ed.; Pearson: New York, 2018.
- (98) Lage, M. J.; Platt, G. J.; Treglia, M. Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. *The Journal of Economic Education* **2000**, *31* (1), 30–43. <https://doi.org/10.1080/00220480009596759>.
- (99) Qualtrics; Qualtrics: Provo, Utah.
- (100) Jurka, T. P.; Collingwood, L.; Boydston, A. E.; Grossman, E.; van, W. RTextTools: A Supervised Learning Package for Text Classification. *The R Journal* **2013**, *5* (1), 6–12.
- (101) Feinerer, I.; Hornik, K.; Meyer, D. Text Mining Infrastructure in R. *Journal of Statistical Software* **2008**, *25* (5). <https://doi.org/10.18637/jss.v025.i05>.
- (102) Hearst, M. A.; Dumais, S. T.; Osuna, E.; Platt, J.; Scholkopf, B. Support Vector Machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13* (4), 18–28. <https://doi.org/10.1109/5254.708428>.
- (103) Mcauliffe, J. D.; Blei, D. M. Supervised Topic Models. In *Advances in neural information processing systems*; 2008; pp 121–128.
- (104) Friedman, J. H.; Hastie, T.; Tibshirani, R. Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics* **2000**, *28* (2), 337–407.
- (105) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, 1984.
- (106) Hothorn, T.; Lausen, B. Bundling Classifiers by Bagging Trees. *Computational Statistics & Data Analysis* **2005**, *49* (4), 1068–1078. <https://doi.org/10.1016/j.csda.2004.06.019>.
- (107) Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.
- (108) Friedman, J.; Hastie, T.; Tibshirani, R. Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics* **2008**, *9* (3), 432–441. <https://doi.org/10.1093/biostatistics/kxm045>.

- (109) Kazama, J.; Tsujii, J. Evaluation and Extension of Maximum Entropy Models with Inequality Constraints. In *Proceedings of the 2003 conference on Empirical methods in natural language processing* -; Association for Computational Linguistics: Not Known, 2003; Vol. 10, pp 137–144. <https://doi.org/10.3115/1119355.1119373>.
- (110) Williamson, D. M.; Xi, X.; Breyer, F. J. A Framework for Evaluation and Use of Automated Scoring. *Educational Measurement: Issues and Practice* **2012**, 31 (1), 2–13. <https://doi.org/10.1111/j.1745-3992.2011.00223.x>.
- (111) American Educational Research Association; American Psychological Association; National Council on Measurement in Education. Fairness in Testing. In *The Standards for Educational and Psychological Testing*; 2014; pp 49–72.
- (112) Cooper, M. M.; Underwood, S. M.; Hilley, C. Z.; Klymkowsky, M. W. Development and Assessment of a Molecular Structure and Properties Learning Progression. *J. Chem. Educ.* **2012**, 89 (11), 1351–1357. <https://doi.org/10.1021/ed300083a>.
- (113) Noyes, K.; McKay, R. L.; Neumann, M.; Haudek, K. C.; Cooper, M. M. Developing Computer Resources to Automate Analysis of Students' Explanations of London Dispersion Forces. *J. Chem. Educ.* **2020**, 97 (11), 3923–3936. <https://doi.org/10.1021/acs.jchemed.0c00445>.
- (114) Dhawan, S. Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems* **2020**, 49 (1), 5–22. <https://doi.org/10.1177/0047239520934018>.
- (115) Adedoyin, O. B.; Soykan, E. Covid-19 Pandemic and Online Learning: The Challenges and Opportunities. *Interactive Learning Environments* **2020**, 1–13. <https://doi.org/10.1080/10494820.2020.1813180>.
- (116) Song, L.; Singleton, E. S.; Hill, J. R.; Koh, M. H. Improving Online Learning: Student Perceptions of Useful and Challenging Characteristics. *The Internet and Higher Education* **2004**, 7 (1), 59–70. <https://doi.org/10.1016/j.iheduc.2003.11.003>.
- (117) Wang, X.; Hegde, S.; Son, C.; Keller, B.; Smith, A.; Sasangohar, F. Investigating Mental Health of US College Students During the COVID-19 Pandemic: Cross-Sectional Survey Study. *J Med Internet Res* **2020**, 22 (9), e22817. <https://doi.org/10.2196/22817>.
- (118) Lantz, B. The Large Sample Size Fallacy. *Scandinavian Journal of Caring Sciences* **2013**, 27 (2), 487–492. <https://doi.org/10.1111/j.1471-6712.2012.01052.x>.
- (119) Royall, R. M. The Effect of Sample Size on the Meaning of Significance Tests. *The American Statistician* **1986**, 40 (4), 313–315. <https://doi.org/10.1080/00031305.1986.10475424>.
- (120) Sullivan, G. M.; Feinn, R. Using Effect Size—or Why the *P* Value Is Not Enough. *Journal of Graduate Medical Education* **2012**, 4 (3), 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>.
- (121) Kararo, A. T.; Colvin, R. A.; Cooper, M. M.; Underwood, S. M. Predictions and Constructing Explanations: An Investigation into Introductory Chemistry Students' Understanding of Structure–Property Relationships. *Chem. Educ. Res. Pract.* **2019**, 20 (1), 316–328. <https://doi.org/10.1039/C8RP00195B>.

- (122) Cooper, M. M. The Crosscutting Concepts: Critical Component or “Third Wheel” of Three-Dimensional Learning? *J. Chem. Educ.* **2020**, *97* (4), 903–909.  
<https://doi.org/10.1021/acs.jchemed.9b01134>.
- (123) Barak, J.; Gorodetsky, M.; Chipman, D. Understanding of Energy in Biology and Vitalistic Conceptions. *International Journal of Science Education* **1997**, *19* (1), 21–30.  
<https://doi.org/10.1080/0950069970190102>.
- (124) Goldring, H.; Osborne, J. Students’ Difficulties with Energy and Related Concepts. *Physics Education* **1994**, *29* (1), 26–32. <https://doi.org/10.1088/0031-9120/29/1/006>.
- (125) Nagel, M. L.; Lindsey, B. A. Student Use of Energy Concepts from Physics in Chemistry Courses. *Chem. Educ. Res. Pract.* **2015**, *16* (1), 67–81. <https://doi.org/10.1039/C4RP00184B>.
- (126) Becker, N. M.; Cooper, M. M. College Chemistry Students’ Understanding of Potential Energy in the Context of Atomic-Molecular Interactions. *J. Res. Sci. Teach.* **2014**, *51* (6), 789–808.  
<https://doi.org/10.1002/tea.21159>.
- (127) Jewett, J. W. Energy and the Confused Student II: Systems. *The Physics Teacher* **2008**, *46* (2), 81–86. <https://doi.org/10.1119/1.2834527>.
- (128) Lindsey, B. A.; Heron, P. R. L.; Shaffer, P. S. Student Understanding of Energy: Difficulties Related to Systems. *American Journal of Physics* **2012**, *80* (2), 154–163.  
<https://doi.org/10.1119/1.3660661>.
- (129) IBM Corp. *SPSS Statistics for Windows*; IBM Corp: Armonk, NY, 2020.
- (130) Krist, C.; Schwarz, C. V.; Reiser, B. J. Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *Journal of the Learning Sciences* **2019**, *28* (2), 160–205. <https://doi.org/10.1080/10508406.2018.1510404>.
- (131) Hammer, D.; Berland, L. K. Confusing Claims for Data: A Critique of Common Practices for Presenting Qualitative Research on Learning. *Journal of the Learning Sciences* **2014**, *23* (1), 37–46. <https://doi.org/10.1080/10508406.2013.802652>.
- (132) Russ, R. S.; Scherr, R. E.; Hammer, D.; Mikeska, J. Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Science Education* **2008**, *92* (3), 499–525. <https://doi.org/10.1002/sce.20264>.
- (133) Hammer, D. Student Resources for Learning Introductory Physics. *American Journal of Physics* **2000**, *68* (S1), S52–S59. <https://doi.org/10.1119/1.19520>.
- (134) Hammer, D.; Elby, A.; Scherr, R. E.; Redish, E. F. Resources, Framing, and Transfer. In *Transfer of learning from a modern multidisciplinary perspective*; Greenwich, CT, 2005; pp 89–120.
- (135) Smith, J. P. I.; diSessa, A. A.; Roschelle, J. Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *Journal of the Learning Sciences* **1994**, *3* (2), 115–163.  
[https://doi.org/10.1207/s15327809jls0302\\_1](https://doi.org/10.1207/s15327809jls0302_1).

- (136) National Research Council. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academies Press: Washington, DC, 2001. <https://doi.org/10.17226/10019>.
- (137) Mislevy, R. J.; Almond, R. G.; Lukas, J. F. A Brief Introduction to Evidence-Centered Design. *ETS Research Report Series* **2003**, 2003 (1), i–29. <https://doi.org/10.1002/j.2333-8504.2003.tb01908.x>.
- (138) Wood, D.; Bruner, J. S.; Ross, G. The Role of Tutoring in Problem Solving\*. *Journal of Child Psychology and Psychiatry* **1976**, 17 (2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>.
- (139) Vygotsky, L. S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press, 1980.
- (140) Bruner, J. S. *Actual Minds, Possible Worlds*; Harvard University Press: Cambridge, Mass, 1986.
- (141) Kang, H.; Thompson, J.; Windschitl, M. Creating Opportunities for Students to Show What They Know: The Role of Scaffolding in Assessment Tasks. *Science Education* **2014**, 98 (4), 674–704. <https://doi.org/10.1002/sce.21123>.
- (142) Caspari, I.; Kranz, D.; Graulich, N. Resolving the Complexity of Organic Chemistry Students' Reasoning through the Lens of a Mechanistic Framework. *Chemistry Education Research and Practice* **2018**, 19 (4), 1117–1141. <https://doi.org/10.1039/C8RP00131F>.
- (143) Graulich, N.; Caspari, I. Designing a Scaffold for Mechanistic Reasoning in Organic Chemistry. *Chemistry Teacher International* **2021**, 3 (1), 19–30. <https://doi.org/10.1515/cti-2020-0001>.
- (144) Graulich, N.; Schween, M. Concept-Oriented Task Design: Making Purposeful Case Comparisons in Organic Chemistry. *J. Chem. Educ.* **2018**, 95 (3), 376–383. <https://doi.org/10.1021/acs.jchemed.7b00672>.
- (145) Bishop, B. A.; Anderson, C. W. Student Conceptions of Natural Selection and Its Role in Evolution. *Journal of Research in Science Teaching* **1990**, 27 (5), 415–427. <https://doi.org/10.1002/tea.3660270503>.
- (146) Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, 96 (9), 1821–1832. <https://doi.org/10.1021/acs.jchemed.9b00455>.
- (147) Jin, H.; Anderson, C. W. *Developing Assessments for a Learning Progression on Carbon-Transforming Processes in Socio-Ecological Systems*; Brill, 2012; pp 149–181.
- (148) Brandriet, A.; Rupp, C. A.; Lazenby, K.; Becker, N. M. Evaluating Students' Abilities to Construct Mathematical Models from Data Using Latent Class Analysis. *Chem. Educ. Res. Pract.* **2018**, 19 (1), 375–391. <https://doi.org/10.1039/C7RP00126F>.
- (149) Cooper, M. M.; Stowe, R. L.; Crandell, O. M.; Klymkowsky, M. W. Organic Chemistry, Life, the Universe and Everything (OCLUE): A Transformed Organic Chemistry Curriculum. *J. Chem. Educ.* **2019**, 96 (9), 1858–1872. <https://doi.org/10.1021/acs.jchemed.9b00401>.

- (150) Cooper, M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116–1122. <https://doi.org/10.1021/ed300456y>.
- (151) Bryfczynski, S. BeSocratic: An Intelligent Tutoring System for the Recognition, Evaluation, and Analysis of Free-Form Student Input. *All Dissertations* **2012**.
- (152) Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **1960**, *20* (1), 37–46. <https://doi.org/10.1177/001316446002000104>.
- (153) Landis, J. R.; Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33* (1), 159–174.
- (154) Cooper, M. M.; Corley, L. M.; Underwood, S. M. An Investigation of College Chemistry Students' Understanding of Structure–Property Relationships. *Journal of Research in Science Teaching* **2013**, *50* (6), 699–721. <https://doi.org/10.1002/tea.21093>.
- (155) Cooper, M. M.; Williams, L. C.; Underwood, S. M. Student Understanding of Intermolecular Forces: A Multimodal Study. *J. Chem. Educ.* **2015**, *92* (8), 1288–1298. <https://doi.org/10.1021/acs.jchemed.5b00169>.
- (156) Williams, L. C.; Underwood, S. M.; Klymkowsky, M. W.; Cooper, M. M. Are Noncovalent Interactions an Achilles Heel in Chemistry Education? A Comparison of Instructional Approaches. *J. Chem. Educ.* **2015**, *92* (12), 1979–1987. <https://doi.org/10.1021/acs.jchemed.5b00619>.
- (157) Henderleiter, J.; Smart, R.; Anderson, J.; Elian, O. How Do Organic Chemistry Students Understand and Apply Hydrogen Bonding? *J. Chem. Educ.* **2001**, *78* (8), 1126. <https://doi.org/10.1021/ed078p1126>.
- (158) Peterson, R. F.; Treagust, D. F.; Garnett, P. Development and Application of a Diagnostic Instrument to Evaluate Grade-11 and -12 Students' Concepts of Covalent Bonding and Structure Following a Course of Instruction. *Journal of Research in Science Teaching* **1989**, *26* (4), 301–314. <https://doi.org/10.1002/tea.3660260404>.
- (159) Jescovitch, L. N.; Scott, E. E.; Cerchiara, J. A.; Merrill, J.; Urban-Lurain, M.; Doherty, J. H.; Haudek, K. C. Comparison of Machine Learning Performance Using Analytic and Holistic Coding Approaches Across Constructed Response Assessments Aligned to a Science Learning Progression. *J Sci Educ Technol* **2021**, *30* (2), 150–167. <https://doi.org/10.1007/s10956-020-09858-0>.
- (160) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid–Base Reactions. *J. Chem. Educ.* **2016**, *93* (10), 1703–1712. <https://doi.org/10.1021/acs.jchemed.6b00417>.
- (161) Crandell, O. M.; Kouyoumdjian, H.; Underwood, S. M.; Cooper, M. M. Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. *J. Chem. Educ.* **2019**, *96* (2), 213–226. <https://doi.org/10.1021/acs.jchemed.8b00784>.
- (162) Torrance, H. Assessment as Learning? How the Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-secondary Education and Training Can Come to Dominate



- Learning. *Assessment in Education: Principles, Policy & Practice* **2007**, 14 (3), 281–294. <https://doi.org/10.1080/09695940701591867>.
- (163) Shepard, L. A. The Role of Assessment in a Learning Culture. *Educational Researcher* **2000**, 29 (7), 4–14. <https://doi.org/10.3102/0013189X029007004>.
- (164) Schwarz, C.; Cooper, M.; Long, T.; Trujillo, C.; Noyes, K.; de Lima, J.; Kesh, J.; Stolfus, J. Mechanistic Explanations Across Undergraduate Chemistry and Biology Courses. In *ICLS 2020 Proceedings*; International Society of the Learning Sciences: Nashville, TN, USA, 2020; pp 625–628. <https://doi.org/10.22318/icls2020.625>.