12-2015

Understanding How Synthetic Organic Chemistry Graduate Students Navigate Scifinder

Lindsey Whitfield Cain
Clemson University

UNDERSTANDING HOW SYNTHETIC ORGANIC CHEMISTRY GRADUATE STUDENTS NAVIGATE SCIFINDER

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemistry

---

by
Lindsey Whitfield Cain
December 2015

Accepted by:
Dr. Gautam Bhattacharyya, Committee Co-Chair
Dr. Steve Stuart, Committee Co-Chair
Dr. Julie Martin
Dr. Rhett Smith
Dr. Michelle Cook

ABSTRACT

Students pursuing a Ph.D. degree are expected to contribute research to their field, for which the success depends, in part, on their ability to find, interpret, and use scholarly information from the primary literature. However, studies from the information sciences show that graduate students from a variety of fields, including the sciences, frequently struggle to comprehensively search their respective dissertation topics because of insufficient prior content knowledge and lack of guidance from their disciplinary community. This body of literature is consistent with the results of my previous research of chemistry graduate students' laboratory decision-making processes. Specifically, that study showed their search and evaluation of the scientific literature to find a research protocol was critical to the success or failure of the students' research. For these reasons, I chose to investigate how synthetic organic chemistry graduate students perform literature searches, using SciFinder, to find protocols for preparing previously unreported compounds. For my study, I used situated cognition and communities of practice (CoP) as my theoretical frameworks in conjunction with an ethnomethodological research design. Five organic chemistry graduate students were interviewed to understand their strategies and sense-making procedures for searching the literature, specifically focusing on how they decide to: 1). input a topic or structural representation, 2). evaluate the search results, and 3). use specific procedures for deciding which of the protocols to carry out in the laboratory.

The findings from my study indicated that the graduate students' information-seeking behaviors and sense-making procedures were directly influenced by their

domain-specific content knowledge and their exposure to the organic CoP. Specifically, the second-year and third-year graduate students heavily depended on the database because their domain-specific content knowledge was not operational at this stage in their training. Comparatively, the sixth-year graduate students could easily use their organic chemistry knowledge—i.e. named organic reactions, functional group chemistry, and the retrosynthetic approach—to propose a research protocol; therefore, they used the database to substantiate their synthesis protocols and/or to find a method to synthesis the proposed starting materials. As a result of their exposure to their Ph.D. research, the graduates had become more proficient with using the database and developed heuristics to evaluate their searches, thereby allowing them to quickly evaluate the often times substantial amount of hits. Finally, the findings indicated that the graduate students were utilizing their CoP in different ways. For instance, the second-year and third-year graduate students would seek their advisor's approval, whereas the sixth-year graduate students would seek their peers' feedback regarding their protocols. Findings from my study can broadly be integrated into the information science field to enhance and improve undergraduate and graduate students' ISB. Furthermore, my findings can be applied to improve how we educate and train organic chemistry students (both at the undergraduate- and graduate-level).

DEDICATION

To my loving husband, mom, and dad, who have supported me throughout this journey.

ACKNOWLEDGMENTS

First and foremost, I would like to thank God for opening this door for me. I would like to thank my family for earnestly praying and supporting me as I pursued my passion for teaching. To Dr. Bhattacharyya, I am truly grateful for the opportunity to not only join your group, but for giving me the freedom to explore my research interests. Also, thank you for helping me see this to fruition. Dr. Martin, thank you for the opportunities you have given me and for accepting me as a member of the SoCap research group. Thank you, Dr. Smith, for encouraging me to take chances and to pursue my passion for teaching. Dr. Cook, thank you for all the wonderful feedback regarding my dissertation. Thank you, Barbara Lewis, for *always* supporting me. To all the former Bhattacharyya group members (Yu Shen, Drea, and Nicole), thank you for all support and constructive feedback. To the *entire* Engineering and Science Education department, thank you so much for accepting me into you community and encouraging me to persist.

TABLE OF CONTENTS

Page

TITLE PAGE i

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

I. INTRODUCTION 1

Personal Background and Research Interest 3

II LITERATURE REVIEW 5

Research on Graduate Students' Information-Seeking Behaviors (ISB) 6
Research on Chemistry Students' ISB 12
Research in Graduate Organic Chemistry Education 14
Summary of Literature Review 23

III. METHODOLOGY 25

Summary of Previous Study 25
Guiding Research Questions 25
Theoretical Considerations 26
Methodology Considerations 27
Participants and Setting 27
Data Collection and Analysis 28
Results and Discussion 29
Dissertation Study 33
Motivation and Research Questions 33
A. Theoretical Frameworks 35
Situated Cognition 35

Table of Contents (Continued)

Page

Communities of Practice 37
Synthesis of the Theoretical Perspective 38
B. Methodological Framework 39
C. Methods 42
Participants and Setting 42
SciFinder Background and Search Key 43
Data Collection 47
Data Analysis 53
Role of the Researcher 54
Validity and Trustworthiness 55

IV. RESULTS AND DISCUSSION 58

Jane 58
Background 58
Ph.D. Research 59
Task 1 Search Chronology 61
Task 1 Search Critique 73
Summary of Jane's Search 92
Peter 93
Background 93
Ph.D. Research 94
Task 1 Search Chronology 95
Task 1 Search Critique 103
Summary of Peter's Search 115
Sophie 116
Background 116
Ph.D. Research 118
Task 1 Search Chronology 119
Task 1 Search Critique 131
Summary of Sophie's Search 148
Bruce 149
Background 149
Ph.D. Research 149
Task 1 Search Chronology 151
Task 1 Search Critique 157
Summary of Bruce's Search 166
Luke 166
Background 166
Ph.D. Research 166

Table of Contents (Continued)

Page

Task 1 Search Chronology 168
Task 1 Search Critique 175
Summary of Luke's Search 186

V. CONCLUSION AND IMPLICATIONS 188

Cross-Case Analysis 188
Answering the Guiding Research Questions 200
Implications 201
Summary 205

APPENDICES 206

A: Understanding Chemical Research: Incorporating Theoretical Constructs into the Laboratory 207
B: Consent Forms for Participation in a Research Study 208

REFERENCES 210

LIST OF TABLES

Table Page

3.1 Options in SciFinder for substance and reaction searches 46

3.2 Options in SciFinder for designating the role of the compound when performing a reaction search (Rx) 46

3.3 Description of data collected 53

LIST OF FIGURES

Figure		Page
3.1	Components of my dissertation study adapted from Crotty (1998)	35
3.2	Screenshot of SciFinder search options	44
3.3	Screenshot of structure editor screen taken from SciFinder	45
3.4	Screenshot taken from SciFinder of an RxS1 search for the Task 1 compound	47
3.5	Interview #1 protocol	49
3.6	Interview #2 and/or #3 protocol	52
4.1	SciFinder screenshot of the SS search results	62
4.2	SciFinder screenshot of reaction #1 from the search for substance #1 from the SS search as a product	63
4.3	SciFinder screenshot of the SSi search results	64
4.4	SciFinder screenshot of reaction #1 from the search for substance #1 from the SSi search as a product	65
4.5	Sketch of starting material	66
4.6	SciFinder screenshot of reaction #16 from the search for the noted starting material as the role of a product	68
4.7	SciFinder screenshot of the transformation search using the starting material from the *JACS* article	70
4.8	SciFinder screenshot of the transformation search using the starting material from the JACS article to a simplified version of the target compound	71
4.9	SciFinder screenshot of the RxS1 search for the simplified target compound	72
4.10	Jane's sketch of synthesis protocol	73

List of Figures (Continued)

Figure Page

4.11 Target compound (a) and substance from the SS search (b) 77

4.12 Target compound (a) and substance from the SSi search (c)..... 79

4.13 Substance #4 from the SE search..... 82

4.14 Starting material discussed from the SSi modified search..... 83

4.15 Complex starting material from reaction #3 from the modified SS search 87

4.16 SciFinder screenshot of the SS search 96

4.17 SciFinder screenshot of the RxS1 search for the modified target compound..... 97

4.18 Sketch of S_N2 reaction 97

4.19 SciFinder screenshot of reaction #3 from the RxS1 search 98

4.20 SciFinder screenshot of reaction #4 from the RxS1 search 99

4.21 SciFinder screenshot of reaction #5 from the RxS1 search 100

4.22 SciFinder screenshot of the transformation search using the simplified starting material 100

4.23 SciFinder screenshot of simplified transformation search..... 101

4.24 SciFinder screenshot of reaction #2 from the RxS search 102

4.25 SciFinder screenshot of reaction #4 from the RxS search 102

4.26 Peter's sketch of synthesis protocol..... 103

4.27 SciFinder screenshot of reaction #1 from the RxS1 search 106

4.28 Target compound (a), product from reaction #1 (b), and product from reaction #2 (c) 107

List of Figures (Continued)

Figure Page

4.29 Screenshot of the complex starting material from reaction/hit #1 109

4.30 SciFinder screenshot of reaction #4 from the RxS search for the simplified target compound 110

4.31 SciFinder screenshot of SSi search results..... 120

4.32 SciFinder screenshot of modified target compound with bromine group 121

4.33 SciFinder screenshot of the SSi search result for the modified target compound with a bromine group 121

4.34 SciFinder screenshot of the modified target compound without the benzyl group..... 122

4.35 SciFinder screenshot of substances from the SSi search for the modified target compound without the benzyl group 123

4.36 SciFinder screenshot of reaction #4 from the search for substance #3 as a product 123

4.37 SciFinder screenshot of reaction #5 from the search for substance #3 as a product 124

4.38 SciFinder screenshot of reaction #3 from the search for substance #3 as a product 125

4.39 SciFinder screenshot of the transformation using the new starting material without the benzyl group 125

4.40 Sketch of ideas regarding how to add the benzyl group 126

4.41 SciFinder screenshot of the transformation search from Step 8 127

4.42 Sketch for the addition of the benzyl group..... 128

4.43 SciFinder screenshot of the compound sketched in Step 11..... 128

List of Figures (Continued)

Figure Page

4.44 SciFinder screenshot of the substances that resulted from the SSi search in Step 12 129

4.45 Ideas regarding performing an alpha, beta unsaturated ketone ring closure step 129

4.46 SciFinder screenshot of the compound sketched in Step 13 for the alpha, beta unsaturated ring closure step 130

4.47 SciFinder screenshot of the transformation input for Step 14 131

4.48 Target compound (a) and substance #1 from the SSi search (b) 133

4.49 Target compound (a) and substance #1 from the SSi search (c)..... 135

4.50 Complex reagent in reaction #3 137

4.51 Target compound (a) and substance #1 from the SSi search (d) 145

4.52 SciFinder screenshot of reaction #1 from the RxS1 search 151

4.53 Sketch of the transformation reported in the article..... 152

4.54 Retrosynthetic analysis with two reaction pathways 153

4.55 SciFinder screenshot of the RxV5 search for starting material A 153

4.56 SciFinder screenshot of the RxV5 search for starting material B..... 154

4.57 SciFinder screenshot of reaction #5 from the RxV5 search for S.M. B 155

4.58 Sketched reaction to make S.M. B..... 155

4.59 SciFinder screenshot of the SE search for the starting material reported in the article..... 156

4.60 Sketch of proposed synthesis route to target compound..... 169

4.61 SciFinder screenshot of the SS search results..... 170

List of Figures (Continued)

Figure Page

4.62 SciFinder screenshot of the SSi search results 170

4.63 SciFinder screenshot of modified target compound with a bromine group 171

4.64 SciFinder screenshot of SS search result from the modified target compound 172

4.65 SciFinder screenshot of the SSi search results from Step 10 173

4.66 SciFinder screenshot of modified target compound 174

4.67 SciFinder screenshot of the SSi search results for the previously modified target compound 174

4.68 Luke's sketch of his synthesis pathway 175

4.69 Target compound (a) and substance from SSi search (b) 179

4.70 Target compound (a) and substance #4 from the SSi search (c) 183

4.71 Target compound (a) and substance #1 from the SSi search (d) 184

CHAPTER ONE

INTRODUCTION

Student engagement in authentic research experiences has garnered considerable interest from STEM education researchers over the last several decades. Their studies have focused on students of all levels – precollege (Barab & Hay, 2001; Bleicher, 1996; Etkina, Matilsky, & Lawrence, 2003; Richmond & Kurth, 1999; Ritchie & Rigano, 1996), undergraduate (Hunter, Laursen, & Seymour, 2007; Kardash, 2000; Lopatto, 2004), and graduate (Feldman, Divoll, & Rogan-Klyve, 2009) – and have been conducted from a variety of theoretical perspectives: cognitive-cultural (Nersessian, 2006), communities of practice, CoP, (Feldman, Divoll, & Rogan-Klyve, 2013; Stucky, 2005), and domain-specific epistemic development (Samarapungavan, Westby, & Bodner, 2006). Among other factors, investigating this aspect of students' education is particularly intriguing since it is one of the few, if not the only, activity in which students are expected to be knowledge *producers* in addition to their more common role as knowledge consumers. Becoming a knowledge producer requires students to develop a variety of skills, particularly the capabilities of reading, interpreting, analyzing, and contributing to the primary literature of a CoP (Feldman et al., 2009).

This acquisition of information, called one's information-seeking behavior (ISB), is defined by Wilson (2000, p. 49) as "the purposive seeking for information as a consequence of a need to satisfy some goal." However, the development of one's ISB has become increasingly challenging with the variety of information resources now available. As a result, becoming a knowledge producer and contributing research to their

respective field has become an increasingly difficult task for students. For instance, the unprecedented amount of information that can be accessed through the World Wide Web can create a complex information environment, making it difficult for students to know what information is applicable to their research topics.

Adding to the complexity, search databases and strategies vary across domains (Association of College and Research Libraries, 2000). For instance, chemists frequently use the database SciFinder, which was developed by Chemical Abstract Services (2015) and "contains more than 101 million unique organic and inorganic chemical substances." SciFinder is an exceptionally powerful database that allows individuals to search for chemical reactions, substances, protocols, patents, and journal references. A typical SciFinder search involves deciding what information to input into the database, which can be accomplished by drawing a representation of the compound, reaction, or entering a topic. Searching by representation may also require an understanding of the substructure of the compound if the compound has not been previously reported. Once the search is applied, the "hits" have to be analyzed based on their relevance to the desired search goal. As such, performing a SciFinder search necessitates that individuals are not only capable of navigating this very complex database, but also that they are able to use their content knowledge throughout each part of the search.

In addition to these considerations, a study conducted prior to my dissertation work (discussed in my Methodology chapter) revealed that chemistry graduate students' ISB were a critical resource for their research development and success. The findings suggest that chemistry graduate students' evaluation of their literature search to find a

research protocol directly influenced the outcomes of their research. For these reasons, the goal of my dissertation research is to explore how organic chemistry graduate students search SciFinder to develop research protocols to synthesize unreported compounds.

Personal Background and Research Interest

My passion for Chemistry Education was sparked during my senior year at Lander University when I had the opportunity to work as a general chemistry tutor for the Academic Success Center. Although my initial goal following graduation was to pursue a career in industry, this experience radically changed my career trajectory. I learned I thoroughly enjoyed interacting with my students and received great joy when I was able to help students reach that "aha moment."

I then applied and was accepted into the Chemistry Ph.D. program at Clemson University and soon began my research endeavors under the guidance of Dr. Bhattacharyya the summer following my first year. At that time, I was unsure about my research interests, so he gave me some chemistry education literature to read. He then gave me the creative freedom to explore my research interests, and the more I read, the more I became intrigued to understand students' methods for accomplishing tasks even though they are not approaching and/or solving them in the most conventional way.

As such, over the course of my Ph.D. training, my goals have been to understand graduate students' methods and cognitive processing for solving research-related tasks. Continuing my research journey to understand this phenomenon, my dissertation study seeks to further explore students' methods for performing a literature search (using

SciFinder) to develop a research protocol and their justifications or sense-making procedures for those steps.

CHAPTER TWO

LITERATURE REVIEW

My research goal is to further understand synthetic organic chemistry graduate students' reasonings and strategies as they perform literature searches (using SciFinder) to develop research protocols for an unreported compound. Therefore, my dissertation study draws from both the information science literature as well as the literature on the development of organic chemistry graduate students.

With respect to the literature on information seeking behaviors (ISB), I purposefully selected articles focusing on Ph.D. students because they are expected to perform comprehensive reviews of their Ph.D. research topics (Gabridge, Gaskell, & Stout, 2008). (For the purposes of this entire dissertation, the term "graduate student" will refer to those seeking a Ph.D. degree.) Related to the previous topic, I will then review the scant literature focusing on undergraduate chemistry students' experiences using SciFinder because my study has implications closely associated to this topic. To complete my literature review, I will then discuss the educational research literature on the experiences of organic chemistry graduate students. I chose to focus on a specific subset of graduate students because the organic chemistry sub-division has a culture that is unique from the other chemistry sub-divisions (Anderson, 2009; Bhattacharyya, 2014; Stucky, 2005), which, like other cultures, necessitate they use a wide variety of domain-specific skills to perform a literature search to find research protocols. Then I will conclude this chapter by explaining how my dissertation study will expand these research areas.

Research on Graduate Students' Information-Seeking Behaviors (ISB)

Information-seeking behavior (ISB) is defined as "the purposive seeking for information as a consequence of a need to satisfy some goal" (Wilson, 2000, p. 49). Since the widespread use of the World Wide Web from the mid-1990s radically changed the way information is accessed and retrieved, I will limit this review to research reported from 1999 onwards. In addition, this review will emphasize research on graduate students' ISB in the context of their Ph.D. training because of the exhaustive and comprehensive searches they are expected to conduct. Based on a search of "ISB and graduate students" in Library Information Databases, Chemistry Databases, and Education Databases, the research on graduate students' ISB, to my understanding, can be organized by the two broad themes that are based on their usage of their disciplinary resources and respective content knowledge.

Motivated to understand if academic discipline significantly influenced the graduate students' ISB, Korobili, Malliari, and Zapounidou (2011) surveyed 175 philosophy and 60 engineering graduate students. The survey consisted of questions regarding the students' background information, information retrieval techniques, frequency with which they used their search techniques, and barriers they faced with retrieving information. Based on the analysis of the questionnaire, the researchers concluded there was no major difference in the graduate students' ISB between the disciplines. However, it was unclear how the researchers made this conclusion because they did not seem to present the statistical differences between the two disciplines. In addition, Sheeja (2010) surveyed 200 social science and science graduate students to

examine which information sources they frequently searched to stay up-to-date on their research areas. As indicated in the survey, science and social science graduate students perceived electronic journals as the most effective method for finding current information (researchers did not distinguish if electronic journals were an exclusively online edition). Science students preferred to use conference proceedings (similar to journal articles) and the World Wide Web; whereas, social science students typically preferred print journals. Similarly, Jamali and Nicholas (2008) surveyed 90 physics and astronomy graduate students and 23 professors to examine which resources they searched to find current information on their respective research concentrations. The researchers classified participants by their research areas (physics or astronomy) and then further sub-categorized them by their research concentrations (theoretical, experimental, or instrumentation). Analysis by research concentration revealed instrumentation physicist and astronomers relied on conference proceedings, whereas theoretical and experimental physicists and astronomers preferred electronic journals. The results from the survey revealed that ISB, specifically the information resources used to stay current in their field, were influenced by the participants' research concentrations. Although the previous research indicated ISB were discipline specific, the focus of these studies was on the graduate students' preference of information resources, not the intricacies of how the graduate students actually searched to obtain these references.

Adding to the disciplinary resources theme, research has indicated that disciplinary communities also influence graduate students' ISB. Brahme and Walters (2010) interviewed ten nonresidential and ten residential education graduate students who

were working on or previously had completed their dissertation literature reviews. The interviews revealed that the nonresidential students were typically isolated from their respective community, resulting in feelings of frustration because they generally lacked the confidence to effectively conduct literature searches. However, the residential students did not display those feelings because they had a network of peers and advisors who were readily available to guide and assist them as they conducted their searches. In addition, Brown (2005) surveyed 25 molecular biology graduate students to determine their choice and justification for using specific bioinformatics databases and scientific journals. The students believed their knowledge of these databases was influenced by their interactions with their advisors and peers in their research group. Brown (2005) suggested the students did not use the resources provided by the library because they preferred to go to their laboratories. As such, these studies suggested that graduate students need the guidance and support of a disciplinary information community.

In addition to the disciplinary resource theme, research has indicated that graduate students' content knowledge significantly influenced their ISB. Barrett (2005) interviewed seven humanities students in various stages of their PhD training to determine if their ISB were comparable to the published norms of the respective community. Not surprisingly, Barrett concluded that the ISB of students in the early stages of their Ph.D. training resembled those of undergraduate students' ISB because they typically lacked the content knowledge to effectively search their research topics. Comparatively, students who were in the later stages of their Ph.D. training more closely resemble the faculty's ISB because the students had developed a deeper understanding of

their research areas. However, students in the later stages reported spending a substantial amount time searching the literature to acquire some background information on their research areas before they could effectively search their research topics. Adding to this topic, Chu and Law (2007) conducted a yearlong study using surveys followed by direct observations (six total) to examine the ISB of six first-year engineering and six first-year education graduate students. As expected, during the early interviews, the students knew very little about their research topics; therefore, they typically conducted broad searches to develop a general understanding of their research area. Although general searches (i.e. broad searches) in relation to a topic often result in a substantial amount of information, the results may have little relevance to the desired topic. As the year progressed, however, the researchers observed that the students were becoming more knowledgeable and articulate about their research topics; thereby, searching the literature was a far less tedious and difficult task. These studies suggested that before graduate students can conduct comprehensive in-depth searches on their research topics, they needed to have significant prior content knowledge.

To determine if there was difference between undergraduate and graduate students ISB, Gabridge, Gaskell, and Stout (2008) interviewed 16 undergraduate and 16 graduate (science and non-science) students over a designated one-week period. The students were asked to take screenshots on their computer of any literature searches they conducted throughout the week. Following that week, the students were shown their previously captured photo screenshots and asked to explain and justify each search. Analysis of the interviews revealed that the undergraduate students successfully and

efficiently conducted course-related searches to find definitions and factual information (much of which can be found in a textbook); however, they frequently used unreliable sources. In contrast, the researchers concluded the graduate students' research-related searches were inefficient and yielded results unrelated to their research topics. Further analysis of the searches revealed that the graduate students were involved in complex research projects that were often times unreported in the literature. For this reason, searching for information related to their research was a challenging and almost impossible task. The results indicate that without the prior content knowledge, the graduate students could not employ a rigorous criterion to comprehensively search their respective research topics.

Brown (1999) conducted a survey to investigate what physical and/or emotional factors influenced how graduate students find, evaluate, and decide to use the information in their research, which content knowledge emerged as the overarching theme. Sixteen biochemistry-chemistry, eight mathematics, and eight physics-astronomy Ph.D. students responded to an open-response questionnaire. As indicated in the survey, the participants spent at least some time each month searching for research-related information; however, they frequently struggled to know what keywords to input into their databases. Some students indicated that they evaluated the results of their literature searches by the number of times the articles were cited and they rarely read the entire articles before they concluded how useful the paper was to their research. Although content knowledge was not explicitly discussed in this paper, it nevertheless contributed to the graduate students' ISB. Without sufficient content knowledge, the graduate students struggled to know

what information to input into the search engine, and this lack of content knowledge can explain why they were not applying a rigorous criterion to evaluate the search results.

George et al., (2006) and Kuruppu and Gruber's (2006) studies revealed that both their respective disciplinary resources and content knowledge influenced graduate students' ISB. George et al. (2006) interviewed 36 masters and 64 doctoral science and non-science students to determine which cognitive and physical factors influenced the students' ISB. The students reported knowing very little about their research topics during the early stages of their Ph.D. training; therefore, they had to rely on their advisors and peers for suggestions on what keywords were applicable to their research projects. Using the search engine Google, they would input the suggested search terms, which would result in a massive amount of general information. From there, the students would spend a substantial amount of time reading the resulting information to develop a very broad foundation of knowledge about their research areas. Using this newly acquired knowledge, they could then search their specific disciplinary databases for more precise and relevant information.

Conversely, Kuruppu and Gruber (2006) revealed the difficulties graduate students encounter with performing a literature search when they do not have sufficient guidance from their advisors. Kuruppu and Gruber (2006) interviewed 23 agricultural and biological sciences graduate students and 14 professors to elicit their research-related ISB. The graduate students stated in the interviews that they typically had little to no guidance from their advisors on how to perform a literature search on their research topics. Additionally, the students believed that learning to search for information relating

to their research topics was part of their Ph.D. training. The students also stated they should have the prior knowledge and information-seeking skills to search their respective literature and they would appear less intelligent if they sought guidance from their advisors. Despite this lack of guidance and instruction from their advisors, the students perceived their ISB to be above average. However, the researchers concluded that the students were failing to conduct exhaustive searches on their topics and relied on information that was easily accessible instead of evaluating the quality of the information. Therefore, these studies once again suggested students, especially in the initial stages of a research project, typically lacked the specific content knowledge to successfully search their respective literature and, therefore, needed direction from their advisors (disciplinary community) on how to search for information related to their research topic.

Research on Chemistry Students' ISB

To my knowledge all of the research-based reports on ISB in the domain of chemistry focused on undergraduate students and the primary database utilized in these studies was SciFinder. According to the American Chemical Society Committee on Professional Training (2008, 2015), ACS-accredited universities are expected to teach undergraduate chemistry students, "chemical literature and information management skills" (p. 17). Responding to these guidelines, researchers and practitioners have created SciFinder exercises and/or assignments to instruct and enhance undergraduate organic chemistry students' information seeking skills (Dawson, Jacobs, & Yang, 2010; Ferrer-Vinent, 2012; Rosenstein, 2005; Swoger & Helms, 2015). Although these studies have

primarily focused on developing or enhancing students' proficiencies using SciFinder and less so on the students' usage and experiences with the database, these studies provide some valuable insights that are related to my study.

A study conducted by Ferrer-Vinent (2012) determined that following explicit instruction on the topic, undergraduate-level organic chemistry students were capable of learning the tools to effectively perform a SciFinder search. Students were given a pre-instruction and post-instruction survey to assess their proficiency in using the database. The pre-survey results indicated that students did not know how to use the SciFinder tools to perform a search; however, the post-survey results showed that students understood many of the search options quite well. As such, this study indicates that the ability to navigate SciFinder is not intuitive for students, and these skills need to be taught so students can optimally use the database.

Additionally, Dawson et al. (2010) developed online SciFinder tutorials instructing undergraduate organic students on the use of the database to perform a search. After the students completed the SciFinder tutorials, they were given a survey to evaluate the effectiveness of the tutorials. Based on the feedback from the survey, students suggested the researchers create a tutorial regarding how to read and interpret the research articles (Dawson et al., 2010).

Unfortunately, these are only two *research* studies reported in the literature; there are, however, many more accounts describing a diversity of contexts and ways in which SciFinder has been incorporated into the undergraduate chemistry curriculum in the United States. As such, explicit training in domain-specific information literacy does not

appear to be a standard part of undergraduate majors' training despite the ACS CPT's call for such preparation.

Research in Graduate Organic Chemistry Education

Since my study will focus on the experiences of organic chemistry graduate students, I will briefly review the education research literature in order to provide some background in this area. For the purposes of relevance, this review is limited to research studies in which graduate students were the primary participants.

The first generation of papers published in this area, from 1990 to about 2005, were centered on problem-solving with respect to synthesis and mechanism tasks. For example, Bowen (1990) reported the range of representational systems used while solving synthesis tasks. Using a think-aloud method, Bowen (1990) interviewed ten students enrolled in a graduate-level synthesis course while they solved 4 synthesis tasks. The results from the study indicated that organic chemistry graduate students used multiple representational systems while solving each tasks. He named the systems: "verbal, pictorial, methodological, principles-oriented, literary, laboratory-oriented, and economic" (p. 356). In a related paper, Bowen and Bodner (1991) concluded that the students' choice of representation was dependent upon their stage in the problem-solving process. During all three stages of the problem-solving process (preparing for the problem, generating a solution, and evaluating the solution) verbal and pictorial representations were the primarily used representations to solve the synthesis tasks. Furthermore, the participants would frequently convert the verbal representations to pictorial representations in order to complete the tasks, indicating that pictorial

representations were a more useful representation when solving synthesis tasks. Bowen (1990) also stated that pictorial representations enable, "rapid communication of structural information and fast translation into other representational systems," (p. 356) which is an essential component when solving synthesis tasks. The participants also stated that methodological representations (named organic reactions and retrosynthetic analysis) were the most useful strategy when solving synthesis problems.

Used less frequently than verbal and pictorial were literacy, laboratory, and economic representations. These representations were incorporated in the "generating the solution" and "evaluation phases" (Bowen & Bodner, 1991). One such representation that was scarcely utilized while the participants were working on the synthesis tasks was the literature (Bowen, 1990). Although only three participants discussed using the literature when solving a synthesis problem, the remaining participants explained that they did not know how to access information because they were unaware of the available resources. Bowen (1990) indicated the students' inability to use the literature was because they would have to convert structural representations (pictorial representations) into words (verbal representation) in order to perform a literature search, which would be a cognitively demanding task. It should be noted, however, at the time this research was conducted, no electronic databases existed. With the development and more recent updates to SciFinder, these students would have been able to, presumably, perform searches using the pictorial representation (structural representations), which appear to be a less cognitively demanding task for organic chemistry graduate students.

In a similar study, Kraft, Strickland, and Bhattacharyya (2010) investigated organic chemistry graduate students' modes of reasoning—rule-based (RBR), case-based (CBR), or models-based (MBR)—while solving organic problems. Sixteen organic chemistry graduate students participated in this study for which they were given three types of organic mechanism tasks to solve, each with several parts. The results indicated that participants primarily used CBR to solve the tasks. However, students lacked a deeper chemical understanding of their modes of reasoning, thereby causing them to "force cases to fit" (Kraft, Strickland, & Bhattacharyya, 2010) while solving the tasks. Students also used RBR when solving the tasks; however, this strategy was proven to be less effective because the rules did not lessen the cognitive load of the tasks. The rules employed by the students were so broad that they could not use the rules to decrease, "the number of potential pathways one could take in attempting to solve the problem" (Kraft et al., 2010, p. 288). Although less frequently observed, the most successful participants used MBR to solve Task 1 in which they were asked to predict the remaining steps using the arrow-pushing formulism. However, this mode of reasoning resulted in less successful responses in Task 2 (predict the product), which the researchers contributed to the students under-developed mental modes that lacked the predictive quality (Kraft et al., 2010).

In a companion piece, Strickland et al. (2010) concluded that organic chemistry graduate students generally lacked the representational competence that is expected of practicing organic chemists. Sixteen organic chemistry graduate students participated in this study for which they were asked to define several chemistry terms (functional group,

nucleophile/electrophile, and acid/base) and were given a reaction mechanism using the arrow-pushing formulism and asked to verbally describe the diagram. The results indicated that the participants could identify species belonging to each of the categories – functional group, nucleophile/electrophile, and acid/base. They could not articulate unambiguous definitions for any of the terms, especially when asked to differentiate nucleophiles and bases. Consistent with their ability to identify classes of substances, the participants used functional group, nucleophile/electrophile, and acid/base in their descriptions when describing the diagrams. However, with one exception, all of the participants limited their descriptions to surface-level features and use the arrows as a way to describe the direction for the change in the reaction without considering the flow of electrons. As such, these researchers concluded that the participants lacked representational competence that is expected of practicing organic chemists (Strickland, Kraft, & Bhattacharyya, 2010).

This lack of a deeper understanding of the representations of reaction mechanisms was also shown in a prior study by Bhattacharyya and Bodner (2005). Fourteen graduate students (ranging from analytical, biochemistry, inorganic, organic, and medicinal) from a first semester graduate-level organic chemistry course were interviewed to elicit their understanding and usage of the arrow-pushing formulism (common tool of the community) to predict the mechanisms for several transformations. The results from this study indicated that, "[t]he curved arrows used in the electron-pushing formalism held no physical meaning for the graduate students involved in this study" (Bhattacharyya & Bodner, 2005, p. 1405). As in the subsequent research, students "forced" their ideas to fit

the task at hand without any consideration of whether their proposed mechanisms were based on chemically-viable processes. The results also indicated that the graduate students could generate the "correct" answers but had no understanding of the underlying principles guiding the process. This study, not surprisingly, suggests that graduate students do not begin their Ph.D. training with the ability to use the tools of this community to function at a practitioner's level.

In addition to students' difficulties to use the tools of the community to solve problems, research has also indicated that organic chemistry graduate students struggle to connect theory to practice. Using a modeling-eliciting activity, Bhattacharyya (2006) interviewed ten organic chemistry graduate students in an attempt to understand students' ability to connect theory to practice. Using a think-aloud protocol, the participants were asked to provide a set of rules they could use to explain the pKa values of various substances. The results from this study indicated that participants' description of the pKa values "did not reflect the maturation that would have been expected from their immersion in the organic chemistry research group environment" (Bhattacharyya, 2006, p. 245) and there was a gap between the students' conceptualization of the pKa values and their research activities. Bhattacharyya (2006) concluded that the chemistry graduate students' disconnect between theory and practice could potentially inhibit their ability to produce authentic scholarly research.

Despite the previous findings, research has also indicated that prolonged exposure to authentic research increased organic chemistry students' problem-solving abilities. Cartrette & Bodner (2010) explored organic chemistry graduate students' problem-

solving strategies as they interpreted spectral data. They interviewed thirteen Ph.D. students and two faculty from the organic chemistry division. Each participant was interviewed twice, with the first interview focusing on their experiences using various spectroscopy instruments and the second interview as a problem-solving session in which participants were given molecular formulas for various compounds as well as IR and ^{1}H NMR data and were asked to draw the structures. Instead of analyzing the data from an expert/novice approach, the researchers categorized the participants' responses based on "more successful," "intermediate success," and "less successful" problem-solvers. The more successful problem-solvers verified that their proposed solutions were consistent with the provided data and were able to articulate their underlying reasoning for their answers; whereas, the less successful problem-solvers rarely checked their solutions and struggled to justify their answers. Furthermore, the data indicated that graduate students who were further along in their Ph.D. training approached the problems similar to the faculty; whereas, the more junior graduate students did not show characteristics similar to the faculty (Cartrette & Bodner, 2010). This study, as well as other studies that will be reviewed, indicate that prolonged immersion in a research group has an effect on organic chemistry graduate students' problem-solving abilities.

Anderson's (2009) study revealed similar results, suggesting that prolonged exposure to research-related (authentic tasks) activities increased organic chemistry graduate students' problem-solving abilities. Eleven organic chemistry graduate students (six first-year students enrolled in a second-semester organic chemistry course, three third-year students, one fifth-year student, and one sixth-year student) were interviewed

between four to eight times throughout the duration of the study regarding their usage of mechanistic reasoning. Results from this study indicated that the synthesis projects on which the first-year students worked throughout the semester required them to engage with the community of practice (i.e. instructors, peers, members of their research group, and the organic literature). Furthermore, this project allowed the first-year students to begin to connect the organic chemistry coursework to their Ph.D. research. Anderson (2009) also observed that by conducting their Ph.D. research (gaining membership and participating in the organic CoP), the first-year students began to view mechanistic reasoning as a tool to accomplish their research goals. For the third-year organic students, continued exposure to research problems necessitated that they continue to use their mechanistic reasoning skills, including the arrow-pushing formulism, to synthesize their target compounds. From these results, Anderson concluded that both the, "direct and indirect interactions with the organic chemistry CoP, via community members and the research literature" (p. 279), increased the organic chemistry graduate students' ability to use the tools of this community to solve mechanism problems.

Similarly, Bhattacharyya and Bodner's (2014) study indicated that early experiences working on authentic tasks and longer exposure to research has an effect on students' epistemic development. Bhattacharyya and Bodner (2014) interviewed two groups of students to understand how organic chemistry graduate students' learn to solve synthesis problems and develop into practicing organic chemists. The first group consisted of four first-year organic graduate students enrolled in a graduate-level organic course, and the second group consisted of two third-year organic graduate students who

were working on their original Ph.D. research proposals. The first-year students were interviewed eight to ten times throughout the semester focusing on their semester long synthesis project, while the third-year students were interviewed six to eight times regarding their oral proposal. When the semester began, the authors reported that "the first-year students seemed unable to critically evaluate data in papers from the primary literature" (Bhattacharyya & Bodner, 2014, p. 707). However, the semester long synthesis project required that the first-year students consult the primary literature, which as a result, gave them a, "sense of being immersed in the culture of organic chemistry" (Bhattacharyya & Bodner, 2014, p. 704). A major theme that emerged from this study was that by the end of the semester, the first-year graduate students were approaching their semester long synthesis project like practicing organic chemists. The authors attributed the epistemic development of the first-year students to the semester long synthesis project, which provided the students with an authentic activity of the organic CoP for which they felt a sense of ownership for their projects. Not surprisingly, the third-year graduate students' approach to the literature was a reflection of being immersed in the organic CoP for a longer period of time. For instance, they were able to critically evaluate the quality of the research reported in journals without simply accepting the author's work as absolute. In addition, they judged the quality of the work based on the journal, realizing that some journals use more rigorous criteria to evaluate the research.

From a more holistic approach, Walls' (2008) study focused on the professional development of chemistry graduate students and their beliefs of what it means to do

chemistry. For this study, she interviewed thirty-one chemistry graduate students from the various chemistry sub-divisions (eight organic, eight inorganic, ten analytical, and five physical chemistry students). The results indicated that before entering a research group, the students did not have a good understanding of the discipline of chemistry. As such, they decided which sub-division to pursue based on their classroom performance or a process of elimination based on their interests. Furthermore, when discussing the chemical concepts they incorporated into their research activities, the students talked as if they were repeating key words they had heard from their research advisors or researchers from their field, suggesting they lacked a deeper conceptual understanding of the terms. However, chemical concepts that were learned in the classroom became real once they could use that knowledge to accomplish their research. The participants, therefore, described developing their identity as a professional chemist as a process of acculturation through interactions with more knowledgeable peers. Wall concluded that, "when one acquires the ability to speak and use the chemistry language, they gain the ability to use the language that allows them to communicate with chemists" (p. 56). As such, their ability to communicate like a practicing chemist was an essential component to becoming a participating member of a community of practice.

Also from a more comprehensive approach, Stucky (2005) sought to characterize the nature of scientific inquiry by studying an organic chemistry research group and found that the literature was a resource the students used to conduct their research. Over a three-year period, Stucky (2005) video-recorded the laboratory activities and the weekly group meetings of an organic chemistry research group. In addition to video-

recording the research activities, she also conducted interviews with the research group members. The results from this study indicated that the researchers in this group encountered and used two primary resources: 1) objects (consisting of artifacts, visible substances, human, and theoretical objects), and 2) processes (consist of chemical reactions and physical process, human action, mental state, and theoretical explanations). Further characterization of the various resources that the researchers utilized suggested that the literature was most useful to gain background information about their research area and/or to find experimental protocols to aid in prepping for specific reactions. In addition, the literature served as a resource that helped students to solve problems after consultation with fellow group members proved to be unfruitful.

Summary of Literature Review

This review indicates that the ability to search and use the CoP's primary literature is an integral component for graduate students to contribute scholarly research to their field. Despite the significance of this resource, students struggle to perform a literature search because they are unable to effectively utilize their disciplinary resources and/or their discipline-specific content knowledge. Although these studies provide a glimpse into the factors influencing graduate students' ISB, to my knowledge, there are no studies which describe the intricacies of *how* students use their discipline-specific content knowledge to decide what information to input into the search database and *how* they use their knowledge to evaluate the resulting information. Adding to this deficit, other than Brown's (1999) study, there is a lack of research on chemistry graduate students' research-related ISB. Although several studies focusing on organic chemistry

graduate students mention the usage of the scientific literature, these studies did not explore the intricacies of their ISB.

As such, my study will explore not only how students navigate SciFinder (disciplinary resource) to find a protocol to synthesize unreported compounds, but also how they use their domain-specific content knowledge to make decisions regarding each step of their search. Specifically, my study will describe the students' ISB as they use their content knowledge to: 1) decide what information to input into SciFinder (topic or structure), 2) evaluate the resulting output of information, and 3) decide which articles would be applicable for the synthesis of the target compound. More broadly, my dissertation will expand the previous literature because my study will provide a deeper understanding, from the Ph.D. students' perspective, about the relationship between their disciplinary resources (SciFinder and their CoP) and their discipline-specific content knowledge (organic chemistry).

CHAPTER THREE

METHODOLOGY

I will begin this chapter by summarizing my previous research that provided the motivation for my dissertation study. I will then discuss my: 1) motivation and research questions, 2) theoretical and methodological frameworks, and 3) methods used for my dissertation study. Also, incorporated into my methods section is a guide to SciFinder and a search key that will be used in my Results and Discussion chapter. To conclude this chapter, I will describe my role as the researcher and describe the validity and trustworthiness of my study.

Summary of Previous Research

The goal of my previous study was to determine what resources chemistry graduate students use to successfully accomplish their research. Previous data indicated that organic chemistry graduate students appeared to conduct their research without explicit or conscious application of the foundational theoretical and conceptual constructs (Walls, 2008). Based on these results I formulated my study to further investigate this finding and to elucidate the knowledge resources used by chemistry graduate students in the course of their research activities.

Guiding Research Questions

My study was guided by three research questions with the overarching goal of understanding what resources chemistry graduate students use to successfully accomplish their research activities. The first research question focused on the developmental stages of their research projects:

How do chemistry graduate students design their research projects?

The second guiding research question was intended to explore the knowledge resources used by chemistry graduate students to accomplish their research activities:

What knowledge resources do chemistry graduate students use in their research?

The third guiding research question pertained to the last stages of a research project:

How do chemistry graduate students determine if their projects are successful?

Since this was an exploratory study, I determined that a qualitative research design was an effective strategy that would allow me to explore this topic in-depth.

Theoretical Considerations

Since the goals of this study were descriptive in nature, I chose ethnography, a theoretical perspective that emerged from cultural anthropology, as the lens for this study. The primary focus of ethnography is to understand or describe a culture (Creswell, 2013; Patton, 2002). A research group can be considered its own culture because, as Patton (2002) suggested, any group of people who are organized together for a period of time tend to develop a set of traditions or standards, that is, a culture. Research groups typically emphasize a specific research concentration and, therefore, employ similar practices and approaches to their research activities (Stucky, 2005). For this reason, I treated the research group as a culture.

Ethnographic data may be collected from two perspectives: emic and etic (Patton, 2002). In the emic tradition, the researcher becomes embedded in the culture to the extent of a full member of that culture. In contrast, the etic tradition keeps the researcher separated from the routine activities of the culture. Given that I did not have the same

research experiences or content background as the students taking part in the study, I chose to adopt the etic orientation.

Methodology Considerations

For chemistry graduate students, the development and execution of research projects typically takes several months to complete. With this considered, I determined that collecting data about what resources they use to successfully complete their research projects would be too difficult to collect during just one interview. I determined that I should perform multiple interviews throughout a designated semester, thereby generating multiple data collection points relating to their research activities. As such, descriptive case studies, which can provide details of an individual or group being studied, were chosen as the data collection method because this approach allowed me to collect rich descriptions about what the graduate students were considering as they progressed through their research projects (Cohen, Manion, & Morrison, 2000). The longitudinal nature of this project also would allow me to adjust and modify my interview protocol as emerging themes developed throughout the semester.

Participants and Setting

To collect rich descriptive cases of chemistry graduate students' research activities, while also limiting the variation of chemical practices to be discussed, I decided to focus on a specific research group. I also decided that selecting a research group that was considered as exceptional within the chemistry department would, thereby, provide many insights regarding how other chemistry graduate students might approach their research. Lastly, I wanted to select a research group with graduate

students in various stages of their Ph.D. training because this would provide multiple perspectives regarding how graduate students approach their research activities.

For these reason, I purposely selected the research group to study because it met the previously mentioned criteria. From interacting with the selected research group, I knew that they were recognized for contributing innovative scholarly research to their research area. Furthermore, it was my impression that the graduate students were exceptionally motivated and managed to work on several projects over the course of a semester. Additionally, the research group had graduate students in various stages of their Ph.D. training, which would further provide multiple perspectives of which resources they were using to successfully accomplish their research activities. Based on these considerations and following IRB approval, three inorganic chemistry graduate students from a large southeastern public university (Charlie, a 2nd year; Lola, a 3rd year; and Morgan, a 5th year) participated in this study. To protect the identity of each participant, I provided each participant with a pseudonym.

Data Collection and Analysis

I interviewed each participant 10 times over the course of a semester, roughly once a week. Each audiotaped interview followed a semi-structured protocol (Appendix A) so that I could further investigate emerging themes in the subsequent weekly interviews. I designed the initial interview to focus on the participants' research backgrounds prior to entering their current research group to determine the existence or extent of their research experiences. For the remainder of the initial and subsequent interviews, I focused on their Ph.D. research, specifically what they worked on that week,

which projects they described as having "worked" (meaning they successfully accomplished their research goals for that week) or "not worked", and what resources they considered utilizing to accomplish their research goals. During these interviews, I collected and/or reviewed and discussed artifacts such as their research notebooks, spectral data (NMR, IR, etc.), and drawings the participants made as part of their explanations or descriptions. I also took field notes during the interviews, which served as another source of data.

Each interview was transcribed verbatim, and the resulting data – transcripts, field notes, and participant artifacts – were organized and sorted by each research project the participants discussed. During the data collection period, Charlie worked on four projects, Lola worked on six projects, and Morgan worked on five projects. I then wrote a detailed case study explaining the progression of each project as it evolved throughout the semester and analyzed each individual case study to determine the major themes (resources) influencing the participant's research activities. To complete my data analysis process, I performed a cross-case analysis to determine the overarching major themes.

Results and Discussion

A major theme resulting from the cross-case analysis was that the problems the graduate students encountered while performing their experiments originated from their evaluation of their literature searches. Searching the literature was the resource they utilized to obtain their research protocols; therefore, if they did not find an applicable protocol, then they would have a difficult time in the laboratory attempting to synthesize

or analyze their compounds. **As such, the data suggested that there was a strong relationship between the students' usage of the primary literature and success in their research projects.** Because of the direction of my dissertation project and the significance of their information-seeking behaviors (ISB), the following presentation of the results will focus on providing evidence supporting the above claim instead of describing in more detail each guiding research questions. Furthermore, I used general descriptions of their research projects in order to protect the identity of the participants.

In project 2, Charlie was attempting to synthesize an ester containing a bulky alkyl group. His first step for this project was to search the literature for a laboratory protocol for its synthesis. Charlie began his search by inputting the exact structural representation of the ester into a chemistry database (SciFinder). The search yielded a paper from 1972 that reported his compound; however, the researchers did not specify the exact reaction conditions used to synthesize the compound. Since the 1972 paper was the only search result reporting his exact compound, Charlie then decided to search for a structurally similar compound for which he could manipulate the reaction conditions to synthesize his target compound. Although Charlie did not explicitly state what information he inputted into SciFinder, his search resulted in a paper reporting a compound with the similar carbon backbone chain and a small alkyl substituent attached. He initially concluded that he could follow this protocol; however, when he applied the protocol in his laboratory, he failed to make the t-butyl compound. He explained:

> *Yeah, but they actually...it's like one of the next steps, the step to make the t-butyl complex. So some people have actually said that, the lower esters like the n-propyl or stuff like that have actually done direct esterification just by putting the [compound], putting the acid in and then with the alcohol in that you are going to*

use. And we reflux that for a little bit, but since you see a t-butyl group is just so bulky, it is not going to work like that. (Interview 3)

As the above quote indicates, Charlie evaluated the search results based on the surface features (surface-level interpretation) of the structure. The ester he was trying to synthesize and the ester reported in the literature had similar carbon backbone chains and appeared, based on the structural representations, to have similar reactivity to his target compound. Even on the basis of the surface-level differences, Charlie was unable to realize that the differences in steric bulk between the alkyl groups of his molecule and the one obtained from the literature search were significant enough to confer distinctly different reactivity under those same experimental conditions. Furthermore, it was not until he failed to synthesize his target compound that he determined the t-butyl group was bulkier than the n-propyl group and thereby had different reactivity under the same conditions. Therefore, Charlie's ineffective usage of this content knowledge to critically evaluate his literature search, specifically the structural representations of similar compounds to find an applicable protocol, resulted in unsuccessful attempts in his research.

As another example, in project 5 Morgan displayed analogous problems as she attempted to develop a new research technique on DNA. Similar to Charlie, she began this project by performing a key word search on a chemistry database to find a research protocol she could use to extract DNA from bacteria. The search yielded a substantial number of articles reporting the one specific extraction technique which led her to initially assume the project would be easy to complete. She explained her literature search by stating:

> *Something like because when I did the [literature search], I saw so many hits like 10,000 hits on it. I thought that will be easy. So you kind of forget about that you have to think. Yeah, because you see so many people doing that, it's like ok I can do it. Yes. So but for something when you type in that so it didn't show too much, you will think, 'Oh, what should I do?' But just sometimes it easy think just make you lost your thought. (Interview 5)*

Furthermore, during her 6th interview she described her process of applying the protocols:

> *Because like the first paper, when that doesn't work at the beginning, I thought, 'Oh, ok, now I will try the second one.' When the second one doesn't work, I was like wait a minute I might do something wrong. So I go back to look at it step by step...So I know from that part one, I did it right because from the paper two that was saying because it was [name of system] so you normally use [experiment 1]...very little cells to do [name of compound] because you can get a lot of [name of compound]. I totally forgot. (Interview 6)*

As the quotes indicate, Morgan's evaluation of her literature search was based on the abundance of research articles reporting her inputted key words. She initially assumed that the protocols would be applicable in her laboratory because of the significant number of research articles using the specific technique on DNA. However, each protocol she applied failed during the extractions phase because the cells she used had distinct properties and would require different experimental conditions. Morgan did not reach the conclusion that the cells were distinctly different until she had repeatedly applied approximately five different protocols, each yielding unsuccessful results. As seen with Charlie, Morgan's inappropriate usage of her content knowledge to critically evaluate her literature search to find an applicable protocol resulted in multiple unsuccessful attempts in her research. Like Charlie's example, Morgan's case demonstrates the necessity of applying domain-specific content expertise when performing literature searches.

Charlie and Morgan's cases indicate that their ISB were an essential component to their research development and outcomes. Moreover, their cases indicate the importance

of applying one's domain-specific content expertise when determining the usefulness of reported protocols. For instance, Charlie focused on articles reporting compounds with "similar" structural features, thereby failing to differentiate the reactivity of tertiary vs. primary alkyl groups whereas Morgan focused on the abundance of articles reporting the specific key words and did not critically examine the papers to determine the effect of the extraction technique on her cells. Neither Charlie nor Morgan were using their content knowledge to critically evaluate the search results, and, as a result, were using inapplicable protocols in their laboratories.

Although the results indicate the significance of their ISB, the results did not describe each step of their literature search. This study, therefore, provided the motivation for my dissertation for which my goal was to further explore how chemistry graduate students use their domain-specific content knowledge to perform a literature search to find an applicable research protocol. Specifically, my goal was to explore how chemistry graduate students decide to: 1) input information into SciFinder, 2) evaluate the resulting information, and 3) keep protocols they determine were useful for the synthesis of a previously unreported compound.

Dissertation Study

Motivation and Research Questions

The literature review as well as the case studies indicate that graduate students' ISB are an important component for contributing scholarly work to their field. Although graduate students are expected to be knowledge producers and contribute to their field (Feldman et al., 2013), research suggests that graduate students frequently have

significant difficulties performing searches and that these difficulties are related to their disciplinary resources and their domain-specific content knowledge (George et al., 2006). As such, graduate students' ability to contribute scholarly research to their field hinges on their ISB, more specifically their proficiency to perform a literature search to find an applicable research protocol. For these reasons, my dissertation study was guided by the following questions:

> *How do synthetic organic chemistry graduate students conduct a literature search using SciFinder to find research protocols?*
>
> *What are synthetic organic chemistry graduate students' sense-making procedures for the different parts of the literature search?*

The results from my study will expand the research that was reviewed (Chapter Two) and provide insights regarding how graduate students use their domain-specific content knowledge to perform literature searches (by navigating a chemical database, SciFinder) to develop research protocols. More broadly, the results of this study will provide evidence for how students, who are members of a CoP, use the tools of the community to solve authentic tasks. As such, my dissertation study focuses on understanding the norms of a specific culture (i.e. the synthetic organic CoP) for which the results can be applied to any field of education. Considering that the guiding research questions are framed to describe the participants' ISB, a qualitative approach (Figure 3.1) was used because of the potential for the "rich description" that Patton (2002) asserts is a hallmark of this methodology.

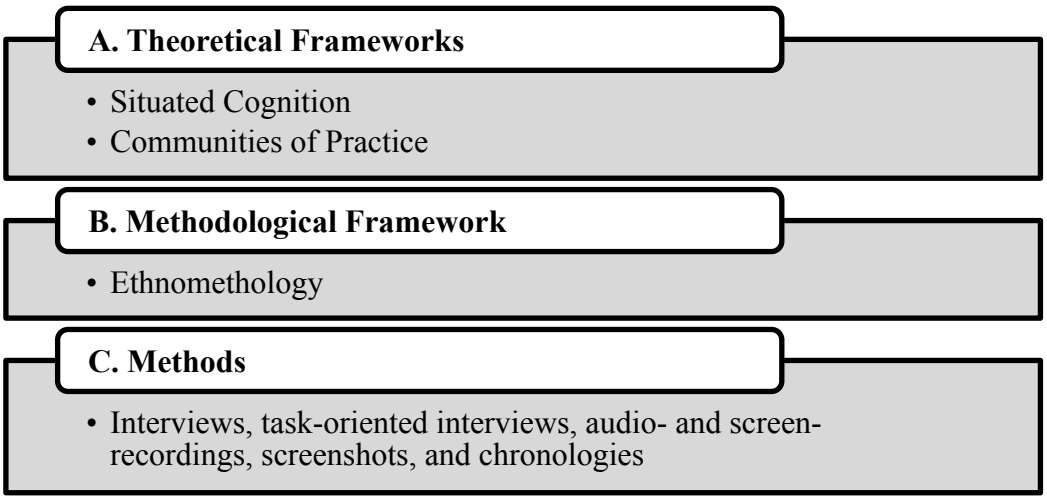


Figure 3.1: Components of my dissertation study adapted from Crotty (1998)

A. Theoretical Frameworks

Theoretical frameworks can be defined as the "assumptions that guide the research, helps the researcher choose appropriate questions for a given study, and directs the researcher toward data collection methods that are appropriate for the study" (Bodner, 2007, p. 11). The guiding research questions emerged from my case studies; therefore, my next step was to determine which theoretical frameworks would be the most appropriate perspective to guide my dissertation study. The theoretical frameworks that I chose are situated within the social learning paradigm, situated cognition and communities of practice.

Situated Cognition

Situated cognition—commonly referred to as situated learning—is a theory concerned with the process by which learning occurs through the dynamic interactions between individuals and their environments (Brown, Collins, & Duguid, 1989). This

theory indicates that knowledge is fundamentally indistinguishable and embedded within the activity and context in which it was learned. Based on this premise, situated cognition suggests that activities are essential for learning and provide the pathway by which the learners continuously and actively develop and/or refine their knowledge (Orgill, 2007).

Within this learning environment, it is more useful to think of knowledge as a set of tools than as abstract concepts. Viewed in this way, individuals *must* use these tools, and as they use these tools they develop a more comprehensive understanding of the context in which the tools are used (Brown, et al., 1989). Learners must use these tools in authentic activities or any tasks considered common practice of the community. Learning to meaningfully and successfully use the tools, therefore, necessitates that the individuals adopt the norms of the culture for which the tools are used (Brown, et al., 1989). Brown et al. (1989) explain that, "To learn to use tools as practitioners use them, a student, like an apprentice, must enter that community and its culture. Thus, in a significant way, learning is, we believe, a process of enculturation" (p. 33).

Furthermore, the situation or context in which learning occurs does not mean that learning only occurs in a specific location. On the contrary, situation cognition, "includes not only the physical objects in the context, but the other people in the context, as well as the social, ethical, and historical norms that guide how people interact with the objects in the environment and how they interact with each other" (Orgill, 2007, p. 189). Within this social learning environment, the authentic activities that provide the context for which the tools are used, are developed through the past and present exchanges and

negotiations by the members of the community about the beliefs and norms of the culture. As such, as individuals become members of a community and use their tools to engage in authentic activities, both the individual and the participating members are co-constructing and/or refining their knowledge. Therefore, according to Brown, et al. (1989, p. 40) "…ideas are exchanged and modified and belief systems developed and appropriated through conversation and narratives, so these must be promoted, not inhibited," with others within the community.

Communities of Practice

Often encountered in conjunction with situated cognition is the social learning theory, communities of practice. According to Wenger (2000) communities of practice (CoPs) are any groups of people who share a common interest in a specific domain area and who work collaboratively to deepen their knowledge and expertise. The core principles for any CoP are that members: 1) engage in a sense of *joint enterprise*, 2) contribute to the community through *mutual engagement*, and 3) produce a *shared repertoire* of the communities' resources (i.e, language, routines, artifacts, tools, stories, styles, etc.) (Lave & Wenger, 1991).

According to Lave and Wenger (1991) learning within a CoP is a process of legitimate peripheral participation (LPP). LPP is concerned with the process by which a newcomer transitions, by developing their identity and knowledge about artifacts and activities, to achieve "old-timer" status within the community (Lave & Wenger, 1991, p. 56). From this perspective, learning occurs as a process of enculturation or increased participation within a CoP. When a newcomer first joins a CoP, through a peripheral

view, they are observing the norms, behaviors, and the forms of communication within the community. However, according to LPP the newcomer must *participate* in an authentic activity in order to eventually transition into an old-timer status within the community. Learning within a CoP, therefore, occurs through social interactions as newcomers discuss their ideas, predictions, or observations about the practices of the community and develop into old-timer.

Synthesis of the Theoretical Perspective

These social learning theories were an appropriate choice for framing my study because the process of becoming a full participating member of a CoP arguably occurs once an individual begins their Ph.D. training (Anderson, 2009; Brown, et. al., 1989; Lave & Wenger, 1991). Once graduate students join their respective research groups, they are not only taking a step towards legitimate participation in the CoP, but they have, "entered into a community of practicing (organic) chemists" (Anderson, 2009, p. 50). As such, developing into a full and accepted member of the organic synthesis CoP necessitates that individuals are able to use the tools of this community (discussed in the methodological section) to perform their Ph.D. research (Anderson, 2009; Bhattacharyya & Bodner, 2014). Therefore, these theoretical frameworks were appropriate for this study because the chemistry department from which the data was collected does not offer any formal information literacy training. For this reason, it is likely that organic chemistry graduate students learn and/or refine how they search and develop research protocols once they join a research group and begin their Ph.D. training. For these reasons, my dissertation study focused on understanding how organic chemistry graduate

students in various stages of their Ph.D. training use the tools of this community to search SciFinder to develop research protocols to synthesize previously unreported compounds.

B. Methodological Framework

Ethnomethodology guided my data collection methods and analysis. Emerging from sociology, ethnomethodology is concerned with understanding the methods by which individuals bring social order to their everyday situations (Leiter, 1980; Patton, 2002). This framework, therefore, focuses on "the norms, understandings, and assumptions that are taken for granted by people in a setting because they are so deeply understood that people don't even think about why they do what they do" (Patton, 2002, p.111).

The goal of this method is to make the implicit (tacit) knowledge explicit as a function of the participants' descriptive accounts. The participants' (commonly referred to as actors) accounts are the method by which they bring order to their routine, taken-for-granted activities. The actors' accounts or sense-making procedures are dependent on the social context or community of practice for which they are used. Once members join a CoP, they are adopting the culture that was socially and historically organized and are beginning to develop methods to function within this community (Leiter, 1980).

Since their personal experiences (knowledge) are embedded within their CoP, the members' explanations have two important attributes: indexicality and reflexivity. Indexicality refers to the notion that the descriptions and the actor's knowledge can only be understood by means of the context in which they are used; whereas, reflexivity links their current or future actions to their past experiences within the culture (Bhattacharyya,

2007). Explanation of how these aspects relate to my study will be presented in the final paragraph of this section.

Ethnomethodology is compatible with the social learning theories because the actor's accounts are situated or indexed within a specific culture. Leiter (1980) stated that the actors' knowledge is socially derived and distributed, meaning individuals learn the culture from interactions with people who are in the community and each individual's knowledge is different from the other members. This perspective, therefore, considers the experiences as perceived and constructed by the participant along with socio-cultural aspect (Bhattacharyya, 2004). From this perspective, the actors, "...organize and render observable the features of society and social settings" by means of accounting for their personal experiences (Leiter, 1980, p. 161).

More recently, ethnomethodology has been used to explore how enculturation into the organic CoP affects organic chemistry graduate students' problem solving strategies (Anderson, 2009; Bhattacharyya, 2004). Because an essential component of the ethnomethodology framework assumes the existence of a culture, for these studies, the sub-division of organic chemistry was treated as a culture. Organic chemists exhibit a specialized language, tools, and artifacts that are unique to their community (Bhattacharyya & Bodner, 2014). They have such a specialized form of communication that people who are not members of this community would have a difficult time understanding the dialogue (Bhattacharyya, 2004). Specifically, synthetic organic chemists frequently use the database SciFinder, which was developed by Chemical Abstract Services (2015). A typical literature search to find an applicable set of protocols

involves using their specialized domain-specific content knowledge, which includes understanding 2D and 3D structural representations, functional group reactivity, named organic chemical reactions, mechanistic reasoning, and the retrosynthetic approach.

Searching the literature not only allows members of the synthetic organic chemistry community of practice to validate that their research topic is authentic to their field, but also can provide the foundation for which they develop a research protocol to use in their laboratory. For example, a typical search (discussed in more detail in the Data Analysis section) involves first deciding what information to input into SciFinder. Often times a search is begun by inputting the 2D structural representation of the target compound and/or transformation (starting material to final product) into the database. Since their research usually involves unreported compounds that are not commercially available, it is very common for organic chemists to refine their search to find a reported substructure. Choosing an appropriate substructure, therefore, requires synthetic organic chemists to understand the chemical reactivity of the target molecules so they can decide which parts are essential for the synthesis. While analyzing the often times substantial number of reactions or substances, these chemists must use their broad organic chemistry knowledge to decide which "hits" are applicable to their target molecule and warrant further investigation. Finally, the choice of a protocol, or set of protocols, must be chosen based on the searchers' understanding of the experimental feasibility in the context of their laboratory resources.

The ethnomethodological accounts discussed in my dissertation study were the synthetic organic chemistry graduate students' descriptions and justifications of how they

search SciFinder to develop research protocols. As such, these accounts were indexical (linked to a specific context) because their search strategies were specific to the type of search they were conducting, which for this study was searching SciFinder to develop a research protocol to synthesize an unreported compound. Furthermore, their accounts were reflexive (action learned from previous experiences) by nature because they were using strategies to search SciFinder that they developed from previously successful searches to find a set of research protocols.

C. Methods

Setting and Participants

This study took place within the chemistry department at a large public research-oriented institution. Before the data collection process began, appropriate documents were submitted and approved by the University's Institutional Review Board (IRB). Following IRB approval, I began the process of selecting my research participants using criterion based sampling in order to obtain information rich cases related to how organic chemistry graduate students search SciFinder to develop research protocols for unreported compounds. This sampling strategy allowed me to select participants who met a certain criteria for my study (Patton, 2002). Since my target population was organic chemistry graduate students, I obtained a document from the chemistry department's Graduate Student Coordinator listing every graduate student who was in the department and contained information regarding their year in graduate school, chemistry subdivision, and research advisor. Since synthetic organic chemistry graduate students develop and carry-out their research protocols within the context of their Ph.D. research

groups, my criteria for selecting the participants included that they were not only members of a organic chemistry research group but were actively engaged in their research. For these reasons, I purposefully excluded first-year graduate students from my study because they typically have not started their dissertation research during this stage of their Ph.D. training. I asked five graduate students to participate in my study who met the criteria. To protect participants' identity, they were each provided with a pseudonym. Jane and Peter were in their second year, Sophie was in her third year, and Bruce and Luke were in their sixth year of their Ph.D. training. No incentive was given to the participants for contributing to this research project.

SciFinder Background and Search Key

This section contains background information and an overview of SciFinder, which will enable the reader to better understand the methods used for my study. This section will also provide a shorthand key that will be used to discuss the participants' search actions in the Results and Discussion Chapter. Although SciFinder provides the users with a wide variety of search tools, the following is not intended to provide an exhaustive review of these tools. The following descriptions are intended to provide the reader with a working summary of the search tools that were primarily used by the participants in this study and will be described in the following chapter.

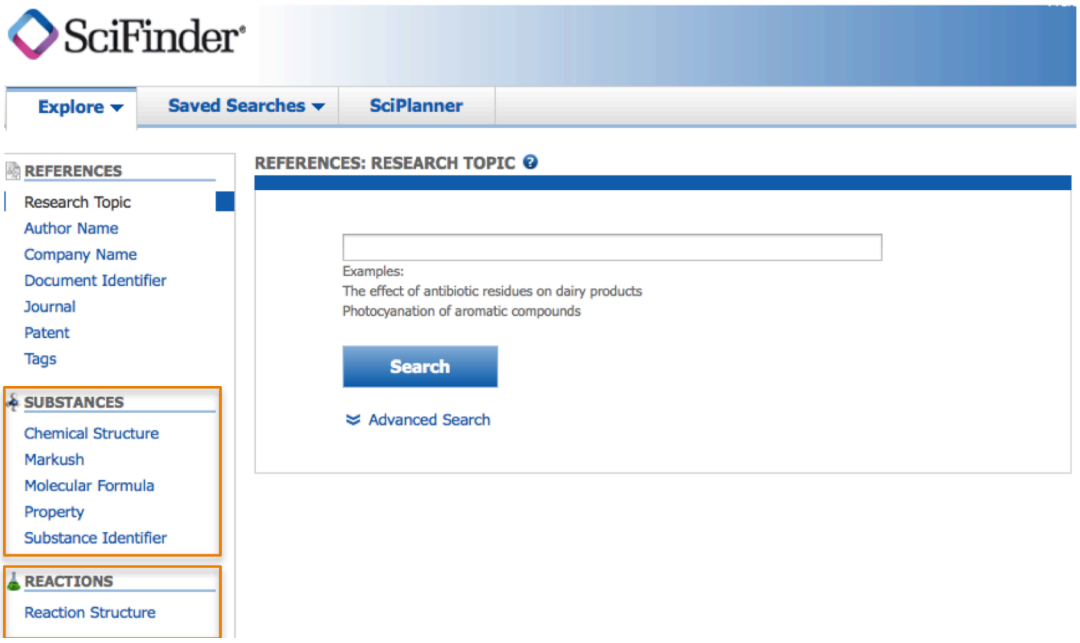


Figure 3.2: Screenshot of SciFinder search options

The participants predominantly used the "substance" and "reaction" search options during the Task 1 search. Note in Figure 3.2, there are several substance options and one option for the reaction searches under the "Explore" heading. Although there are several options under the substances, the participants frequently would select the "Chemical Structure" option to access the structure editor screen as seen in Figure 3.2. Once they accessed the structure editor screen they would draw the structural representation of the target compound and/or substructures.

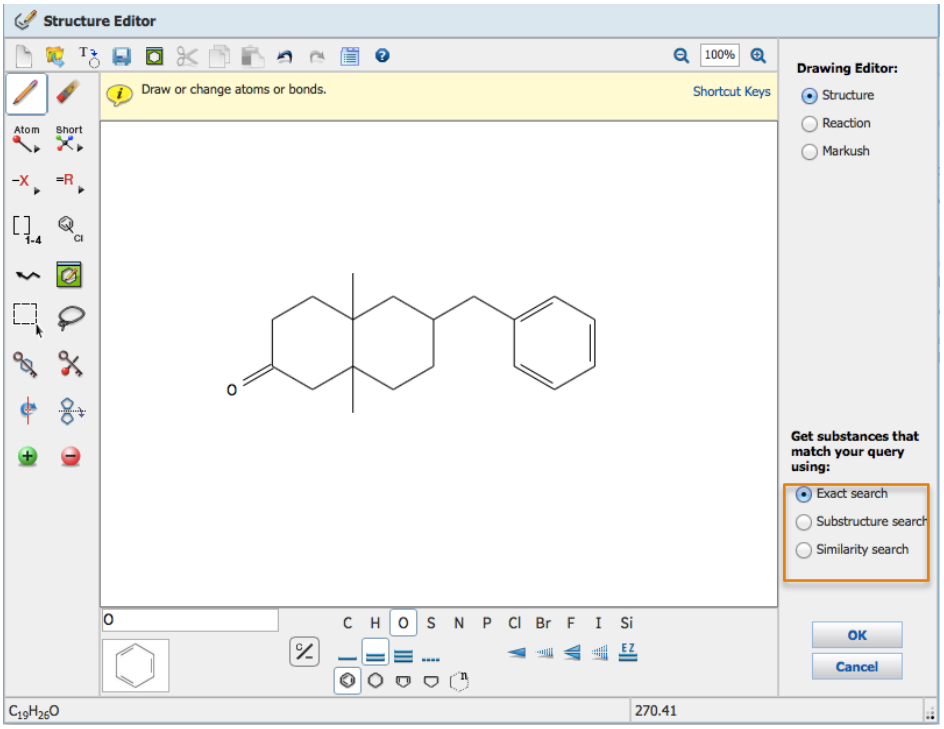


Figure 3.3: Screenshot of structure editor screen taken from SciFinder

After they input the structural representation of the target compound, there were several types of searches they could conduct (Figure 3.3). The participants could perform an exact search (SE) for which the database would search for the exact structural representation of the original compound. The participants also had the option to perform a substructure search (SS) for which the database would search for substances with the original structure embedded within other substances. The last substance search the participants could conduct was a similarity search (SSi) for which the database would search for substances based on how similar they were to the original compound (Chemical Abstract Services, 2015).

Table 3.1: Options in SciFinder for substance and reaction searches

Reaction (Rx)	Substance (S)
Variable only at the specified positions (RxV)	Exact Search (SE)
Substructure of more complex structures (RxS)	Substructure Search (SS)
--	Similarity Search (SSi)

From the structure editor screen, the participants could also perform a reaction search (Rx) by either inputting a transformation (starting material yielding a final product) or the structural representation of the target compound Table 3.1. If they input a transformation, they could perform a "variable only at the specified positions" (RxV) search, which will search for the exact transformation or they could perform a "substructure of more complex structures" (RxS) search, which will search for reactions that were embedded within a larger or more complex transformation (Chemical Abstract Services, 2015).

Table 3.2: Options in SciFinder for designating the role of the compound when performing a reaction search (Rx)

Reaction (Rx)				
1. Product	2. Reactant	3. Reagent	4. Reactant/Reagent	5. Any Role

Another option was to input the structural representation of a substance and perform a reaction search (Rx). For this type of search they still had to perform either a "variable only at the specified positions" (RxV) or "substructure of more complex structures" (RxS) search, but now they had to select "the role" for which they wanted to

search for the substance within a reaction. Now, they have the option (Table 3.2) to search for the original structure as any role in the reaction or designate the substance as the role of a: 1) Product, 2) Reactant, 3) Reagent, 4) Reactant/Reagent.

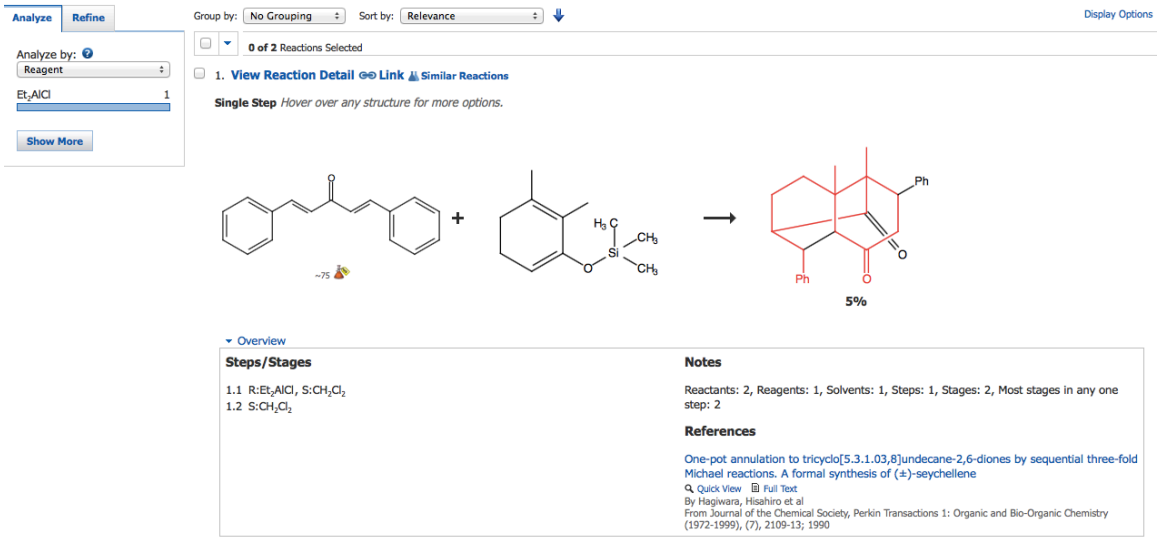


Figure 3.4: Screenshot taken from SciFinder of an RxS1 search for the Task 1 compound

As seen in Figure 3.4, the results from a reaction search provide a scheme representing the overall transformation (starting materials to final product), an overview of the steps and stages, and information regarding the journal article. From this view, the participants could see the product yield of the reaction and if the starting materials or product(s) were reported as commercially available (flask displayed below the substances). In addition, if they chose to access the journal article reporting the transformation electronically, they could select the "full text" link displayed under the references.

Data Collection

I asked each graduate student to participate in two and/or three interviews, each

lasting 45-90 minute. Considering the ethnomethodology framework, I used a semi-structured interview protocol to elicit their sense-making procedures of how they search via SciFinder to find a set of research protocols to synthesize previously unreported compounds. The semi-structured interview protocol provided me with the flexibility to ask individually tailored follow-up questions regarding their responses.

Before each interview, I informed the participants that I was going to frequently ask questions that might seem very obvious to them and these questions were designed to have them account for how they decided/knew to perform each action of their searches. Specifically for the second and/or third interview, I asked each participant to verbally describe his or her thoughts while performing the SciFinder searches. Lastly, I told the participants that I was going to take notes during the interview but that I was not grading their responses.

Interview #1:

Interview #1 was subdivided into two parts as seen in Figure 3.5. The first part of the interview consisted of questions regarding the participants' research experiences prior to graduate school. During that time, I asked them to describe their undergraduate research experiences and to elaborate on how they obtained their synthesis protocols. Following the background questions, I asked each participant to discuss their current Ph.D. research; specifically, I asked them to explain how they obtained their research protocols. During this part of the interview, I provided the participants with paper and a pen, and I asked them to draw the compounds they were researching and explain how they developed their research protocols. Asking them to draw their compounds allowed

me to ask specific questions about their process of searching the literature to obtain a synthesis protocol.

Interview #1 Protocol

Part 1- Background Information and Research:

- What year of graduate school are you currently in?
- What is your area of emphasis?
- Prior to coming here, what types of experiences did you have with scientific research? How long did those last?
- Can you please explain the current research you are working on?

Part 2- SciFinder Experience:

- Do you have any past experiences using SciFinder? Can you please explain those experiences?
- Do you currently use SciFinder for your research? If so, how do you use it?
- If the molecule that you are researching has never been reported before, how would you search to find a protocol that you could apply to synthesize the molecule?
- Once you find articles how do you know which articles are appropriate to your research?

Figure 3.5: Interview #1 protocol

During the second part of the interview, if they had not previously discussed using SciFinder, I asked them to explain if they had any experiences during their undergraduate

degree using the database as well as how they used the database to perform their Ph.D. research. Specifically, I asked them to explain step-by-step how they used the database to search for protocols if their compounds were previously unreported in the literature.

Interview #2 and/or #3:

For the second interview (Figure 3.6), I asked each participant to perform a research-related literature search that was based on the routine practice of the synthetic organic CoP. During this interview, I gave each participant an unreported organic compound (Task 1) and an unreported transformation step (Task 2), which consisted of a starting material and final product. These tasks were designed and approved by several expert synthetic organic chemists to ensure that the activities reflected the routine practices of the CoP. After each activity was designed and approved, I searched SciFinder to confirm that the compound and transformation was, in fact, unreported and to become familiar with possible search results. Also, I interviewed a chemistry graduate student prior to the data collection period to pilot the Interview 2 protocol and to test the screen-recording software (Blueberry Software©).

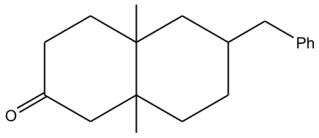
Although the second protocol was designed so that the participants would perform the two authentic searches (Task 1 and Task 2) during a single interview, during the data collection period, I observed that the participants were becoming fatigued during the Task 2 search. I made this conclusion after I interviewed my sixth-year participants, who both performed the Task 1 and Task 2 searches during a single interview. Because the participants were becoming fatigued during the Task 2 search, during the data collection period, I decided the Task 1 and Task 2 searches should be performed in two separate

interviews. Therefore, the sixth-year participants participated in two interviews and second-year and third-year participants participated in three interviews.

For the second and/or third interview, I provided the participants with a LiveScribe pen (a device that simultaneously records what they draw and say), paper, and access to SciFinder (via my computer). I then asked each participant to verbally describe their thought processes as they performed the search. Then, I gave the participants a sheet of paper that displayed the Task 1 compound (or Task 2), and I asked them to walk me through how they would use SciFinder to find a set of protocols that they could use in their laboratories to synthesize the compound. During each step of their search, if they did not verbalize their actions, I would ask them how they knew to perform each action of their search.

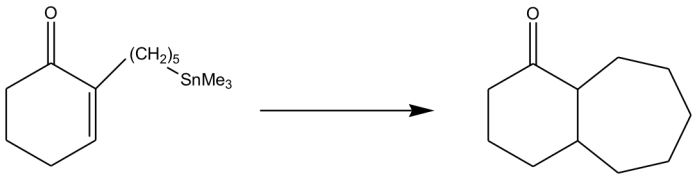
Interview #2 and/or 3

Task 1: Image you were given this molecule and were asked to synthesize this in your lab. Your goal is to search SciFinder to find a set of protocols that would allow you to synthesize the following molecule:



- Can you please explain what you observe when you look at this molecule?
- Using SciFinder, can you walk me through, verbalizing each step on how you would find an appropriate article/protocol to apply to synthesize this molecule? Please explain each step.
- Which articles would you choose to keep and apply? Why?
- How do you know the protocol would work for this synthesis? Please explain.

Task 2: Now, image you were given this transformation and asked to perform this in your lab. Your goal is to search SciFinder to find a set of protocols that would allow you to perform the following transformation:



- Can you please explain what you see when you look at this transformation?
- Using SciFinder, can you walk me through, verbalizing each step on how you would find an appropriate article to apply to perform this transformation? Please explain each step.
- Which articles would you choose to keep and apply? Why?
- How do you know the protocol would work for your research? Explain.

Figure 3.6: Interview #2 and/or #3 protocol

Data Analysis

I audio-recorded and/or screen-recorded (Interview #2 and/or #3 using Blueberry Software©) each interview, and then I transcribed the interviews verbatim. In accordance with the ethnomethodology framework, the unit of analysis for my study was the individual (Anderson, 2009; Bhattacharyya & Bodner, 2014). As such, the data analysis consisted of multiple stages to develop individual case records for each participant (Table 3.3).

Table 3.3: Description of data collected

Participant Pseudonym	Number of Interviews	Total Hours of interview	Total Pages of transcript	Total Number of Screenshots
Jane	3	3:04	60	160
Peter	3	1:34	36	70
Sophie	3	3:28	59	121
Bruce	2	2:30	41	27
Luke	2	3:25	43	43
Total	**13**	**14:02**	**239**	**421**

I made several iterative passes through the data to develop the individual case records. First, I described each participant's background experiences performing undergraduate research, which consisted of describing their previous experiences performing literature searches to develop research protocols and/or experiences using SciFinder. Second, I wrote a detailed chronology about their Ph.D. research (Interview 1 data) describing each step of their research-related literature searches using SciFinder. (Note: the detailed Ph.D. research chronologies were not included in the Results and Discussion chapter due to confidentially issues.)

My third step was to write a detailed chronology of their Task 1 and Task 2

SciFinder searches. Before I could write these search chronologies, several steps were taken to prepare the data. During this stage, I watched the screen-recorded literature searches (Task 1 and Task 2) and input a written description of the action into the interview two and/or three transcript. I also took screenshots of the action, (including any drawings that were recorded using the LiveScribe pen) as it was observed and discussed by the participants, and I input those screenshots into a separate document. I then watched the recorded search for a second time and wrote a detailed search chronology for each participant. Each chronology consisted of a description of the search action, quotes from the participants' transcripts to justify the action, and the screenshot (drawing using the LiveScribe pen and images from SciFinder).

Once the case records were organized, I analyzed each individual case record from a holistic perspective to determine the major emerging themes. Finally, I performed a cross-case analysis for all the participants to determine the major over-arching themes (Conclusion Chapter).

During my analysis I noticed some unforeseen limitations of the Task 2 (transformation) search. After I transcribed and wrote the chronologies for the Task 2 searches, I realized that during the interview I had to encourage the participants to deviate from their normal search behavior because the final product was commercially available for which they would just purchase. I felt that some of the authenticity of the task was lost and, as a result, my role as the researcher changed. Therefore, the data is neither included nor discussed in the results section.

Role of the Researcher

In qualitative research, the researcher is considered an instrument (Creswell, 2013; Patton, 2002). As an instrument, I was responsible for designing the study, and then collecting and preparing the data. My role during the interviews was as a "participant observer" (Patton, 2002). That is, I was a participant in the sense that during the interviews I would ask probing questions in order have the participants account for their actions during the literature search. I was an observer in the sense that I was not a member of the synthetic organic CoP for which I adopted the etic orientation. I was also a member of the chemistry department for which the data was collected. Therefore, I was responsible for maintaining the participants' confidentiality by striping any information that could link them to this study. In addition to performing the interviews, my role was to transcribe, analyze, and interpret the data. As an instrument in regard to interpreting the data (Patton, 2002), I had to be aware of my biases throughout this stage of my study.

Validity and Trustworthiness

The steps I took to establish the validity and trustworthiness of my study included: 1) ensuring transferability, 2) establishing credibility, 3) maintaining methodological consistency, 4) conducting peer debriefing, 5) performing triangulation, 6) executing member checking, and 7) using low inference descriptors.

Transferability (Lincoln & Guba, 1985): Qualitative research does not make claims to generalize findings, however, qualitative research does seek to maximize transferability. Therefore, I was transparent throughout each stage of my dissertation study and used "thick descriptions" so that my methods and results can be transferred to other contexts. For instance, I purposeful selected my participants who were in various stages of their

Ph.D. training to capture a variety of perspectives.

Credibility (Patton, 2002): Credibility or the rigor of my study was addressed in both the design of my study as well as during data analysis stage. Task 1 was designed and approved by several expert synthetic organic chemists to confirm that the task was, in fact, a reflection of the common practices of the synthetic organic CoP. I also addressed the credibility of my dissertation study by piloting the Interview 2 protocol with a chemistry graduate student before the data collection period began. This provided me with a chance to become comfortable with asking the participants to verbalize their thoughts and to use the screen recording software. Lastly, during the data analysis stage, I periodically met with several expert chemists to discuss my data.

Methodological Consistency (Strauss & Corbin, 2008): During each interview, I followed the same semi-structured interview protocol, and each participant was given the same target compound (Task 1) in Interview 2, thereby ensured methodological consistency. Furthermore, I used to the same method to develop and analyze the individual case records.

Peer Debriefing (Creswell, 2014; Lincoln & Guba, 1985): Throughout the course of this study, I discussed my research with expert organic chemists and STEM education researchers. During these peer-debriefing sessions, they would ask me questions and I would discuss and/or validate my methods and findings.

Triangulation (Creswell, 2014): Triangulation was achieved in several ways for this study. I used two different methods (two interviews) to collect data on how organic chemistry graduate students search SciFinder to develop research protocols. The first

interview involved the participants accounting for how they use SciFinder to develop research protocols to accomplish their Ph.D. research. During the second interview I observed and recorded their behaviors while they searched SciFinder to develop a research protocol for an unreported target compound. Therefore, I used multiple methods to gather data regarding their ISB. Also, multiple sources of data were analyzed to develop the individual case records. The sources of data consisted of my written field notes taking during the interviews, screenshots, transcripts, and the chronologies (Table 3.3).

Member checking (Lincoln & Guba, 1985): Following the second interview, I asked my sixth year participant (Bruce) questions regarding the authenticity of the Tasks. These questions allowed me to further confirm that Task 1 was authentic to the organic synthesis CoP. I also contacted my second-year participant to confirm that my interpretations of their ISB were consistent with the participants' understandings.

Low Inference Descriptors (Johnson, 1997): The goal of my study was to capture the participants' sense-making procedures of how they search, via SciFinder, to develop research protocols for unreported compounds. I used low inference descriptions that included thick descriptions and verbatim quotes in the Results and Discussion chapter to describe the participants' ISB, which grounded the results in the participants' accounts.

CHAPTER FOUR

RESULTS AND DISCUSSION

In this chapter, I will present the case records for each participant, beginning with the second-years (Jane and Peter), third-year (Sophie), and sixth-years (Bruce and Luke). Each case record will consist of the participant's background including prior experiences with SciFinder, Ph.D. research, Task 1 search chronology, Task 1 search critique, and a summary of their searches. The quotes included in this chapter were taken from the transcripts without correction of their grammar and/or syntax. Annotations that were added to the quotes were placed within square brackets. When there is an exchange between the interviewer and the participant, the interviewers will be prefaced with an "I".

JANE'S CASE

Background:

Jane was a second-year Ph.D. student who obtained her undergraduate chemistry degree from a large southeastern research institution. During that time, she had several research experiences in which she synthesized nanoparticles, worked in a forensics laboratory, and worked for a pharmaceutical company. For each research project, her advisor or employer gave her the experimental protocol, and then she was responsible for performing the experiments. She explained that the forensic internship was a great experience because she learned the process of how to develop a research method. In addition to carrying out the experiments, her employer wanted her to be knowledgeable about the compound with which she was working. She did not have access to SciFinder,

so she had to search by inputting the key words related to the compound into the Google search engine to obtain background information.

She had experience in her undergraduate degree using SciFinder to complete an organic laboratory assignment and an inorganic class assignment. For both assignments she participated in a SciFinder training session with the university librarian that was a requirement for each course. During the tutorial, she was taught how to set-up an account, search by topic and author, and was briefly instructed on how to perform a structure search. She explained that SciFinder was, "our primary source for our papers" for those assignments. To complete the organic laboratory assignment, she primarily searched by topic and performed some structural searches. For the topic search, she would input the type of experiment she was performing or she would try to name the compound and input that into SciFinder. For the inorganic assignment she was given a reported, yet not easily accessible, compound and she had to device a synthesis route. The instructor gave her a starting paper and she had to finish the project and find other articles to support the methodology. Specifically for this project, she would primarily search by the structural representation of the compound. She noticed that the results from the structure search displayed a reoccurring author that was primarily reporting the compound, so she then refined the search by the author's name. When deciding what paper to use, she stated that she used the rule: "if I found [the method] in two papers then I would use it." Lastly, her analysis for finding a research method was based on experimental conditions (specifically, concentration and reflux time).

Ph.D. Research

For her Ph.D. research, Jane was working on a project for which she was trying to synthesize a ligand, attach it to a cyclical organic backbone structure, and then complex the compound to a transition metal ion. In the early stages of this project, she was performing SciFinder structure searches to develop a method to attach the ligand to the reported cyclical organic backbone structure. Because the ligand with the backbone structure was unreported in the primary literature, her next step was to review her textbooks to familiarize herself with organic reactions. She explained:

> *A lot of just looking at the chemistry to remind myself. Like going through the textbooks. And then we realized this is just a basic, an easy reaction that you can add [ligand to backbone structure]. So I got a lot of papers downloaded with the backbone and we looked through the synthesis but they are all like convoluted and difficult for some reason. So we decided to try just using a basic one that wasn't reported. Just throwing them in a pot and boiling them basically vs. what they did.*

Following her textbook review, Jane's advisor proposed that they perform a series of common organic reactions to add the ligands. She attributed that decision to her advisor's knowledge of organic reactions.

> *It is off of [advisor] organic knowledge. It is just something you would do in an organic lab it is not hard chemistry.*

They had developed a method to attach the ligand to the backbone structure, but she was still performing SciFinder structure searches to determine how to perform the complexation to the transition metal ion. For this search, she would input structural representations, such as the complexed final product and also the backbone structure with the ligand. Because the complexed final product (backbone structure with ligand, complexed to the transition metal ion) was not reported, she would judge "similar" structures based on the backbone structure, length of the ligands, and if the ligands were

complexed to the same class of transition metal as the one of interest. Once she found a similar structure, she would then retrieve the article and read the abstract for key words relating to her project, skims the articles by looking at the reaction schemes, and read the experimental sections to evaluate the complexation step. With her advisor's help and the assistance of a colleague familiar with the project, they were able to modify the reported protocols to make the final complex.

Additionally, she used this opportunity to determine how researchers in the organic CoP characterize and report similar compounds. Her goal was to publish the data, so she needed to determine the common practices of the community. For instance, she did not know if the organic CoP reported the characterization of each step or just the final product. For this reason, she also evaluated the articles based on the researchers' methods for characterizing and reporting their data so she could disseminate her data to the community in a comparable way.

Task 1 Search Chronology

For the Task 1 search, Jane was given the structural representation of the target compound (Figure 3.6 in the Chapter Three: Methodology) on a sheet of paper and asked to describe her initial interpretation of the compound. Then she was provided with access to SciFinder and asked to walk me through each step as she searched to find research protocols she could use to synthesize the target compound from commercially available starting materials. The following is a chronology of her search:

- **Step 1**: Jane input the exact structural representation of the target compound and performed an RxS1 search (reaction search as the role of a product as a substructure of more complex structures).
    - **Step 1A**: She quickly viewed the two transformations resulting from the RxS1 search.
- **Step 2**: She then decided to discard the results from the previous step and performed a SS (structure search for substructures) for the structural representation of the target compound.

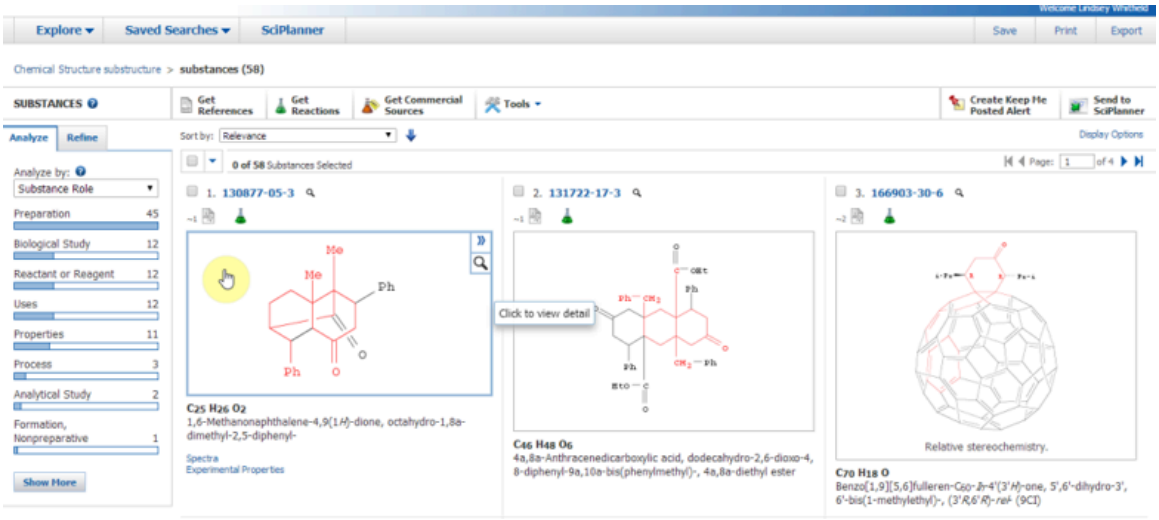


Figure 4.1: SciFinder screenshot of the SS search results

- **Step 2A**: The SS search yielded 58 substances (Figure 4.1). After viewing each substance, she then clicked on substance #1 and modified her search to retrieve the reaction for substance #1 as a product (product in Figure 4.2). (She clicked the "get reaction flask" for substance #1 and selected the role as a "product").

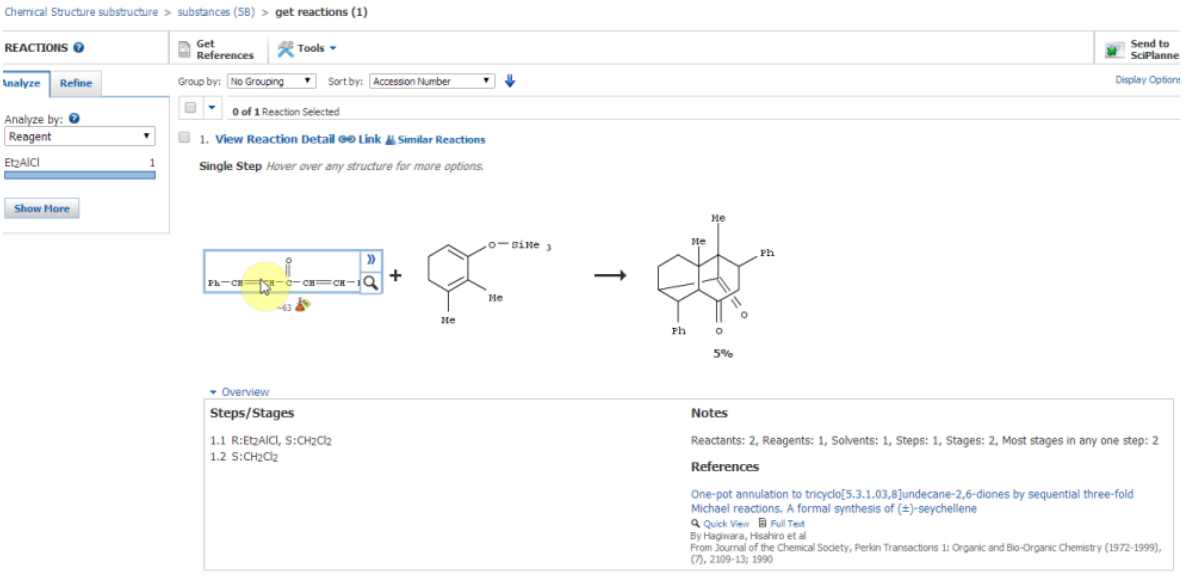


Figure 4.2: SciFinder screenshot of reaction #1 from the search for substance #1 from the SS search as a product

- **Step 2B**: The modified search yielded one reaction. She then clicked "full text" to retrieve the article reporting the transformation (Figure 4.2).
- **Step 2C**: She moved her attention to the downloaded paper and then scrolled to find the compound for which she was looking in the schemes. As she was viewing the schemes she noticed the named organic reactions, Michael and Diels-Alder, displayed in the figures above the reaction arrow. At this point, she also stated she would speak to her advisor as well as google the reactions.
- **Step 2D**: She concluded that the method reported in the article was not useful for the synthesis of the target compound.

- **Step 3**: She then proceeded to perform a SSi search for the structural representation of the target compound using, once again, the same representation that was used in **Step 1** above.
    - **Step 3A**: She retrieved the fourteen substances with 85-89% and 80-84% similarity matches, respectively. She viewed all the substances and decided to modify her search to retrieve the reactions for substance #1 as a product (Figure 4.3).

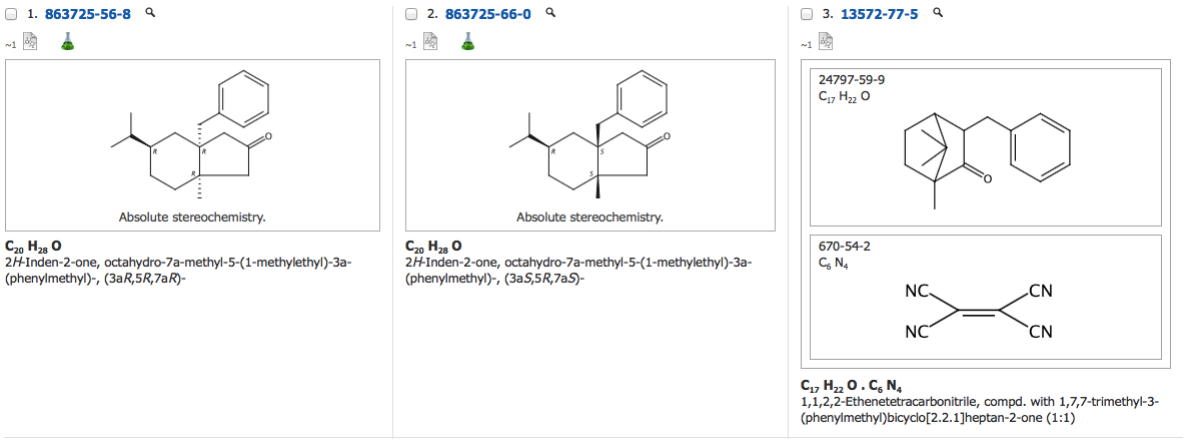


Figure 4.3: SciFinder screenshot of the SSi search results

- **Step 3B**: Once the seven reactions reporting substance #1 were retrieved, she then decided to sort them by percent yield (highest to lowest). She continued to evaluate the transformations by percent yield, number of steps, and number of starting materials. She concluded that the transformation did not provide her with a method she could adopt to synthesize the target compound (Figure 4.4).

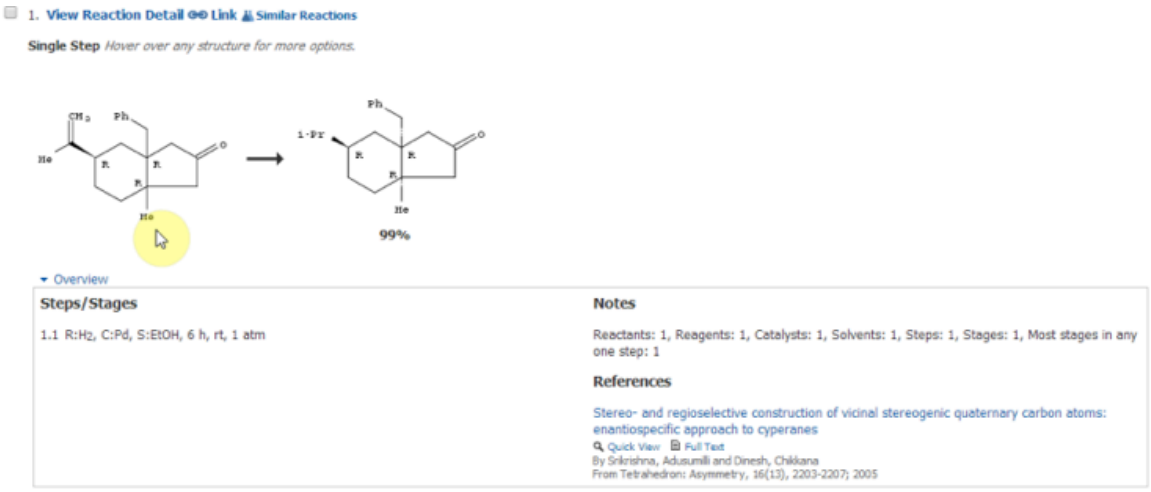


Figure 4.4: SciFinder screenshot of reaction #1 from the search for substance #1 from the SSi search as a product

- **Step 4**: Based on these results she decided to remove the benzyl substituent and perform a SE search for the backbone structure.
    - **Step 4A**: This search yielded 10 substances with the core structural representation of the target compound with different stereochemistry. She decided to retrieve reactions (7 reactions) reporting substance #1 (with undefined stereochemistry) as a product.
    - **Step 4B**: Using the same criteria as before, she evaluated the seven reactions based on the percent yield, number of steps, number of products, and simplicity (ring vs. chain) of starting materials and reagents. In addition to these considerations, as she was comparing each reaction she noticed that reactions #1-6 were reporting the same starting material.

- **Step 4C**: Therefore, she decided to sketch/note the starting material on the sheet of paper that was provided (Figure 4.5). (*Note*: she did not notice that the starting material was reported as commercially available in each of the reactions.)

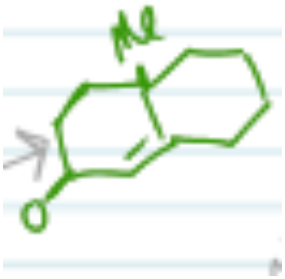


Figure 4.5: Sketch of starting material

- **Step 4D**: She retrieved the full text for reaction #5. After viewing the schemes, she then skimmed through the article to determine how the researchers either synthesized or purchased the noted starting material. She could not find that information, so she returned to the reaction results in SciFinder to continue to evaluate the reactions.
- **Step 4E**: She tried to pursue the full texts, most of which she could not access electronically, for several reactions reporting the simplified target compound.
- **Step 4F**: She continued to pursue the remaining articles until she was able to retrieve an abstract reporting the experimental conditions using the starting material.

- **Step 5**: Since Jane was still unable to determine how the starting material was obtained, she decided to perform a SS search for the previously noted starting material as seen in Step 4C.
    - **Step 5A**: The search resulted in 51,533 substances. She quickly viewed the first page of substances and then modified her search to retrieve the reactions for substance #1 as the role of a reactant.
    - **Step 5B**: The search yielded 279 reactions. After viewing the reactions, she immediately returned to the substance results and returned to the previous substances and modified the search to retrieve reactions for substance #1 as the role of the product.
    - **Step 5C**: The modified search yielded 66 reactions. Once again, during this part of the interview, she tried to access several articles (five different reactions) in an attempt to determine how the starting material was either synthesized or purchased, but she was unable to obtain access electronically.
    - **Step 5D**: After multiple unsuccessful attempts to retrieve several articles, she continued to the next page of results and found a *JACS* (*Journal of the American Chemical Society*) paper which she could access electronically (Figure 4.6).

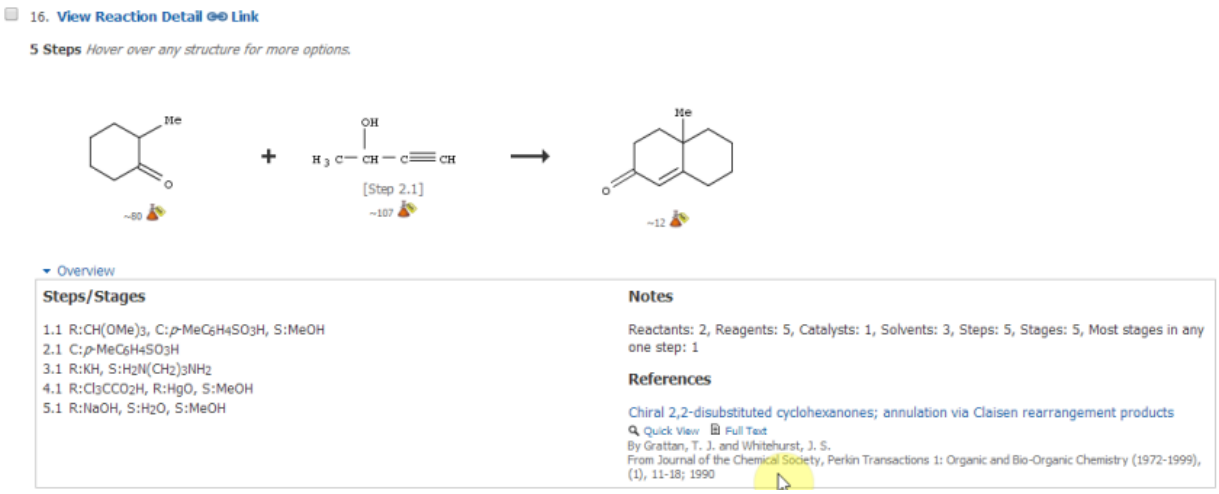


Figure 4.6: SciFinder screenshot of reaction #16 from the search for the noted starting material as the role of a product

- **Step 5E**: Once the article was retrieved she scrolled to the schemes to find the picture of the starting material. As she was evaluating the schemes, she stated that the reaction reported in the article looked like a catalyzed Diels-Alder reaction. She initially thought that the ketone on the final product was coming from the ring. However, she struggled to understand the reaction steps in the article, specifically how the alkene (chain) was added to the ring. For this reason, she would flip between the screen with the reactions in SciFinder and the schemes reported in the article to determine where the ketone was coming from (the chain or ring).

- **Step 6**: Because she thought that the ketone group had come from the chain, Jane decided to perform a SE search for the chain because she thought if she could find a carbon chain with a benzyl group attached, then she could use the method in the article to make the target compound.

- **Step 6A**: The search yielded 75 substances. She then retrieved the reactions reporting substance #1 (the chain) as the role of the reactant.
- **Step 6B**: The search yielded 6,986 reactions. She concluded the search for the chain was not useful because she still did not know how to add the benzyl group.

- **Step 7**: Returning to the *JACS* article, she continued to evaluate the schemes and the experimental section and discovered the researchers were using different carbon chains (alkene and alkyne). This discovery allowed her to conclude that the ketone group was resulting from the cyclization of the chain. Therefore, she concluded that the cyclohexanone starting material would need the benzyl group on it or she would need to add the benzyl group in an additional step.
- **Step 8**: Because she determined that the benzyl group would need to be on the cyclohexanone starting material, she then attempted two reaction searches. For the first search, she input the transformation from the *JACS* paper (Figure 4.7) and performed an RxS search.

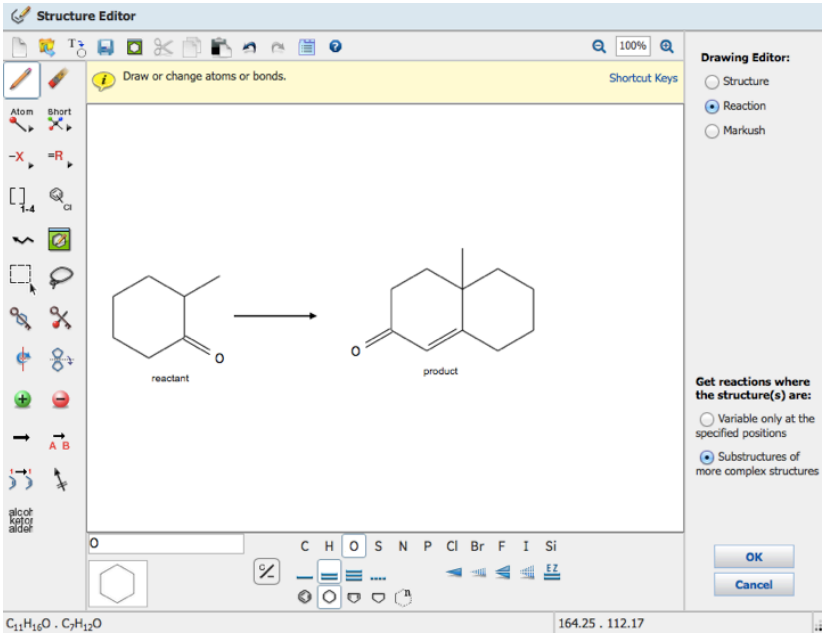


Figure 4.7: SciFinder screenshot of the transformation search using the starting material from the JACS article

- **Step 8A**: The search yielded 11,147 reactions. She evaluated the first two reactions and concluded she could not access the articles electronically.

- **Step 9**: Because the previous search yielded articles that were not available electronically, she performed another RxS search as seen in Figure 4.8.

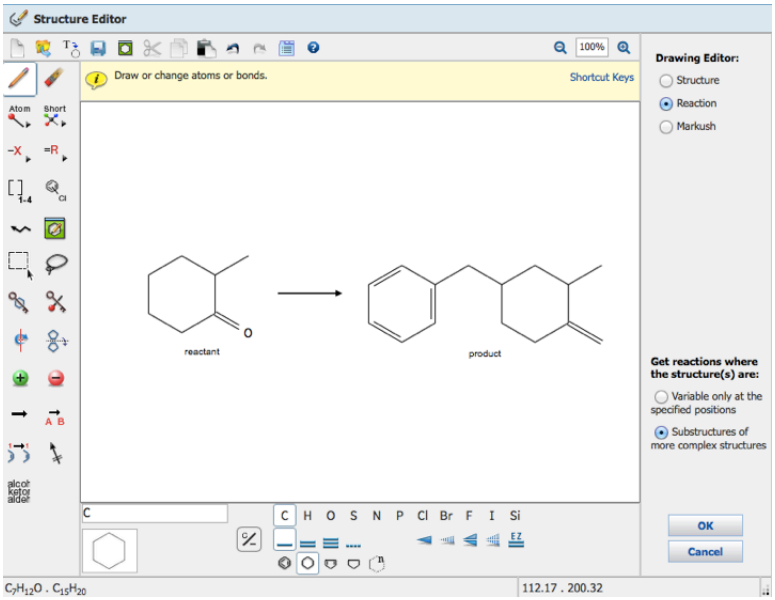


Figure 4.8: SciFinder screenshot of the transformation search using the starting material from the JACS article to a simplified version of the target compound

- **Step 9A**: The search yielded 131 reactions. As she was evaluating the results, she decided to sketch/note the first transformation and the reaction conditions.

- **Step 10**: She was still unable to use the information from the previous reaction searches to determine how to add the benzyl group. For this reason, she performed a third RxS1 search, now for the simplified cyclohexanone with the benzyl group (Figure 4.9).

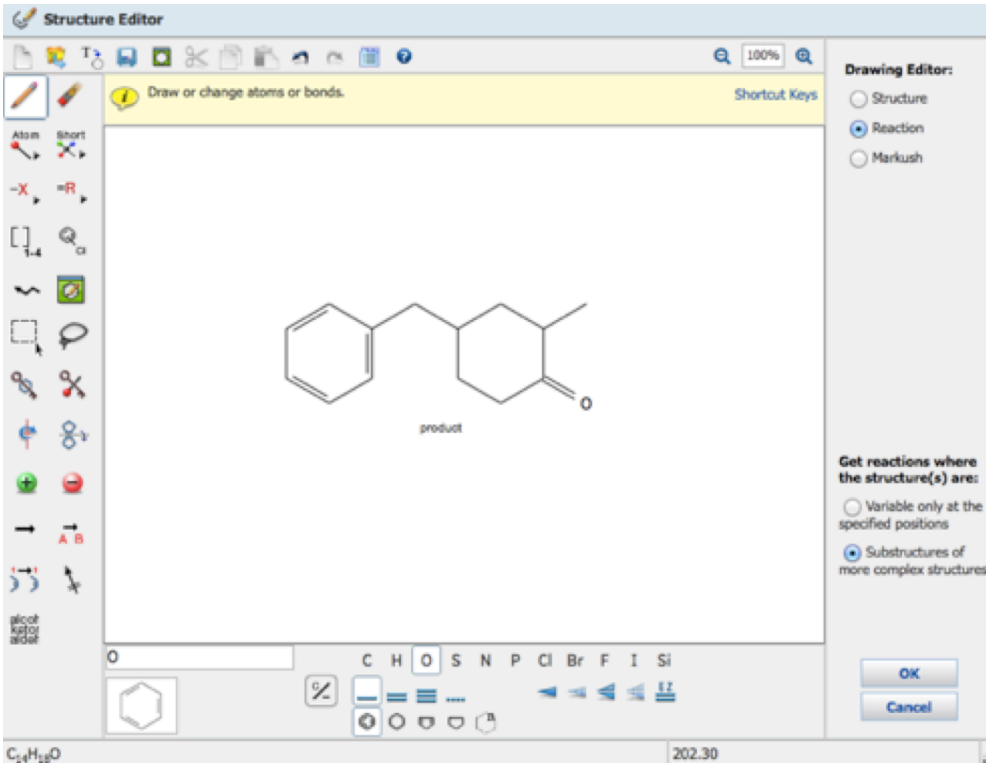


Figure 4.9: SciFinder screenshot of RxS1 search for the simplified target compound

- **Step 10A**: The search yielded 1,467 reactions. She quickly concluded that the reaction search for the simplified benzyl cyclohexanone compound yielded useful results because the first reaction had a high yield, the solvent was available in her lab, and the starting material was similar to the starting materials reported in the previous search results.
- **Step 10B**: She then decided to pursue the full text. She was able to access the article for which she then viewed the schemes and concluded she had all the information she need to propose a synthesis route to her advisor.

- **Step 10C**: She then sketched her steps (Figure 4.10) and stated she would need to discuss the synthesis with her advisor because she did not know how to selectively reduce the ketone located near the phenyl group.

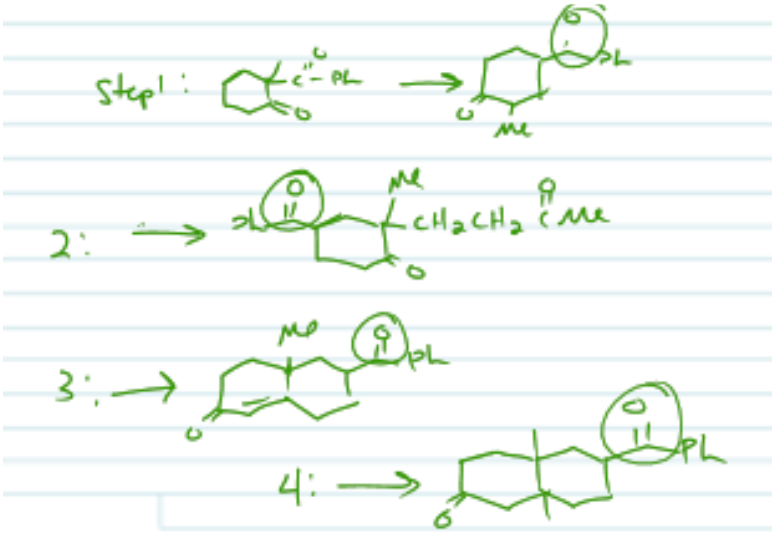


Figure 4.10: Jane's sketch of synthesis protocol

Task 1 Search Critique

One of the more significant characteristics about Jane was her recent entry into the organic synthesis CoP. This issue even permeated her initial reaction to and interpretation of the target compound:

> *So I don't have any ideas on how I would make it. Like this is completely new to me, like I can't even...like I would have to sit there and think of the name of it unfortunately. Yeah so I would immediately just search the structure of it. (Interview 2)*

In addition to the sense of concern she expressed, Jane was the only participant who thought about trying to name the target molecule. Since Jane had experienced success in performing in structure searches on SciFinder, the tone of the final sentence in this data

clip seem curious; her tone almost feels like doing a structure search is a *less* desirable strategy than, perhaps, searching the target by its name. Unfortunately, we were unable to follow up with Jane to get a clearer sense of what she was communicating to me at the time of the interview.

This strong sense of the unknown that were among Jane's first impressions of the target structure, she indicated that she was actually take a step "back" from immediately doing the search if this task were to be performed in the context of her research group,

> *I will pull my book out and look at it, my undergrad, organic book. I will look at it and talk to [group member] who is like a senior level organic grad student. So he has a lot more expertise in it. So I have ways of finding the knowledge but it is not just there for me, which can slow me down. But once I, like if I was synthesizing stuff like this for a year then I feel like I would be able to go through and find this information pretty quick. Having something completely new it would take me; I would probably spend like a day or two reading through and finding information before I actually start doing. Well I mean it would be longer than that before I would start doing the actual reaction. But I would spend a day or two by myself researching before I go and talk to somebody to try to figure out if I am even going in the right direction. (Interview 2)*

Consistent with her relative newcomer status in the organic CoP, the role of authority figures, such as senior graduate students or her advisor, was, not surprisingly, a significant aspect of Jane's overall strategy.

Despite these issues, Jane, like all the other participants, input the structural representation of the target compound and performed an RxS1 search (**Step 1** in her search chronology). She justified her decision to perform a substructure search by stating:

> *Just that is how [SciFinder] has it defaulted and it is a little more vague so it brings up more structures. Like some of my products that I make are new so if I search by substructure sometimes I can find something similar enough that it would have the same protocols. Like a different length amine chain or something*

like that, versus if you do the exact [structure search] then sometimes you won't find anything at all.

Even though Jane may have been a new member of the CoP, she did, however, have a clear strategy for searching for information in SciFinder. Although the RxS1 search resulted in two reactions (**Step 1A**), she quickly discarded the results and proceeded to modify her search to perform a SS search (**Step 2**). She justified this decision by explaining:

Yeah so when I do these searches the first thing I want to do is try to come up with the actual structure, and if I can't then I will go back and kind of settle with this substructure search.

Although she began her search by performing a broad substructure search to find compounds with similar structural features (specifically focusing on the substituents attached), her search behavior indicates that she would not consider adopting protocols for the purposes of synthesizing the target compound until she had determined that the target compound was unreported in the literature.

Moreover, she learned through her prior experiences using SciFinder that navigating the database from various routes would provide different information (RxS1 followed by a SS search).

Yeah cause I don't know like how they have them tagged but sometimes if you do structure search and then you find the structure and then go in set it to be the product structure you will actually find something even though it didn't come up with the reaction.

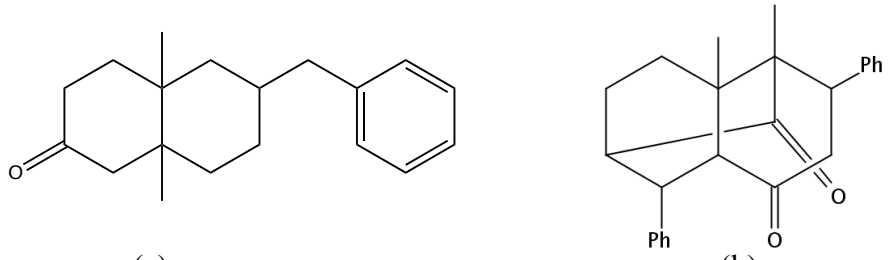
The SS search did broaden her results to 58 substances; however, she was still searching to determine if the target compound was reported.

> *I don't see the exact, like the exact structure is not popping up. I don't know if that means it is just not something that has been published or I don't know. It doesn't look like it would be a novel compound but it is not popping up.*

Interestingly, even though the RxS1 **(Step 1A**) and SS **(Step 2A**) searches did not yield the target compound, Jane continued to search for the target compound rather than conclude that it was likely to have been unreported. As the previous quote indicates, Jane appeared to rely more on her tacit assumption or intuition that the target compound "doesn't look like it would be novel" instead of using the search results to make that conclusion. As such, she proceeded through the series of search options.

Jane quickly viewed the 58 substances that resulted from the SS search (Step 2A), which were becoming increasingly more complex with additional rings and substituents (Figure 4.1), and returned to evaluate the first substance (Figure 4.11**b**). Although the SS search (**Step 2A)** yielded the same final products from the previous RxS1 (**Step 1A)** search, she decided to retrieve the reaction for substance #2 (Figure 4.11b) as the role of a product (seen in **Step 2A**) because the target compound did not result from either of the two searches. Her reasoning for pursuing this strategy was:

> *I like that this has the same backbone with two rings and it has your side phenyl group. But what worries me is that double bonded oxygen is not in the same location but I feel like it would still be kind of similar, a similar reaction. So then I would go and see what the reaction is, setting it as the product.*



(a) (b)

Figure 4.11: Target compound (a) and substance from the SS search (b)

Comparable to the target compound (substance #1), Jane observed that substance #2 had similar structural features, such as a bicyclic ring with a phenyl group and a ketone group. Because of the structure similarities, she thought the protocol might provide chemical information for which she could adopt/modify in order to synthesize the target compound (substance #2). Furthermore, at this stage in her method development, she was not concerned about the positioning of the ketone group; however, this consideration could affect the chemistry she could accomplish by adopting the protocol. This tactic is particularly significant since synthetic routes are determined by the nature of the structural features of a target *along with* their relative orientation.

The modified search for substance #2 as a product (**Step 2A)** only retrieved one reaction, so she decided to access the journal article (**Step 2B**) to further evaluate the reaction schemes. Once the article was retrieved, she viewed the schemes and stated:

> *But I would look through but just to see if they have something like a basic scheme of their reaction that even would be helpful. Like that [final product] almost looks like it could be useful because it has got your double bonded oxygen and they are putting on the side chain. So I might glance through the experimental section just to see what they have got in here and then maybe take it to [her advisor]. So they are doing a Michael reaction and then a Diels-Alder [reaction]. So those are both well-known organic reactions so I think I would go and talk to [her advisor] since he is the organic Professor and he would very well know. And I might even do some Google searches just on the basic Michael and*

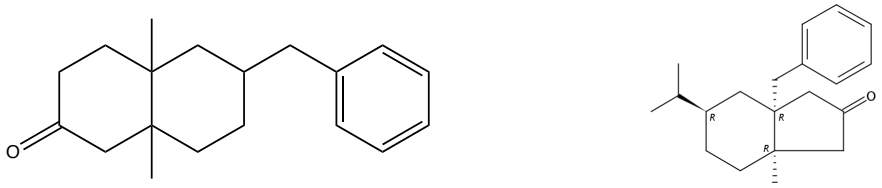
Diels-Alder reaction so I can kind of compare what I would be wanting to do [for the purposes of the target compound]. But that is really all I would do with that search.

The previous quotes suggest that her initial response when encountering chemical information for which she could not either comprehend or put into practice, such as the Michael or Diels-Alder reactions, was to seek her advisor's assistance so that he could communicate if the information was useful for the purposes of synthesizing the target compound. In addition to seeking help from her advisor, she would also use Google to search for those reactions instead of using SciFinder, the most widely utilized database by practicing chemists. This indicates a clear disconnect between how Jane utilizes the chemical database, which was for the purposes of developing protocols and not necessarily for gaining basic chemical knowledge.

Ultimately, she concluded the article provided her with named organic reactions, but she could not put her knowledge of those reactions into practice to propose a synthesis protocol for the target compound. In other words, she did not know which bonds to break on the target compound to propose starting materials (retrosynthetic approach) that would undergo either the Diels-Alder or Michael reactions. For this reason, she continued to rely on the database and proceeded to perform an SSi search for the structural representation of the target compound (**Step 3)** because she thought this search might provide new information. The search did yield different substances for which she quickly concluded that substance #3 (Figure 4.12**c**) would potentially provide useful information due its structural similarities to the target compound.

Like this actually looks quite useful cause it is a 5-membered ring instead of the 6-membered ring that you would want. But it does have the double bonded oxygen

in the same location. And then this would just need your side product or side chain down on a different spot and of course a different side chain. But that is still similar enough that I could probably look at that and see what their reaction is, and I would just glance though.



(a) (c)

Figure 4.12: Target compound (a) and substance from the SSi search (c)

Jane gravitated to substance #3 because, in comparison to the target compound, it had a bicyclic ring structure with a ketone group and an iso-propyl group. Once again, her actions were indicative that she was more concerned that compounds had similar structural features and less concerned about the location of these groups. More importantly, she was concerned that the starting materials had the ketone group already attached. For instance, she explained:

Jane: I don't feel like you would be able to easily go back through and add [the ketone group] after you have this synthesized. Like you would want [the ketone group] at the start. But I think adding like your different side chains might be a little easier than the ketone group, if that makes sense.

I: And why do you think that?

Jane: I don't know, just like when I have done reactions like this, I've felt that having my starting materials always would have like that part of the structure already there. Because that is something you would have, like you can get rid of a ketone really easily, but I feel like you can't make ketones quite as easily.

I: Do you know why you would think that you can, you said you can make ketones...what did you say?

Jane: I said you can get rid of them, like you can throw on a reduction group and just get rid of that but I feel like...so if you are wanting to use it in your starting product then you need to also remember to avoid those.

I: And you know that because?

Jane: Basic organic knowledge. Remembering it back from sophomore year forever ago.

Jane's inclination to pursue substances with a ketone group was based on her previous experiences for which the group was already attached. Additionally, the previous exchange indicates that she was basing her knowledge of ketone groups on her undergraduate experiences and not her research experiences. Although Jane had been teaching undergraduate organic chemistry labs and performing organic research for about a year, her justification for wanting the ketone group on the compound was based on her undergraduate experiences. More importantly, her justifications were based on her chemical intuition and not chemical reactions that yield ketones.

She retrieved the reactions for substance #3 as the role of a product (**Step 3A**) and decided to sort the reactions by the percent yield (**Step 3C**). She justified her action by stating:

I've played around with SciFinder a lot. I like to look at the sort by button because it can be quite helpful because sometimes you will search something and have like 2,000 products and just looking through this I really don't care about publication year or accession number. So these three [number of steps, product yield, experimental procedure] would be the more helpful except I don't know enough about the procedure to know what kind of procedure I would want. So these two [number of steps and product yield] are a little bit more helpful. Like you would want to avoid ideally something with 20 steps, but I would start with product yield because that is kind of the most important thing to me is having a good yield.

During the Task 1 search, Jane was constantly evaluating the reaction results by the percent yield. Her search behavior indicates that she had developed heuristics regarding her approach to evaluating of the reactions results. Her first step was to find a reaction with a high yield, and then she would evaluate the remaining information, such as the number of steps, complexity of starting materials, and number of products. These heuristics seemed to lessen the cognitive load thereby enabling her to quickly evaluate the reactions.

Following her evaluation of the reactions from the modified SSi search, she concluded the results were not useful because:

> *We are still not making the product that we want which could be an issue. Because I am not seeing, with my organic knowledge, I am not seeing how they made the five-membered ring with the double bonded oxygen on the side. So that would be something else that I would take to [her advisor] and talk to him about it.*

Jane once again encountered information for which she could not conceptualize and would seek help from her advisor so he could communicate the usefulness of the information. Although Jane seemed to encounter another roadblock with her method development, she continued to perform another search for the modified target compound (**Step 4**). At this stage in her search, she had exhausted her search options for the exact structural representation of the target compound, so she decided to perform a SE search for the modified target compound (without the benzyl group attached). She justified this action by stating:

> *Jane: Just since [the benzyl group] is like a side product or side ring, a side chain. I feel like that is something that could probably be changed, even if it is changed on the starting material, it is probably not something super major.*

Because that phenyl group is not going to be too much of an issue I don't think with your reaction.

I: Adding it on?

Jane: Yeah. Because if they had like some form of a halogen group right here [referring to the benzyl position] it could probably be a doable thing to like throw on a side chain.

I: You know that?

Jane: Basic knowledge.

I: I guess some of the stuff that I am asking, is what's kind of, what are you going back to know that?

Jane: A lot of it is actually; I like teaching because it kind of reminds me things that I have forgotten. But I remember that halogens are used a lot to add things to organic compounds just like because they make the area more reaction so you can swop out the halogen for something. I remember that basic reaction.

Despite that Jane was immersed in an organic research group, her justification to remove the benzyl group was situated within her Ph.D. experiences teaching undergraduate organic labs. Through these teaching experiences, she had gained some chemical intuition regarding "basic reactions" to add the benzyl substituent; however, she never put her idea of these reactions into practice during the Task 1 search.

The SE search for the modified target compound yielded 10 substances (Step 4A) for which she then retrieved the reactions reporting the simplified target compound (Figure 4.13) as a product (**Step 4A)**.

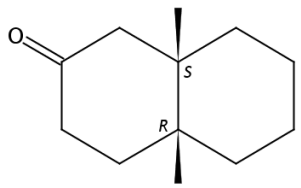


Figure 4.13: Substance #4 from the SE search

She continued to use her same criteria to evaluate the reactions (**Step 4B**); however, during this time she would frequently skip reactions reporting starting materials that appeared expensive to either purchase or make.

> *This looks ridiculous [*Figure 4.14*]. I don't know what this is so I am going to ignore that one. I do like to look at the easier reaction obviously especially if something looks like it is going to be a cheap starting material. So methyl lithium and methyl copper I am not sure like for instance how expensive that would be or that [*Figure 4.14*]. These are probably catalysts I would image. So I might even go over it with, we have a computer inventor so I would go and type these in and see if we have methyl lithium in the lab. And if we don't, it doesn't mean I am not going to look at it, I might still come back and look at it.*

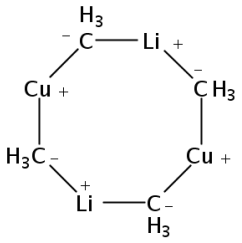


Figure 4.14: Starting material discussed from the SSi modified search

The prior quote reiterates a common theme for Jane's Task 1 search, which was to either seek her advisor's help or to completely ignore chemical information which she did not understand. In this case, she did not know the price of the start material (Figure 4.14), so she would proceed to evaluate the other reactions. Furthermore, she was specifically discussing methyl lithium which was a compound listed on SciFinder as a commercially available substance. Interestingly, she would search her laboratory inventory for the compound or avoid the reaction completely, instead of using the links SciFinder provides to determine the cost of the compound. Although cost is an important criterion to consider when developing a research protocol, she was avoiding reactions

based on her assumptions and not on actual reliable information from chemical suppliers, such as Aldrich.

She proceeded to evaluate the remaining reactions and noticed a re-occurring staring material, which she decided to sketch on the paper provided (**Step 4C**).

> *I'm looking kind of at the similarities between the reactions and one that I am seeing is they seemed to have started from this [noted starting material]. So at this point, I am kind of almost working backwards through the reaction. So I normally actually just keep a notepad next to me when I do searches and I have a lot of random notes from me going through.*

This was a critical step for Jane's Task 1 search because she was able to use the information that SciFinder provided to take one step back and propose a potential starting material. More importantly, she used the organic CoP, accessed via SciFinder, to perform the retrosynthetic approach because she was able to propose a starting material. Therefore, her search objective was now directed at determining how the starting material was either purchase and/or synthesized. Jane, however, did not notice that SciFinder was reporting the starting material as commercially available (indicated by the flask underneath the starting material), and as a result, she continued to search for either a synthesis protocol or the purchasing information for the noted starting material. As she was searching for the information regarding the starting material, she attempted to access several articles reporting different reaction with the noted starting material, and once the articles were retrieved she would notice that she had retrieved the same article. (The paper was reporting several methods to make the simplified target compound.) For instance, after she had clicked to retrieve a "different" reaction, she realized she had retrieved the previously accessed article.

But unfortunately this is the exact same paper. Um so that is something that I probably wouldn't notice because I am a little oblivious but I would notice as soon as I opened it up that it is the same paper.

Her actions indicate that she did not read the information regarding the journal article very carefully, but was primarily relying on the structural representations when deciding to pursue an article. Because she tried to access several transformations that were reported in the same article, over the course of the interview, she altered her behavior to read the information regarding the journal articles before she clicked to retrieve the papers.

During this part of her search, Jane was unable to obtain access to the electronic version of several articles she was pursuing, but she was persistent to determine how the starting material was either synthesized or purchased. As such, her inability to obtain the electronic version of the articles was a roadblock, but Jane refused to deviate from her search objective and continued to search for the information for the starting material. For this reason, she decided to perform a SS search for the structural representation of the noted starting material (**Step 5**). She justified this step in the following exchange:

Jane: Because I do want the reaction of how to turn it into that final structure [modified target compound] and since I couldn't pull up a paper for it except for this one, I'm hoping maybe it will bring up some more [reactions with the starting material] just because it is listed differently. I've never understood, occasionally it will do it though.

I: So you have had previous instances where you have that same product?

Jane: Yeah you kind of have to like sneak around SciFinder sometimes. You have to trick it into giving you the answer you want it to give.

Again, Jane was trying to manipulate the database in an attempt to obtain reactions using the starting material for which the previous search did not provide. As seen in **Step 5** in

the search chronology, she then decided to pursue the reactions for the starting material as the role of a reactant. She initially concluded the search yielded useful information for the synthesis of the target compound; however, she finally concluded that the results were not useful at this stage in her method development. She stated:

> *But I think it would be more important to figure out if I used this [noted starting material] so I think I will go back and figure out how they made it to see if that is something.*

Her quote and search behavior suggests that she was able to reflect on her actions and adapt her strategy. Therefore, she modified her search for the starting material as the role of the product (**Step 5B**), which she justified by stating:

> *Because I am going to see if I could just change the starting material to where I would end up with [the starting material] having the side chain that I want…But I am focusing more on simple starting materials than anything else because I want to be able to either easily make my starting materials or buy them without a lot of money.*

Earlier in the interview, Jane was trying to retrieve articles to determine how the starting material was either purchased or synthesized; however, now she was broaden her approach in an attempt to determine if there was method reporting starting materials with a benzyl group or similar alkyl group already attached.

The modified search as the role of the product yielded 66 reactions for which she quickly evaluated the reactions based on the previously mentioned criteria. As she was evaluating the reactions, she would attempt to access several articles, but could not access the papers electronically. Furthermore, she would, once again, disregard reactions reporting compound for which she did not know how much effort it would take to make the starting materials.

*It looks a little complex [*Figure 4.15*] so it would definitely require me making it. It looks like they are using this same starting material but it looks like they adding something to it and I don't know how much that would take effort wise. So if I can avoid it, I'm going to avoid it.*

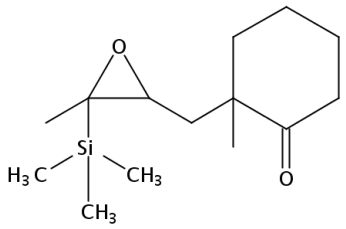


Figure 4.15: Complex starting material from reaction #3 from the modified SS search

Jane's approach to avoiding reactions that appeared complicated was a practical approach because she had several pages of reactions that she could potentially adopt. For instance, as seen in **Step 5D**, she proceeded to the next page of results and noticed a reaction reported in *Journal of the Chemical Society* (*JACS*). She justified her decision to pursue the article by stating:

*Jane: [The reaction in **Step 5D** in the chronology] has similar starting materials, which is useful and also that it is JACS.*

I: So they gave you similar starting materials to some of those other papers that you were looking at?

Jane: Yeah but it is actually accessible. So I should have just gone to the next page.

The previous exchange indicates Jane's familiarity with *JACS* articles and more importantly that she immediately decided to retrieve the article because it was available electronically. Thus, this suggests Jane had a tendency to avoid reactions with "complex" starting materials or reagents and to pursue articles that were readily

accessible. Once the article was retrieved, she proceeded to find the compound for which she searched in the schemes.

> *Jane: So I'm looking at it I'm just scrolling through and I like to look at their schemes first, like their images.*
>
> *I: Why is that?*
>
> *Jane: Because if I can find it, like a lot of times when they make organic papers they will make like a bunch of different things and they will just label them like numbers and letters. And a lot of times they don't explain it very well but sometimes in the pictures you can find it. So I can see what I want is number 7. So then it is a little bit easier in their experimental section to figure out.*

Jane's strategy to evaluate journal articles allowed her to quickly link the compounds in the schemes to the experimental procedures. Although she was able to connect the labels/compounds to the experimental section, she still needed to add the benzyl group, either to the chain or the cyclohexanone starting material. She elaborated on her thought process by stating:

> *But it looks like they have start with like these chains [alkene and alkyne groups]. I think that is what all of these are. Yeah so they all have the chain plus this ring product. So I would probably just try to have my chain either added with that phenyl group or already on there if I don't have to add it myself that is one less step. I just don't understand. So the one other thing I might search is, because I don't understand this reaction that I was trying to find. So the one other thing I might search is how to add that phenyl side chain onto it wherever it needs to be. Which I imagine, I'm trying to figure out this reaction because they don't explain it very well in their paper. That is the only issue.*

She was still trying to devise a method to add the benzyl group, either to the chain or the ring, so she performed a search for the chain (**Step 6**). Her behavior was very surprising during this part of the interview. She was "trying to figure out this reactions" reported in the article, but she never attempted to use the arrow-pushing formulism (a tool used by practicing organic chemists) as a way to conceptualize the mechanism. Instead, she

performed a structure search for the chain (**Step 6)** to gain additional insight regarding how the substance was reported as a reactant in other reactions.

> *I am just looking to get an idea of like what [the chain] it is used for. So like right here I'm seeing they kind of added more chain to it. It is really not super helpful but I don't know where the phenyl group needs to be added.*

Although she was attempting to gain additional perspectives regarding how the chain was used in different reactions, the search did not provide her with any useful information about the reaction reported in the *JACS* article. For this reason, she then decided to return to the *JACS* article and once again evaluate the scheme (**Step 7**). She was still struggling to comprehend the mechanism (i.e. how the chain was being added to cyclize the ring) because the researchers were using several hydrocarbon chains (alkene and alkyne) and were labeling them in a separate location. She finally noticed that the researchers were using multiple hydrocarbon chains and concluded:

> *The ring with the oxygen and the methyl group on it and they are adding their side chain to it and then they are closing the ring. I think that is what they are doing. So in that case they probably...but so I am looking at what they have in the scheme, which they didn't number because apparently they didn't do that in 1990. I don't know. So it looks like what they are doing is taking the starting material [cyclohexanone starting material], adding their side chain like I said, and closing their ring. So this side chain I think is actually going to be where this double bonded oxygen comes from. I think that is what is happening. Which would mean in that case I would want my phenyl group on this ring. Not on the chain that I add. So it is backwards from what I thought if that is the reaction that they are doing which I think it is looking at it. So it looks like they stick that on here.*

Although she was not using the arrow-pushing formulism to conceptualize the mechanism, she was able to process the reaction schemes and conclude that the benzyl group would need to be on the cyclohexanone starting material. Also, during **Step 7**, as she was reading the experimental section and stated:

> *So that is talking about the Claisen reaction which is another named organic reaction. So that would be something else that I could Wikipedia just to look at the basic reaction to see if that is something that would be useful.*

Again, instead of using the most widely used database by practicing chemists to search for the Claisen reaction, she would use Wikipedia to determine the usefulness of the reaction for the purposes of synthesizing the target compound.

She was still trying to determine how to add the benzyl group to the cyclohexane starting material, so she performed a reaction search for the cyclical starting material from the JACS paper (**Step 8)**. She first thought the search yielded new, useful information, but she quickly realized the transformation was the one for which she previously could not access electronically.

> *So this looks promising just because it is 95% is what I see immediately. That looks like it might be something pretty straightforward to make but I'm pretty positive I'm going to have to make it since it has that chain coming off of it. However it is the same reaction. Oh yep Tetrahedron.*

The quote suggests that Jane's search strategy had evolved over the course of the interview. For example, in the beginning of the Task 1 search she had a tendency to gravitate to reactions with high product yields. However, as the interview progressed, she started to adapt her behavior to consider the information regarding the journal article as well.

Because the previous search yielded the same reactions, she then tried another transformation search (**Step 9**). Although the search yielded 131 reactions, she quickly concluded the results were not useful for the purposes of synthesizing the target compound.

Jane: So I'm not seeing what I need to see. So that would probably be something that I talk to [advisor] about.

I: On how to get the benzyl group on there?

Jane: Yeah I feel like it would be a basic reaction but it might be more difficult than I'm thinking.

Earlier in the interview, she stated that she could perform a nucleophilic substitution reaction to add the benzyl group. She remembered this reaction from her experiences teaching undergraduate organic chemistry labs, but she was unable to find precedence for this transformation in the literature. As such, Jane had reached another roadblock for which she would need to seek her advisor's assistance.

Despite that Jane had reached a roadblock in her search, she was persistent to find a method to add the benzyl group. For this reason, she performed a search as seen in **Step 10**, which yielded information that she quickly thought would be useful to add the benzyl group because reaction had a high yield, the solvent was available in her lab, and the starting material was similar to the starting materials reported in the previous search results. For these reasons, she decided to retrieved the full text and concluded she had all the information she needed to propose a method to her advisor. She then sketched her proposed synthesis protocol (**Step 10C**) and explained:

Yeah so this would give it with my phenyl group that I want on it. Which I believe I would then be able to react into that right here. So I would have step one, I would have this thing. I'm just going to draw it. I think step one would be. So this would give me my side chain that I need and 79% yield seems pretty good. If I could get that. So that would happen [step 2]. And then the only thing I don't know about this is that I would have to get rid of that double bonded oxygen [circled in Figure 4.10*]. And I am not entirely sure how I would do that without having to get rid of that one as well. So that would be something I might have to talk to [her advisor] about to see if it would be an issue. Cause it might be that reacts differently than that one because this one is not part of a ring is the only*

> *thing I can think. So I would do that and then you could take it to two would then be reacting it with this side chain. So I would want to make that somehow. But now it has my phenyl group on it. Then I would react it to make...[sketching step 3] I don't know when this [circled in* Figure 4.10*] would go away is the only thing that I would need to worry about, this double bonded oxygen. And then finally I would react it into that final product. [sketching step 4] But I still don't know about that double bonded oxygen. So that would be something I would have to talk to [advisor] about.*

She was able to compile all the reactions/information she retrieved from the SciFinder search and proposed a synthesis protocol for the target compound. Despite that she was able to compose a research protocol, she would still need to discuss with her advisor about a potential protecting group that would enable her to selectively reduce the ketone located near the phenyl group. Furthermore, she explained:

> *I: So before you actually go in the lab, although you have all these protocols, you would say ok this is what I am thinking what do you think?*
>
> *Jane: Yeah cause I would probably walk in there with these five papers and [her advisor] would be like oh you just have to throw it all in a pot and boil it. That is what would happen inevitably. But I would at least have an idea to walk in and not sound like an idiot, which is the goal.*

Although she was persistent to compose a feasible method, she was more concerned with compiling a series of potential reaction steps that she could at least propose to her advisor. In other words, she seemed satisfied that she had something to propose and was not going to her advisor empty-handed. As such, Jane could discuss her protocol with her advisor and he could ultimately provide her with feedback regarding the feasibility of the method before attempting the chemistry.

Summary of Jane's Search

Although Jane was a newcomer to the organic CoP, her prior experiences performing SciFinder searches had enabled her to develop the dexterity to use the

database as a vehicle to develop a research protocol. Using a reductionist perspective, she relied the database to perform the retrosynthetic approach to develop a research protocol for the target compound. Furthermore, she frequently encountered roadblocks, but used her chemical intuition to persistently modify and change her search strategy. Her strategy ultimately enabled her to compose a series of reaction steps that she would discuss with her advisor and he could suggest potential protecting groups and/or alternate methods.

PETER'S CASE

Background

Peter was a second-year Ph.D. student who obtained his undergraduate chemistry degree from a small university in the northwest United States. During his undergraduate degree he participated in a research experience for which he collected and analyzed a variety of water samples. Peter had experience in an undergraduate organic chemistry course using SciFinder to complete a semester-long assignment for which he was given a target molecule and his goal was to propose a synthesis pathway. Before he began this assignment, a librarian attended two class lectures and taught them how to access SciFinder, create a password, draw compounds in the JavaScript structure editor screen, retrieve articles, and access the supplementary information. He explained that she just showed them the basics and did not explain how to evaluate the articles to determine which protocols would be applicable for the synthesis of the assigned compound. As a result of the class assignment, he became more effective at performing a SciFinder search and develop a strategy to evaluate the search results, which was based on the percent

yields, compounds similar in structure to the compound for which he was assigned (functional groups and alkyl moieties), and commercial availability. Similar to the undergrad class project, Peter took a graduate level organic chemistry course for which he was given a target compound and asked to propose a synthesis pathway by considering the most feasible reactions, *i.e.* those with the best yields and lowest costs.

Ph.D. Research

For his Ph.D. research, Peter worked on two different synthesis projects. For one project, he halogenated a specific class of compounds. He began this project by inputting the exact structural representation of one of the target compounds into SciFinder; however, the search did not yield any results because the target compound was unreported. He then searched by inputting the structural representation of the core structure and inserting an –X (SciFinder recognizes an –X as any halogen) to find similar compounds.

> *First you would put the [halogen] and you would see that no one has ever made, [target compound] has never been reported. So then you would go back, and SciFinder has this really good thing where instead of putting this [halogen] there you can put –X, meaning any halogen, or you can put an –R group meaning any alkyl group. And it will come up with different relevant things, reporting many different –R groups and –X groups. And so that helped out big time. Kind of narrowed things down. And then I guess I finally got lucky by finding a paper that had something similar to this.*

He was able to find a compound with a halogen attached to the backbone of the core structure of the class of molecules in which he was interested. Ultimately he slightly modified the protocol that he found in the retrieved citation from the search and successfully made his final product with a high yield.

His second project was still in the preliminary stages, so he stated that, "SciFinder [was] huge for this" project. He was trying to synthesize a "huge" biologically-relevant, unreported organic compound, so he was breaking the compound into parts in an attempt to design a synthesis for the final product. He explained that each piece of the compound might be reported but the steps to attach each piece was not. When he would find a potentially useful article for the synthesis of each fragment of the compound, he would focus on the schemes because the pictures displayed the chemistry.

> *So they will give you different schemes that will show you the reactions. It will tell you what is going on in that reaction. So maybe it pertains to the molecule you are making. Maybe it will give you some sort of enantiomer selectivity so you will know which enantiomer you are making. Um cause like I said the percent yields that you need, because you don't want to have something that gives you a really low percent yield. So most people focus on, I would say schemes and the text just helps you learn the schemes*.

During the data collection period, he was still trying to develop a method to make the target compound.

Task 1 Search Chronology

The following is a chronology of his search:

- **Step 1**: Peter input the exact structural representation of the target compound into SciFinder and performed a SS search.

- **Step 1A**: The SS search retrieved 58 substances (Figure 4.16**)** for which he quickly scrolled down the page to view the substances.

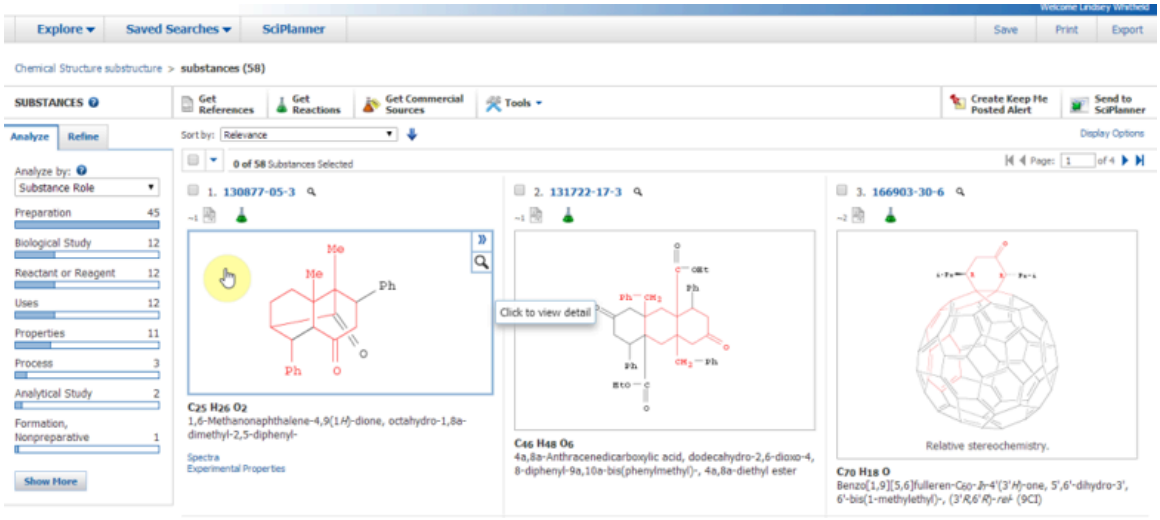


Figure 4.16: SciFinder screenshot of the SS search

- **Step 2**: He then discarded the results from the previous step and decided to perform an RxS1 search (search for substructures) for the structural representation of the target compound.
    - **Step 2A**: The RxS1 search yielded two reactions. He quickly viewed the two reactions, but he did not decide to pursue the full texts.

- **Step 3**: Returning to the structure editor screen, he modified/simplified the target compound and performed an RxS1 search as seen in Figure 4.17.

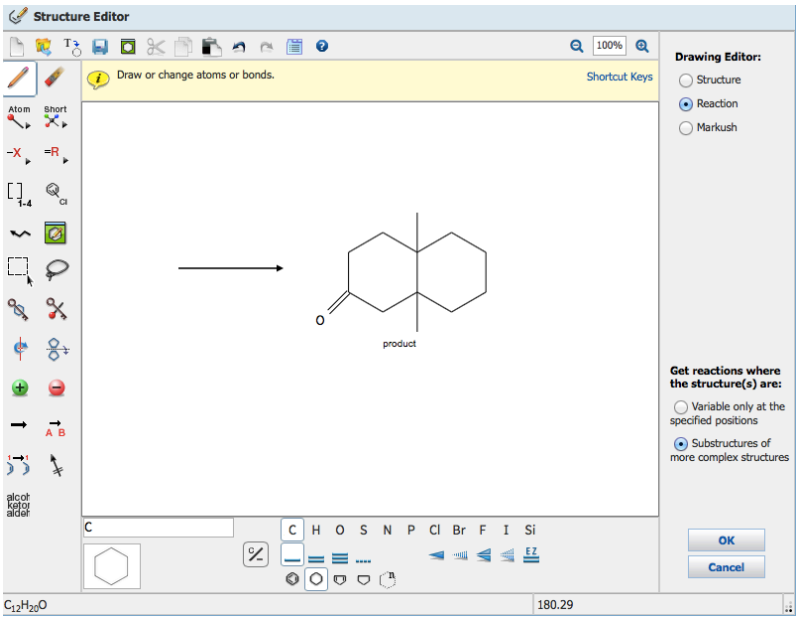


Figure 4.17: SciFinder screenshot of the RxS1 search for the modified target compound

- **Step 3A**: During this time, he also sketched his thoughts regarding an S_N2 reaction pathway to add the benzyl group (Figure 4.18).

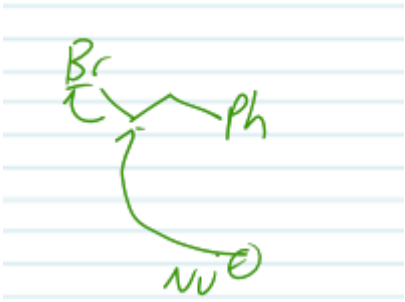


Figure 4.18: Sketch of S_N2 reaction

- **Step 3B**: The RxS1 search yielded 2,667 reactions. He proceeded to evaluate reactions one through six by focusing on the product yield.
- **Step 3C**: He also evaluated the starting material in reaction #1 and decided to click "view substance detail" to take a closer look at the compound.
- **Step 3D**: Returning to the SciFinder results, he continued to evaluate the reactions for which he then retrieved the "full text" for reaction #3 (Figure 4.19).

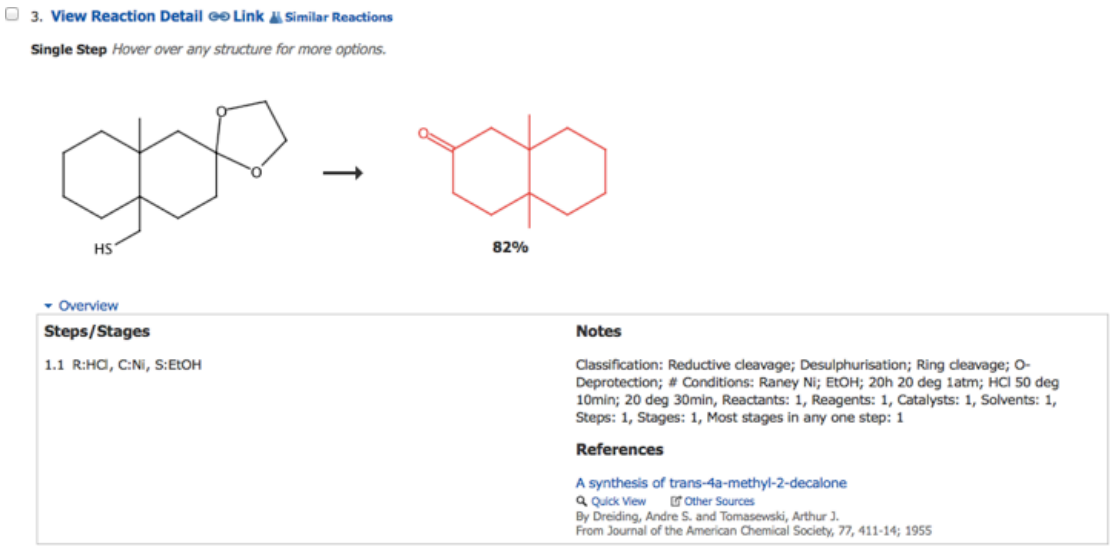


Figure 4.19: SciFinder screenshot of reaction #3 from the RxS1 search

- **Step 3E**: Once he retrieved the article, he stated the publication year and continued to evaluate the schemes to find the modified target compound. He could not find the methylated version of the modified target compound in the schemes.

- **Step 3F**: He returned to the SciFinder reaction results and retrieved the "full text" for reaction #4 (Figure 4.20).

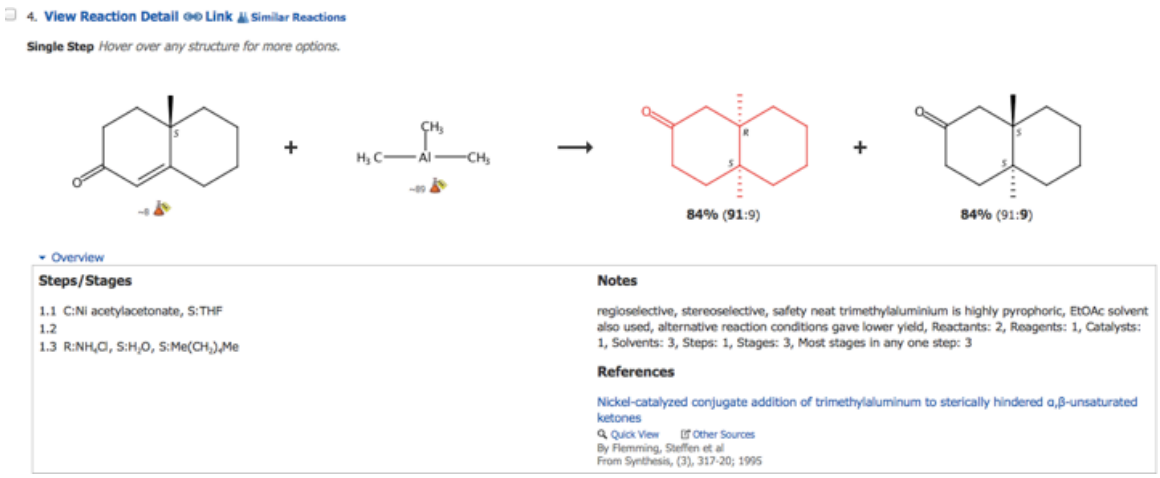


Figure 4.20: SciFinder screenshot of reaction #4 from the RxS1 search

- **Step 3G**: He now moved his attention to the downloaded article and scrolled to find the compound for which he was looking in the schemes. Once he found the compound, he then scrolled to read the experimental section.
- **Step 3H**: He concluded the previously retrieved article was useful, but he returned to the SciFinder results and continued to evaluate the reactions. After further evaluation of the transformations, he concluded that reaction #5 (Figure 4.21) would be a "better" option.

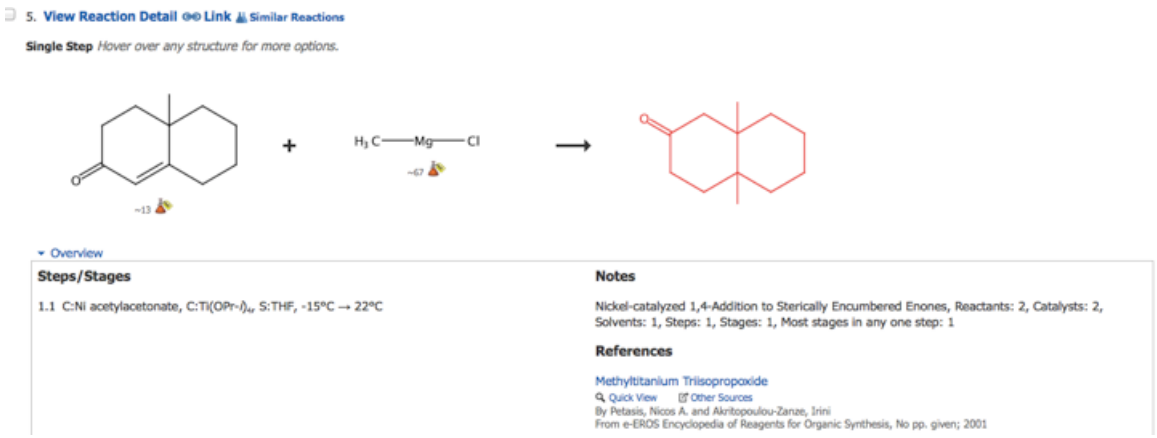


Figure 4.21: SciFinder screenshot of reaction #5 from the RxS1 search

- **Step 4**: Peter then decided to perform an RxS search for the transformation of the simplified starting material (without benzyl group) to the target compound, seen in Figure 4.22 below.

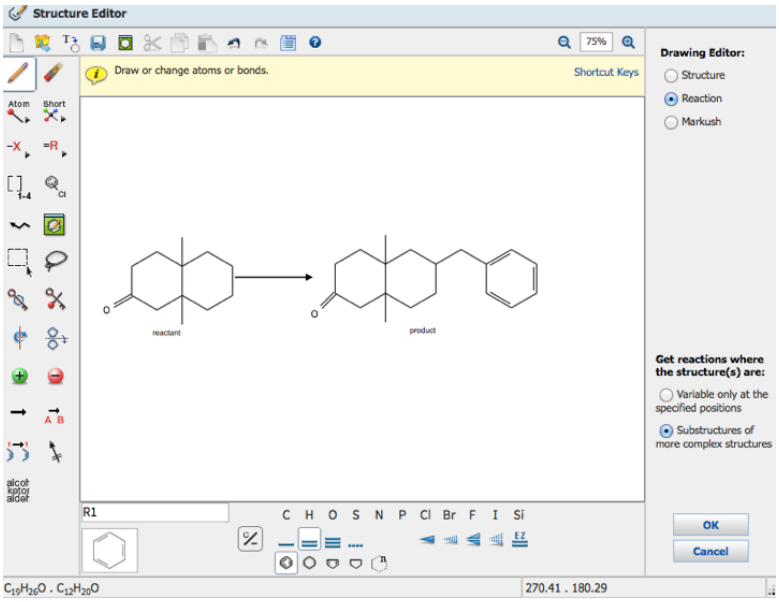


Figure 4.22: SciFinder screenshot of the transformation search using the simplified starting material

- **Step 4A**: The RxS search for the previously searched transformation yielded no results.

- **Step 5**: He then decided to input the following transformation (Figure 4.23) and he performed an RxS search.

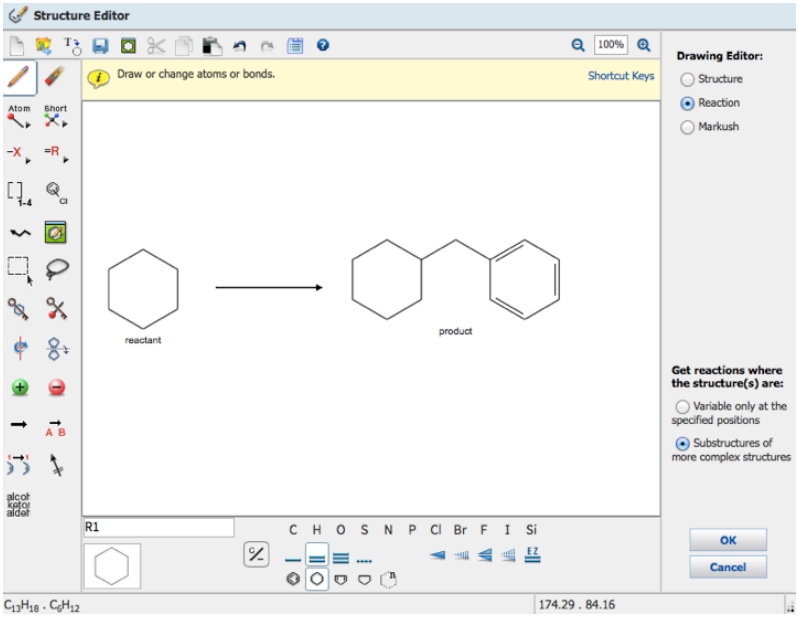


Figure 4.23: SciFinder screenshot of simplified transformation search

- **Step 5A**: The RxS search yielded 290,651 reactions. He proceeded to evaluate the reactions based on the product yield and simplicity of the starting materials.

- **Step 5B**: He then retrieved the "full text" for reaction #2 (Figure 4.24), but was unable to access the article electronically.

Figure 4.24: SciFinder screenshot of reaction #2 from the RxS search

- **Step 5C**: Returning to evaluate the remaining reactions, he then clicked to retrieve the "full text" for reaction #4 (Figure 4.25**)**

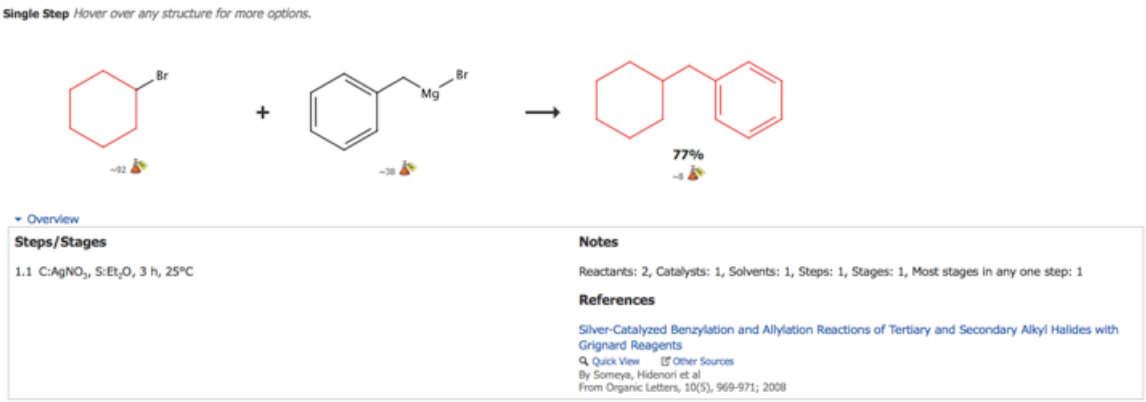


Figure 4.25: SciFinder screenshot of reaction #4 from the RxS search

- **Step 5D:** He then directed his attention to the article and scrolled to find the compound for which he was looking in the schemes.
- **Step 5E**: Once he found the compound, he then scrolled through the article searching for the experimental section. He could not find the experimental

section, so he returned to the SciFinder results and retrieved the Supporting Information (SI) document for reaction #4.

- **Step 5F**: The SI document did not contain the experimental procedures, so he returned to the article. He continued to scroll through the article and he finally found the experimental section.

- **Step 6**: He concluded that he had all the information to synthesize the target compound. Using the previously accessed information, he then sketched his proposed protocol for which he would discuss with his advisor before proceeding into his laboratory to attempt the synthesis (Figure 4.26).

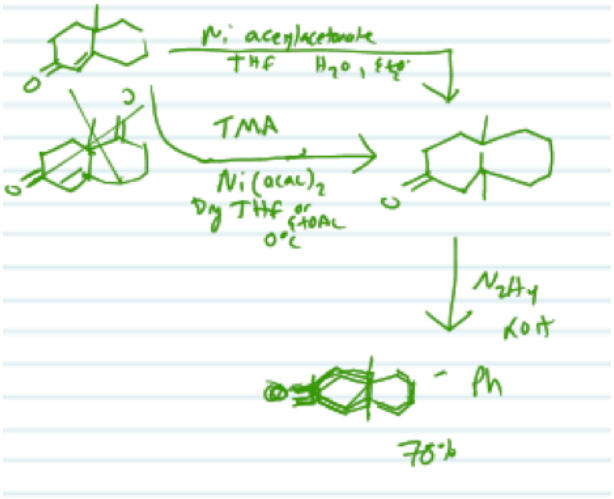


Figure 4.26: Peter's sketch of synthesis protocol

Task 1 Search Critique

Before Peter began the Task 1 search, he shared his initial reaction to and interpretation of the target compound:

Peter: Well you see a ketone for sure. So that always gives you a hint on some of the things that you can do. So you have a carbonyl, there is a lot of carbonyl chemistry out there, so that doesn't really bring it down to anything. But then you also have this phenyl group. So it is not a ridiculously hard problem, it is just, does it matter what these two [methyl groups] too?

I: So does that matter to you [referring to the orientation of the bridged methyl groups]?

Peter: I mean it would yes, depending on obviously what it was used for. It can change big time what that molecule does for you, whether it is R or S, if they are both sticking up or both sticking down, or if one is up and one is down. That is really it though. The ketone is the big thing for me. Obviously the phenyl group is something that you are going to have to put on there. But even if you look half the molecule like that, this is usually something that you can find pretty easily, something that is commercially available.

After discussing his initial interpretation, he proceeded to input the exact structural representation of the target compound and performed a SS search for the structural representation of the target compound (**Step 1**), which he justified by stating:

I would initially go to chemical structure just to initially see if this is actually commercially available. That is like one of the first things you usually want to do because obviously if it is already made or commercially available you just buy it. So that is the first thing I would do.

He was approaching the task from a practical perspective for which he would first determine if the compound was reported. Moreover, he had developed a clear strategy for the type of search he would perform to determine if a compound was commercially available.

So reaction structure that is more of a reaction, so something plus something gives you this. But I just want to see if this is just commercially available. See if this compound, if someone has made it already, like that is the reason why I go to this [substance search].

The SS search for the structural representation of the target compound yielded 58 substances (**Step 1A**) for which he was able to quickly conclude from the output of substances that the target compound was unreported.

> *Peter: Obviously right away, you can tell that it hasn't been made yet. The exact structure hasn't been made.*
>
> *I: How can you tell that?*
>
> *Peter: Yeah just because usually that will show up as one of the first couple of things right here. As you can see, I mean, it has partially some of the structure here [substance #1] then you can see this other carbonyl group [substance #2].*

Although he stated that the substances had *"partially some of the structure,"* he quickly decided to modify his search and performed an RxS1 search for the structural representation of the target compound (**Step 2**). Peter retrieved the same substances (two tricyclic cyclical products) from the previous SS search, and he, once again, decided to not pursue the full text (Figure 4.27).

> *So this is something similar [reaction #1]. Right, this is the first thing that they showed us in the last one [previous substance search] and it kind of stinks because even if I wanted to make something like this, this is only 5% yield. So this is probably not the first thing that I would go to.*

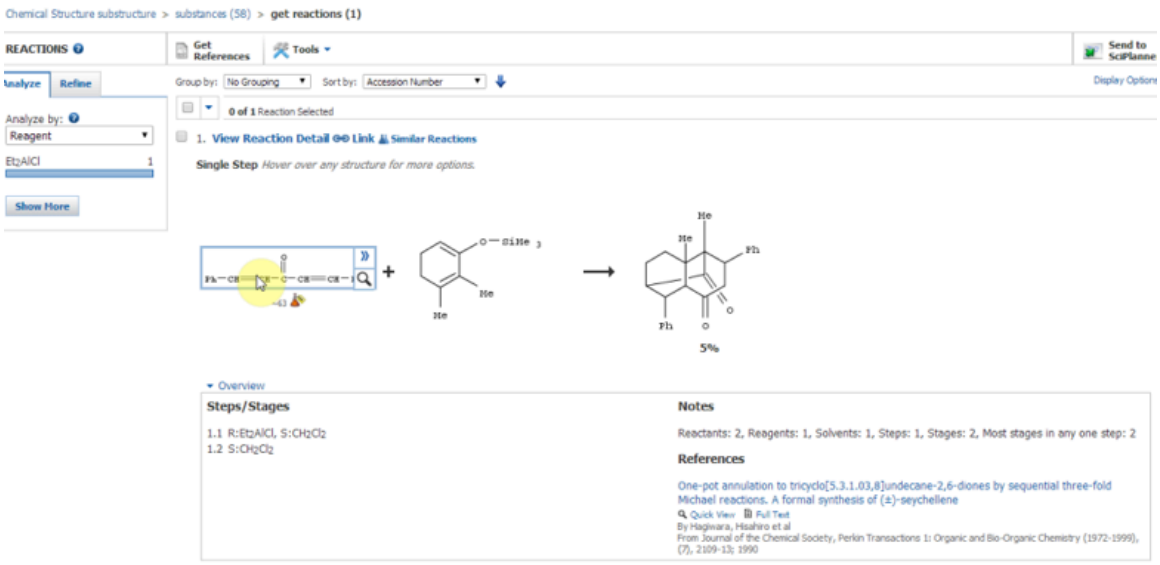


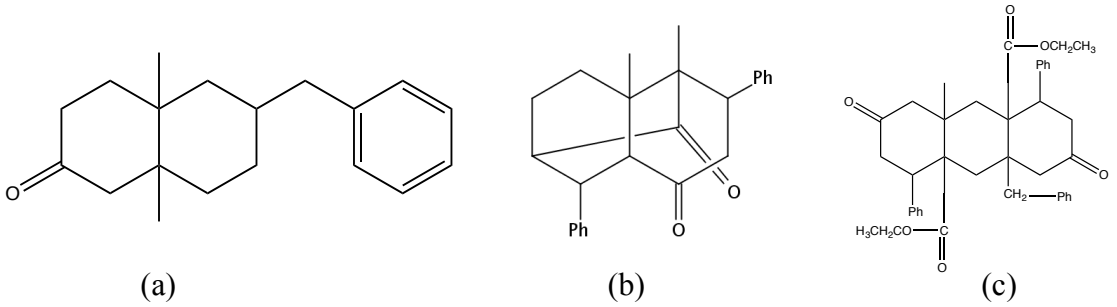
Figure 4.27: SciFinder screenshot of reaction #1 from the RxS1 search

As seen in the previous quote, he immediately concluded the experimental protocols for reaction #1 were not useful because of the very low product yield of 5%. Throughout the Task 1 search, percent yield was one of the major criteria he was considering while evaluating reaction results. Although percent yield is indicative of the efficiency of the experimental procedure, placing such a high emphasis on product yield could cause him to miss information for which he could adopt/modify for the synthesis of the target compound.

After he concluded that the reactions were not useful, I asked him to further explain his rationale, which he then stated:

> *Peter: They weren't useful because it wasn't really, it had some parts of the molecule that I wanted but it didn't have; it had a lot of extra stuff. So instead of making something more complicated I just want to go back and see if I can make something a little less complicated.*
>
> *I: Extra as in?*

Peter: More parts of the molecule. There were actually like another ring added on in there and everything.



(a) (b) (c)

Figure 4.28: Target compound (a), product from reaction #1 (b), and product from reaction #2 (c)

The previous exchange suggests that early in his search, he would disregard methods with complex structural features (Figure 4.28b,c) and reactions reporting low yields, and seek substances with simpler structures, such as fewer rings and substituents. His strategy for seeking a simpler compound was also reflected during **Step 3** as he input a structure without the benzyl group and performed an RxS1 search. During this time he explained:

> *Because if you, just from prior knowledge, you can easily do like an S_N2 with something like this. If you have this and then if there was a bromide, so if I had...if I had something like that [drawing from **Step 3A**], something could easily, this thing right here could easily come in as a nucleophile and just displace this bromide. And then you would actually have what you needed. So I am going to try to dumb it down to something like this.*

Peter was proceeding to put into practice the ideas he discussed regarding his initial interpretation of the target compound. During that time, he stated, "*Obviously the phenyl group is something that you are going to have to put on there."* Moreover, he easily justified his decision to remove the benzyl group because he knew of a reaction for which he could later add the group. As such, he was able to work retrosynthetically because he

knew reactions for which he could add the benzyl group. Also observed during this time was his usage of the arrow-pushing formulism, a tool commonly used by practicing organic chemists. While he was justifying his thoughts about the nucleophilic substitution reaction, he was concurrently using the arrow-pushing formulism to explain both verbally and pictorially the S_N2 reaction to add the benzyl group.

Since he had an idea regarding how to add the benzyl group, he removed the benzyl group and performed an RxS1 search for the structural representation of the simplified target compound (**Step 3**). The search resulted in 2,667 hits reporting the synthesis for the simplified target compound (with varying stereochemistry). As he was viewing the results, he stated:

> *But so to me that would be a big hit already. That is basically most of the molecule already made up. So yeah like I said the first thing I would do is look at the percentage.*

Although Peter displayed a tendency to gravitate towards higher yields, he was also evaluating the reactions based on the complexity and/or simplicity of the starting material. For instance, reaction #1 reported a titanium compound that he described as complex and potentially expensive to synthesize or purchase. Because he could not see the details of the compound (Figure 4.29) from the way it was listed in the reaction results, he clicked "view substance detail" to gain a closer look (**Step 3B**). From that perspective he explained:

> *Peter: It is this metal that, something that we would have to buy probably. Just because I know in the lab we don't have something like this.*
>
> *I: Oh so you would either have to make it or buy it?*

Peter: Exactly. I don't know if I would want to go through that. Just because some of the other ones I looked at have pretty similar yields. One was 85, and one was 84 so it is not a huge cut off percentage. So I probably wouldn't use this one in particular. So I would go back to that same page.

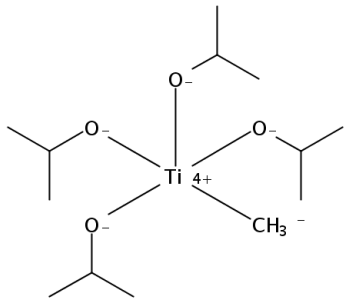


Figure 4.29: Screenshot of the complex starting material from reaction/hit #1

Peter had a significant number of other articles containing other high-yielding pathways to consider, so he decided to not pursue the article and proceeded to evaluate the remaining schemes. Peter skipped the reaction from the article from hit #2 because it reported the same titanium compound and continued to evaluate the reaction from the article from hit #3 for which he concluded the reaction was useful because the reagents (hydrochloric acid and organonickel compound) were simpler than the titanium complex. Thus, he clicked to retrieve the full text.

This one [reaction #3] seems a little easier just because it has hydrochloric acid and maybe nickel. So what I would do is I would go down here where it says full text just click on that.

Once the article was retrieved he immediately stated the year (1955), which he described as somewhat worrisome. Although he was apprehensive because of the publication year, he continued to evaluate the schemes in the article. During this time, he explained:

Obviously you can go through and read [the entire paper] but just for a quick glance, what I would normally do is just go down and find where that molecule is that I am looking for.

In this particular instance, finding the final product in the schemes was an advantageous strategy because he could not find the methylated compound reported in the schemes. However, this did not seem to surprise Peter because the paper was published in 1955.

> *Yeah how they actually did and I am not seeing the experimental. Again this paper is from 1955. I don't even see the exact molecule actually. Because this one has a hydrogen instead of the methyl group. Yeah so I don't even see the exact thing so I will probably go back.*

Because he could not find the methylated compound, he returned to the results and continued to evaluate the other reactions. Peter had over 2,000 transformations to consider, so he returned to the results and clicked to retrieve the full text for hit #4 (Figure 4.30).

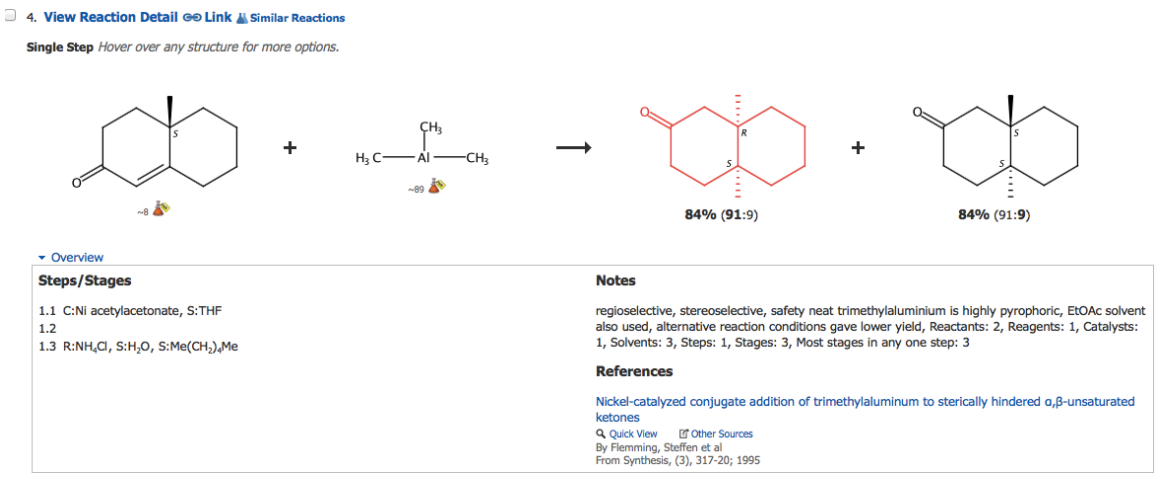


Figure 4.30: SciFinder screenshot of reaction #4 from the RxS search for the simplified target compound

He justified his decision to pursue the full text for hit #4 by stating:

> *Similar yield, 84% and then the starting materials looked pretty similar to the last one [hit #3], pretty simple. And they actually you can tell, this symbol right here just means that they are commercially available [referring to the starting materials]. So, now I know that these two things, I could buy.*

Peter's approach to evaluating reaction results suggests that he had developed heuristics, thereby enabling him to quickly evaluate the information. For instance, first he would find reactions reporting high yields (>80%), and second he would determine the method reporting "simpler" starting materials and reagents. The simplicity of the starting materials and reagents seemed to be based on the structural features, commercially availability, and/or his familiarity with the substance.

He retrieved the full text for hit #4 for which he then found the compound in the schemes and concluded he could use the protocol to make the modified target compound. However, he still proceeded to evaluate the reactions reported in SciFinder and stated that hit #5 might be a better option because the starting materials looked cheaper. Then he stated that he would check the price for each compound because his advisor would want him to pursue the cheaper options.

> *Obviously you would go through all of these and try to get it down, but the farther you go down, probably the less useful these papers are. [He was scrolling through the remaining SciFinder results]. But just because I initially got a good hit I would stick with that [reaction #5] at first until I actually ran the reaction and it didn't work. Then I would go back to square one again.*

Since he had a protocol to make the simplified target compound, *i.e.* without the benzyl group, he directed his attention to determine how to add the benzyl group. Peter decided to perform an RxS search using the simplified target compound (without benzyl group) as the starting material and the target compound (with the benzyl group) as the product, seen in **Step 4**. The RxS search for the transformation yielded no results; so he performed another RxS search as seen in **Step 5**, which he justified by stating:

> *So now I just go back to the reaction structure again and I would [simplify] my reactant side now. I will try to find something like this...I would get rid of*

probably this side right here…and I will just try to find a reaction that...just an alkylation like this. So it is just kind of taking half the molecule really and seeing if I can just do a reaction like this.

Peter was once again applying his initial interpretation of the target compound to persist in developing a synthesis protocol for the target compound. As Peter explained before he began the Task 1 SciFinder search, he viewed and discussed the features of the compound from a symmetrical perspective. Consequently, he decided to reduce the problem to the side of the compound for which he was adding the benzyl group. His perspective allowed him to reduce the complexity of the compound and input a simplified transformation that yielded 290,651 reactions. Displaying the same heuristics as before, once the results were retrieved, Peter evaluated the reactions by their percent yields. Although the first reaction reported commercially available starting materials and a high yield, he decided to ignore the reaction because of the "complicated structure" of the zinc complex reported in step 1. Proceeding to evaluate reaction #2, he decided to retrieve the full text for which he explained:

So this one seems really easy so this is probably something that I would use right here. [referring to reaction #2] Both of these things are commercially available and then these things right here are commercially available as well. And this seems like an easier, easier thing and you still get a 95% yield. So what I would do is just go into this again and do full text of it.

He was unable to retrieve the electronic version of the article, so he returned to the SciFinder results and continued evaluate the remaining reactions. Returning to the reaction results, he quickly scrolled past hit #3, and proceeded to hit #4. He decided to retrieve the full text because the reported yield (77%) was relatively high, and using

commercially available starting materials (**Step 5B**). Once the article was retrieved, he stated:

> *Peter: And sometimes you, I mean this is usually not a bad thing but looking at what journal it is in too.*
>
> *I: So you look at the journal?*
>
> *Peter: Yeah if it is, if it is a journal that I haven't really heard of before I am kind of skeptical of whether the work is 100%. Like what they did, if the percent yields were correct like that, even if the reaction worked.*
>
> *I: So right away you saw that was Organic Letters?*
>
> *Peter: Yeah I knew that is probably going to be something pretty good.*

Peter's statement indicates that he knew some journals employ more rigorous standards for the data they publish. As such, Peter's familiarity with the journals that were widely accepted within this community suggests his growth and epistemic development as a member of this CoP.

After he discussed the importance of the journal, he scrolled to find the schemes reporting the transformation and then quickly skimmed the experimental section. However, Peter could not find the relevant experimental procedures. For this reason, he concluded the experimental procedures were probably reported in the Supporting Information document. Upon retrieving those documents, he realized that it only contained the spectral data and not the experimental procedures. He stated:

> *This is just telling me the IR and NMR and stuff, and it shows the pictures too. So once you make the molecule it is great because you can do to this and compare your stuff to it. I think that was the molecule. Yeah this was the molecule that we are trying to make right here, 2f I guess it is. I just don't see their procedure. It is only giving me the information like once you make it. It is not giving me the exact procedure for it yet. So I'm going to go back to that paper right quick to see if they just give a general scheme.*

Returning to evaluate the article, he found the general reaction conditions and concluded he had all the information he needed to at least attempt the synthesis of the target compound. Therefore, during this part of the interview he sketched the synthesis steps and explained:

> *Peter: And I would just go ask my [advisor] how feasible it is to do it or if I should look at something else.*
>
> *I: Would this be something that you would print out the papers and then go show your advisor?*
>
> *Peter: Exactly.*
>
> *I: So you would get his approval before you actually carry this out?*
>
> *Peter: Absolutely! Yeah. So this is something that I would normally do, even with my [advisor]. I would just kind of get the general scheme of what they were doing as well.*

Because he did not note the previous reaction conditions, he had to perform the previous searches again to retrieve the procedures. During this time he sketched a reaction to add the benzyl group; however, the reaction was not the transformation from the simplified search. As such, his actions indicate the importance of taking notes and saving the search results as to avoid retrieving the wrong information.

Once he had sketched the synthesis, I asked him how confident he was that the synthesis protocol would yield the target compound, for which he then explained:

> *Because a lot of the stuff is commercially available I think it is definitely pretty feasible. And then also there are a lot of different hits like when I did the search, so there were a lot of options that I could choose from. So if one didn't work you know you could try another one. If there was only like one search that came up then I would kind of be a little worried. Cause if it didn't work then obviously I would have to try a whole new scheme, a whole new synthesis to do it. And then*

> *also how easy it kind of seemed and the percent yield they were getting, it definitely seemed like it would work.*

Based on his experiences as a synthetic organic chemist, Peter knew that reactions as written on paper often do not work in the intended manner when carried out in the lab. Consequently, his process for solving this task was not to just find one synthetic route but to have multiple alternatives for each step should there be any problems. He justified this strategy by stating:

> *So I could do something like this as well where I go through a lot of these things and give my boss a lot of options in order to make this first thing. And then do the same thing, a lot of options to make this second thing. But again I would try the first, the first one that we did just because how simple it seemed. So and after I give him like a bunch of options too he can probably by prior knowledge, or if we have the chemicals in the lab already, he would say which route to try first. And then if that didn't work then you could try a different route or you could just try to do something different from what you did.*

Although he had developed a synthesis pathway, he stated that he would still consult his advisor regarding the feasibility of his proposed protocol. Ultimately, he sketched multiple pathways so that his advisor could decide on the best route and provide him with suggestions as to ways to modify the experimental conditions if he failed to make the target compound.

Summary of Peter's search

Peter's prior experiences using SciFinder allowed him to know how to strategically and methodically navigate the database to find research protocols he could potentially use to synthesize the target compound. Although Peter relied on the database to develop a research protocol, his Task 1 search was very purposeful and he could articulate his strategy for inputting, evaluating, and deciding to use the information.

Specifically, his strategy was guided by his underlying assumption of the symmetrical approach, which strongly biased his subsequent perception of the target molecule. Utilizing the database and his symmetrical perspective, he was able to propose a research protocol that was composed of multiple reactions for each step, which he would then seek his advisor's (more knowledgeable member) approval.

SOPHIE'S CASE

Background

Sophie was a third-year Ph.D. student who obtained her undergraduate chemistry degree from a small institution in the southeastern United States. During that time, she had two undergraduate research experiences, one at her undergraduate institution where she obtained her degree and the other was a summer undergraduate research experience at a large, Ph.D. granting institution. At her undergraduate institution, she investigated reusable inorganic solvents and also synthesized a couple of small molecules. Although Sophie recalled that her undergraduate research advisor used SciFinder to retrieve journal articles related to their research, Sophie did not have any direct experiences with the database at her home institution. However, it is important to note that her undergraduate research advisor regularly met with Sophie to discuss the retrieved articles. These meetings, which occurred after Sophie had read the article, included discussions of the chemistry represented in the synthetic schemes as well as reviews of the experimental techniques and the interpretation of the authors' spectral data. Sophie felt these meetings helped solidify her understanding of organic chemistry.

During the summer research experience at the Ph.D.-granting institution, she worked on a total synthesis of a small target compound. When she first joined the group, her advisor gave her some articles reporting a fifteen-step synthesis protocol for the compound. However, during that time, she was unsure about the chemistry reported in the method, so her advisor told her, "*if you don't know how it works then find precedence for it*." Before this research experience, she had never used SciFinder, so she was unfamiliar with the structure search tools. For this reason, she performed a topic search using the name of the most prominent structural feature of the target compound as the key words. She chose that structural feature because her internship advisor kept on referring to that aspect of the target compound multiple times during each conversation, "*[name of structural feature] was what he just kept saying over and over again*." Sophie recalled that the key word search resulted in a substantial amount of results, so she refined the search by inputting the name of the author who wrote the articles her advisor had given her when she joined the group. She justified her search strategy because she believed that most researchers would focus on a certain research area. Eventually, by refining her search using the author and the target's prominent structural feature, the results narrowed from around 16,000 to 600. From that point, she would review articles based on the titles, which she said, "*looked similar to the project we were working on*." She then explained that as she was evaluating each article, she stumbled upon an article reporting the synthesis for the carbon backbone structure. Also, during this time she discovered that SciFinder had an option for which she could perform structure searches. She then began to perform searches for the structural representation of the feature whose

name she had previously used for a key-word search. She was comfortable using the structure search option in SciFinder because she had prior experiences using ChemDraw at her undergraduate institution. Because she was ultimately searching SciFinder for a more practical protocol, she was able to find a more recent protocol that reported a three-step protocol for the starting material instead of the 15-step protocol her advisor provided.

Ph.D. Research

For her Ph.D. research, Sophie was working on two different projects at the time of the data collection. The first project was in collaboration with a research group from another department at the university. Her role in the first project was to analyze the substances that were synthesized by the collaborators. In time, Sophie began to synthesize some of these substances as well.

Sophie's second project was to develop a more efficient synthetic route to a compound that had already been synthesized in the lab, albeit at a less-than-hoped- for overall yield. Sophie's modified method resulted in an 80% product yield, but they wanted to increase the yield so they could publish their results in a more prestigious journal within their CoP. Therefore, she was performing SciFinder structure searches in an attempt to figure out how to increase the product yield. She justified her strategy for performing a structure search in the following quote:

> *I found that is the best way to start is just to do a structure search, if you have, if it is a small molecule. If it is a big molecule you usually have to do more of a subject line first.*

To begin her SciFinder search, she would input the structural representation of the starting materials. However, she quickly realized that searching for the starting materials would not provide useful information because the results were unrelated to the type of reaction she was performing, specifically the final, target compound.

> *It just would give me…it wasn't focusing on the [functional group] portion of it. It was just focusing on the carboxylic acid and what they used in other methods. I needed to specifically look for the [functional group]* transformation.

Based on her summer research experiences, Sophie then began a search using the structural representation of the target compound. Again, using her previous knowledge gained during the summer research internship, Sophie narrowed her initial set of hits by inputting another functional group of the target molecule. Doing so, led her to some articles reporting her desired class of compounds for which she would then:

> *Read through [the experimental section] and see how they clean this up. What kind of characterizations they did it on. That kind of stuff. But I usually don't focus on so much on the reagents that they used, just that final product to see. Because I can make the product I just have to figure out where it is going and that kind of stuff.*

During the time of the interview, she was still performing SciFinder searches in an attempt to determine how to increase the product yield.

Task 1 Search Chronology

The following is a chronology of her search.

- **Step 1**: Sophie input the exact structural representation of the target compound and performed an SSi search.

- **Step 1A**: She decided to retrieve the 14 substances that were reported with an 85-89% similarity match to the target compound. She quickly viewed the substances (Figure 4.31) and decided to modify her search.

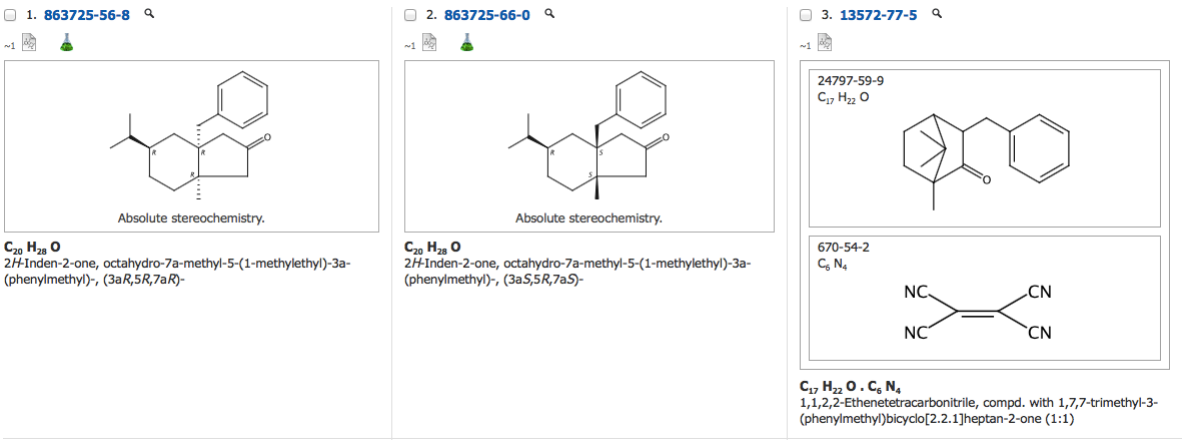


Figure 4.31: SciFinder screenshot of SSi search results

- **Step 2**: She decided to disregard the previous results and returned to the structure editor screen. She then modified the target compound by removing the benzyl group and inserting a bromine group (Figure 4.32). She then performed an SSi search.

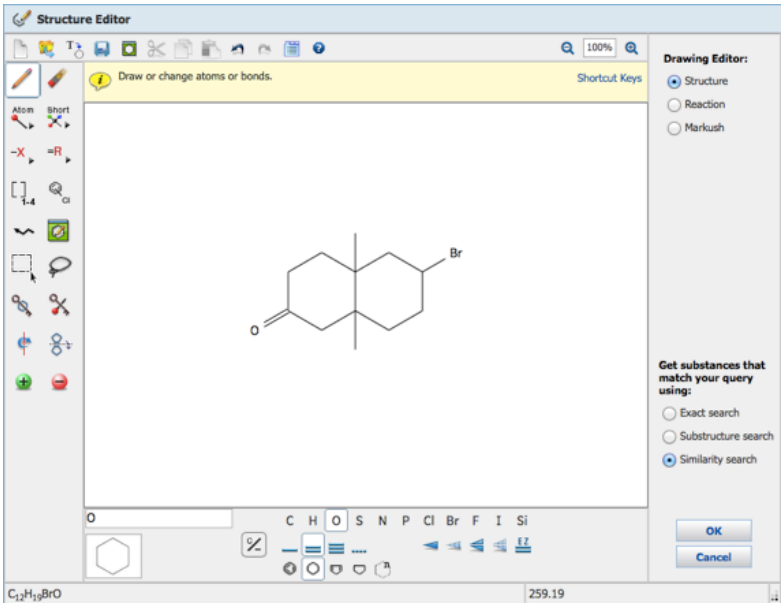


Figure 4.32: SciFinder screenshot of modified target compound with bromine group

- **Step 2A**: She retrieved the one substance with a 95-98% similarity match to the modified target compound.
- **Step 2B**: She quickly evaluated the substance (Figure 4.33) and decided to modify her search.

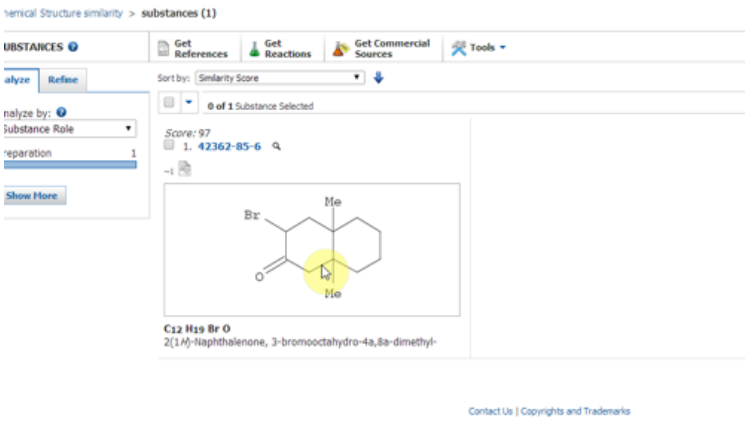


Figure 4.33: SciFinder screenshot of the SSi search result for the modified target compound with a bromine group

- **Step 3**: Returning to the structure editor screen, she modified the structure by removing the bromine group and performed an SSi search (Figure 4.34).

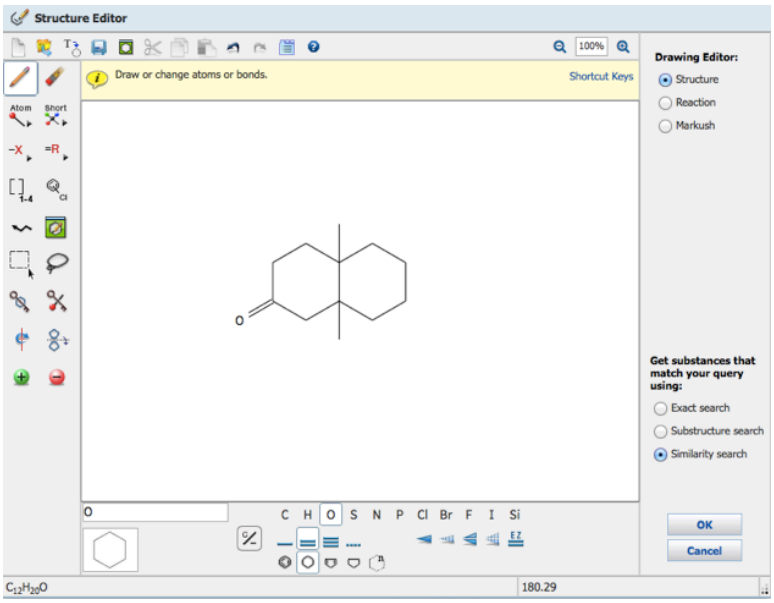


Figure 4.34: SciFinder screenshot of the modified target compound without the benzyl group

  - **Step 3A**: She then retrieved the 8 substances with a greater than 99% similarity match to the compound in Figure 4.34**.**
  - **Step 3B**: After viewing each substance, she then clicked on substance #3 and modified her search to retrieve the reactions for substance #3 as a product (Figure 4.35).

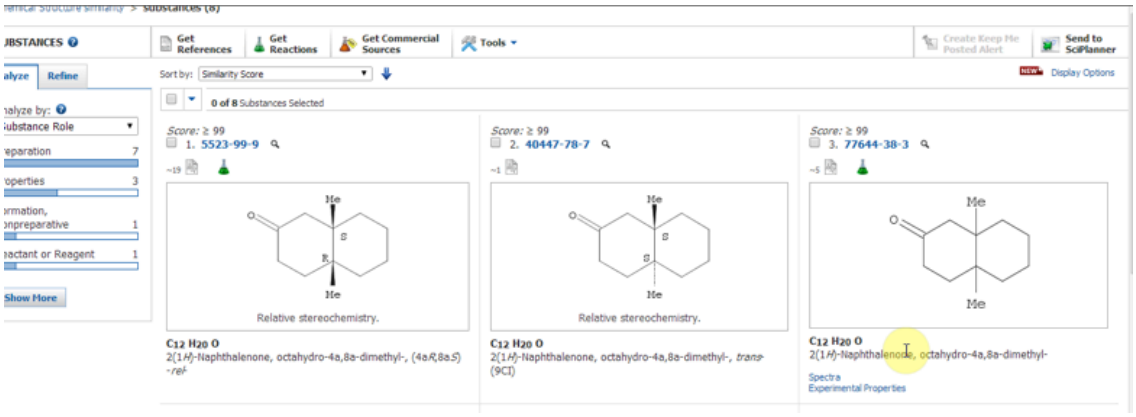


Figure 4.35: SciFinder screenshot of substances from the SSi search for the modified target compound without the benzyl group

- **Step 3C**: The modified search yielded five reactions for which she proceeded to evaluate the transformations by the percent yield, complexity of the starting materials, experimental conditions, and number of steps.
- **Step 3D**: After evaluating the reactions, she clicked to retrieve the "full text" for reaction #4 (Figure 4.36).

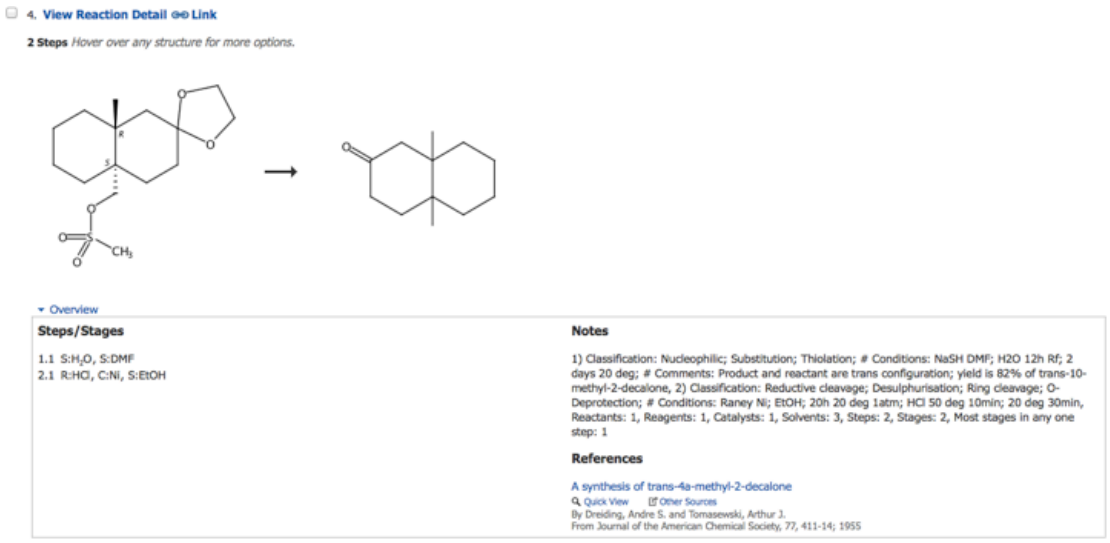


Figure 4.36: SciFinder screenshot of reaction #4 from the search for substance #3 as a product

- **Step 3E**: She now directed her attention to the article and scrolled to find the compound in the schemes. She could not find the methylated version of the modified target compound in the schemes.
- **Step 3F**: She returned to the SciFinder results and clicked to retrieve the "full text" for reaction #5 (Figure 4.37).

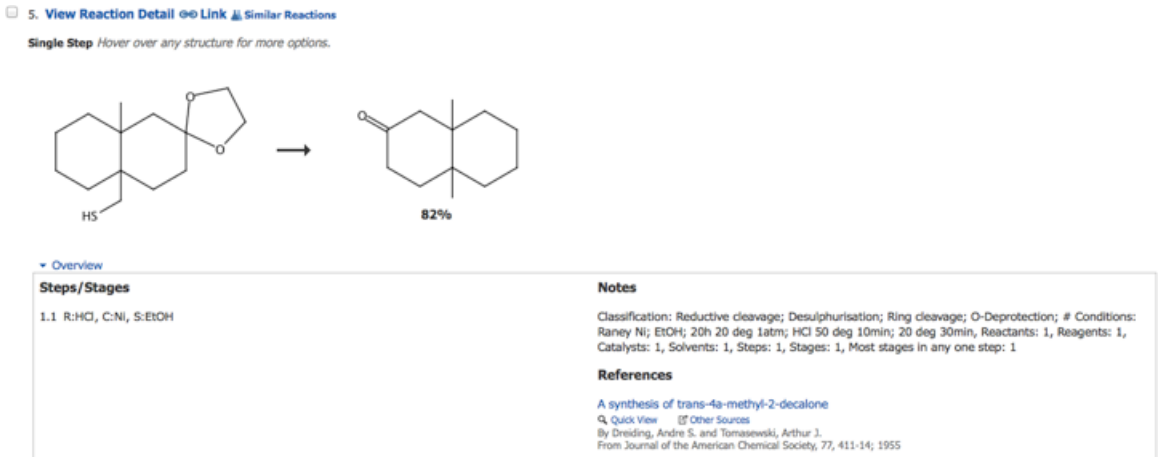


Figure 4.37: SciFinder screenshot of reaction #5 from the search for substance #3 as a product

- **Step 3G**: Once the article was retrieved, she realized it was the same article that reported reaction #4 from **Step 3D**.
- **Step 3H**: She then attempted to access the "full texts" for reactions #3 (Figure 4.38) and was linked to an abstract reporting the experimental conditions.

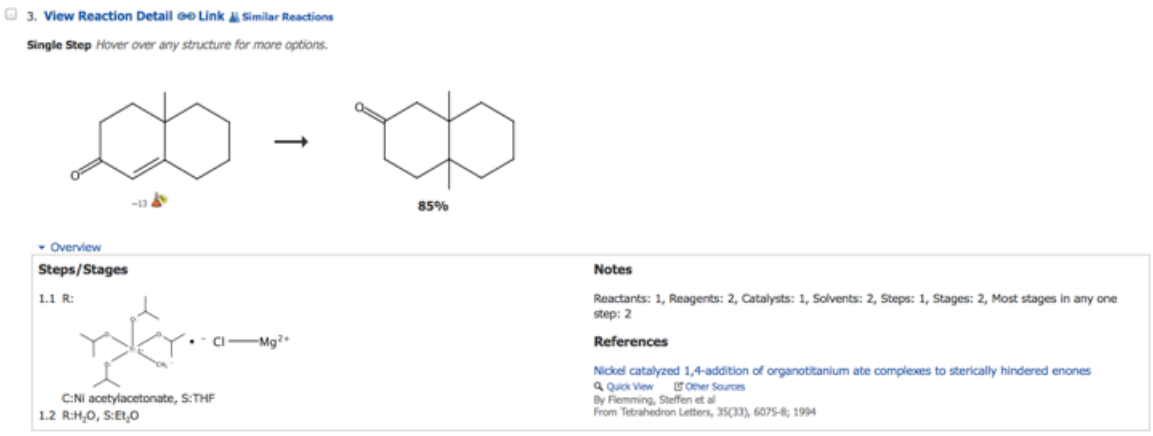


Figure 4.38: SciFinder screenshot of reaction #3 from the search for substance #3 as a product

- **Step 4**: She left the tab open with the abstract and then proceeded to input the following transformation (Figure 4.39) and performed an RxV search

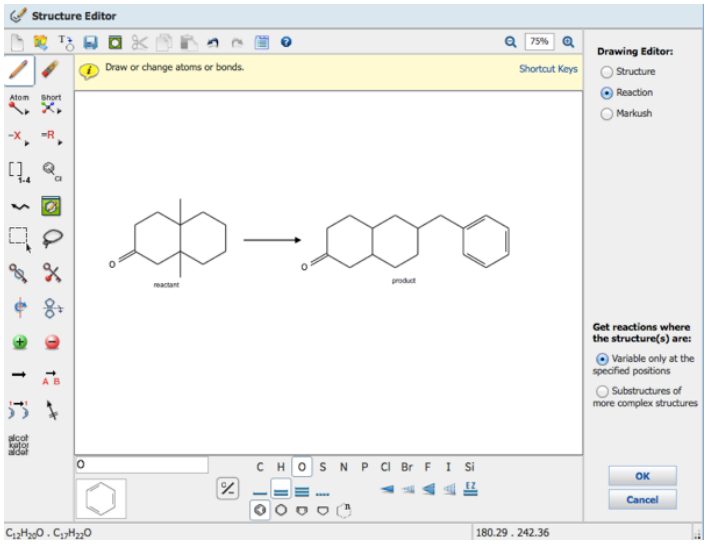


Figure 4.39: SciFinder screenshot of the transformation using the new starting material without the benzyl group

- **Step 4A** The RxV search for the transformation yielded no results.

- **Step 5**: Returning the structure editor screen, she performed an RxS search for the previous transformation (seen in Figure 4.39).
    - **Step 5A**: The RxS search yielded no results.
- **Step 6**: She decided to sketch her ideas regarding how to add the benzyl group (Figure 4.40).

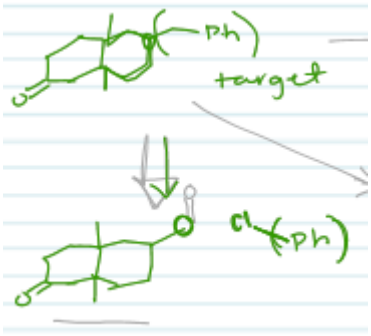


Figure 4.40: Sketch of ideas regarding how to add the benzyl group

- **Step 7**: She decided to change her strategy and performed a research topic search for the key words: "Lewis acid mediated phenyl substitution."
    - **Step 7A**: The topic search retrieved six articles containing all the key words.
    - **Step 7B**: She decided to retrieve article #4 with the title, "Lewis Acid Mediated Reactions of N-Arylsulfonimidoyl Chlorides with Alkenes, Some Steric Effects of Alkene Substitutions" (Harmata & Kahraman, 1998).
    - **Step 7C**: Turning her attention to the article, she then scrolled to evaluate the reactions schemes. She decided to keep the article open, but returned to perform another structure search.

- **Step 8**: Returning to the structure editor screen, she input the following transformation and performed an RxV search (Figure 4.41).

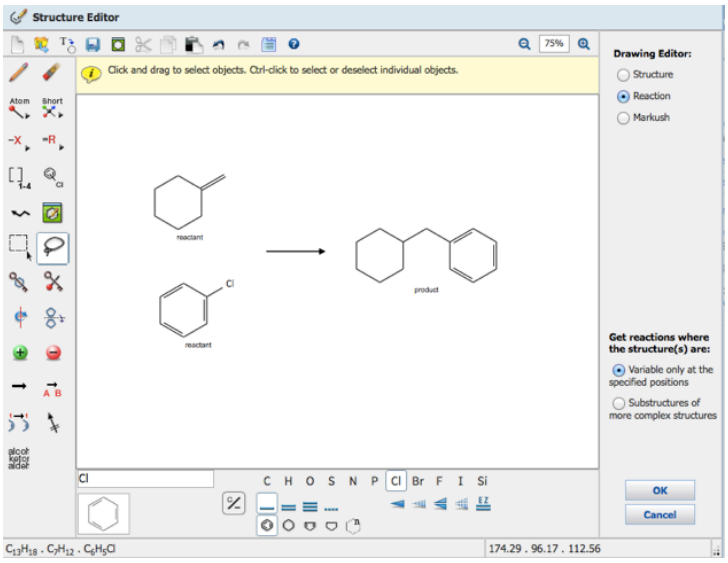


Figure 4.41: SciFinder screenshot of the transformation search from Step 8

  - **Step 8A**: The RxV search for the previous transformation did not yield any results.
- **Step 9**: She then proceeded to adjust her strategy and performed an RxS search for the previous transformation.
  - **Step 9A**: The RxS search yielded 3,230 transformation; however, she quickly viewed the top four reactions and concluded the results were not useful.
  - **Step 9B**: Before she completely disregarded the results, she searched for a reagent aluminum chloride by clicking "analyze" followed by "show more."
- **Step 10**: She concluded the previous results were not useful, so she performed another SSi search for the exact structural representation of the target compound.

- **Step 10A**: This time, she retrieved the 14 substances with an 85-89% and the 86 substances with 80-84% similarity match to the target compound. She once again concluded the results were not useful for the synthesis of the target compound.

- **Step 11**: During this part of the interview, she had another idea for which she decided to sketch on paper (Figure 4.42).

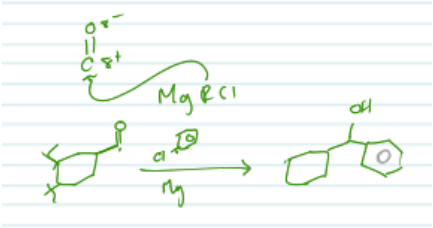


Figure 4.42: Sketch for the addition of the benzyl group

- **Step 12**: After sketching her idea, she then input the following compound (Figure 4.43) and performed an SSi search.

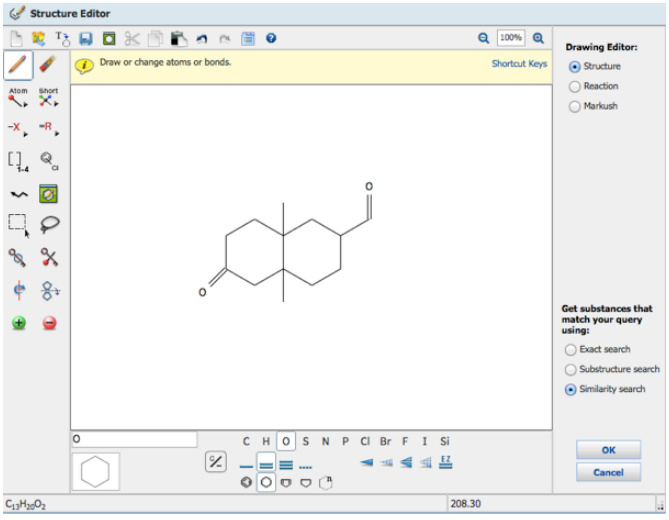


Figure 4.43: SciFinder screenshot of the compound sketched in Step 11

- **Step 12A**: She clicked to retrieve the two substances with a 95-98% similarity match to the compound seen in Figure 4.44.

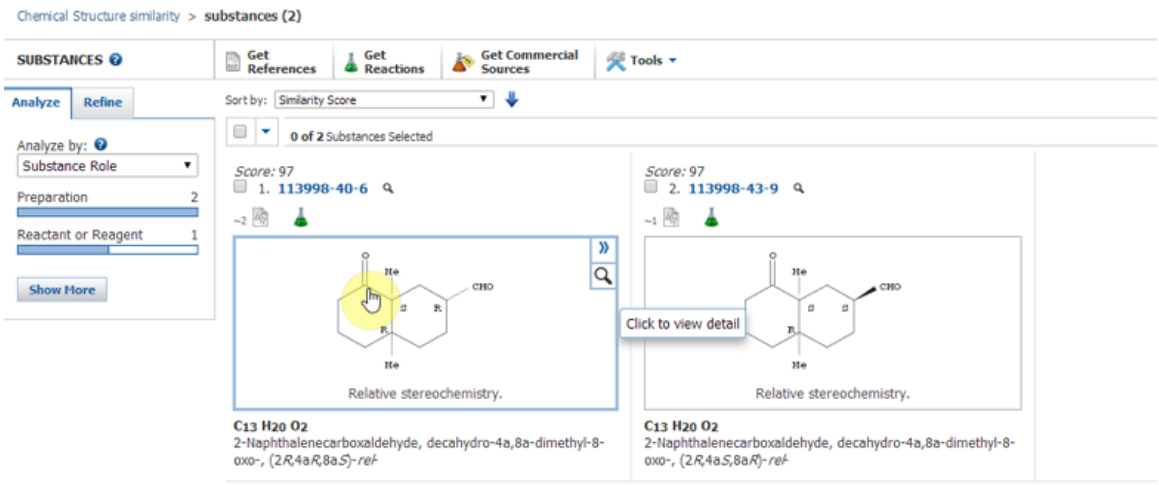


Figure 4.44: SciFinder screenshot of the substances that resulted from the SSi search in Step 12

- **Step 13**: She decided to disregard the previous substances and once again sketched her ideas and proposed performing an alpha, beta unsaturated ketone ring closure (Figure 4.45).

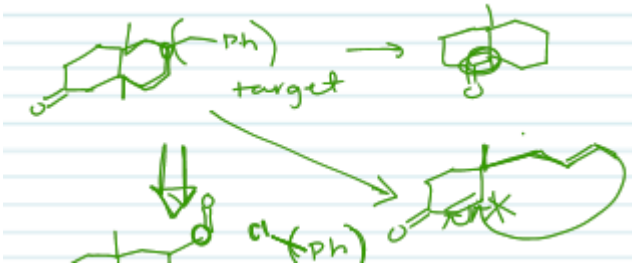


Figure 4.45: Ideas regarding performing an alpha, beta unsaturated ketone ring closure step

- **Step 13A**: After she sketched her ideas, she input the following compound and performed an SSi search (Figure 4.46)

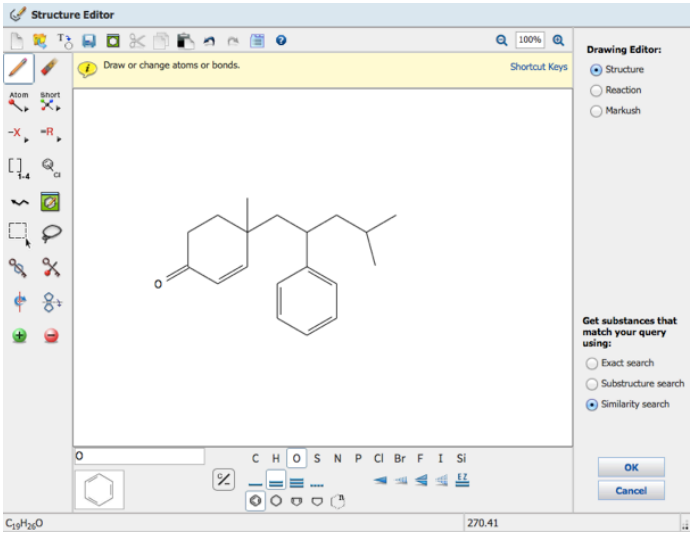


Figure 4.46: SciFinder screenshot of the compound sketched in Step 13 for the alpha, beta unsaturated ring closure step

- **Step 13B**: She retrieved the one substance with a 90-94% similarity match.
- **Step 13C**: Once the substance was retrieved, she realized the substance was reported in the previously retrieved article from her topic search (**Step 6**). Since the article was still open in the tab, she then redirected her attention to re-evaluate the paper.
- **Step 14**: She was unable to use the information in the article for the synthesis of the target compound, so she performed one last RxS search for the following transformation (Figure 4.47).

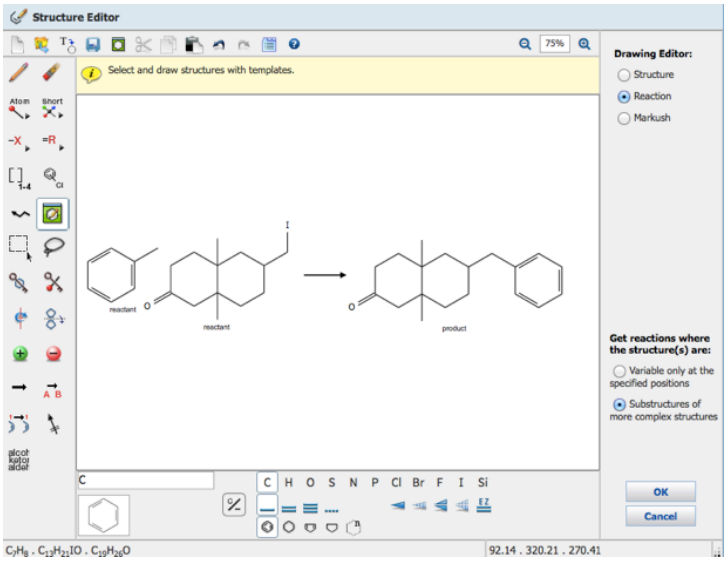


Figure 4.47: SciFinder screenshot of the transformation input for Step 14

- **Step 14A**: The reaction search yielded no results. At this point in the interview, she could not find precedence for any of her ideas regarding the addition of the benzyl group, and she would have to seek help from her advisor.

Task 1 Search Critique

Before Sophie began to perform the Task 1 search, she shared her initial reaction to and interpretation of the target compound:

> *The first thing, when I see, when I am given a molecule the first thing I do is I look at the structure itself to see what kind of functional groups I am working with, what kind of reactivity that can mean for me. And then I just take the molecule and I go straight into the chemical structure.*

Sophie began by inputting the exact structural representation of the target compound into the structure editor screen and performed an SSi search (**Step 1**). She justified her strategy for performing an SSi search in the following exchange:

Sophie: If you get too specific sometimes it can narrow down the search so far to where you don't get any hits. So you have to be very generic at first and then get more in-depth.

I: And how did you learn that?

Sophie: Mostly trial and error. Actually, even some embarrassing moments in group meetings where [my advisor] would be like, 'no I have synthesized that before.' It would be situations where I would be like, 'I couldn't find the experimental literature values for this compound by searching through SciFinder,' and he would be like, 'that is not true because I can find it in a specific paper.'

The previous exchange is another example of how interactions with her CoP, specifically with her advisor, significantly influenced her ISB. As such, Sophie's change in her ISB is likely due to reinforcement through multiple interactions like this one.

She was able to very quickly conclude that the target compound was unreported by simply reviewing the percentage values. She explained:

Sophie: Based on chemical working, I have been in the lab for 3 years now; I have been searching on SciFinder for 3 years. If it is not within a 90% similarity, it means that it is probably a multiple step synthesis. But that again just comes from me dealing with synthesis on a daily bases. I didn't know that to begin with.

I: How did you learn it?

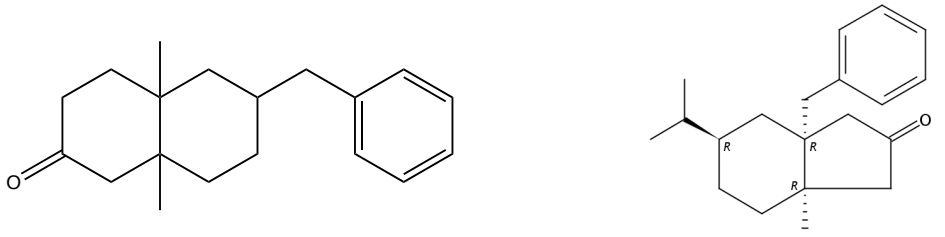
Sophie: Trial and error. Mostly from just trial and error. Or some of the molecules that I can find that are really fast, I know that they are usually commercially available because they are easy to make. They are out there a lot, so they have multiple similarities in the molecule. Sometimes that can be misleading though. Just because the similarity is the same doesn't mean it is easy to make similar molecules.

Sophie's experiences performing SSi searches had thus enabled her to develop a heuristic for judging the complexity for synthesizing a compound and sometimes even predicting the commercial availability of that substance. Although she could deduce some information from the percentage values, she was still skeptical about the usefulness of the

results. From prior experiences, she knew that, "*Just because the similarity is the same doesn't mean it is easy to make similar molecules."*

She retrieved the 14 substances (the highest hits) that were reported with an 85-89% similarity match to the target compound (**Step 1A**). She then proceeded to evaluate the substances based on the bicyclic backbone structure. She explained her thoughts about the substance in the following exchange:

> *Sophie: Because we are going for a 6-membered ring here [*Figure 4.48*b]. We have two cyclohexane rings just fused together and that is given me a pentane and a cyclohexane…So there is a 5-membered ring within the 6-membered one with the bridge over the one and four position. So again that tells me that is not what I am looking for.*
>
> *I: So what are focusing on right now?*
>
> *Sophie: I look at just basic structure. I don't look at functional groups at this point I am just straight up looking to see if the backbone is similar. The meat of the molecule, that is what I look at and for at least these 14 "hits" that they gave us, I am not seeing anything similar at all actually.*



(a) (b)

Figure 4.48: Target compound (a) and substance #1 from the SSi search (b)

Before she began the SciFinder search Sophie focused on the functional groups; however, during this stage of her search she was primarily concerned about the bicyclic backbone structure. Since the SSi search did not result in the specific backbone structural

representation of the fused cyclohexane ring, she then decided to change her search strategy. She explained that she was going to "cut out the fat" on the compound (i.e. benzyl group). She justified her strategy in the following exchange:

> *Sophie: This ketone and this phenyl, so that to me is what is excess to the molecule. To me the center of the molecule is the bicyclic ring system. So what I do usually is just cut out the fat. Kind of how I would say it. I just cut out the stuff that makes it pretty and what we want.*
>
> *I: And how do you know to cut out the fat?*
>
> *Sophie: The only reason I do that is because I know if I cannot find the exact structure that has the functional groups that we are looking for, what I usually end up doing is trimming away pieces that I know I can add. Usually you can, benzene rings are a little more reactive, they are extremely stable. We know that from general organic chemistry, but there is an easier way to transform them into something else than this original non-unsaturated bicyclic system. But again most of that just comes from general organic, what I have seen from literature and then just being in the lab and working with molecules. I usually understand which parts of the molecules I can add or subtract.*

On the face of it, Sophie's strategy seems quite reasonable. However, it is unlikely to be practicable. Although saturated hydrocarbons can react under certain conditions, those reactions are hard to control, especially with respect to regioselectivity, *i.e.* the specific area(s) where it is desired for the transformation to take place. Similarly, reactions of aromatic compounds often result in undesired products with respect to the number and location of the moieties added on. As such, this seemingly rational idea did not lead to a fruitful path because it is not possible to analyze one aspect of a molecule while disregarding, even momentarily, all others.

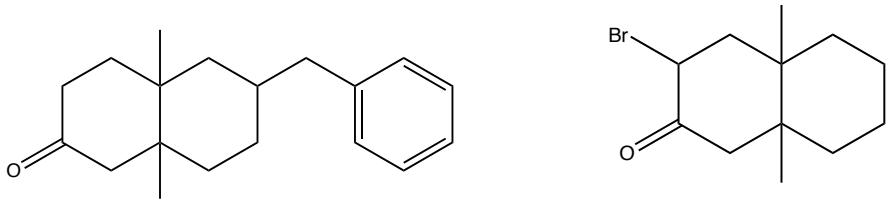
Proceeding with her strategy, Sophie removed the benzyl group and inserted a halogen (bromine) group and performed an SSi search (**Step 2**), which she explained:

> *Cause we know just from chemical working, I know I have to add something to it, so something is going to have to leave. I am going to put any halogen on that molecule.*

In contrast to her previous steps, Sophie planned this step keeping the ultimate assembly of the target compound in mind.

She retrieved the one substance with a 95-98% similarity match to the modified target compound (**Step 2A**); however, she didn't pursue the full text because the bromine group was not located in the desired position (Figure 4.49c). She explained:

> *It is not, the bromine is not in the position I want, but it does have the dimethyl at the bridged carbon so that is good. So that is something that I don't want to mess with. It is not easy. It is not easy and I don't want to mess with it. The only thing that disturbs me about this molecule [substance #3] is the bromine isn't where I want it to be to add in the phenyl group. So that being said, I go back to the drawing board.*



(a) (c)

Figure 4.49: Target compound (a) and substance #1 from the SSi search (c)

Because the bromine was in the wrong position in relation to the target compound, she then decided to modify her target compound by removing the bromine group. She then performed an SSi search for the backbone structural representation of the target compound without the benzyl group (**Step 3**) because she felt the previous searches were too specific.

> *Just because I know from being in chemistry for a couple years, that has been made before. Like I know that this has and I am like, ok I know that this bicyclic system with the ketone and these dimethyl groups has got to be out there somewhere and it is.*

She had been immersed in an organic research group for three years and from this experience she had gained knowledge regarding types of compounds that had been reported, thereby giving her an "intuition" about the current target's likely existence. Furthermore, in a way her method for broadening her search was an effective strategy because she was able to retrieve the backbone structure for which she searched. However, its utility may be questionable given the monumental task it would have taken to modify that backbone to the target.

She retrieved the 8 substances with a greater than 99% similarity match to the compound in **Step 3**. After quickly viewing each substance, she then clicked on substance #3 and modified her search to retrieve the five reactions for substance #3 as a product (**Step 3B**). Once the reactions were retrieved, she first evaluated the transformations to find the reaction reporting the highest product yield. After considering the percent yield, she then evaluated the reactions based on the complexity of the starting materials/reagents. For instance, reaction #3 reported a titanium compound that appeared complex, which she associated with additional cost and an unfavorable transformation.

> *So there is this titanium center [*Figure 4.50*], which has a plus four charge on this whole molecule, which is offset by this. So I just don't even want to deal with that. I don't know if it is pure when I get it. You have to test the purity of it because reactive reagents like that can decompose very quickly. There is a lot that goes into that and I just don't want to mess with it.*

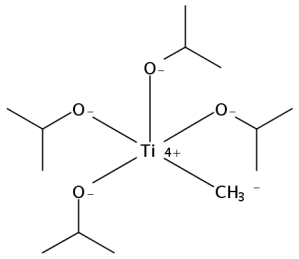


Figure 4.50: Complex reagent in reaction #3

At this stage in her method develop she had several other reactions to consider, so she simply continued to evaluate the remaining reactions. Proceeding to evaluate reaction #4, she state:

> *I am looking at the yield. So the yield is 82%. I can read that in here, the notes from the reaction. It actually tells me what type of reaction this is considered. All the list of reagents that I will need, two days at 20 degrees, 20 degrees is usually room temperature. The only reason I know that is from dealing with just protocol, after protocol, after protocol of synthesis. Usually 20-25 degrees, if you see that, it is common for room temperature. The reagents looks good, they look cheap and easy to use, so I don't have to worry about explosions. Like I don't have to worry about anything flaming up on me…Ok so let's get the "full text".*

Once the paper was retrieved, she immediately stated the publication year (1955); however, this was not a concern. She explained:

> *Sometimes there can be easier or better ways to do it. What I have found a lot of times is, if it is not broke don't fix it. And chemists try to. I mean there can be extremely beautiful catalyst that will do this and do it great, but I just need it as a starting material for another reaction, I don't need something that's beautiful and all that stuff. I just need the molecule in high yield where I can use it for something later.*

Sophie's analysis in this case shows a fair amount of maturation in that knowing the *purpose* of one's doing at any given stage of the research process. Since this reaction is

not a centerpiece of a research problem and is nothing more than a means to an end, Sophie realized that there was not a compelling need to find a more contemporary method, as long as she could easily and safely perform the transformation in high yield.

However, her opinion of the article quickly changed following her evaluation of the schemes.

> *Sophie: I am looking at the schemes. Just the pictures. I don't want to read it yet. I just want to make sure it has what I am looking for first.*
>
> *I: How did you learn to focus on the paper like this?*
>
> *Sophie: Time. Let's see. At first I would try to read them. I would try to go through them and learn, and learn, and learn. But I just don't have time to read all of them. Especially if I don't even know if I can make it yet. If I don't even know if this paper has the molecule in it that I am looking for. SciFinder said it did sure, but like for me, right now I am not able to track down the dimethylated version.*

Prolonged exposure to her Ph.D. research had influenced her approach to evaluating journal articles. Her approach to first skimming through the schemes to find the modified target compound was an advantageous strategy because she could not find the dimethylated version of the modified target compound. She then concluded:

> *This is a bad paper. It is not the one I want. So we will go to another [article]. And that happens a lot actually. You will find something and you will be like, I really like those reagents but then you get to the paper and you are like what?*

Because she could not find the compound in the article, she returned to the reactions and continued to evaluate the other transformations (**Step 3F**). Proceeding through the reaction results, she quickly stated that reaction #5 was potentially better than the previous article because it was reporting only one step with an 82% product yield.

However, once she retrieved the full text, she realized the transformation was reported in the previously retrieved article (**Step 3G**).

> *That's sad because they got that in 82% yield, which is pretty good. But it is the same paper.*

Her actions suggest that she primarily focuses on the product yield and the experimental conditions, and does not evaluate the information regarding the journal publication. The information regarding the journal should be considered because certain journals within the organic CoP employ more rigorous standards for the quality of the data they publish.

Since she was unsuccessful in finding the compound reported in the previously accessed articles, she returned to the SciFinder results to evaluate the remaining reactions. Because she had encountered the previous roadblocks (i.e. the articles did not report the methylated modified target compound), she returned to the reactions and decided to access the article for reaction #3, which she initially avoided because of the complex titanium reagent (**Step 3H**). She justified her decision to now pursue the full text in the following quote:

> *I don't like it but sometimes you do have to end up doing things that you wouldn't necessarily want to do. Since I am noticing that this transformation is more difficult than I originally thought you probably would end up going with this catalyzed reaction since you see it multiple times from different papers [reaction #2 and #3]. That is the difference. Yeah the titanium is used in multiple sources.*

Although she previously avoided the reaction due to the complex titanium complex, she changed her perspective because she had reached several roadblocks, and, more importantly, the reagent was reported in multiple reactions. As such, she associated the usefulness of the method with the frequency with which the complex was reported in other reactions. She could not access the full text electronically; however, she was able

to obtain the abstract reporting the reaction conditions. She stated that, "*the next thing I would do is just find, get the paper itself, find the conditions*."

At this point in the interview, she had established a method to synthesize the backbone structure without the benzyl group. Therefore, she now turned her attention to finding a method to add the benzyl group. Using the modified target compound as the starting material, she performed an RxV search for the transformation seen in **Step 4**, to determine if the reaction of adding the benzyl group to the modified target compound was reported.

> *So [the modified target compound without benzyl group] is now going to be my reactant and I want to see if I can make, if there is any known transformations out there that already induce that...induce that phenyl group onto that backbone.*

The RxV (**Step 4**) search for the transformation yielded no results, so she returned to the structure editor screen and performed an RxS (**Step 5**) search for the transformation (backbone structure yielded target compound). Both searches yielded no results, so Sophie concluded that the addition of the benzyl group could not be readily accomplished and she was not approaching this step from the "*right*" perspective as she explained:

> *If this was something that I was actively going after and had to spend time looking for like this, it would be like ok, something is really...I am not looking at it right. That is what it would be tell me. It tells me that maybe instead of focusing on this as the backbone, maybe I should change my original idea up. If it is that hard to find usually there is something easier out there. Or if there is not, usually there are at least some precedence of it being out there at some point. I am not seeing anything.*

She concluded that she needed to change her approach to adding the benzyl group, because the previous transformation searches yielded no results. Changing her approach to the task, she tried to use her prior knowledge of organic reactions to sketch and reason

through how she was going to add the benzyl group (**Step 6**). During this time, she stated:

> *So really this is just my bane, this phenyl group. If I could figure out how to attach [the benzyl group] onto this molecule, that is really all I have to do. So now I am thinking in my mind of different reactions that I have learned in general organic and just in grad school to attack at that position. Adding straight-up phenyl groups is not common unless it is attached to like an acyl chloride or something like that where it can pop off the chlorine and have like a Lewis acid aluminum tetrachloride or something come in and they form big complex and they do this great chemistry. But that is one way people do generally add in phenyl rings just by themselves. That could be a possibility because I have this extra carbon to play around with. It is not like this phenyl ring is directly attached at this position. I have got a carbon to play with so I might as well use it. We are going to keep that backbone structure the same because we already found it and we know how to make it theoretically. Now I just need to figure out what functional group I need to put in this position to make this happen.*

She had a method to make "most" of the target compound, and, for this reason, she was very apprehensive to deviate from that path. Moreover, at this stage in her method development, she was trying apply all the knowledge she had constructed from her undergraduate and graduate organic chemistry experiences to determine what functional group she needed in order to add the benzyl group.

> *Yeah it is a good thing I am learning, this is literally how I would deal with a problem. I like carbon-carbon double bonds. Like I said the phenyl group is not always the easiest thing to attach unless it has got a chlorine attached to it where the chlorine can be the leaving group, pop off, do an $AlCl_3$. What synthon needs to be there? Well let's look.*

She was unable to recall which functional group needed to be at the benzyl position; therefore, she decided to perform a topic search to find precedence for her ideas. She input/searched for the key words (**Step 7**): "Lewis acid mediated phenyl substitution." She stated:

> *So I have an idea, I think that I can do this addition of the phenyl ring if I have the chlorine attached to it that can act as a leaving group in the presence of a Lewis acid. So I am going to go to my research topic. Let's learn about it.*

Her chemical intuition was suggesting she could easily add the benzyl group to the modified target compound, but she could not support her ideas using the chemical structure search option. For this reason, she was changing her search strategy and using the database, specifically the topic search, as a vehicle to connect the missing pieces.

When evaluating the subject search she was specifically focusing on the key words listed in the titles. After reading the titles for each article, she decided to pursue the full text for an article (**Step 7B**) because she stated it had all the important key words. (There were, however, several other words in the title, namely a specific class of heterocyclic compounds that were completely unrelated to the task at hand.) She then proceeded to evaluate the article in the previous manner by scrolling down to the reaction schemes. During this time, she stated:

> *The only issue I have with this paper is that again it is closing rings. I wish I could see examples of non-ring closures and simple additions. This is just not what I am after. But if we are closing rings I don't want to make a new ring. But there are multiple rings that I could play around with. So this would be a paper that I would hold on to.*

Following her review of the schemes, she concluded that they were performing a ring closure and the paper would not be useful at this stage in her search because she was focusing on substitution reactions. Although she did not want to deviate from her substitution idea, she would keep the article open in another tab just in case she needed to change her strategy and perform a ring closure. Returning to the topic results, she further

read the titles and concluded that the remaining articles would not be useful at this stage in her method development because they were also reporting methods for ring closures.

Returning to the structure editor screen, she attempted to simply the structures by focusing on adding the benzyl group to a cyclohexane ring.

> *What I am going to do is go back to the drawing the substance drawing and I'm going to go super generic. I am getting to that point to where...you get to that point where you are like alright I just need to see if this is going to happen. I have got to see if this is even possible. So I'm going to take the bare basic minimum of what I've got. I would like for this to be a double bond because I know that it is probably out there somewhere you know with a primary carbon-carbon double bond. Now this is going to be my general form of this, of our bicyclic system. Now we are going to go for the phenyl. Now we are going to see if I can track down another reactant. I really want to do this acyl chloride displacement.*

She was trying to simplify her structures in an attempt to find precedence for the chemistry she was proposing. As seen in **Step** 8, she input the simplified structures and performed an RxV search. However, the RxV search yielded no results, so she returned to the structure editor screen and broadened her search to perform an RxS search for the transformation (**Step 9**). Although the substructure reaction search yielded 3,230 transformations, she only reviewed the top four reactions (**Step 9A**) and concluded they were not useful because, "*They are reducing the bond is all it looks like to me. Ok so that is not what we are after*." Before she completely disregarded the results, in **Step 9B** she attempted to refine the results based on the reagent aluminum chloride, but the reagent was not listed as an option.

Because the transformations did not provide any useful information and she could not find the reagent, she performed another SSi search for the structural representation of the target compound (**Step 10**). She justified her actions by stating:

Sophie: Sometimes I go back and see if I was just not in the right mindset.

I: If you missed something?

Sophie: Yep I am not too proud to admit that many times, especially if you have been doing this for hours you are like no, no I am done.

Although she performed the SSi search for the structural representation of the target compound to verify that she did not miss any potentially useful information, she came to the same conclusion, that the ring sizes were different and the substances would undergo different chemistry (**Step 10A**).

In an attempt to once again apply her organic knowledge to propose a method to add the benzyl group, she decided to sketch and explain her ideas (**Step 11**):

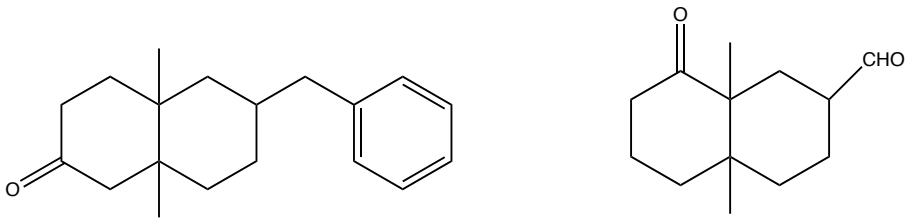
So I have an idea. Grignard reagents with you know you can do you know the general reactivity of those is to insert a nucleophile at the carbon oxygen double bond. So we know that [the carbon] is partially positive, [the oxygen] is partially negative with a Grignard reagent can act, it can come in and do carbon-carbon double bond. What my thought is, is taking any...that cyclic system that we were talking about before but put an aldehyde on the end of it. And then we add in our phenyl that has the chlorine on it. Then we add in the Grignard reagent, it is going to do the insertion, and then it is going to add in this bond to here, but it will give us the ring. And then usually the work-up of that is an alcohol. Alcohols are a lot easier to remove than a lot of other things. It would be a multiple step process but it wouldn't be unrealistic. So in theory, I haven't found any literature to support this, but we know we have the starting material of this. It should find now is to go and see if we can find an aldehyde on the other end.

Once again, her chemical intuition directed her to use a Grignard reagent to add the benzyl group, which she justified as follows:

Just chemical, dealing with general organic and grad school. These are just general reactions that for people that have dealt with it and read a lot of papers these are just simple basic organic chemistry transformations that we teach them over and over again. That is why I am comfortable with it and that is why that is the first thing that I want to gravitate towards. That is something that I have been exposed to longer and feel more comfortable with.

After she sketched the proposed transformation, she input the compound as seen in **Step 12** and performed an SSi search. She then retrieved the two substances with a 95-98% similarity match to the compound for which she searched (**Step 12A**). She quickly evaluated the SSi search results and stated that the substances were not useful because the ketone was in the wrong position.

> *[The percent similarity] was really high, it wasn't 100%, but it was very similar. And again the only difference is the position of that ketone [*Figure 4.51*d]. And neither of those are commercially available. So again we would have to find the synthesis for them*.



(a) (d)

Figure 4.51: Target compound (a) and substance #1 from the SSi search (d)

She encountered yet another roadblock because she could not support the chemistry she was proposing. This roadblock caused her to deviate from her initial approach to finding a method to add the benzyl group to the backbone structure. Changing her strategy, she then proposed and sketched her thoughts for performing a ring closure on an alpha, beta-unsaturated ketone (**Step 13**).

> *So essentially we are just going to break the molecule apart. We can have a methyl group here, [top methyl group] no problem. The methyl group here [bottom methyl group] can either be in the original molecule, most likely not, it is going to hinder the reactivity at the fourth position. So you don't really want to do that we want to make it as extremely favorable to react there as possible. I*

wouldn't put a methyl group there just steric hindrance will cause it to not react as well.

Although she could justify her decision for breaking the carbon-carbon bonds, she was skeptical that she could find the structure the sketched with a benzyl group attached.

I am concerned yeah that I won't be able to find a material that has a chain that long on it. So it is not so much will it not work, it is can I find something out there that will be a good possibility to react. Odds are finding a ring with that long of a chain out to the side, absolutely they are out there, throwing on a phenyl group too, probably not. So we are asking a lot of that backbone. But I do know specific examples where they do ring closures with alpha, beta-unsaturated ketones and then reduce, and it just reduces that bond there. So no big deal there, but causing to do that one, four, not so easy.

Despite her skepticism, she continued to input the structural representation she previously sketched in **Step 13** into SciFinder. As she was inputting the structure, she did not notice that the benzyl group needed an additional carbon to yield the target compound. Unaware of the missing information, she proceeded to perform a similarity search for the compound (**Step 13A**), and as she was reviewing the substances she remembered that substance #1 was reported in the previously retrieved article from the topic search. Since the article was still open a tab, she clicked to continue reviewing the schemes in the article (**Step 13C**). She was becoming more frustrated because she had ideas about the chemistry she could perform, but was unable to put her knowledge into practice to complete her synthesis protocol for the target compound.

I know where the reaction supposed to occur. My chemical intuition is telling me that the phenyl ring is what is supposed to be easily attached in some way shape or form. What is frustrating is for me, I am just not able to make the connection or pull from all the reactions that I have seen to say ok that is the one.

Although she was struggling to find information in the literature to support her ideas, she would not seek her advisor's help at this stage in her method development.

I: Would you go talk to your advisor?

Sophie: Not yet. Um I would sit down with it some more. I would sit down with it some more and try and think or go back to my organic books and you know try some other situations first.

Sophie persisted to use the database, in conjunction with her organic knowledge, to develop a research protocol. Despite the frequently encountered roadblocks, she was continuing to use all her resources to solve this cognitively demanding task. Furthermore, she was reluctant to quickly seek her advisor's assistant because she had ideas regarding the chemistry she could perform, and therefore wanted to exhaust all her options before asking for help.

She then decided to attempt one last search before she officially stopped working on the task. For this reaction search, she input the transformation seen in **Step 14** and performed an RxS search. She explained:

There is one other thing I want to try real quickly. I want to reverse this and put the leaving group on the backbone that we want and see if I can find an electrophile on the phenyl group to kick it off. So I did it the other way, I did it where the nucleophile was going to attack, or the electrophile was going to be on the big portion of the molecule and have it attack the phenyl group. Now I am going to try to reverse that and see if I can find anything.

Unfortunately, the search did not yield any hits; therefore, she concluded that she had exhausted all her ideas and strategies. She explained:

So in theory sure the reaction should occur just like with the Grignard with removing the proton and all that, but in a legitimate reaction you can't say that the palladium won't also maybe reduce some of the phenyl ring or....it shouldn't reduce that because it is palladium but you know. There is never a set...ok the theory that works there is definitely going to work there. It is never that black and white…At that point you go to your advisor and be like alright [advisor] these are the ideas I have; I haven't been able to find literature to back any of it up. So that right there is red flags that you are not approaching it the right way. And I understand that. So that being said, 'what is the appropriate way to

approach it.' And that is where I have to go to him and say, 'alright I have all these ideas but it is not matching up with what is in the literature do you have any other suggestions that I am not seeing'.

She had used the database in conjunction with her organic knowledge to develop a research protocol for the target compound, but she was unable to find precedence in the literature that her ideas would work. At this point, she realized that she would need help from her advisor for which he could suggest alternative ways to either perform a SciFinder search or different organic reactions that might yield the target compound.

Summary of Sophie's Search

Unlike some of the other participants in this study, Sophie did not spend much time analyzing potential pathways to synthesize the target molecule. As such, her overall search appeared less organized than those of some of her peers. Interactions with her advisor, especially during group meetings, influenced Sophie to develop what she perceived to be "a broad approach" to searching the literature. However, her version of this "broad approach" resulted in a reductionist strategy in which she would first generate the hydrocarbon skeleton of the target and then add the requisite functionality. Even though she eventually found a potential method for synthesizing that backbone she soon would run into a series of roadblocks in searching to transform that backbone structure into the target compound. On the way, she even found that even an analog with 99% similarity to of one of her intermediates gave no insight into finishing the task at hand.

Most interesting, however, that in light of all of this evidence, Sophie was still quite confident that her basic strategy would work, even if she may not be able to. Despite her relatively brief experience as a member of the organic CoP, Sophie gave far

more weight to her intuition that a compound as simple in appearance as the target must exist, thereby almost ignoring the repeated results of her numerous searches suggesting otherwise. Perhaps the anticipated interaction with her advisor would have served a metacognitive role by which she would approach the task with a completely new strategy thus increasing the likelihood of successfully completing the task in this interview.

BRUCE'S CASE

Background:

Bruce was a sixth year Ph.D. student who obtained his undergraduate degree in chemistry from a foreign university. He participated in an undergraduate research experience in organic total synthesis during his senior year that lasted for three months. Since his undergraduate institution did not have access to SciFinder or any other major electronic database, Bruce used his undergraduate organic chemistry textbook as his primary reference resource.

Bruce also had a research experience during his Master's degree, which he obtained from a small university in the southwestern United States. Based on the demands and rigor of his present Ph.D. program, however, Bruce was very reluctant to call his Master's degree experience a "*research experience*" because his advisor provided him with a compound and asked him to purify it using analytical separation techniques. As such, he felt like he did work more in line with a teaching lab than what he considered to be research.

Ph.D. Research

Bruce's advisor asked him to synthesize an organic compound and to selectively attach different nucleophiles to specific regions on the compound. His first step for this project was to search SciFinder to determine if someone had previously reported the main target compound. To begin his search, Bruce performed a reaction search for the carbon backbone structure with a generic -R group attached (SciFinder recognizes –R as any alkyl group). The search produced hits reporting methods for synthesizing the carbon backbone structure with various alkyl groups attached. Bruce would then retrieve the respective articles and compare the starting materials, focusing on the functional groups and substituents, which he would be using to those reported in the articles. He would also evaluate the protocols based on the experimental conditions for feasibility with respect to factors such as commercial availability and cost of the reagents. For instance, he would initially skip reactions reporting complex experimental conditions, such as running a column to purify the product, but would returned to consider those reactions if he was limited by his options. Once Bruce found a set of protocols that he believed would result in making the target compound, he tried the procedures in his laboratory without further consultation with his advisor.

Although Bruce had previous experiences conducting research, he had not used SciFinder prior to entry into his current Ph.D. research group. He learned how to use SciFinder because he felt that searching the database was the fastest way to obtain scientific information. He explained that his research group members showed him how to: 1) access SciFinder on his computer, 2) draw the structural representation of various compounds, and 3) perform a structure search.

Task 1 Search Chronology

The following is a chronology of his search:

- **Step 1**: Bruce input the exact structural representation of the target compound and performed an RxV1 search.
    - **Step 1A**: The RxV1 search yielded no results.
- **Step 2**: Returning to the structure editor screen, he modified his search strategy and performed an RxS1 search for the structural representation of the target compound.
    - **Step 2A**: The RxS1 search yielded two reactions. He proceeded to evaluate the reactions by focusing on the starting materials and percent yield. He then decided to retrieve the full text for reaction #1 (Figure 4.52).

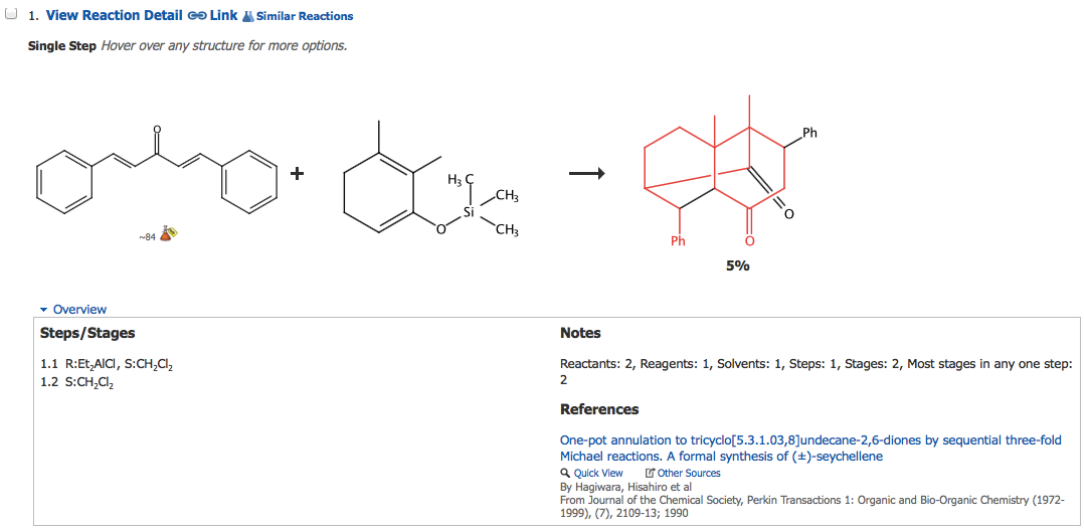


Figure 4.52: SciFinder screenshot of reaction #1 from the RxS1 search

- **Step 2B**: He then directed his attention to the downloaded paper and scrolled to find the compound for which he was looking in the schemes. While looking

at the reaction scheme, he concluded that the researchers were performing a Diels-Alder reaction to make the final product.

- **Step 2C**: He was struggling to comprehend specifically how the carbon-carbon bonds were constructed on the final compound. For this reason, he decided to sketch the transformation on the paper that was provided. As he was sketching the transformation, he would look at the transformation in the article and at the SciFinder reaction results (Figure 4.53).
- **Step 2D**: He concluded that adopting/modifying the protocol would not produce the exact structural representation of the target compound.

Figure 4.53: Sketch of the transformation reported in the article

- **Step 3**: He decided to disregard the previous results, and then performed the retrosynthetic analysis and proposed two synthesis pathways with four different starting materials, (Figure 4.54).

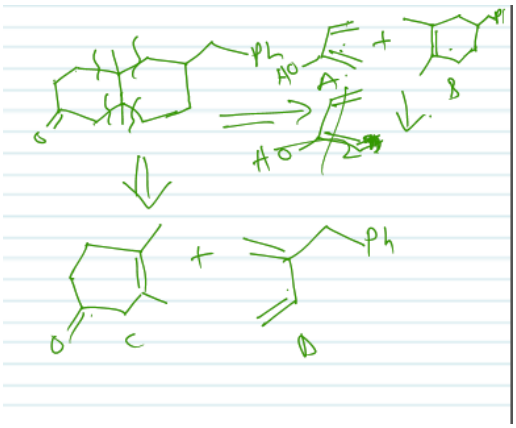


Figure 4.54: Retrosynthetic analysis with two reaction pathways

- **Step 3A**: He concluded the reaction pathway using starting materials A and B (Figure 4.54) would be a more desirable transformation.

- **Step 4:** He then input the structural representation of starting material A (S.M. A) in the structure editor screen and performed an RxV5 search (Figure 4.55).

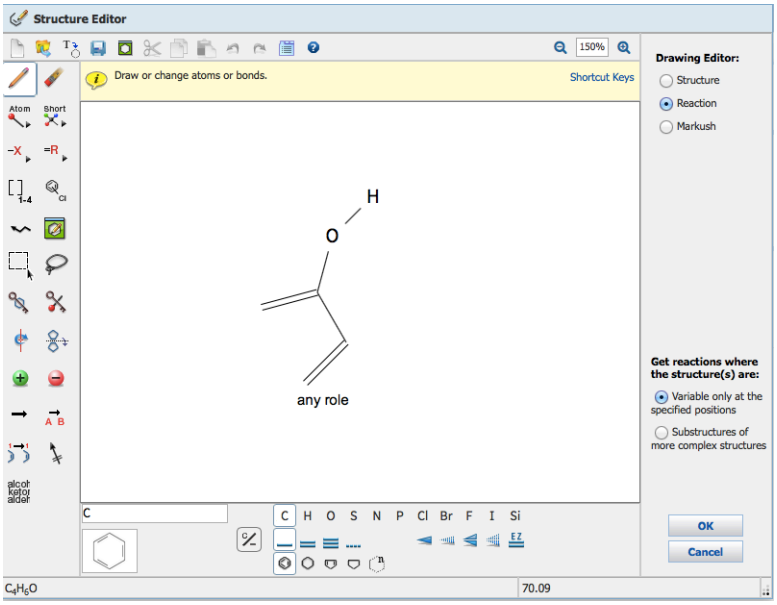


Figure 4.55: SciFinder screenshot of the RxV5 search for starting material A

- **Step 4A:** The RxV5 search resulted in 23 reactions. From the results, he concluded that he had a protocol to make S.M. A.

- **Step 5:** He then shifted his attention to starting material B (S.M. B) and input the structural representation of S.M. B into the structure editor screen and performed an RxV5 search (Figure 4.56).

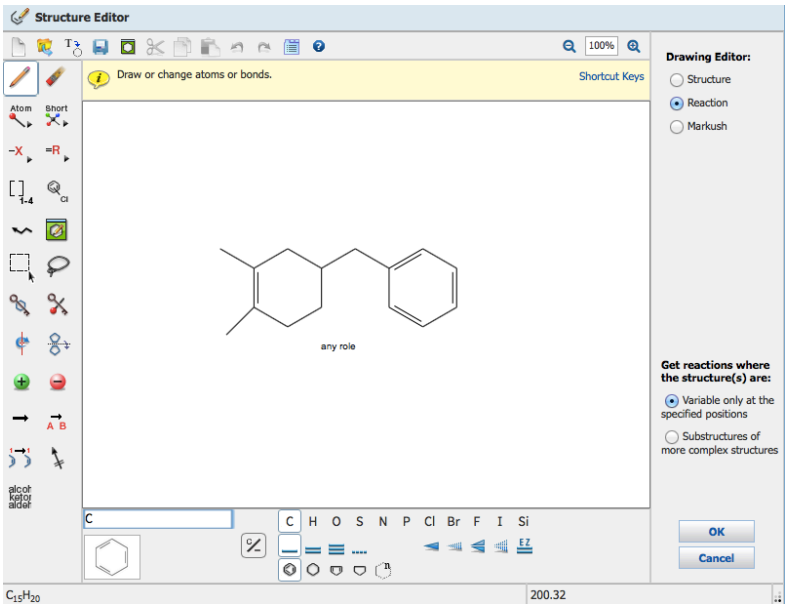


Figure 4.56: SciFinder screenshot of the RxV5 search for starting material B

- **Step 5A**: The RxV5 search yielded five reactions reporting the synthesis of S.M. B. He then evaluated the reactions according to the percent yield, simplicity of starting materials, number of steps, and reaction conditions.
- **Step 5B**: He then decided to retrieve the full text for reaction #5 (Figure 4.57).

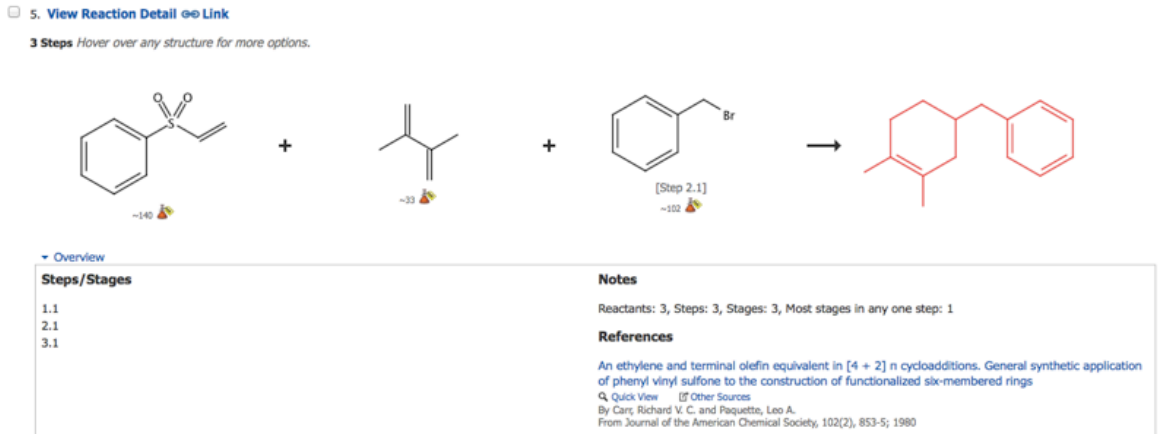


Figure 4.57: SciFinder screenshot of reaction #5 from the RxV5 search for S.M. B

- **Step 5C**: Turning his attention to the article, he quickly scrolled to the schemes to find the reaction reporting S.M. B. He then sketched the steps to make SM B (Figure 4.58).

- **Step 6**: He now directed his attention at determining if the noted starting materials in Figure 4.58 were commercially available. (He did not notice that SciFinder reported the starting materials as commercially available.) He input the structural

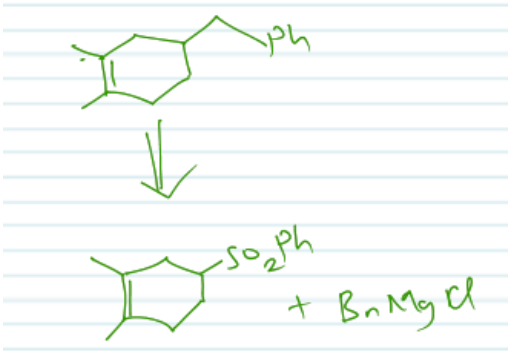


Figure 4.58: Sketched reaction to make S.M. B

representation of the starting material reported in the article and performed an SE search (Figure 4.59).

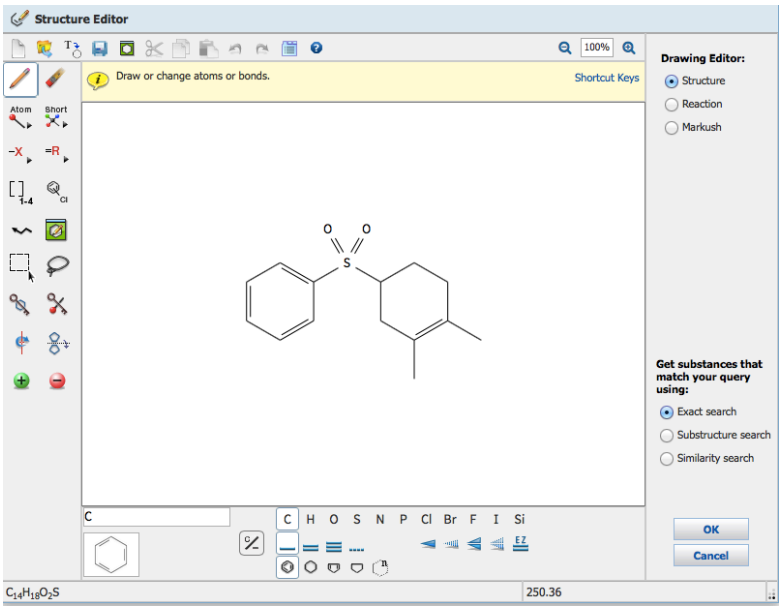


Figure 4.59: SciFinder screenshot of the SE search for the starting material reported in the article

- **Step 6A**: The SE search resulted in the exact structural representation of the compound for which he searched in the Figure 4.59. He then clicked the commercially available flask, which directed him to a page that displayed sources to obtain the purchasing information.

- **Step 7**: He stated that he would purchase the starting material (as seen in **Step 6**) and he would make the benzyl Grignard starting material as sketched in **Step 5C**. Therefore, he concluded that had all the information he needed to make the target compound.

Task 1 Search Critique

Before Bruce began to perform the Task 1 search, he shared his initial reaction to and interpretation of the target compound:

> *For the structure that you have here it is not like a complete structure. In terms of, you are not showing any stereochemistry. Ok it just a general. What you have here is, it has three stereogenic. One, so let me see, yeah one, two, three. So three stereo. Yeah, so these are the three stereogenic centers. So this is just a plane, if you are showing stereocenters then you have to use the wedge and dashed lines to show where these methyl groups are pointing. But it is ok.*

Although Bruce noticed the undefined stereochemistry, this observation did not deter his search strategy. After noting the undefined stereochemistry, his quickly began to use the database to determine if the compound was reported. As such, he input the exact structural representation of the target compound and performed an RxV1 search (**Step 1**). He justified his search action by stating:

> *This problem at hand is about synthesis, and I am going to do a reaction search and find out if somebody has synthesized it before. But it depends on the problem, because I can also do, let's say I want to buy, maybe I have made this compound, and let's say I do a reaction and I don't see any synthesis, and then maybe I go ahead and make it. I also want to find if it is known at all, if it has been reported at all.*

Further elaborating on his search strategy, he explained that he had developed a strategy for using a reaction verses substance search option:

> *So I can do a substance search. And when I do a substance search, I thought maybe someone mentioned and used it somewhere and it is commercially available you know. The substance search will be useful if I synthesize it and I do substance search and it is out there, I can compare my characterization results. If I characterize it and want to compare my results to what is out there to establish the stereochemistry. You know like the stereocenters you have to find out if you have the right stereocenter. So I can do a substance search if it is out there, then I can find information about maybe its melting point, how it's IR or NMR looks*

like and compare it to mine. If I compare and they are the same, then I can say they are maybe you know.

From Bruce's explanation of a reaction verses substance search options, he seemed to initially conclude that the target compound was unreported, so he proceeded to perform a reaction search in order to develop a research protocol. Also indicated within his previous justification was the way in which he used the database and the organic CoP.

But if it hasn't been reported then I have to go through all that and characterize it you know fully just so I can put the information out there in the literature so somebody will know that I am the one who did it.

His statements suggest the cyclical nature of his CoP. For instance, he would first use SciFinder to determine if his CoP reported the compound, but if it was unreported, then he would synthesize, characterize, and report the data so other members could use the information.

The RxV1 search yielded no results (**Step 1A**), so he concluded that, "*It hasn't been done*." He then returned to the structure editor screen and performed an RxS1 for the structural representation of the target compound (**Step 2**), which he justified in the following exchange:

Bruce: Ok so a substructure search, usually when you are synthesizing a compound, usually you synthesize them in parts, and then you put them together. The molecule as you have it, if you are going to synthesize, you are going to synthesize it in parts. Like you build the molecule part by part. Just like say you are constructing or you are building, putting up a building, you have to construct the various parts and then it all comes together as a whole building you know. You work on each part at a time.

I: How did you learn to work in that way?

Bruce: That is the easiest way to do it because there are no reactions that will synthesize a big molecule like at a time. This is a giant molecule and there is no reaction, because in giant molecule you need several reactions to make them.

Then he was asked to further explain how he acquired that knowledge for which he stated:

> *I learned it; part of it was undergrad and part in graduate school.*

Bruce's reasoning for performing a substructure search was to broaden his search by retrieving more complex compounds for which he could potentially modify to make different pieces of the target compound. He proceeded to evaluate the RxS1 search results (**Step 2A**) by specifically focusing on the entire transformation (starting materials and final product) and comparing it to his target compound. During this time, he explained:

> *The yield they got here is 5% but the point is, I'm not going to do exactly what they are doing so I may have to look into this a little bit. So here, [the starting materials] they have one ring and then they fuse it to form two rings just like I have.*

Although the percent yield does matter in organic synthesis, he knew that he would modify the protocol, so the very low percent yield was not a significant factor at this stage in the method development. His approach to evaluating the reaction indicates that early in the method development he was more concerned about making the carbon-carbon bonds and less concerned with the percent yield. More importantly, he knew that he would modify the protocol, so he was more concerned with the over-all transformation and not the percent yield. Bruce decided to pursue the full text for reaction/hit #1 (**Step 2A**) and he justified this action by stating:

> *I had only two options. I had two reactions here. When I look at the two, [reaction #1] is the best. In the literature this is what is available to me. Like as far as building the structure from simple material, this is what is available to me. So I have only two and so I choose the best of them. Like I said, in this one*

> *[reaction #2] they have this tricyclic as the starting but and so they didn't build a new ring. So here I have two rings I want to build so I have to get an example that actually use the strategy that builds a new ring, bicyclic ring. And that is the only one. So that is why I chose that.*

Despite that Bruce was limited to only two reactions, he gravitated towards reaction #1 because it reported a method for building a fused bicyclic ring. Once the article was retrieved, he scrolled to find the scheme reporting the final product (**Step 2B**). He explained, "*well you have to focus on the schemes and see if it is useful before you read everything*." Relating his statement to Interview #1, he explained that, "*You know chemistry, equation sums up everything.*" Organic chemists use structural representations to communicate the underlying chemistry involved in a transformation. Being in the organic CoP for six years now, Bruce naturally gravitates towards the schemes in order to understand the chemistry.

Bruce spent several minutes evaluating the schemes to determine if he could modify the reactions for the synthesis of the target compound. During this time, he further justified his desire to pursue the article by stating:

> *Because one, that is already available information and I want to use that first to solve my problem you know. If it doesn't solve my problem then I can sit down and do the hard work of going through and starting everything from scratch. That is the hard work. But if you have the problem, then you want to do the easiest way to solve it. That is why we consult the literature first. Otherwise if I want to do from scratch from my own knowledge then I have to sit down and apply all that I know in organic and to make sure if I can solve it. That is really tough way of solving it.*

Bruce could apply his organic knowledge to propose a synthesis; however, he was reluctant to do the "hard work" because he could easily access the information that was available by the organic CoP via. SciFinder.

While looking at the reaction scheme, he concluded that the researchers were performing a Diels-Alder reaction to make the final product. Although he could deduce the type of reaction from the scheme, he was struggling to understand specifically how the carbon-carbon bonds were constructed to make the final compound. For this reason, he decided to sketch the transformation on the paper that was provided (**Step 2C**). As he was attempting to conceptualize the transformation, he would look at the transformation in the article and at the SciFinder reaction results in an attempt to visualize the reaction from different perspectives. Moreover, he used the arrow-pushing formulism to conceptualize the mechanism reported in the scheme. This is one of the tools that practicing organic chemists use to understand the flow of electrons from source to sink. As such, he used this tool to, "*see where the bonds are constructed*," and by utilizing this tool, he was able to conclude that the reaction was not transferable for the purposes of synthesizing the target compound (**Step 2D**). Following his analysis using the arrow-pushing formulism, he stated:

> *That doesn't give us exactly what we are looking for because you get a fused ring but not a bridge ring. So this one does not give us exactly the construction that we are looking for. You get a fused ring but it doesn't have the bridged junction. These reported bridged junction doesn't give you, you want a bridged junction. But this one may not, does not work, doesn't give us the compound, the exact construction that we need.*

Because he determined the previous RxS1 search results were not useful for the synthesis of the target compound, he decided to change his approach to the task and used his organic knowledge to perform the retrosynthetic analysis and proposed two synthesis pathway (**Step 3**).

So if this [protocol] does not work, what we have to do is just do a Diels-Alder synthesis reaction. So I am going to do the reaction. So I will just have to do what you call the retrosynthetic analysis. So what I will do is to break this that way and see what I get. So this one will give me, if I break it over there.

Bruce seemed to effortlessly apply his knowledge of named organic reactions to perform the retrosynthetic approach. He was able to "break" the bonds in the target compound and performed the retrosynthetic approach to propose four different starting materials with two different reaction pathways. Named organic reactions, such as the Diels-Alder reaction, and the retrosynthetic approach are common tools used by the organic CoP. Moreover, he justified his decision for performing a Diels-Alder reaction in the following exchange:

Bruce: We are going to construct a ring. Now construction of six-[membered] rings usually Diels-Alder reaction is well known for...Yeah Diels-Alder is well known for constructing 6-membered rings. The reaction [starting materials A and B] yield might be very, it might be higher with some functionality. So looking at this, the Diels-Alder was constructed from like a diene, so this is a diene. This is a plane compound, double bond. This is the only one that will have some functional groups. And I know these two compounds [starting materials C and D], two starting materials maybe harder to make. If you look at it, it will be harder to make something like this. I know that this one might be readily available.

I: How do you know that?

Bruce: This is one, two, three, four. It depends on my prior knowledge, but we can also work...that doesn't mean this one is completely discarded [reaction C and D]. I may try this one [reaction A and B], this step may not work and then I may come back to this one. If I find this one then I can use it…Actually as I am going through here I may consider this again and see if the yield is not satisfactory here I may consider this route.

After he proposed the two possible pathways, he reflected on these options and decided that the reaction using starting materials A and B would be a more feasible transformation (**Step 3A**), for which he now needed to:

Go back to SciFinder again and look for this, if they exist. If they exist, commercially available then I can just buy them and do this one step to get that.

Although he had determined which starting materials were more feasible for the synthesis of the target compound, he once again needed to use SciFinder to determine if the starting materials were reported.

Now directing his search strategy to the proposed starting materials, he then input the structural representation of starting material A (S.M. A) and performed an RxV5 search (**Step 4**). He justified this search by stating:

So I just want to see if it exists out there if someone has ever used it. So I click any role and search and see.

The RxV5 search for S.M. A yielded 23 reactions (**Step 4A**); however, he did not pursue the full text for any of the articles because the results indicated that S.M. A was reported. He verified that S.M. A was reported in the literature; therefore, he then redirected his attention to S.M. B. He justified his actions by stating:

The point is that the reason I forgot about [S.M. A] and put this on hold. Once it is in the literature, whether it is commercially available or someone made it, and if someone has made it the procedure that they used is out there so I can just go and find. So that is why I stopped. So I am focusing on [S.M. B]. Ok so now I draw the next one. Let's say search.

Redirecting his search strategy to S.M. B, he then input the structural representation of SM B and performed an RxV5 search (**Step 5**). The RxV5 search for S.M. B retrieved five reactions reporting the synthesis of S.M. B. He then evaluated the reactions to find the most practical protocol (**Step 5A**), which he explained in the following exchange:

I: What are you focusing on?

Bruce: How much yield did they obtain from their transformation? What is their yield? What is the ease of starting material? It is a long process so you have to

go through and analyze all them and you know look at. First you have a quick scan of which one might be good. And once you have an idea of which one might be good then you want to sit down and go through the ease of starting material, what is the difficulty of the reaction itself?

Proceeding to evaluate the reactions by the aforementioned criteria, he decided to retrieve the full text for reaction #5 (**Step 5B**), which he justified by stating:

So you see this one [reaction #5], this molecule would be easier to make. They have constructed the whole ring all together. This one is practical. They also used the Diels-Alder to make it and it is from the "Journal of American Chemical Society." If this one seems practical I would just look for the paper. Because they constructed a new ring from simple starting material.

In addition to considering the practicality of the method, Bruce was considering the journal. Therefore, Bruce's approach to evaluating the transformation by also considering the journal was a good strategy and was indicative of his experience with conducting organic research, and his familiarity with more prestigious journals.

Once the full text was retrieved, he quickly scrolled to the schemes to find the reaction (**Step 5C**). As he was evaluating the schemes he decided to sketch the reaction and stated:

*So this compound is obtained from this, and again a reagent benzyl and Grignard reagent. So right now what you need to know if this one [starting material in **Step 6**] is commercially available.*

Although he had a method to synthesize S.M. B, he still needed to determine if the starting materials reported for the synthesis of S.M. B were commercially available. Interestingly, Bruce did not notice that SciFinder had the starting materials listed as commercially available. Because he did not notice the labels for the starting materials, he then searched to determine if they were commercially available (**Steps 6-7**). As such, he input the structural representation of the starting material reported in the article and

performed an SE search (**Step 6**). The search resulted in the structural representation of the substance, which he then clicked the commercial available flask (**Step 6A**). At this point in the interview, he stated that he had all the information he needed to make the target compound. He explained:

> *Bruce: You want to reduce your problem to commercially available. But like I said you want to see if it has already been done. But once it is not then you have to do the problem.*
>
> *I: So after you determine that it is commercially available what would you do then? Go back and find each [article]?*
>
> *Bruce: So if you already have the compound in your lab then you just go ahead and start your work. If you don't then you have to order it and then once you get it then you start the reaction.*
>
> *I: Do you go back and get those papers though?*
>
> *Bruce: Oh yeah…yeah. You get those papers.*
>
> *I: Do you read them?*
>
> *Bruce: Well it depends I mean what information. You can read all of it or you can read just get the portion that you need. But usually it is good if you read all of it so that if there are any anomalies and stuff, any problems they encountered during their synthesis that will help you to avoid.*

Then, he further explained for the synthesis of S.M. B (**Step 7**):

> *Bruce: So that is commercial available I buy it and make it, benzyl Grignard. So that I can make in the lab and I can combine them and do this.*
>
> *I: And you are finished.*
>
> *Bruce: And I am finished.*

At this point in the interview he stated that he had all the information that he needed in order to synthesize the target compound and he was prepared to attempt the synthesis in his laboratory.

Summary of Bruce's Search

Bruce's Task 1 search was reflective of his six years of experience performing organic research. Bruce quickly learned that SciFinder was a valuable resource once he began his Ph.D. training, and, as a result of increased usage, developed a strategy for searching for protocols for the synthesis of an unreported target compound. He approached the initial search from a practical perspective, which was to first verify that the target compound was unreported. Following this search, he then used his knowledge of named organic reactions and the retrosynthetic analysis as tools to develop a synthesis route to the unreported target compound. With respect to his organic knowledge, he was able to determine which bonds he could break in the target compound that would subsequently undergo the Diels-Alder reaction. As such, he then utilized the database to find precedence for his proposed starting materials, which he then felt confident to perform the experiment without his advisor's approval.

LUKE'S CASE

Background

Luke was a sixth-year Ph.D. student who obtained his undergraduate and Master's degree in chemistry from a foreign institution. He did not have a research experience during his undergraduate education. During his Master's degree, he had experience with natural products research, specifically extracting and characterizing substances from natural sources, which was the only research experience he described prior to beginning his Ph.D. training.

Ph.D. Research

Luke worked on three major projects throughout his graduate school experience. For each project, his research goal was to develop more efficient synthetic methods with respect to better yield and fewer steps. For the first project, he attempted to use an organometallic reagent to synthesize his target compound. He explained:

> *And I don't know whether this compound exist or not. First I will go in and draw this molecule in SciFinder.*

Luke's first step for this project was to input the target compound and perform an SE search. Since he did not get any hits for the target compound he concluded that it was not reported. He then performed the retrosynthetic analysis and used SciFinder to perform an SS search for the structural representation of the target compound. He knew the organometallic reagent he was going to use, and a general idea of the type of starting material he wanted, so he refined his search to retrieve organometallic-based methods using substrates similar in structure – *i.e.* functional groups and alkyl moieties – to his anticipated starting materials. Using the results of the refined search, he was able to modify the protocols and developed a research protocol for which he "*discuss[ed] in our group meetings.*" Going beyond the scope of the project, he modified the starting material he was using and was able to synthesize several derivatives of the target compound, which he published in three different papers.

For the second project, his goal was to develop an enantioselective general method en route to an important class of compounds. Although the target compound was reported, no one had reported a method to selectively synthesize derivatives of the target compound. After Luke proposed this project, his advisor became very excited and

encouraged him to carry out the research. He was still working on this project during the data collection period.

The third project was an extension of a research project on which one of his group members had previously published. His goal was to use the starting material published in the article and develop a new synthesis route to a target compound.

Luke did not have any experiences using SciFinder prior to entering graduate school. In the beginning of his Ph.D. training, his advisor helped him use SciFinder by guiding him through the steps to search for and find research protocols. Luke's advisor explained that performing a SciFinder search can result in a substantial amount of results, and he would need to decipher through the results to determine which methods would be relevant for his final product. His advisor would also sometimes devise a synthesis protocol and tell Luke to search SciFinder to determine if the methodology was previously established and/or if there was a better method reported. However, Luke did not discuss the specific details of what his advisor taught him regarding how to use the tools to perform a SciFinder search.

Task 1 Search Chronology

The following is a chronology of his search:

- **Step 1**: As Luke described his initial interpretation of the target compound, he decided to sketch a potential synthesis pathway to the target compound (Figure 4.60).

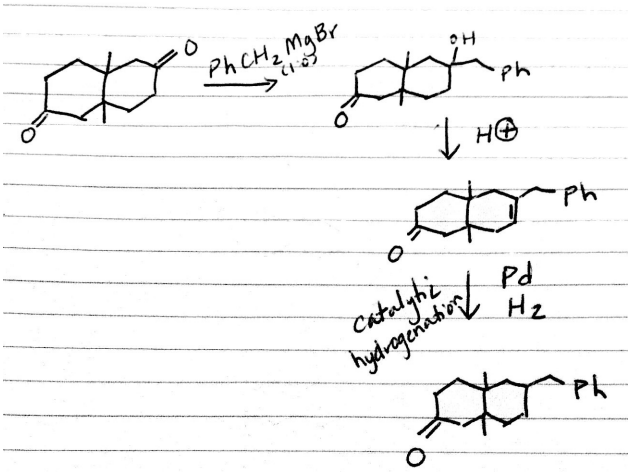


Figure 4.60: Sketch of proposed synthesis route to target compound

- **Step 2**: Once he finished sketching his thoughts, he then input the structural representation of the target compound and performed an SE search.
    - **Step 2A**: The SE search yielded no results.
- **Step 3**: Returning to the structure editor screen, he then performed a SS search for the structural representation of the target compound.
    - **Step 3A**: The SS search yielded 58 substances (Figure 4.61). He proceeded to evaluate the substance but decided to continue to search for other options.

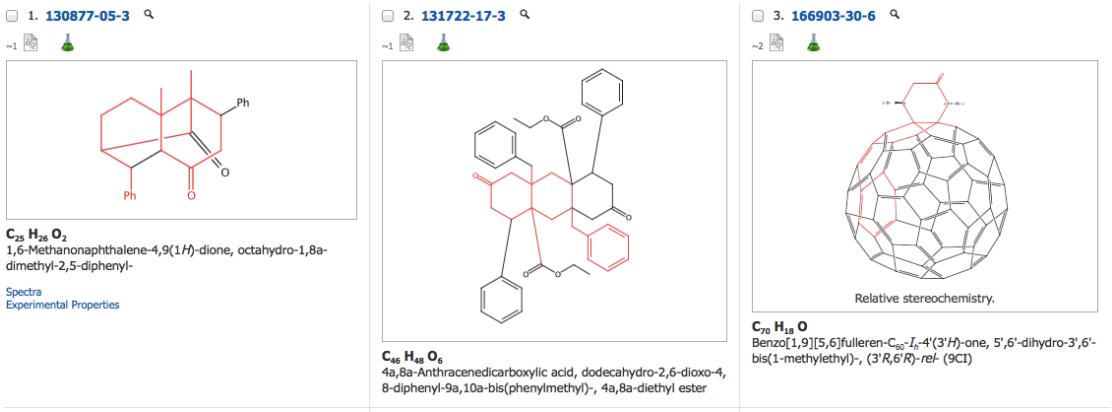


Figure 4.61: SciFinder screenshot of the SS search results

- **Step 4**: Continuing through his search options, he performed an SSi search for the structural representation of the target compound.
    - **Step 4A**: He then clicked to retrieve the 14 substances with an 85-89% similarity match to the target compound (Figure 4.62).

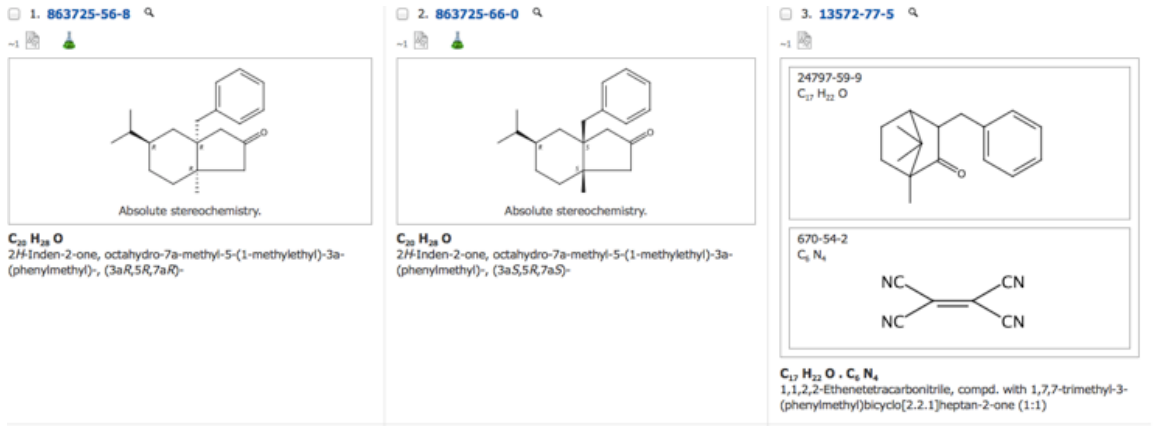


Figure 4.62: SciFinder screenshot of the SSi search results

- **Step 4B**: After he viewed the first page of results, he decided to retrieve the references for substance/hit #1 (Figure 4.62).
- **Step 4C**: Then he decided to retrieve the "full text" for substance #1.
- **Step 4D**: Turning his attention to the article, he scrolled to evaluate the scheme reporting the compound for which he searched.
- **Step 4E**: He decided to keep the reference, but concluded the protocol would not be the best method to adopt.

- **Step 5**: He then input a modified version of the target compound, as seen in the Figure 4.63, and performed an SE search.

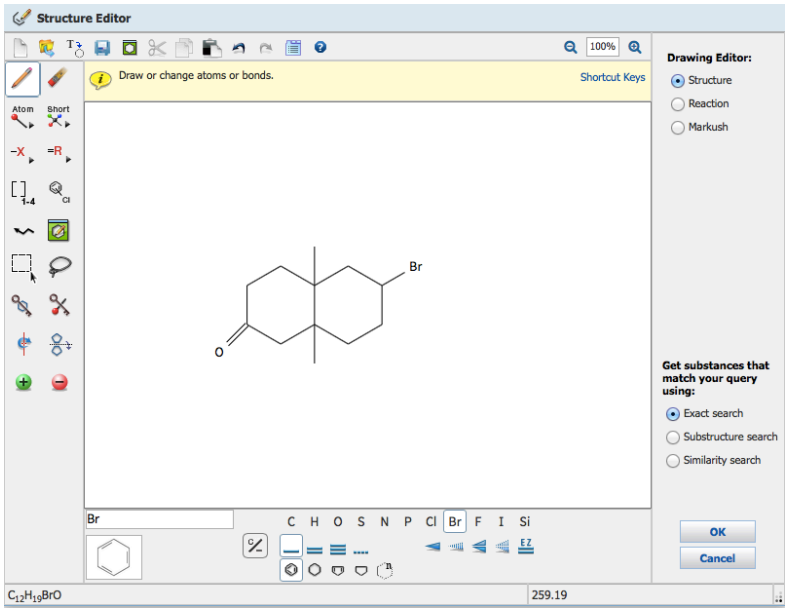


Figure 4.63: SciFinder screenshot of modified target compound with a bromine group

- **Step 5A**: The search yielded no results.

- **Step 6**: Returning to the structure editor screen, he removed the bromine, inserted an iodine group and performed an SE search.
    - **Step 6A**: The SE search yielded no results.
- **Step 7**: Returning to the structure editor screen again, he removed the iodine group, inserted a chlorine group and performed SE search.
    - **Step 7A**: Once again, the search yielded no results.
- **Step 8**: He returned to the structure editor screen and performed an SS search for the modified target compound with a chlorine substituent.
    - **Step 8A**: The SS search yielded no results.
- **Step 9**: Following this similar pattern, he returned to the structure editor screen, removed the chlorine, and inserted a bromine substituent. Then he performed a SS search.
    - **Step 9A**: The SS search yielded one substance (Figure 4.64), which he then clicked on the structure to further evaluate the compound.

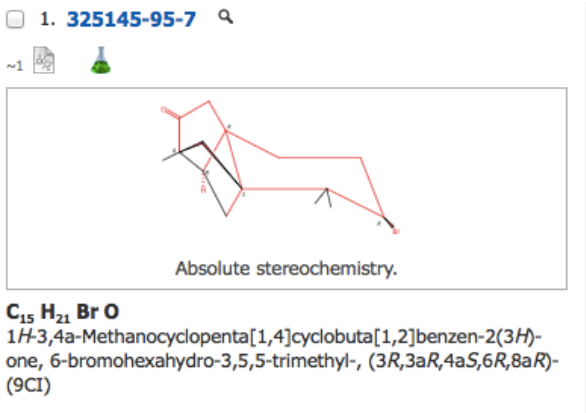


Figure 4.64: SciFinder screenshot of SS search result from the modified target compound

- **Step 10**: He decided to disregard the results and returned to the structure editor screen. He changed the bromine to iodine and performed an SSi search.
    - **Step 10A**: The search resulted in one substance with a 90-94% and six substances with an 85-89% similarity match to the compound (Figure 4.65).

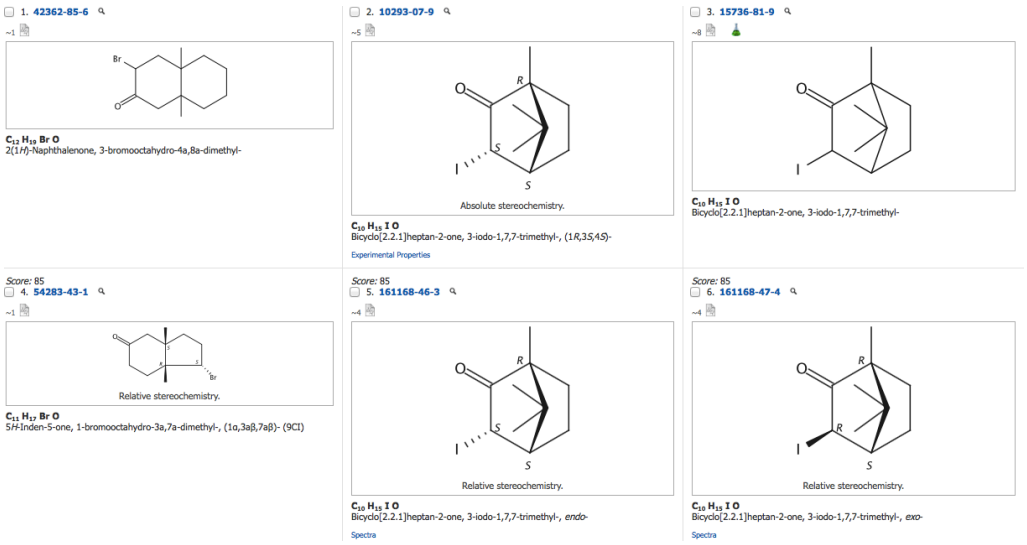


Figure 4.65: SciFinder screenshot of the SSi search results from Step 10

    - **Step 10B**: After viewing the substances, he decided to disregard the results.
- **Step 11**: He returned to the structure editor screen and input the following compound (**Figure 4.66**). At this point, he performed both an SE and SS search but retrieved no results. Continuing through his search options, he then performed an SSi search.

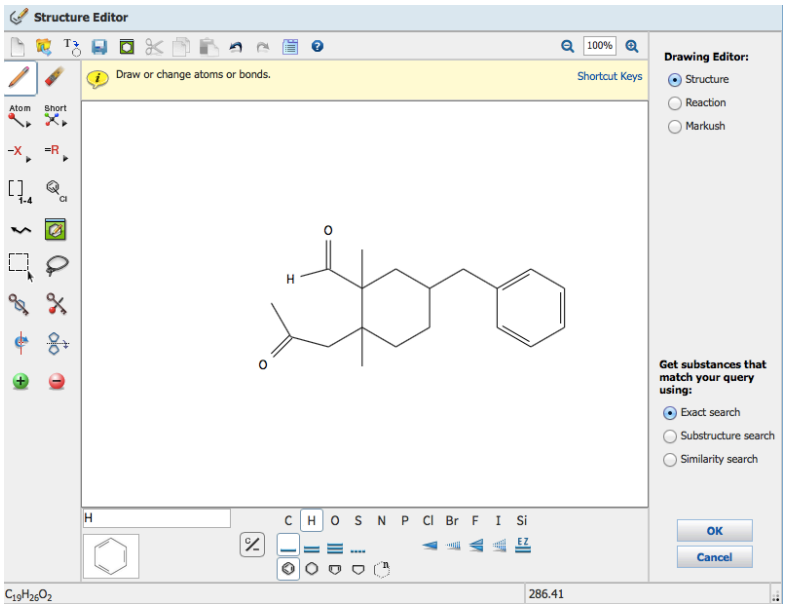


Figure 4.66: SciFinder screenshot of modified target compound

- **Step 11A**: He clicked to retrieve the four substances with an 80-84% similarity match to the compound, seen in Figure 4.66.
- **Step 11B**: He clicked to retrieve the reference for substance #1 (seen in Figure 4.67), and then retrieved the full text.

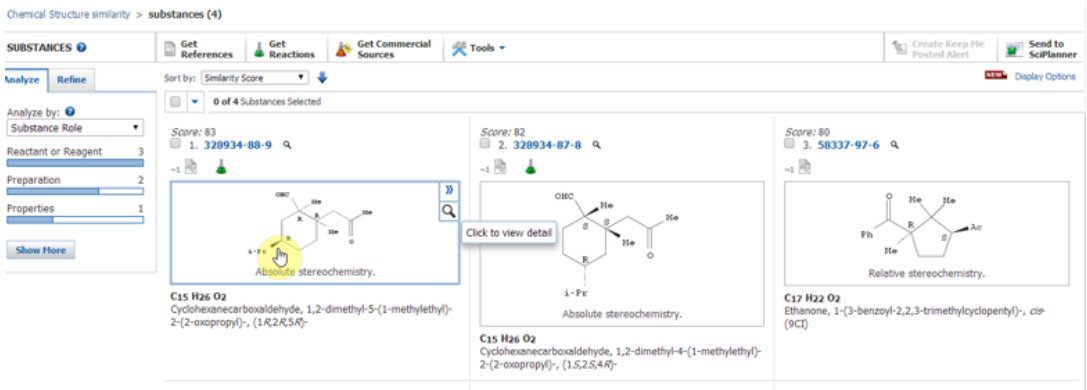


Figure 4.67: SciFinder screenshot of the SSi search results for the previously modified target compound

- **Step 11C**: Once the article was retrieved, he quickly scrolled to the schemes. During this time, he concluded he had all the information he needed and he sketched his synthesis protocol (Figure 4.67).

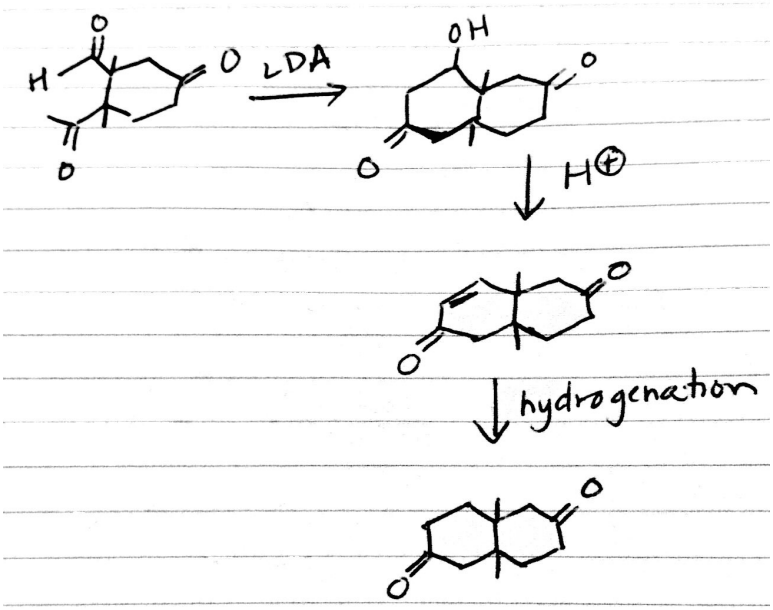


Figure 4.67: Luke's sketch of his synthesis pathway

Task 1 Search Critique

Before Luke began to perform the Task 1 search, he shared his initial reaction to and interpretation of the target compound:

> *I mean I saw that this is a carbonyl functional group, right. And the carbonyl functional groups can be in a bicyclic compound. Two cycle right. So I want to have the carbonyl group. And one kind of thing that we can do here is, you know we can do the Aldol condensation. One of the principles that we use is the Aldol condensation. And the second thing is, you are putting the second benzyl group here, right. And this benzyl group can be generated by the maybe S_N2 reaction. And the other possibility is we can put them into the corresponding carbonyl compound and you know we can use the Grignard reagent we can go for the dehydration and catalytic the hydrogenation of the alkane to get this compound.*

To further explain his ideas, he decided to sketch (as seen in **Step 1**) a potential synthesis route and stated:

> *This is the molecule ok. So we can selectively react only one carbonyl group. Ok, if we use one equivalent of $PhCH_2MgBr$…it should be one equivalent. If it is two equivalent then that is going to react with both of those. But if we use one equivalent of this one, then you are going to end up with…And so it means that...and if you go and dehydrate this one in the presence of acid you are going to end up with this kind of compound. And if you do the catalyzed hydrogenation in the carbon and hydrogen reaction, the catalytic hydrogenation is going to reduce only this one.*

As seen in the previous exchange, Luke could easily use the tools of the community, such as organic reactions, to propose a synthesis route to the target compound. However, he explained during interview 1 that he did not begin his Ph.D. training with the ability to use this tool.

> *Yeah that is the beauty of the retrosynthetic analysis. You need to have broad knowledge on the synthetic reactions. Which kind of reactions are going to work, which will not work, you know. Choosing based on the existing you know literature knowledge. At the beginning you know before I start the research I did not have a good idea about the retrosynthetic analysis.*

His previous quote suggests that through legitimate participation in the organic CoP, he developed the dexterity and competency to use the tools of this community, such as organic reactions, to perform the retrosynthetic analysis and propose synthesis protocols. Although he proposed a synthesis pathway, he explained that he still needed to search SciFinder and simultaneously be thinking/performing the retrosynthetic approach in his mind.

> *Basically before starting the search, the first thing is always go in, and try to find this compound in the SciFinder and at the same time you can think about…this is a thinking process. If you have to use the SciFinder you need to stay out of certain places and think about certain structure is going to have certain reaction*

and those kinds of things. But if you have some molecule in your mind that you need to synthesize you can think those kinds of things anyways.

The previous quote suggests his ability to deduce the reactivity of the target compound based on the structural features, and, therefore, he could use that knowledge to "*stay out of certain place*" in SciFinder. This was indicative of his ability to link structural features of the target compound to certain chemical properties, which for a practicing organic chemist relates to the reactivity. As such, his comments suggest that he was able to deduce the implicit information from the explicit structural representation of the target compound.

With the reactivity of the target compound in mind, Luke input the exact structural representation of the target compound and performed an SE search (**Step 2**). As he was in the process of performing the SE search, he explained how he learned to use the database:

Luke: Like in the beginning when I start grad school research, I did not know what SciFinder was. And I have not used SciFinder before I joined [my Ph.D. research group]. And then once I joined the [Ph.D. research group], I heard like we can get the reference and those kinds of things from SciFinder. Then I talked with some of my seniors and I asked my senior how to use SciFinder, how to open the SciFinder, and those kinds of things. And my friends helped me to open the account and then he draw one molecule, and he said, 'to go try by myself.' And then I start to draw very small molecules like ethanol, benzene, all those kinds of things.

I: Why were you drawing those?

Luke: Because those are the common compounds that I use for the reactions in my lab. Those are some of the common compounds that I am using in the lab. It means that those compounds are commercially available and just like I was thinking if the compounds are commercially available, who are selling the compounds? And during that process, you know I start to, when I use the SciFinder I start to use different options in the beginning of SciFinder.

I: Different what?

Luke: Different options, different substance type. When you go for substance type, these [exact structure, substructure, and similarity] are the three different types. So the exact structure should be the same structure, substructure means maybe that is a part of the complete structure. And similarity means that should be some similarity in the molecule or those kinds of things. I use my own judgment to figure out which will be the best options to go and search. I mean this is just what it is called intuition or something like that you know just trying. Hit and trial method. Like if you go this way, how it will go? Because when we are starting the synthesis of any compound, you need to make sure that you know all possible sites need to be open. Just like if it is does not come by one way you need to go by another way.

In absence of formal training, Luke's comments in this exchange indicate that he used his fellow group members for initial instruction in using SciFinder; not directions produced by the creators of the search engine or some other internet- or library-based reference. Additionally, frequently using the database for his research helped Luke construct knowledge regarding each search option (SE, SS, and SSi).

When the SE search yielded no results (**Step 2A**), Luke proceeded to do an SS search using the given structural representation of the target compound (**Step 3**). The SS search yielded 58 substances (**Step 3A**), however, he quickly stated that he did not know if they would be useful for the synthesis of the target compound.

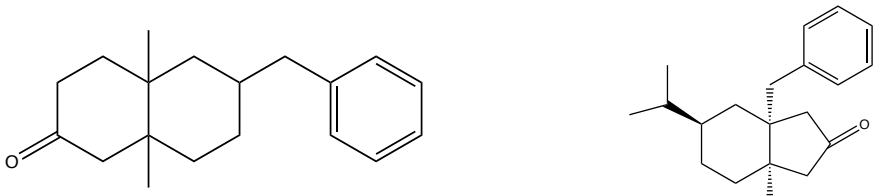
Luke: So I may not necessarily need to use these protocols but I want to keep this, record it somewhere in my textbook or notebook so I don't need to go there again and again and find the same thing.

I: So you would write this down and you would just write it down for what reason?

Luke: Because these are the two protocols that we might can adopt for our methodology.

Although he would note hit #1 and #2 (substances from the SS search) as potential options, he continued to explore his other search option and performed an SSi search for the structural representation of the target compound (**Step 4**). He clicked to retrieve the 14 substances with an 85-89% similarity match to the target compound (**Step 4A**). He then decided to retrieve the references, followed by the full text for substance #1 (**Steps 4B-4C**), which he justified in the following quote:

> *So now I'll go and take a look at the reference because if I can adopt this methodology, provided if this reference has used isopropyl magnesium bromide or isopropyl Grignard reagent for you know adding that substituent there. We have alternated for this one, just like we can use the benzyl Grignard reagent there. So one of the thing is the position is on the right position, 2 position and they have different substituents that we have and I am going to go and find out how they put the substituents there and generally adopt similar methodology for our reaction system or not.*



(a) (b)

Figure 4.68: Target compound (a) and substance from SSi search (b)

Based on the structural features of substance in Figure 4.68b, specifically the comparable functional groups, he thought that he could potentially adopt the methodology. For this reason, he pursued the article to further investigate the researcher's methodology to determine if they reported using a Grignard reagent.

Once the article was retrieved, he scrolled straight to the schemes reporting substance #2 (**Step 4D**). He justified this action by stating:

Luke: Because the scheme will give you the visual presentation of the reaction that they did there. And try to figure out how they used their methodology for their reaction.

I: So do you remember how you learned to focus?

Luke: Basically at the beginning you know I used to go through all the papers and later on I realized like whatever was writing in the discussion part will be the part of their scheme. They will just describe the scheme on the narrative way. And later on you know I just decided to go through the scheme.

Luke's approach to evaluating research articles had evolved as a result of being exposed to the literature. He now relied on the pictorial representations of the reactions in the schemes, which was a more efficient strategy to assess the chemistry reported in the articles. In this case, Luke was able to quickly evaluate the reaction schemes to determine how he would modify the procedure for the purposes of synthesizing the target compound. He stated:

Just like here [on the target compound] we have to put the benzyl group and they start out with isopropyl group here and they end up with isopropyl group. Now my concern will be if we can go and make the compound that has the benzyl group in the starting material or not. The other part is we need the methyl group here, right. The methyl group and they put the ethyl group here. So what they did is they did, they used the ethyl bromide lithium, THF, and do the reaction on that way. So instead of using the ethyl bromide we use have to use the methyl bromide because we need the methyl bromide. So we are just modifying the procedure. And the important part is they are making the 5-membered ring and we have to make the 6-membered ring. So this is the most challenging stuff on this thing.

Although Luke concluded that he could modify the protocol for the purposes of synthesizing the target compound, he finally concluded that this would not be an advantageous method to adopt because the protocol listed 7 steps, and the first step reported a 70% product yield (**Step 4E**). Therefore, he determined that adopting this method would not be the best choice.

> *We need to figure out some other effective method that will give a larger amount of the product in a shorter time by using cheaper reagent. So it needs to be an efficient process. So I am going to go through that one. So this you reference, what I know is we can adopt this methodology for the synthesis of this compound but this is not the best choice.*

His trial and error approach to performing a literature search was evident in his search strategy for which he was exploring all his possible search options for the structural representation of the target compound. Although he would note the previous article, his chemical intuition indicated that he could find a better, more efficient protocol. Since he had exhausted all his search options for the structural representation of the target compound, he decided to continue his search by modifying the structural representation of the target compound. As seen in Figure 4.63, he removed the benzyl group and inserted a bromine group (halogen) and performed a SE search (**Step 5**). He justified this action by stating:

> *And then I will go and do the retrosynthetic analysis on the molecule. I put halogen there, because if there is halogen, we can kick that bond with the Grignard reagent and carryout the synthesis of that reaction…that is we can yield this benzyl magnesium bromide to kick out the halogen and put the substituent there.*

Once again, Luke's prior knowledge of chemical reactions allowed him to know which parts of the target compound he could modify to add the benzyl group. Furthermore, he was concurrently using the tools of this community, such as organic reactions, the retrosynthetic analysis, and SciFinder, as a vehicle to develop a research protocol for the target compound.

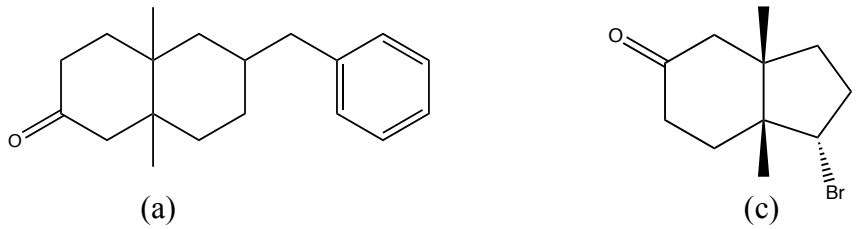
He received no results from the SE search for the bromine compound (**Step 5A**), so he removed the bromine, inserted an iodine group and performed an SE search (**Step**

6). The SE search yielded no results (**Step 6A**), so he removed the iodine group, inserted a chlorine group and performed SE search (**Step 7**). Once again he received no results (**Step 7A**), so he then performed an SS search for the chlorine compound but received no results (**Steps 8-8A**). Following his similar pattern, he went back to the structure editor screen and removed the chlorine group and inserted a bromine group. Then, he performed a SS search (**Step 9**) and finally obtained one result for the bromine compound (**Step 9A**); however, after he clicked to further evaluate the structure, he stated:

> *So this substructure is complex but it shows that this is the cyclohexanone moiety...*

Because he concluded the substance from the previous SS search was too complex, he returned to the structure editor screen and changed the bromine to an iodine group and performed an SSi search (**Step 10**). He then retrieved one substance with a 90-94% and six substances with an 85-89% similarity match to the modified target compound. As seen in **Step 10A**, he disregarded substances #1-3 because the halogens were in the wrong location for the purposes of synthesizing the target compound. He proceeded to evaluate the remaining substances and explained:

> *Luke: We need two cyclic rings and one of the choice might be this [substance c* in Figure 4.69*].*
>
> *I: Why?*
>
> *Luke: Because this one is the cyclohexane ring with cyclopentene it will have the methyl on both sides and you can go here [get reference] and you can just like note down this reference for your future correspondence too.*



(a) (c)

Figure 4.69: Target compound (a) and substance #4 from the SSi search (c)

Although he stated that substance #4 (Figure 4.69c) was potentially be a viable method to adopt because of the bridge methyl groups and positioning of the bromine group, he did not pursue the full text.

At this point in his search, he returned to his initial thoughts about the Aldol condensation reaction. As he was drawing the structure in **Step 11**, he explained:

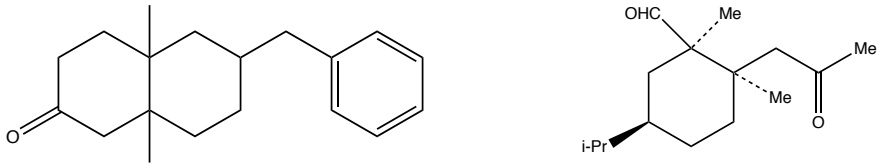
> *And I am going to go and try to figure out this kind of system. So what we can do is we can deprotonate this hydrogen and make the enolate. And this enolate is going to kick out this one, or I think that is not the best option because you do need the carbonyl on this carbon. If you do need the carbon on this one, I mean if the enolate is on this carbon it will go into the alcohol. So that will be going into the alcohol and oxidizing it back into the ketone. Let's go and find this kind of system where you have the ketone here. So you are going to make the enolate here and that is going to reach here to make the carbonyl group. Carbonyl group will convert into the alcohol.*

Again, Luke was justifying his decision to modify the target compound based on the chemistry for which he could perform. Once again, during this instance of his search, he was applying his knowledge of the Aldol condensation reaction, which was based on his initial interpretation of the structural representation of the target compound.

As seen in **Step 11** in the search chronology, he performed multiple searches (SE, SS, and SSi searches) for the modified target compound in Figure 4.66. After several

unsuccessful searches, he finally retrieved four substances with an 80-84% similarity match to the compound as seen in **Step 11A**. During this time, he stated:

> *And we had just like an isopropyl group instead of that one we need to put the benzyl group there. The other framework is quite similar like if you go for this one [substance #1 from Figure 4.67] or you go for this one [substance #2 from Figure 4.67]. Because you do need that two methyl group on the bridge carbon. You are going to have two methyl group and you are going to have just like you need to make six member ring. So if you go and deprotonate this one on this carbonyl that is cyclized. So one, two, three, four, five, six membered ring. So you need to figure out if this ring will be formed or not.*



(a) (d)

Figure 4.70: Target compound (a) and substance #1 from the SSi search (d)

He was able to find a structure similar to the one for which he searched, so now his goal was to determine if the protocol reported the cyclization and formation of a six-membered ring. For this reason, he pursued the full text for substance #1 as seen in Figure 4.70d (**Step 11B**). Once the article was retrieved, he quickly scrolled to the schemes and concluded that the author from the previous article he retrieved in **Step 4C** utilized the same concept. Once the full text was retrieved, he elaborated on his thoughts about the method by stating:

> *Luke: If we didn't find a better method than this one, this would be the best method to try.*
>
> *I: So this would be the one you would try?*

Luke: Right. But usually in grad school life, usually when we come up some problem, and if the problem can be solved in one or two day, you know 24-hour problem, or 48-hour problem, I usually go and try by myself.

I: So you would actually take?

Luke: Try.

I: And you would actually try this?

Luke: But if the problem will take longer step, a lot of reagents, you know longer step means of course it is going to take a lot of time, right? Then I am going to write the scheme, you know pen and paper and I'll go for the scheme with my group members' discussion and I going to take help from them. You know if any of those steps are not feasible when I try a reaction or not. And sometimes doing the research, the single idea does not work all the time. You need to go to some other people, and other people will you know think the same thing on a different way. And you may use his idea and your idea together, put both of those ideas together and from that one you may come up with completely new idea. And that is how we used to do in our group meeting also. And sometime we going to have individual meeting with boss, and on that one, just like once we agree within a group we go and talk with boss and whether if he is going to accept our idea or reject our idea. And if he rejects our idea, then we are going to ask, 'what are the reasons for rejecting our idea?' Probably sometimes the experience on the lab and you know experience in the knowledge will make all of the difference on those kinds of things.

Being that Luke was in his sixth year in an organic research group, he knew there were various ways to approach a synthesis task. From his Ph.D. experience, he knew that members of his research group/organic CoP might suggest other methods and/or reactions that would enable him to synthesize the target compound. For that reason, he would also discuss his approach with his research group, not necessarily to approve his method, but as a way to obtain constructive feedback regarding his method and/or alternative routes. His openness to feedback was something that he learned by being immersed in an organic research group and by performing authentic research. Then he further elaborated about his openness to constructive feedback, by stating:

> *Long story short just like when you do the research, there is not a single way to you know gather the knowledge or something like that. When you are chatting with your friends you can go for those kinds of things. You can Google some of the things. Not only the SciFinder, some of best answers you can find in the Google. Like people may have some problem and they put the problem in Google and then you know the other guys you know comment on that one and they will suggest some of the literature paper or some of the things.*

Although Luke was capable of independently developing a research protocol, at this stage in his training he still saw the process of developing a synthesis protocol as a collaborative effort. Every opportunity, Internet searches and interactions with other members of the organic CoP, was a chance to obtain feedback regarding his synthesis protocol, specifically with respect to the efficiency of the method.

To conclude the Task 1 search, he sketched his ideas and explained:

> *If you cyclize that one just like, if you cyclize this one just by LDA that is the base usually used to make the enolate and we are going to end up with this kind of compound [step 1 in* Figure 4.67*]. Then if you go and treat with acid that will give you this compound [step 2 in* Figure 4.67*]. And again if you go the hydrogenation you are going to selectively reduce the double bond [step 3 in* Figure 4.67*], right. And you can do the same protocol on the substrate. Now if we have this compound then we can make this one. So this substituent will be already there this methyl group will be here all the time.*

At this point in the interview, he determined that he could modify the protocol from the article and he had all the information he needed to attempt the synthesis for the target compound. However, Luke did not explicitly explain how he would modify the protocol so that he would synthesize the target compound. (Note: The final synthesis that he sketched would be one carbon short.)

Summary of Luke's Search

Luke's search strategy reflected the development that would be expected of a sixth-year graduate student. From his experiences performing his Ph.D. research, he had

constructed knowledge regarding how to use SciFinder as a resource to develop a research protocol. Furthermore, the Task 1 search revealed that the retrosynthetic analysis in conjunction with named organic reactions were now tools for which he could use to propose a research protocol for the synthesis of the target compound. With respect to his domain-specific content knowledge, Luke used his knowledge of organic structures to judge the parts of his target molecule that were essential to preserve in a substructure search or in evaluating similarity matches versus those moieties that could be more variable in their structures. Although he was able to independently propose a research protocol by using the database and his organic knowledge, he would still use his colleagues as an additional resource to provide constructive feedback about other, potentially more feasible methods.

CHAPTER FIVE

CONCLUSION AND IMPLICATIONS

I will begin this chapter by presenting the major themes that resulted from the cross-case analysis. I will then attempt to address my guiding research questions and conclude this chapter by discussing the implications for my study.

Cross-Case Analysis

Consistent with the research in the literature review the data showed that the mere process of performing a search of the primary, scholarly literature in a field – successfully or not – is, cognitively, a highly demanding one. The participants in my study needed to draw and interpret several representations of complex organic structures so that they could input *and* interpret the results of their queries; based on those understandings they had to decide whether or not to pursue any of the hits; and at some point they had to read titles, abstracts, and bodies of multiple papers and evaluate the extent to which any of the information may have been relevant to their current task. In cases where they found a promising article most of the participants chose to consider experimental viability of the protocol in making the final determination of whether to retain or discard the hit, realizing that discarding meant going through the bulk of the process for yet another cycle. With respect to one's domain knowledge, this act of searching the literature may be, therefore, one of the most comprehensive tasks an individual may perform.

Another prevalent characteristic of the participants' searches seems to be that the greater challenge in performing literature searches may be the process of *excluding*

information more so than the process of actually finding it. Considering the many and varied circumstances in which the participants would get dozens, if not hundreds or thousands, of hits in response to a single query. In each of these situations, the participants had to make relatively quick decisions regarding the number of those hits to review, acknowledging the possibility that the best hit could have been embedded in the middle of those scores of hits. Aside from the relevance to this study, this second observation shows how the challenge of our so-called "Information Age" is more about evaluating information than it is just about finding it. In the remaining of this section, I will discuss some of the most important themes that I believe emerged from my data.

The data indicated the interrelation between the participants' discipline-specific content knowledge, their CoP, and the ways in which they utilized SciFinder. To elaborate on this theme, I will begin by discussing the relationship between their discipline-specific content knowledge and their usage of the database. Then I will describe how their CoP—a social learning environment—influenced their ISB.

Although each participant would began the Task 1 SciFinder search by performing a search for the exact structural representation of the target compound, the participants displayed distinct ISB that were directly related to their content knowledge. For example, consider their ability to use organic reactions and retrosynthetic analysis (tools of this CoP).

Each participant began by performing a search for the structural representation of the target compound. Regardless of their stage in their Ph.D. training, this is a practical

first step for any practicing organic chemist or, furthermore, any individual beginning their research endeavors. In this way, the participants were using the database to access the organic chemistry CoP to establish if the target compound was reported. More specifically, Bruce and Luke (a sixth-year graduate students) viewed this action as a way to lessen the cognitive load of the Task. Accordingly, they would not use the tools of the CoP, specifically organic reactions and the retrosynthetic analysis, to develop a research protocol until they had established that the target compound was unreported.

Furthermore, the initial stages of their searches provided insight into the participants' ontological beliefs. Following their first search, each participant, with the exception of Jane, was able to immediately conclude that the target compound was unreported. Despite that SciFinder did not retrieve the target compound after Jane's first search, she continued to perform several searches for the compound and was surprised when she could not find precedence for it in the literature. Evident in Jane's case, her "chemical intuition" led her to believe that the target compound *had to have been* reported somewhere in the literature. Comparatively, the other participants' behaviors suggest their ontological beliefs, which were if SciFinder did not quickly retrieve the compound; then it did not exist in the literature.

Once the participants verified that the target compound was unreported, they then began to deviate and use the database in different ways. The newcomers to the organic CoP, Peter and Jane (second-year students) were heavily scaffolded by the database, and used it as a vehicle that enabled them to propose a research protocol. Sophie (third-year Ph.D. student) utilized the database in a comparable way; however, she used the database

and tried to use her domain-specific content knowledge to find precedence for the chemistry she proposed. Whereas, the old-timers Luke and Bruce (sixth-year students) were able to use their discipline-specific content knowledge (named organic reactions and retrosynthetic analysis) to propose a synthesis protocol for the target compound, and were, subsequently, using the database to support their ideas. In the following paragraphs I will first discuss the second- and third-years ISB, and then elaborate on the sixth-year ISB.

Jane, Peter, and Sophie used the database as a scaffold that enabled them to develop a research protocol, or in Sophie's case at least compile a series of potential reactions, to potentially synthesize the target compound. So when they could not find the target from the database, they developed a reductionist strategy to search for compounds that closely resembled the target compound. In this case they kept removing components of the target, such as functional groups or substituents, and inputting the resultant compound into the database. Although searching for substances that have comparable structural features (such as similar functional groups and substituents) could be a useful strategy when developing a synthesis protocol, they were frequently evaluating the output of results (reactions/transformations) without considering how they could add back on the "removed" parts of the target compound.

As an example, they input the target compound without the benzyl group and were able to find a method to make "most" of the target compound. Redirecting their attention to the addition of the benzyl group, they then needed to find a method to transformation the simplified compound into the target compound. They each tried to

search for the aforementioned transformation but obtained no results, indicative that the transformation to the target compound was not easy or practical using the simplified compound. As such, they continued with their reductionist strategy and simplified the complexity of the structures they input to find precedence for the addition of the benzyl group. By reducing the complexity of the substances they input, Jane and Peter were able to find a method for which they believed they could add the benzyl group. Although Jane and Peter's ISB did not indicate that they could perform the retrosynthetic analysis on the target compound, their approach to using the database suggest they could use it as a resource to think retrosynthetically.

The data also indicated that Peter and Jane were able to propose a synthesis protocol for the target compound, but they struggled to justify their protocols based on their chemical understanding. For instance, they mentioned performing a nucleophilic substitution reaction to add the benzyl group, however, their search actions did not reflect their ideas. Furthermore, Jane would notice named organic reactions but she could not use her knowledge of those reactions to propose any starting materials that would undergo the Diels-Alder reaction to yield the target compound. This suggests that the second year graduate students had not reached a phase for which these tools were operational. Similar to this finding, Bhattacharyya and Bodner (2005) suggested that students who are in the earlier stages of their Ph.D. training could not effectively use the tools of the community to solve problems. This finding was furthermore reiterated as Peter and Jane explained their justifications for their final synthesis protocols, which were based on the frequency of methods that were reported in the literature, such as the

number of transformations reporting the synthesis of a specific compound, and their chemical intuition. This suggests that students early in their Ph.D. training can generate a solution without having a deeper conceptual understanding of their solutions (Bhattacharyya & Bodner, 2005). Their actions provide insight into the type of knowledge, procedural or conceptual, they had constructed. According to Hatano and Inagaki (1984), conceptual knowledge enables the individual to give, "meaning to each step of the skill and provides criteria for selection of possible alternatives for each step within the procedure" (p. 28). Whereas, procedural knowledge encompasses the, "decision rules as well as executive strategies, along with the skills necessary for applying the knowledge" (p. 28), Jane and Peter appeared to not have a meaningful understanding of these procedures/protocols. As such, the tools of this community cannot become fully operational and adaptable for other problem-solving situations until they developed a meaningful understanding of the chemical principles and concepts for which the reactions are based.

Comparatively, throughout the Task 1 search, Sophie was trying to explain each step of her method development process based on her conceptual understanding, but was unable to find precedence in the literature to support her ideas. Like Jane and Peter, Sophie had a method to make the "backbone" structure, so she continued to search for a method to add the benzyl group. During this time, her ISB reflected her attempts to support her ideas regarding the chemistry she could perform to add the benzyl group. However, she was unable to find a method to add the benzyl group using the structure search tools. In an attempt to find precedence for her ideas in the literature, she

performed a topic search, but was unable to use the information (specifically a ring-closure step) for the purposes of synthesizing the target compound. Although she did not find the information she was searching for, Sophie was trying to use the topic search tool as a scaffold to bridge the problem-solving gap. Sophie's actions were particularly interesting because she relied on SciFinder, a more reliable resource, to find chemical information, whereas Jane would use Google or Wikipedia to find chemical information.

Despite that Sophie encountered several roadblocks indicating that the addition of the benzyl group to the hydrocarbon backbone structure was not an advantageous transformation, her chemical intuition suggested otherwise. As such, she seemed to have developed tunnel vision and was initially reluctant to deviate from her thoughts of performing a substitution reaction to add the benzyl group. However, the repeatedly encountered roadblocks forced her to consider alternative methods. Therefore, she searched for a method to perform a ring closure step, but she was unable to transfer that information from the article for the purposes of developing a research protocol for the target compound. Her ISB suggests she was attempting to use her content knowledge, but was unable to find precedence in the literature for the chemistry she proposed. Her Task 1 search actions and justifications suggest she was beginning to construct conceptual knowledge but it was not transferable at this stage in her Ph.D. training.

In contrast, Luke and Bruce's ISB displayed the development and sophistication that is expected of sixth-year graduate students. Their six-years of experience performing organic research was evident by their ISB for which they were able to use the tools of this community to propose a synthesis protocol for the target compound. At this stage in their

Ph.D. training, they used the literature (accessed via SciFinder) to substantiate their synthesis protocol and/or to find a method to synthesis the proposed starting materials. Although Luke performed the retrosynthetic analysis on the target compound, he stated that performing a literature search was a cognitive process and he should use his knowledge in conjunction with the database to develop a research protocol. Accordingly, their Task 1 search behaviors indicate that their domain-specific content knowledge had become a tool for which they could actively use to not only propose a research protocol but also evaluate the SciFinder results (Brown, et al., 1989) to find applicable protocols. As such, the results indicate that the sixth-year graduate students had constructed conceptual knowledge that was transferable to other problem-solving situations.

The data also indicates that the sixth-year graduate students had developed multi-variate thinking strategies (Kraft, Strickland, & Bhattacharyya, 2010) whereas the second- and third-years' actions suggested they had not. The sixth-year graduate students seemed to approach the Task 1 compound by processing the chemical system as a whole and were able to easily propose a synthesis protocol. Comparatively, Jane, Peter, and Sophie's actions did not suggest that they were able to think about the complexity of the system, i.e. the target compound, as a whole entity. Guided by their chemical intuitions, they arbitrarily removed pieces of the target compound without a clear strategy for adding the piece in a subsequent step, proving to be a very difficult step. As such, the data suggests the necessity to train students to develop multi-variate thinking strategies, where they not only consider the functional groups, but also the location and the chemistry that can be performed on that target compound.

Also depicted by their ISB was the influence of the social learning that occurs as students engage in the authentic practice of the CoP. **The cross-case analysis indicated that they had learned the value of using the database from being immersed in the organic CoP, and over-time became more proficient with performing a search.** Bruce and Luke had no experience with the database before they began their Ph.D. training; however, being immersed in their research group, they quickly learned that SciFinder was a powerful resource for conducting their research. For this reason, they asked their group members to show them how to access and use the database. Utilizing SciFinder for their research, they eventually became accustomed to using the database, thereby developing an understanding of the search tools. Comparatively, Jane, Peter, and Sophie had experiences using SciFinder prior to their Ph.D. training. Although Sophie had prior experiences using SciFinder, interactions with her Ph.D. research advisor had influenced her ISB, specifically her "broad strategy" to perform searches. Despite Jane's undergraduate SciFinder training, she still was unaware of the information that the database provides, such as the commercially available flask. As a result, Jane, like Bruce, continued to search for commercially available information that was readily displayed underneath the compounds. This is indicative that students can begin to learn how to use the database from their research group members but would still benefit from formal training.

The data also provided insight into how they make decisions regarding structural representations. **Throughout the Task 1 search, the participants used the structure tools (both reaction and substance) to: 1) perform a search, 2) evaluate the**

usefulness of substances or transformations, and 3) evaluate the usefulness of articles/research protocols. Although the participants were using pictorial representations throughout the Task 1 search, the results indicated that the second- and third-year students were focusing on the structural changes without a deeper chemical understanding of the transformations and/or structural features. Furthermore, they would disregard chemical information, such as complex structures, for which they could not comprehend. Jane, Peter, and Sophie's tendency to disregard chemical information/substances that appeared "complex" was primarily a result of their inexperience within the organic CoP. For instance, Jane quickly disregarded a method that reported a lithiumdialkylcuprate starting material and Peter and Sophie disregarded a titanium complex, which are common substances within the organic CoP. Once again this is an issue of the newcomers relying too much on their experience. Furthermore, their actions suggests that they were frequently unable to deduce the implicit information indicated by the explicit structural representations (Bowen & Bodner, 1991). These results were also evident in Strickland et al. (2010) for which the researchers concluded that the organic chemistry graduate students lacked the representational competence that was expected of practicing organic chemists.

Comparatively, Luke had over six years of experience performing organic research and was capable of deducing the chemical information from the structural representation of the substances. In a similar way, Bruce was evaluating a scheme in an article he retrieved but was unable to understand the reaction mechanism. Instead of disregarding the information, he proceeded to use the arrow-pushing formulism as a way

to conceptualize how the bonds were constructed. As a result of increased immersion in the organic CoP, their knowledge had become more than just abstract concepts, but a tool that the sixth-year graduate students used to understand organic transformations and substances (Brown et al., 1989). As such, the previous theme is furthermore linked to the need to deviate from the surface-level approach and to facilitate the development multi-variant thinking skills.

Also related to their increased exposure to their research activities, the participants had developed heuristics to narrow down their search to find practical experimental protocols to execute in their labs (Maeyer and Talanquer, 2013). (Luke did not perform reaction searches, so he was not included within this section.) Jane, Peter, and Sophie had developed heuristics—intended to lessen the cognitive load of a task—to evaluate the search results. For instance, they would first look at the percent yield of the reaction, and in many cases disregard a reaction if the transformation did not report a high product yield. Then, they would evaluate the starting materials and/or reagents to find "simple" and/or commercially available substances. Bruce (sixth-year) had developed a similar heuristics, but seemed to be more open to consider protocols with low yields. For instance, Bruce pursued a reaction with a 5% product yield because he knew he would have to modify the protocol, whereas Peter did *not* pursue the same article because of the low yield. Although this heuristic lessens the cognitive load and enables them to narrow down their search, this strategy could potentially cause them to miss information that could be incorporated into their research protocols. Despite the

potential drawbacks of these heuristic, the participants were able to quickly evaluate the search results to find practical, more efficient methods.

Furthermore, several participants stated that their approach to evaluating a research articles had evolved since they began their Ph.D. training. For instance, in the beginning of their training they would try to read the entire article; however, their strategy had evolved because they realized the schemes/pictures were indicative of the chemistry that was discussed in the written sections. They attributed this change to their experiences performing organic research, which as a result, caused them have more exposure to the primary literature.

In addition, the second- and third-year participants had a tendency to emphasize the structural features and changes, and to not thoroughly evaluate the information regarding the journal article. On several instances, Jane and Sophie's would decide to retrieve an article, and realize after it was retrieve, that they had previously accessed the same article. Their actions suggested that they were not evaluating the journal reference. Interestingly, over the course of the interview, Jane had adapted her behavior to read the information that pertained to the journal before she decided to retrieve the full text. Furthermore, Peter did not discuss the journal reference until after he had retrieved the article. Their actions suggest they did not evaluate the quality of the journal, which is intrinsically linked to quality of the data published.

Also related to their exposure and experience within the organic CoP, was that the participants were utilizing their organic CoP in distinct ways. The newcomers (Jane, Peter, and Sophie) would, not surprisingly, seek the approval and/or

help from their advisors regarding their proposed synthesis pathway. This suggests they were relying on predominantly more knowledgeable others for additional scaffolding to develop a synthesis method. Luke, however, was utilizing the CoP in a different way. To conclude his Task 1 search, he stated that he would seek feedback from his peers and colleagues regarding his synthesis proposal. Being a sixth-year graduate student, he felt confident that his protocol would work, *but* he still saw the value of discussing his method with his group members. This indicates his epistemic development for which his advisor and fellow research group members were seen as colleagues instead of simply sources of information. According to Wenger (2000), members of a CoP can engage by negotiating meaning, which suggest he had transitioned into a full member of the organic CoP. This comparison suggests that students who have not been conducting organic research need more support to bridge this problem-solving gap so they can transition into a full participating member of the CoP.

Addressing the Guiding Research Questions

In this section, I will answer the two research questions that guided my study. The questions will be italicized and bulleted, and my answers will follow.

- *How do synthetic organic chemistry graduate students conduct a literature search using SciFinder to find research protocols?*
- *What are synthetic organic chemistry graduate students' sense-making procedures for the different parts of the literature search?*

The graduate students displayed a continuum ISB and sense-making procedures that were directly influenced by their domain-specific content knowledge and their

exposure to the organic CoP. With respect to their sense-making procedures, the newcomers had developed procedural knowledge that was less adaptable for this task. For this reason, they relied on the database and struggled to explain their chemical reasoning for their proposed protocol. In contrast, the old-timers had developed conceptual knowledge that was adaptable and transferable for the purposes of synthesizing the target compound. As such, they could use the tools of this community to not only develop a research protocol, but they could explain their chemical reasoning for each step of their protocol.

Exposure to the organic CoP conducting authentic research influenced their ISB. Specifically, interactions with their research group influenced their perception of the utility of the database, and their approach to performing a literature search. For instance, exposure to authentic research caused them to develop heuristics, which enabled them to quickly evaluate their searches to find more efficient protocols. Furthermore, the newcomers relied on their more knowledgeable others to approve their protocols, whereas the old-timers were confident that their protocols would work.

Implications

In this section, I will discuss the implications for my study that are relevant to both researchers and practitioners across all disciplines, particularly the chemistry education community.

Implication: Educators need to be aware that performing a literature search is a complex, cognitively demanding task. The results from this study indicate that performing a literature search is a cognitively demanding task. Performing a literature

search requires students to use the tools of the community through each stage of the search.

Research has shown that following instruction students can learn how to utilize the database (Ferrer-Vinent, 2012). Therefore, instructors can begin to incorporate assignments that integrate students' domain-specific content knowledge and the database. For undergraduate instruction, professors should try to use the SciFinder tutorials to guide students through the process of not only how to use the database but also to link their domain-specific content knowledge to their literature search to develop a protocol. Used as a teaching resource, students can learn the heuristic of the culture from different perspectives and can gain exposure to scientific literature.

Implication: Graduate students from the various chemistry subdivisions would benefit from undergoing a SciFinder training session during their first-semester of their Ph.D. training. Previous studies suggest that chemistry graduate students, especially those graduating from ACS-accredited undergraduate institutions, are expected to have learned information seeking skills during their undergraduate degrees (American Chemical Society Committee on Professional Training, 2015). However, even following undergraduate SciFinder instruction, students deciding to pursue a graduate degree may not have developed information seeking skills prior to beginning their graduate training. Although Ph.D. student are expected to be knowledge producers and to contribute scholarly research to their respective fields, this is a challenge task especially if they have not developed the skills to find and interpret chemical information (Feldman, Divoll, & Rogan-Klyve, 2009).

Although my results indicate that prolonged exposure to authentic research will gradually enable the graduate students to more easily use the tools of the community to accomplish their research activities, students frequently do not know what resources are available or how to use the database once they begin their Ph.D. training. Moreover, graduate school can be a competitive environment that makes student reluctant to seek help (Kuruppu & Gruber, 2006). For this reason, graduate students across all disciplines would benefit from information-seeking training their first semester of graduate school.

Additionally, advisors need to provide additional scaffolding during the early stages of their Ph.D. training because some graduate students may not be able to use the tools of the community to develop a research protocol. To make the transition more fluid, tasks such as the synthesis proposal used for the class assignment in Bhattacharyya and Bodner (2014) could allow students to more easily use the tools within the research environment.

Implication: Students should be encouraged to construct conceptual knowledge. Deeper, more meaningful construction of conceptual knowledge will thereby enable the students to use the tools of this community to not only solve problems, but to justify their solutions. To facilitate the construction of conceptual knowledge, Hatano and Inagaki (1984) suggested that students have a framework of prior knowledge to build upon and learn to apply the skill from a variety of perspectives. As such, educators and practitioners should first teach the chemical principles and theories. Then, they need to encourage, not inhibit failures or errors, students to experiment with various variables and to manipulate the constraints so they can observe the outcomes. During the latter years of

their undergraduate training and the first year of the Ph.D. training, students should be asked to articulate the underlying reasoning behind their knowledge, which educators can act as a scaffold to assist the students as they construct conceptual knowledge that can be transferred and adapted to new problem-solving situations.

Furthermore, my study provided a peripheral perspective into the "spontaneous" construction of knowledge that occurs through the accumulated experiences as one engages in the authentic practices of the CoP. At this stage in their Ph.D. training, the second-year and third-year students should be asked to articulate the processes guiding their procedural knowledge. Once they are able to articulate the underlying reasoning with which the processes are guided, they can begin to develop and construct their meaningful conceptual knowledge.

Implication: Instructors and professors need to encourage students to become more metacognitive about their search behaviors. Although designing activities that encourage the development of metacognitive skills are important, educators and advisors need to emphasize the importance of spending more time in the early stages developing research protocols. Students should understand that spending time searching SciFinder to develop a research protocol is an extremely important aspect of their research activities, and, more importantly, developing a synthesis protocol based on the underlying chemistry can save you days or even months in the laboratory.

Implication: We need to develop authentic activities that encourage and necessitate the actions of practicing organic chemists. Practitioners need to design activities that require students to not only solve the problem but to consider the

practicality of their experimental protocols and the execution thereof in the lab. Furthermore, practitioners should emphasize the importance of developing heuristics that are based on the practices of synthetic organic chemists. Synthesis assignments should teach the students the nature of organic chemistry, which is that students cannot divorce the laboratory portion from the theoretical portion. As such, students need to realize the paper assignment is unequivocally linked to the laboratory feasibility. In this way, the activities should be evaluated based on the experimental feasibility of the methods, i.e. number of steps, complexity of the experimental set-up, and reaction conditions. From these assignments, students can begin to appreciate the true nature of the synthetic organic culture and develop the heuristic of the community.

Summary

My goal was to understand how graduate students search a database to find a research protocol for an unreported compound. The results of my study provide a deeper understanding into the dynamic ways in which a specific culture influences students abilities to use their discipline-specific content knowledge to navigate a database, SciFinder, to solve an authentic task. Findings from my study can broadly be integrated into the information science field to enhance and improve undergraduate and graduate students' ISB. More specifically, these findings can be applied to improve how we educate and train organic chemistry students (both at the undergraduate- and graduate-level).

APPENDICES

Appendix A

Understanding Chemical Research: Incorporating Theoretical Constructs into the Laboratory

The Interview Protocol

Background

- In what year of graduate school are you currently?
- What is your area of emphasis?
- Prior to coming here, what types of experiences did you have with scientific research? How long did those last?
- Before coming to Clemson, did you work in a scientific field? If so, can you please explain what you did and for how long?

Research

- Can you please explain the research that you are doing?
- How was (were) your current research project(s) developed?
- What resources do you use when you:
    - Set up an experiment?
    - Analyze and interpret data?
    - Troubleshoot experiments?
- Why do you use these particular resources?
- How do you think you have learned to do all of these activities?

SUBSEQUENT INTERVIEWS

- Can you please explain what you did in the laboratory this week?
- How did you decide to do the work that you did this week?
- Which experiments "worked"?
- What didn't "work" this week? Why do you think it didn't "work"?
- How have you gone about dealing with the things that didn't "work"?
- Overall, how would you rate this week?

Appendix B

Consent Form for Participation in a Research Study
Clemson University

Gautam Bhattacharyya and Lindsey Cain – Department of Chemistry

Understanding How Inorganic and Organic Graduate Students Navigate SciFinder

Purpose of Research
The purpose of this research is to explore how inorganic and organic chemistry graduate students conduct a literature search using SciFinder.

Specific Procedures to be Used
The research will consist of three audio-recorded interviews. The first interview will probe your research background, current research experiences, background using SciFinder, and your current experiences using SciFinder to conduct a literature search. During the next interview you will be given two molecules and asked to conduct a literature search using SciFinder to find a synthetic protocol. During the last interview, you will be asked to elaborate on some of your responses. The interviews will take place in your office or in a room of your choice in which your confidentiality can be protected. The interviews will be audio recorded for the sole purpose of the accurate transmission of the interview. Any written artifacts will be kept by the researcher for the sole purpose of data analysis. Federal regulations require all research records to be maintained for at least 3 years after the completion of the study. After this time, all tapes and written artifacts will be destroyed. You are 1 of approximately 12 participants in this study.

Duration of Participation
You are asked to participate in three interviews. Interviews are expected to last 30 to 45 minutes in length.

Benefits to the Individual
There will be no tangible benefits to **you as a participant**.

Risks to the Individual
The risks **to you, as a participant,** will be minimal. **You are free** to terminate **your** participation at any time during the interview. **This interview is not a test. The researcher is not concerned with your ability to correctly respond to questions. If, at any time, you feel uncomfortable, you are absolutely free to terminate your participation or skip a particular part of the interview without any penalty or risk to your standing in the Division of Inorganic or Organic Chemistry, the Department of Chemistry, or Clemson University. Volunteering to participate does not obligate you to the researcher or the research in any manner.**

Confidentiality

Data collected from the interviews will be kept confidential. A **pseudonym will be used for you** throughout this study and in the dissemination of the results. **Only the researchers** will know the identity of the student participants. The interviews will be tape recorded for the sole purpose of the accurate transmission of the interview. The tapes will be destroyed upon completion of the project. **The researchers will complete the transcriptions.**

Voluntary Nature of Participation

You do not have to participate in this research project. If you do agree to participate you can withdraw your participation at any time without penalty. **Furthermore, you may decline to answer or address any question or any set of questions. Declining to answer a question or withdrawing participation will, in no way, affect your standing in the Division of Inorganic or Organic Chemistry, the Department of Chemistry, or Clemson University. You are not obligated to the researcher or the research in any manner.**

Contact information

If you have any questions about this research project or if any problems arise, please contact the Co-Investigator, Lindsey Cain at: Department of Chemistry, Clemson University, Clemson, SC 29634; phone: 864.656.1593. If you have any questions or concerns about your rights as a research participant, please contact the Clemson University Institutional Review Board at 864.656.6460.

Consent

I have read this consent form and have been given the opportunity to ask questions. I give my consent to participate in this study.

Participant's signature: _____ Date: _____

Participant's Name: _____

Researcher's signature: _____ Date: _____

A copy of this consent form should be given to you.

REFERENCES

American Chemistry Society Committee on Professional Training. (2008). *Undergraduate Professional Education in Chemistry: ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs*. Retrieved from http://www.acs.org/content/dam/acsorg/about/governance/committees/training/acsapproved/degreeprogram/2008-acs-guidelines-for-bachelors-degree-programs.pdf.

American Chemistry Society Committee on Professional Training. (2015). *Undergraduate Professional Education in Chemistry: ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs*. Retrieved from http://www.acs.org/content/dam/acsorg/about/governance/committees/training/2015-acs-guidelines-for-bachelors-degree-programs.pdf.

Anderson, J. (2009). Learning the Language of Organic Chemistry: How do Students Develop Reaction Mechanism Problem-solving Skills? (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (3379296)

Association of College and Research Libraries. (2000). *Information Literacy Competency Standards for Higher Education*. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf.

Barab, S. A., & Hay, K. E. (2001). Doing science at the elbows of experts: Issues related to the science apprenticeship camp. *Journal of Research in Science Teaching*, *38*(1), 70–102.

Barrett, A. (2005). The Information-Seeking Habits of Graduate Student Researchers in the Humanities. *The Journal of Academic Librarianship, 31*(4), 324–331.

Bhattacharyya, G. (2004). *A recovering organic chemist's attempts at self-realization: How students learn to solve organic synthesis problems*. (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (3166598)

Bhattacharyya, G. (2006). Practitioner development in organic chemistry: how graduate students conceptualize organic acids. *Chemistry Education Research and Practice, 7*(4), 240-247.

Bhattacharyya, G. (2007). Ethnography and Ethnomethodology. In G. M. Bodner & M. Orgill (Eds.), *Theoretical frameworks in chemistry/science education.* (pp. 172-186). Upper Saddle River, NJ: Pearson Education, Inc.

Bhattacharyya, G. (2008). Who am I? What am I doing here? Professional identity and the epistemic development of organic chemists. *Chemistry Education Research and Practice, 9*, 84-92.

Bhattacharyya, G. (2014). Trials and tribulations: student approaches and difficulties with proposing mechanisms using the electron-pushing formalism. *Chemistry Education Research and Practice, 15*, 594–609.

Bhattacharyya, G., & Bodner, G. M. (2005). "It Gets Me to the Product": How Students Propose Organic Mechanisms. Journal of Chemical Education, 82(9), 1402–1407.

Bhattacharyya, G., & Bodner, G. M. (2014). Culturing reality: How organic chemistry graduate students develop into practitioners. *Journal of Research in Science Teaching, 51*(6), 694-713.

Bleicher, R. (1996). High school students learning science in university research laboratories. *Journal of Research in Science Teaching, 33*(10), 1115–1133.

Bodner, G. M. (2007). The Role of Theoretical Frameworks. In G. M. Bodner & M. Orgill (Eds.), *Theoretical frameworks in chemistry/science education.* (pp. 3-27). Upper Saddle River, NJ: Pearson Education, Inc.

Bowen, C. W. (1990). Representational systems used by graduate students while problem solving in organic synthesis. *Journal of Research in Science Teaching, 27*(4), 351-370.

Bowen, C. W. & Bodner, G. M. (1991). Problem-solving processes used by graduate students while solving tasks in organic synthesis. *International Journal of Science Education, 13*, 143-158.

Brahme, M., & Walters, L. (2010). While Technology Poses as the Great Equalizer, Distance Still Rules the Experience. *Journal of Library Administration, 50*, 484-514.

Brown, C. M. (1999). Information Literacy of Physical Science Graduate Students in the Information Age. *College & Research Libraries*, *60*(5), 426–438.

Brown, C. M. (2005). Where Do Molecular Biology Graduate Students Find Information? *Science & Technology Libraries*, *25*(3), 89-104.

Brown, J. S., Collins, A., Duguid, P. (1989). Situated Cognition and the culture of learning. *Educational Researcher. 18*(1), 32-42.

Cartrette, D. P., & Bodner, G. M. (2010). Non-mathematical problem solving in organic chemistry. *Journal of Research in Science Teaching*, *47*(6), 643–660.

Chemical Abstract Services. (2015). CAS Registry: The gold standard for chemical substance information. Retrieved from https://www.cas.org/content/chemical-substances.

Chu, S. K.-W., & Law, N. (2007). Development of information search expertise: Research students' knowledge of source types. *Journal of Librarianship and Information Science*, *39*(1), 27-40.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education* (5th ed.). New York, NY: RoutledgeFalmer.

Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (3rd Ed.). Thousand Oaks, CA: Sage Publications.

Crotty, M. (1998). The Foundations of Social Research: Meaning and perspective in the research process. Thousand Oaks, CA: Sage Publications.

Dawson, P. H., Jacobs, D. L., & Yang, S. Q. (2010). An Online Tutorial for SciFinder for Organic Chemistry Classes. *Science & Technology Libraries*, *29*(4), 298–306.

Etkina, E., Matilsky, T., & Lawrence, M. (2003). Pushing to the edge: Rutgers astrophysics institute motivates talented high school students. *Journal of Research in Science Teaching*, *40*(10), 958–985.

Feldman, A., Divoll, K., & Rogan-Klyve, A. (2009). Research education of new scientists: Implications for science teacher education. *Journal of Research in Science Teaching*, *46*(4), 442–459.

Feldman, A., Divoll, K. A., & Rogan-Klyve, A. (2013). Becoming Researchers: The Participation of Undergraduate and Graduate Students in Scientific Research Groups. *Science Education*, *97*(2), 218–243.

Ferrer-Vinent, I. J. (2012). Teaching SciFinder Basics to Organic Chemistry Students. *Science & Technology Libraries*, *31*(2), 164-179.

Gabridge, T., Gaskell, M., & Stout, A. (2008). Information Seeking through Students ' Eyes : The MIT Photo Diary Study. *College & Research Libraries*, *69*, 510–523.

George, C., Bright, A., Hurlbert, T., Linke, E. C., Clair, G. S., & Stein, J. (2006). Scholarly use of information: Graduate students' information seeking behaviour. *Information Research, 11*(4), 1-23.

Harmata, M., Kahraman, M. (1998). Lewis acid mediated reactions of N-arylsulfonimidoyl chlorides with alkenes. Some steric effects of alkene substitution. *Journal of Organic Chemistry, 63*(20), 6845-6851.

Hunter, A. B., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education*, *91*(1), 36–74.

Jamali, H. R., & Nicholas, D. (2008). Information-seeking behaviour of physicists and astronomers. *Aslib Proceedings*, *60*(5), 444–462.

Johnson, R. B. (1997). Examining the Validity Structure of Qualitative Research. *Education*, *118*(2), 282–292.

Kardash, C. M. (2000). Evaluation of undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. *Journal of Educational Psychology, 91*(1), 191-201.

Korobili, S., Malliari, A., & Zapounidou, S. (2011). Factors that influence information-seeking behavior: The case of greek graduate students. *The Journal of Academic Librarianship, 37*(2), 155-165.

Kraft, A., Strickland, A. M., & Bhattacharyya, G. (2010). Reasonable reasoning: multi-variate problem-solving in organic chemistry. *Chemistry Education Research and Practice, 11*, 281-292.

Kuruppu, P. U., & Gruber, A. M. (2006). Understanding the Information Needs of Academic Scholars in Agricultural and Biological Sciences. *Journal of Academic Librarianship*, *32*(6), 609–623.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

Leiter, K. (1980). *A Primer on Ethnomethodology*. New York, NY: Oxford University Press.

Lincoln, Yvonna S., Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Lopatto, D. (2004). Survey of Undergraduate Research Experiences (SURE): First Findings. *Cell Biology Education*, *3*(4), 270–277.

Maeyer, J., Talanquer, V. (2013). Making predictions about chemical reactivity: Assumptions and Heuristics. *Journal of Resesarch in Science Teaching, 50*(6), 748-767.

Nersessian, N. J. (2006). The Cognitive-Cultural Systems of the Research Laboratory. *Organization Studies*, *27*(1), 125–145.

Orgill, M. (2007). Situated Cognition. In G. M. Bodner & M. Orgill (Eds.), *Theoretical frameworks in chemistry/science education.* (pp. 187-203). Upper Saddle River, NJ: Pearson Education, Inc.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Richmond, G., & Kurth, L. A. (1999). Moving from outside to inside: High school students' use of apprenticeships as vehicles for entering the culture and practice of science. *Journal of Research in Science Teaching*, *36*(6), 677–697.

Ritchie, S., & Rigano, D. (1996). Laboratory apprenticeship through a student research project. *Journal of Research in Science Teaching*, *33*(7), 799–815.

Rosenstein, I. J. (2005). A Literature Exercise Using SciFinder Scholar for the Sophomore-Level Organic Chemistry Course. *Journal of Chemical Education*, *82*(4), 652-654.

Samarapungavan, A., Westby, E. L., & Bodner, G. M. (2006). Contextual epistemic development in science: A comparison of chemistry students and research chemists. *Science Education*, *90*(3), 468–495.

Sheeja, N. K. (2010). Science vs social science: A study of information-seeking behavior and user perceptions of academic researchers. *Library Review*, *59*(7), 522–531.

Strauss, A., & Corbin, J. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3rd Ed.). Thousand Oaks, CA: Sage Publications.

Strickland, A. M., Kraft, A., & Bhattacharyya, G. (2010). What happens when representations fail to represent? Graduate students' mental models of organic chemistry diagrams. *Chemistry Education Research and Practice*, *11*, 293-301.

Stucky, A. P. (2005). *Empirical grounding of the nature of scientific inquiry: A study of developing researchers* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (3196070)

Swoger, B. J. M., & Helms, E. (2015). An Organic Chemistry Exercise in Information Literacy Using SciFinder. *Journal of Chemical Education*, *92*(4), 668–671.

Walls, B. (2008). Becoming a Chemist: Graduate Students' Perspectives on Chemist and Chemistry (Master's thesis). Retrieved from ProQuest Dissertations and Theses. (1461005)

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, *7*(2), 225-246.

Wilson, T. D. (2000). Human Information Behavior. *Informing Science*, *3*(2), 49–55.