



Lost in Translation

Upper Secondary Students' Challenges with Protein Synthesis



Sara Wahlberg

Faculty of Health, Science and Technology

Chemical Sciences

DOCTORAL THESIS | Karlstad University Studies | 2026:20

Lost in Translation

Upper Secondary Students' Challenges with Protein
Synthesis

Sara Wahlberg

Lost in Translation - Upper Secondary Students' Challenges with Protein Synthesis

Sara Wahlberg

DOCTORAL THESIS

Karlstad University Studies | 2026:20

urn:nbn:se:kau:diva-108965

ISSN 1403-8099

ISBN 978-91-7867-686-6 (print)

ISBN 978-91-7867-687-3 (pdf)

<https://doi.org/10.59217/uqzj6480>

© The author

Distribution:

Karlstad University

Faculty of Health, Science and Technology

Department of Engineering and Chemical Sciences

SE-651 88 Karlstad, Sweden

+46 54 700 10 00

Print: Universitetstryckeriet, Karlstad 2026

WWW.KAU.SE

“For now we see through a glass, darkly...”
(1 Corinthians 13:12, KJV)

Abstract

Learning complex systems in the molecular life sciences is challenging for students, in large part because of the domain-specific language. This thesis provides insights into how domain-specific language is structured and used in upper secondary school to communicate complex molecular systems, using protein synthesis as an example. Across four studies, students' conceptual structures and mechanistic reasoning about protein synthesis, and how chemistry and biology textbooks describe it through metaphors and domain-specific concepts were investigated. The results show that students use domain-specific concepts in fragmented clusters rather than formulating descriptions showing an integrated understanding of protein synthesis. Students and textbooks follow the central dogma of a linear process from DNA to protein and compartmentalise concepts with mRNA as a key bridging idea. Few students reach an emergent mechanistic level of reasoning, and only in isolated sub-processes. Textbooks use two main metaphor systems: Information-based and construction-based metaphors. Teaching should emphasise links between clusters of concepts and make concepts and metaphors explicit to help students develop a coherent understanding of protein synthesis.

Keywords: Domain-specific language, protein synthesis, students' understanding, upper secondary school

Sammanfattning

Elever möter utmaningar i att förstå komplexa system inom molekylär livsvetenskap, delvis på grund av det ämnesspecifika språk som används. Denna avhandling ger insikter i hur ämnesspecifikt språk är strukturerat och används i gymnasieskolan för att kommunicera komplexa molekylära system, med proteinsyntes som exempel. I fyra delstudier analyserades elevers begreppsanvändning och mekanistiska resonemang om proteinsyntesen samt kemi- och biologiläroböckers beskrivning av proteinsyntes genom metaforer och ämnesspecifika begrepp. Resultaten visar att eleverna använder ämnesspecifika begrepp i fragmenterade kluster snarare än som en sammanhängande helhet. Elevers och läroböckers beskrivningar följer det centrala dogmats linjära struktur från DNA till protein och avgränsar användningen av ämnesspecifika begrepp till isolerade delar. mRNA visade sig ha en central roll i att brygga över dessa olika delar. Få elever nådde ett utvecklat mekanistiskt resonemang, och då endast i begränsade delprocesser. Läroböckerna använde två metaforsystem: informationsbaserade och konstruktionsbaserade metaforer. Lärare bör tydliggöra samband mellan begreppskluster och synliggöra ämnesspecifika begrepp och metaforer för att stödja elevers utveckling av en sammanhängande förståelse av proteinsyntes.

Nyckelord: Elevers uppfattningar, gymnasieskolan, proteinsyntes, ämnesspecifikt språk

List of articles

- I** Clusters of Concepts in Molecular Genetics: A Study of Swedish Upper Secondary Science Students' Understanding*
Niklas M. Gericke and Sara J. Wahlberg
Journal of Biological Education, 2013, 47(2), 73-83.
<https://doi.org/10.1080/00219266.2012.716785>
- II** Conceptual Demography in Upper Secondary Chemistry and Biology Textbooks' Descriptions of Protein Synthesis - a Matter of Context?
Sara J. Wahlberg and Niklas M. Gericke
CBE—Life Sciences Education, 2018, 17(3), ar51.
<https://doi.org/10.1187/cbe.17-12-0274>
- III** Metaphors on Protein Synthesis in Swedish Upper Secondary Chemistry and Biology Textbooks – A Double-Edged Sword
Sara J. Wahlberg, Jesper Haglund and Niklas M. Gericke
Research in Science Education, 2025, 55(2), 425-444.
<https://doi.org/10.1007/s11165-024-10197-y>
- IV** Upper Secondary Students' Application of Domain-Specific Concepts and Mechanistic Reasoning about Protein Synthesis
Sara J. Wahlberg, Jesper Haglund and Niklas M. Gericke
Manuscript

* This is an Accepted Manuscript published by Taylor & Francis in *Journal of Biological Education* on 12 September 2012, available online: <https://doi.org/10.1080/00219266.2012.716785>

Author's contribution

Across all four studies, I:

- provided and developed the initial idea and plan of the studies' designs
- executed the ethical vetting processes
- provided the initial designs for the data collection methods and collected all of the data
- planned the data analyses and conducted main data analyses
- chose conceptual frameworks
- constructed visual presentations of results
- constructed initial drafts of the articles, wrote on all parts of Article I, and wrote most of the text in Articles II-IV.

Specifically in Study II, I:

- developed the idea of 'conceptual demography'
- provided the main idea for the data analysis software and its construction in collaboration with a programmer
- validated the data analysis software

Specifically in Studies II-III, I:

- executed the submission and review processes
- was the corresponding author

Preface

This doctoral thesis is submitted for a PhD in Chemistry with the specialisation in Chemistry Education. The research was conducted at Karlstad University, Sweden. The overall project began in 2009 when I joined the Erna and Victor Hasselblad Foundation Research School in Molecular Science Education until the licentiate degree in 2019. Thereafter the project subsequently progressed within the local research centre of SMEER (Science, Mathematics and Engineering Education Research) at Karlstad University. The project has thereby been ongoing for an extended period. Such a long timeframe naturally brings certain disadvantages such as intermittent workflow and setup-times, but it has also offered significant advantages, for instance the opportunity for ideas to mature, the chance to follow developments in the field of education over time, and a space to grow into the role of a researcher. The project started when I noticed just how much students at various educational levels struggled to understand molecular life sciences, where genetics seemed to be particularly challenging. As I paid closer attention to the matter, it became increasingly apparent that many of these challenges seemed to be related to a highly domain-specific language that the students were expected to manage. At that point, however, much was still unclear, and my curiosity about this uncertainty ultimately became the driving force behind the project.

The aim of the licentiate thesis was to contribute to the understanding of teaching and learning protein synthesis through domain-specific language in upper secondary education. The licentiate thesis comprises two studies: Gericke & Wahlberg (2013) and Wahlberg & Gericke (2018). The data comprised semi-structured interviews (Gericke & Wahlberg, 2013), and a content analysis of chemistry and biology textbooks (Wahlberg & Gericke, 2018). From the results, I argue that there are four facets of students' learning of protein synthesis in upper secondary education: mechanistic or conceptual descriptions, compartmentalisation, mRNA as core concept, and canonical representation. These facets synthesised the findings in the licentiate thesis and evoked the interest in further exploration of the topic in this doctoral thesis.

The doctoral thesis is a compilation of four studies, and the comprehensive summary that you are now reading. The aims and research questions of the four studies were:

In Study I, we examined how eleventh graders in upper secondary school reason about concepts in molecular genetics that are associated with protein synthesis. The research questions were:

- To what extent are upper secondary science students able to explain concepts of the protein synthesis correctly, using the accepted nomenclature?
- In what ways do upper secondary students relate concepts in molecular genetics to each other?

In Study II, the aim was to explore the importance of context as a differentiator between the meaning conveyed of the same life science phenomenon in biology and chemistry texts. Specifically, the conceptual demography of written protein synthesis descriptions in Swedish upper secondary biology and chemistry textbooks were compared to reveal implications for teaching and learning protein synthesis. The research questions were:

- With what frequency is each sample term used?
- What are the distribution patterns of each sample term?
- What relationship structures can be found between the sample terms?

In Study III, the aim was to provide insight into the use of metaphors in protein synthesis descriptions in upper secondary chemistry and biology textbooks. The research question was:

- What metaphor systems can be identified in descriptions of protein synthesis in Swedish upper secondary biology and chemistry textbooks?

In Study IV, the aim was to provide insights into how mechanistic reasoning is reflected by upper secondary students' use of domain-specific concepts when they describe protein synthesis. The research questions were:

- What domain-specific concepts and sub-processes of protein synthesis do student descriptions focus on?

- What characterises upper secondary students' level of sophistication in mechanistic reasoning when they describe protein synthesis?

Given the research trajectory across the licentiate thesis and this doctoral thesis, some sections are inevitably similar, overlapping, or even identical. All text has nevertheless been carefully reviewed and adapted to align with the broader scope of a doctoral thesis.

Karlstad in March of 2026

Acknowledgements

Over the years many, many paths have crossed, and countless insights and thoughts have been shared. As a beginning researcher, I owe each of you, my gratitude. Some I will thank by name, while others will have my lasting appreciation regardless.

I want to express my heartfelt appreciation to my supervisors over the years. Above all, I am deeply grateful to professor Niklas Gericke for his steadfast guidance throughout these years. Your profound expertise and high expectations have shaped not only the direction of this work but also my development into being a researcher. Your resilience and unwavering dedication have left an everlasting impact, and I remain profoundly thankful and humbled for all you have invested in me and in this project.

My sincere gratitude also goes out to associate professor Jesper Haglund. Your sharp insights and gentle approach have been invaluable. Your ability to challenge with clarity while supporting with compassion is something I have highly cherished. I am deeply appreciative of the thoughtfulness and encouragement you have brought to the project.

I also want to express my appreciation to my co-supervisors professor Eva Bergman, and associate professor Nina Christenson for your enthusiasm and expertise across different areas. I am truly grateful for the positivity and sharpness you brought to the later stages of the project.

I wish to acknowledge professor emerita Lena Tibell, who first supervised this project and welcomed me into the work with such generosity. Thank you for taking me aboard and for guiding me through the early stages with commitment, clarity and encouragement. I remain truly grateful for the opportunity you gave me.

My gratitude also extends to dr. Michal Drechsler, who mentored me in research within the department during the early stages of my journey.

I want to acknowledge the financial support provided by the Erna and Victor Hasselblad Foundation, the research centre for Science, Mathematics, and Engineering Education Research (SMEER) at Karlstad university, and the Department of Engineering and Chemical Sciences at Karlstad university.

A special thanks goes to the discussion partners at previous seminars. Professor Marcus Hammann, University of Münster, Germany, offered an engaging and thoughtful discussion which helped shape the project. I also want to acknowledge associate professor Inger Edfors for insightful comments on my early work.

This project would not have been possible without the voluntary participation of upper secondary students, and access to written textbooks. I thank each student for your participation, the teachers and headmasters for providing opportunities in schools, and the textbook writers, editors, and publishers for making materials available.

To all my close colleagues at Karlstad University and other universities, I owe a great deal of thanks. Special appreciation goes to everyone at the division of chemistry at Karlstad University in all its various form over the years. I especially want to acknowledge associate professor Stefan Frodeson, head of department, and dr. Gunilla Carlsson Kvarnlöf, deputy head of department, as well as professor Jan van Stam, head of subject, and dr. Anna Smedja Bäcklund, director of studies, for providing opportunities for research. I am grateful to associate professor Jörgen Samuelsson for steadfast encouragement. I also want to thank all colleagues at SMEER for input, inspiration and discussions over the years, as well as close colleagues at the teacher education programmes. My profound thanks to dr. Anna-Sofia Preece for your unwavering support and cherished friendship.

Last but not least, I wish to extend my deepest gratitude to my friends and family. I am especially grateful for love and support from my father, Tommy, my mother, Ingela, my sister Anna and her family, my grandparents Inger and Jörgen, my partner Markus, and our children Matilda and Hjalmar.

Table of contents

ABSTRACT	2
SAMMANFATTNING	3
LIST OF ARTICLES	4
AUTHOR’S CONTRIBUTION	5
PREFACE	6
ACKNOWLEDGEMENTS	9
TABLE OF CONTENTS	11
INTRODUCTION	13
AIM AND RESEARCH QUESTIONS	15
BACKGROUND	16
Life sciences and the challenges to understand it.....	16
Central dogma and canonical representation	17
Models of protein synthesis, and their limitations	18
Students’ understanding of complex life sciences systems	19
Domain-specific language, and the Swedish context.....	20
Compartmentalisation of language	22
Conceptualisation through metaphors.....	23
Mechanistic reasoning on abstract phenomena.....	25
Resources: The physical textbook.....	26
CONCEPTUAL AND THEORETICAL FRAMEWORKS	28
METHODOLOGY	31
The educational context – Swedish upper secondary school	31
Study I.....	33
Sample of respondents in Study I.....	33
Data collection and analysis in Study I	33
Studies II and III.....	35
Sample of textbooks and texts in Studies II and III.....	35
Data collection and analysis Study II	36
Data collection and analysis Study III.....	37
Study IV	38
Sample of respondents in Study IV	38
Data collection and analysis for Study IV	38
Trustworthiness, reliability and transferability	39
Ethical considerations.....	41

RESULTS	42
Main results of Study I.....	42
Main results of Study II	44
Main results of Study III.....	45
Main results of Study IV	48
DISCUSSION	50
Facet on mechanistic reasoning	50
Facet on compartmentalisation	53
Facet on mRNA as a central core concept	54
Facet on metaphors.....	55
Facet on the canonical representation in molecular biology	56
Methodology discussion.....	57
CONCLUSIONS	61
Students' structure and use of domain-specific concepts.....	61
Students' sophistication of mechanistic reasoning.....	61
Domain-specific language in textbooks.....	62
IMPLICATIONS FOR TEACHING COMPLEX MOLECULAR SYSTEMS	64
FURTHER STUDIES	66
REFERENCES	67

Introduction

In this doctoral thesis, the role of domain-specific language in upper secondary school science education is examined, with a particular emphasis on complex molecular processes. Such processes involve numerous components that interact across space and time at different hierarchical levels, that require precise timing and coordination to ensure a successful outcome. Complex phenomena emerge at higher levels of organisation and cannot be deduced solely from the interactions of the individual components of systems in isolation. To illustrate a complex molecular process, protein synthesis in a eukaryotic cell is used as an example, where proteins are synthesised in a multicomponent, multicompartiment biological system. A solid understanding of protein synthesis is essential to molecular life sciences and to understanding life itself, as proteins carry out vital regulatory, structural, and catalytical functions in biological systems (Voet & Voet, 2010). Precise knowledge of protein synthesis therefore offers crucial insights in the mechanisms that govern biological systems.

Research in science education over the past decades indicates that students experience substantial difficulties in understanding molecular life processes, including protein synthesis (Tibell & Rundgren, 2010). These challenges have often been attributed to the inherent conceptual complexity of these processes, as well as an extensive reliance on a domain-specific language (Knippels, 2002). Domain-specific language refers to the specialised language used within a particular field or community (Halliday & Martin, 1993; Pearson & Hughs, 1988). This language consists of technical terms, expressions and communication strategies that carry shared meaning for those working in that field (Tibell & Rundgren, 2010). In the molecular life science field, domain-specific language has been shown to be concept-dense, contain a vast number of specialised words and expressions, and to lack everyday referents, which can make it challenging to navigate (Tibell & Rundgren, 2010). Professional cell biologists and biochemists are experts in their field and in how to use the domain-specific language (Tibell & Rundgren, 2010). Students that are learning molecular life science encounter a tailored, specialised domain-specific language in

learning material such as written textbooks. In a school context, a key challenge is to transform science as it is practiced in the scientific community into school science without losing the essential features of that domain-specific language (Olander & Ingerman, 2010). At the same time, domain-specific language can hinder students' understanding of a content. It is therefore important to understand how domain-specific language functions in a school setting. The findings presented in this thesis can thereby provide a conceptual foundation for developing educational support and instructional frameworks designed to facilitate teaching and learning complex molecular processes.

Aim and research questions

The aim of this thesis is to provide insights into how domain-specific language is structured and used in upper secondary school to communicate complex molecular systems, using protein synthesis as an example.

The research questions are:

1. In what ways do upper secondary students structure and use domain-specific concepts to explain protein synthesis?
2. How do upper secondary students' use of domain-specific concepts reflect the sophistication of their mechanistic reasoning about protein synthesis?
3. In what ways do upper secondary chemistry and biology textbooks employ domain-specific language to describe protein synthesis?

Background

Life sciences and the challenges to understand it

Life sciences is an umbrella term for scientific disciplines that investigate living organisms and the mechanisms of life (e.g. Tibell & Rundgren, 2010). The field has emerged from the increasingly extensive and diverse growth of research in chemistry and biology. It spans multiple levels of organisation from ecology, which examines entire ecosystems, to the molecular aspects of life explored in biochemistry, molecular biology, and biotechnology. Studies on life processes at the molecular level, such as cellular mechanisms, are now integrated within the interdisciplinary field known as molecular life sciences. Molecular life sciences constitute essential parts of the broader disciplines of chemistry and biology. For instance, protein synthesis can be explored in either of these disciplines, requiring a robust disciplinary knowledge across multiple domains.

People who are trained in the molecular life science field develop the ability to navigate a complex scientific landscape and communicate using a shared disciplinary knowledge. Entering the molecular life sciences has, however, been found challenging (Höst et al., 2013), partly due to its domain-specific language (Knippels, 2002; Tibell & Rundgren, 2010). An extensive and abstract body of disciplinary knowledge combined with its highly specialised language (Orgill & Bodner, 2007) can overwhelm learners (Wood, 1990) and create substantial challenges to effective learning. Learning can be facilitated when the domain-specific language, reflecting underlying domain-specific concepts, is identified and explained with sufficient depth and detail (Stadig Degerman & Tibell, 2012). Some of these concepts constitutes the core concepts that form the essential foundation for communication in the field (Driver et al., 1996). Mastering the specialised language, especially its underlying conceptual foundations, is therefore a crucial cornerstone for understanding the life sciences at any level (Fang, 2004).

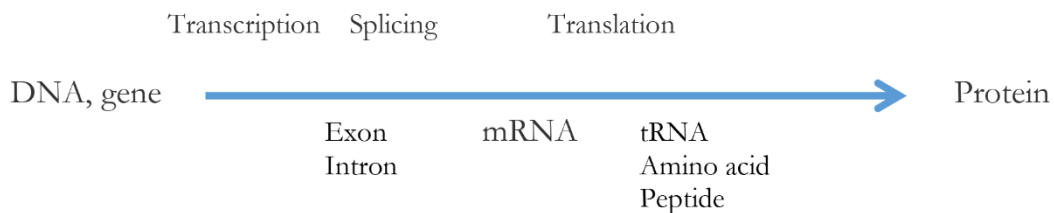
Central dogma and canonical representation

In this thesis, protein synthesis is used as an abstract multi-step, multi-compartment process through which living cells synthesise proteins. Understanding protein synthesis has been a cornerstone in the life sciences (Reinagel & Speth, 2016) ever since the discovery of the structure and function of DNA, to which Franklin, Wilkins, Levene, Chargaff, Watson, Crick, and others made foundational contributions (Cobb, 2017; Crick, 1958; Franklin et al., 1958; Hamilton et al., 1959; Pray, 2008; Watson & Crick, 1953). Heredity had long been recognised through similarities between relatives (Condit, 1999), and through selective breeding of livestock (Feliuss et al., 2014), even though the underlying mechanisms were unknown. The idea that a specific structure within the cell might be responsible for inherited traits had also been proposed several years before the discovery of the double helix in the 1950s. The elucidation of the structure of DNA triggered a paradigm shift from viewing traits as abstract, phylogenetic characteristics to understanding a concrete pathway that links genetic material to a mature protein product. This pathway from DNA to protein became known as the central dogma in molecular biology (Crick, 1958), schematically represented in Figure 1 below.

Figure 1

The Central Dogma of Molecular Biology

Studies



Note. The image is from Article I. Reprinted with permission from the journal.
Original image: <https://doi.org/10.1080/00219266.2012.716785>.

The main idea as presented by Crick (1958, p. 138) was that the “main function of the genetic material is to control (not necessarily directly) the synthesis of proteins”. From this idea, a dogmatic view of protein synthesis emerged, grounded in a linear understanding of how proteins are synthesised. This view was widely accepted even though Crick later

acknowledged that the processes involved in protein synthesis were likely more complex than first explained (Crick, 1970), which research later has clearly shown (Jaenisch & Bird, 2003). It is also well established that epigenetic factors, for instance chemical modifications in the cell or environmental stress, play an important role in protein synthesis, which the original formulation of the central dogma did not account for (Koonin, 2012). Hence, teaching the dogma uncritically therefore promotes an outdated idea of protein synthesis. The term dogma is however inherently problematic, as it implies unquestionable truths (Dogma, 2026; Shapiro, 2009). Within a scientific discourse, such authoritative connotations are unfortunate (Ladyman, 2002). Consequently, the updated idea is now more commonly referred to as the canonical representation of molecular biology or the canonical representation of protein synthesis, designations that avoid the truth-claims associated with the term dogma. In this thesis, I use both dogma and canonical representations but for different purposes. Dogma is employed when historical framing of the linear idea of protein synthesis is addressed, whereas canonical representations refer to contemporary perspectives in education and current scientific ideas of protein synthesis.

Models of protein synthesis, and their limitations

Protein synthesis is typically divided into three major processes: transcription, splicing, and translation. In each of these sub-processes, molecular entities interact in defined regions of the cell, where intermediates are carried forward to the next stage of the pathway. To illustrate a specific molecular entity, an interaction, or what occurs at a particular point in time, researchers use models (Gericke, 2009). In the context of complex systems such as protein synthesis, models make it possible to isolate specific components, explore how they interact, and predict outcomes (Haskel-Ittah & Yarden, 2017). Within the literature, three overarching models of protein synthesis are commonly described: the Translation model (T-model), the Transcription–Translation model (TT-model), and the Transcription–Splicing–Translation model (TST-model), as explained in my licentiate thesis (Wahlberg, 2019). The T-model defines protein synthesis as translation only. Key concepts in the T-model include amino acid, mRNA, polypeptide, protein, and tRNA. The TT-model adds domain-specific

concepts that are associated with transcription (DNA, and gene) to the concept cohort of the T-model. The TST-model extends the TT-model by incorporating domain-specific concepts that are associated with mRNA maturation (intron and exon).

In this thesis, I adopt the most inclusive modelling approach where the TST-model provides the fundamental network, yet the analysis remains open to more intricate extensions. None of the three models include post-translational modifications, such as the mechanisms involved in protein folding, nor do they include epigenetic mechanisms such as DNA methylation. Although more comprehensive models exist that account for these additional layers of regulation, they fall outside the scope of the present thesis project, as the project commenced prior to the introduction of epigenetics into the relevant curriculum. In contemporary upper secondary biology textbooks, epigenetics is increasingly included but in an old way that only our genes shape our future (Gericke & Mc Ewen, 2023).

Students' understanding of complex life sciences systems

Thinking about complex systems has increasingly become a necessary skill in understanding the world around us (Focant et al., 2025). For instance, the understanding of photosynthesis for carbon dioxide harvest in sustainability issues, or the role of bioenergetics in understanding training and nutrition are global issues that are in everyday attention in media. Substantial amounts of research have been conducted in students' understanding of these issues.

Many studies have examined learning challenges in different areas of molecular life sciences, particularly students' understanding of central biological concepts, and the challenges associated with them (Allchin, 2000; Duncan & Reiser, 2007; Gericke & Hagberg, 2007; Gericke et al., 2013; 2014; 2017; Lewis & Kattmann, 2004; Smith & Gericke, 2015; Thörne & Gericke, 2014; Thörne et al., 2013; Venville & Treagust, 1998; 2002). Over the past 25 years, considerable attention has also been given to students' understanding of gene structure and function (Duncan & Reiser, 2007; Lewis et al., 2000; Marbach-Ad, 2001; Machová & Ehler, 2023; Marbach-Ad & Stavy, 2000; Thörne & Gericke, 2014; Wood-Robinson et al., 2000). For example, students

have been found to believe that amino acids are synthesised in the translation process (Fisher, 1992), or that genes are equal to chromosomes (Lewis & Wood-Robinson, 2000). Collectively, this body of work shows that students struggle with understanding core concepts needed to communicate biological processes.

Protein synthesis is a complex molecular system in life sciences and serves as the focal example in this thesis. Teaching and learning protein synthesis have, however, received limited attention in educational research over the past 20 years. Some aspects have been explored, however. For instance, Rundgren (2006) examined how students used metaphors and ‘help-words’ in their meaning-making about proteins, while Guzman and Bartlett (2012) investigated the teaching and learning of the translation process. One of few studies focused on communicating protein synthesis (Wright et al., 2014) critiques common representations of the central dogma, particularly arrow diagrams, which can lead students to believe that DNA is converted into RNA and RNA into protein. They argue that such representation may be counterproductive unless teachers explicitly clarify their meaning. The scarcity of research on how protein synthesis is taught and understood highlights the need for further investigation.

Domain-specific language, and the Swedish context

When knowledge is communicated within a disciplinary community, spoken and written language follow characteristic patterns (Halliday & Martin, 1993). Each community develops a domain-specific language where a specialised vocabulary plays a central role (Fang, 2004; Tibell & Rundgren, 2010). This vocabulary consists of technical terms used to describe entities, processes, relationships, activities, and organisational structures relevant to the shared subject matter in the community. Within domain-specific language, certain words are essential for describing phenomena and label elements of the domain. In semantics, such words are referred to as technical terms. A term can also convey abstract meaning, which can be understood as a concept. Thus, when a technical term is assigned meaning, that meaning constitutes its associated concept (Ogden et al., 1923). Domain-specific concepts that are fundamental for understanding a phenomenon can be regarded as core concepts (Driver et al., 1996). Concepts that are

relevant but less central to communication about the phenomenon at hand are referred to in this thesis as peripheral concepts (Gericke & Wahlberg, 2013).

Scientific advances often require new linguistic structures to articulate emerging concepts. For example, the terminology used to describe early insights into protein synthesis drew heavily on the developing field of informatics (Kay, 2000). Earlier theories had proposed that cells could learn from experience through a form of ‘cellular memory’, implying that acquired traits might be inherited. The scientific discoveries of the first half of the 20th century, later refined by contemporary research in areas such as epigenetics, challenged this view (Gericke & Mc Ewen 2023; Thörne et al., 2026). Sequences of DNA came to be understood instead as templates or messages that direct the growth, development, responses, and functions of organisms. These sequences were increasingly conceptualised as ciphers requiring a code key for full interpretation.

In a Swedish context, linguistic conventions differ in ways that may affect students’ understanding of molecular life science phenomena. Swedish often adopts direct equivalents of English terminology, yet these terms lack everyday associations that could support intuitive understanding. For instance, ‘transkription’ (Eng. transcription), exists in Swedish, but outside the molecular life sciences it primarily refers to dictation. This gives a narrower and more specialised meaning than the English counterpart. ‘Splitsning’ (Eng. splicing) is another example. In everyday Swedish it refers to the splicing of ropes, an increasingly uncommon activity, which may offer little intuitive connection to the biological process. Translation presents a different challenge. Although the proper Swedish term ‘översättning’ is widely used in everyday language for communicating translation between different languages, molecular life sciences contexts use a ‘Swedified’ version of the English term. Consequently, Swedish students may be able to describe the underlying molecular life science phenomenon but are challenged to grasp the conceptual essence of the terminology by drawing on everyday Swedish.

Compartmentalisation of language

A central dimension of learning in the molecular life sciences is the capacity to recognise, integrate, and articulate relationships among domain-specific concepts (e.g. Novak & Gowin, 1984). These relationships can be visualised through concept maps that depict the conceptual knowledge through the organisation, structure, and relationship of domain-specific concepts as represented by individuals, groups, or text sources (Kinchin et al., 2000; Novak & Gowin, 1984). Gagatsis (2012, p. 665) explains that “the term ‘compartmentalisation’ is used in the sciences of learning and cognition to designate the phenomenon that occurs when an individual has two or more different, potentially conflicting conceptual schemas concerning a particular domain in his/her cognitive structure.” A certain schema is triggered in specific situations and when a situation does not activate the most relevant schema, another schema that is less relevant can be activated, and compartmentalisation becomes evident (Gagatsis, 2012).

A characteristic pattern in students’ conceptual organisation is the emergence of isolated ‘concept islands’ which is identified as one type of compartmentalisation. This phenomenon involves identifying relationships between a limited subset of domain-specific concepts while maintaining conceptual boundaries from other conceptual groups, thereby forming a compartment, or cluster. Marbach-Ad (2001) documents the tendency of students to compartmentalise concepts, and additional studies indicate that although students may reason competently about individual specific domain-specific concepts, they often perceive only limited relationships between distinct clusters of domain-specific concepts (Bahar et al., 1999; Lewis & Kattmann, 2004). To develop a coherent conceptual understanding of cellular processes requires that students grasp the interrelationships among central domain-specific concepts (Lewis et al., 2000). Collectively, these findings suggest that more integrated representations of conceptual relationships, both within and between clusters of concepts, may support deeper learning in molecular life sciences. Concept maps provide a means of making such relationships explicit.

Concept maps serve multiple pedagogical and analytical purposes, including the visualisation of students' conceptual understanding through the ways that they cluster concepts (van Zele et al., 2004). The structural patterns that emerge in concept maps can be interpreted as indicators of students' conceptual understanding of a topic, reflected in their selection and deployment of domain-specific concepts, and their capacity to articulate meaningful relationships between them (Mintzes et al., 2005; Novak & Gowin, 1984).

Conceptualisation through metaphors

A central component of understanding abstract phenomena is the ability to articulate them through more familiar terms (Duit, 1991). In conceptual metaphor theory (Lakoff & Johnson, 1980), the unfamiliar concept is designated as the target domain, whereas the familiar concept used to illuminate it, which is grounded in previous, embodied experiences, is referred to as the source domain. The meaning of the metaphor emerges in the interplay between these two domains. Over time, however, the link between the different domains may weaken. When the metaphorical association either becomes obscure or is no longer recognised, the expression is considered a dead metaphor (Rundgren et al., 2009). In educational contexts, dead metaphors can be problematic because the etymological and conceptual links that once supported understanding have faded. For instance, the biological term cell has largely lost its historical connection to the monastic cell from which it originated (Rundgren et al., 2009). Consequently, if the relationship between source and target domains, along with the historical meanings embedded in the domain-specific terminology, is made clear, students' conceptual development could be supported.

The rapid advances in molecular life sciences during the early 20th century led scientists to borrow ideas from emerging fields such as cryptography and informatics to describe the informational nature of DNA (Kay, 2000). In the search for a vocabulary to describe protein synthesis, metaphors from informatics did more than provide convenient linguistic bridges and shortcuts (Aubusson et al., 2006), they shaped how the molecular processes were conceptualised. The copying of genetic material into immature mRNA, for instance, was compared to the conversion of speech to text and came thus to be

known as transcription. Likewise, the ribosomal conversion of genetic information encoded in RNA into proteins was framed as the translation of a text from one language into another (Kay, 2000). Consequently, the process was named translation, and the RNA that carried this information became known as messenger RNA (mRNA) (Voet & Voet, 2010).

Newly introduced terms described and imported an entire semantic framework in which DNA functions as a template (Taylor & Dewesbury, 2018) that carries a code (Schrödinger, 1944), RNA as a messenger, and polypeptides or proteins as the executed output. In this sense, the scientific description of protein synthesis became metaphorical at its formation. The metaphors borrowed from informatics did not simply illuminate existing processes but shaped the conceptual boundaries of the field (Kjosavik, 2007). As a result, mid-20th century molecular genetics emerged not only as a technical enterprise but also as a linguistic and epistemic undertaking, fundamentally conditioned by the metaphors through which it was articulated.

Metaphors are widely used in science education (Amin, 2009; Eriksson, 2025). There is a broad agreement that figurative language can support conceptual understanding (Haglund, 2012; Eriksson, 2025). Metaphors can help students grasp difficult scientific ideas by linking them to familiar experiences (Lakoff & Johnson, 1980). In addition, genuine metaphors create novel, indispensable connections between ideas and cannot be paraphrased without loss of meaning (Knudsen, 2005). However, this mechanism is not inherently reliable. If students are unfamiliar with the everyday expression used as the source domain, the meaning of the metaphor may not be understood. Conversely, students who do know the expression may be misled if the relationship between its everyday and scientific meaning is unclear. Metaphors also require explicit guidance when the familiar idea functions differently from the scientific concept. A concept without everyday referent, such as lipid or polypeptide, may therefore be less confusing (Tibell & Rundgren, 2010). By contrast, the energy concept carries multiple everyday meanings and is therefore more easily misunderstood (Amin, 2009).

Metaphors such as blueprint or drawings are pervasive in explanations of gene function (Taylor & Dewsbury, 2018). These comparisons, however, suggest a direct one-to-one correspondence between genes and phenotypic outcomes resembling the ideas of the central dogma of molecular biology (Crick, 1958; Kjosavik, 2007). Such framing may inadvertently reinforce genetically deterministic thinking (Carver et al., 2017). To counter such portrayals, other metaphors have been suggested, such as comparing protein synthesis to following a recipe (Condit, 1999), but this shift has instead been criticised to oversimplify gene function. Less thoughtful use of metaphors risks fostering public misconceptions that may enter the classroom and undermine key ideas of science education (Venville & Donovan, 2006).

Mechanistic reasoning on abstract phenomena

Protein synthesis can be understood as an abstract and complex system composed of many interacting components. Because these interactions cannot be observed in real time, students must rely on mediated representations, such as figures, images, teaching material, or other resources, to construct an understanding of the underlying processes. To understand molecular systems, including cell structures and functions, students are required to engage in spatiotemporal reasoning across multiple hierarchical levels of biological organisation, spanning the cell, its organelles, and the involved molecular entities (Haskel-Ittah, 2023). This entails simultaneously tracking the sequence of events, their spatial location, and the interactions among constituent components, all within an abstract, model-based understanding of the cell. In a mechanistic description, entities engage in spatiotemporal organised activities that occur in ordered segments with defined starting points and endpoints (Machamer et al., 2000). The capacity to integrate these segments into a coherent causal explanation reflects a well-developed mechanistic reasoning.

When students are unable to understand the spatiotemporal dynamics that characterise intracellular phenomena, their understanding of the content decreases (Duncan & Reiser, 2007). As a result, students often rely on memorisation strategies, which tend to yield superficial explanations that overlook the molecular mechanisms that underpin biological phenomena (Haskel-Ittah, 2023; Haskel-Ittah & Yarden,

2018). Accordingly, students' capacity to engage in mechanistic reasoning about entities and processes involved in protein synthesis constitutes a critical indicator of the depth and coherence of their understanding. Thereby, the students need scaffolding in making causal connections explicit (Duncan & Reiser, 2007; Haskel-Ittah, 2022).

Sometimes students provide mechanistic reasoning that are fragmented or contain gaps (Machamer et al., 2000). These gaps are often referred to as black boxes where each represents a part of a mechanism that is unspecified or left out, leaving a missing piece in the explanation (Haskell-Ittah, 2023). Black boxes can also serve as temporary placeholders that students use until deeper understanding of the mechanism is developed (Machamer et al., 2000). Importantly, black boxes in explanation are not always detrimental. They may reflect attempts to manage cognitive load by avoiding the introduction of too many components at once. In this way, black boxes can support learning by allowing students to unpack one part of the mechanism at a time. Furthermore, the use of black boxes mirrors authentic research practise, where not all aspects of a mechanism are known from the start (Livini Alcasid & Haskel-Ittah, 2025).

Resources: The physical textbook

Textbooks constitute a central medium for transforming scientific knowledge into teachable school knowledge (Mikk, 2000). They also serve as an important learning resource for students (Ekvall, 2001). In Sweden, the textbook market is dominated by a small number of commercial publishers that supply materials to both schools and the public. In many countries, printed textbooks or their digital equivalents continue to dominate the market (Danielsson & Selander, 2021). Teachers generally retain considerable autonomy to select, purchase, and implement textbooks that they consider appropriate (Richenberg, 2016), although such decisions are often coordinated within individual schools. Textbooks are typically loaned to students and reused across multiple cohorts, sometimes for many years.

Despite the growing availability of digital alternatives, studies have indicated that students prefer printed textbooks (Woody et al., 2010), and such materials have been particularly influential in secondary science education (Bergqvist, 2017). Research further demonstrates that science textbooks tended to be densely packed with facts (Nelson, 2006) and characterised by an extensive domain-specific vocabulary (Smith-Walters et al., 2016). This substantial linguistic load has been identified as a potential barrier to students' comprehension of the scientific content (Edling, 2006). In the early 2000s, however, it was proposed that learning domain-specific vocabulary is most effective when such vocabulary is embedded within meaningful contexts where domain-specific concepts were selected for their broad applicability (National Institute of Child Health and Human Development, 2000). Teaching domain-specific concepts across multiple contexts similarly enhances conceptual understanding (Butler et al., 2010; Gilbert, 2006).

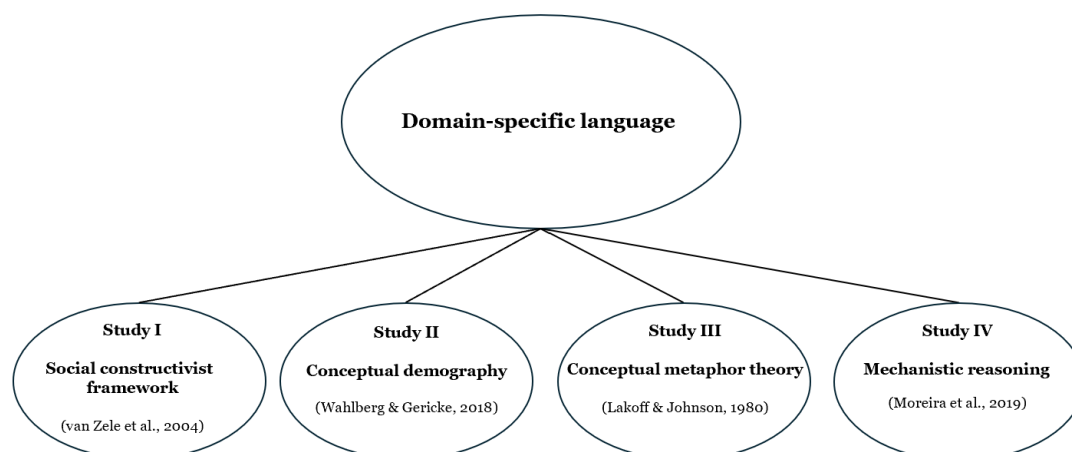
Textbook studies have been conducted in relation to the molecular life sciences, most of which have dealt with genetics education. For instance, Martínez-Gracia et al. (2006) found that Spanish secondary biology textbooks describe procedural ideas but do not support students in learning the central ideas of genetics. In addition, Gericke et al. (2014) reported that upper secondary chemistry and biology textbooks struggle to clearly present the relationship between genes and traits. Textbooks across subdisciplines present gene function in varying ways, and they often rely on simplified explanations that avoid biochemical detail (Gericke & Hagberg, 2010). Gericke et al. (2013) showed that students' understanding of the gene–trait relationship is hindered by the inconsistent ways this relationship is presented across textbooks from different subdisciplines. To sum up, textbooks are among the most important teaching materials and knowledge mediators in life science education. However, upper secondary students often have problems understanding the life science content in their chemistry and biology textbooks because of the use of scientific language in different contexts.

Conceptual and theoretical frameworks

This doctoral thesis comprises four interrelated studies that are situated in the upper secondary school context. While each study uses a distinct conceptual framework, together they form a coherent research project founded in a social constructivist tradition. A unifying theme across the four articles is how the use and structure of domain-specific language of molecular life sciences shape students' opportunities to understand protein synthesis as an example of a complex process. In Figure 2 below, the different frameworks are outlined.

Figure 2

Overview of the Four Study-Specific Frameworks and Connection to the Overarching Framework of Domain-Specific Language



In Study I, students' *conceptual networks* are examined in a *social constructivist framework* by exploring how students cluster domain-specific concepts (Gericke & Wahlberg, 2013). Knowledge of specific aspects of a subject, along with the development of connections between these aspects, is ultimately integrated into a student's existing knowledge framework (Glaser & Bassok, 1989; Van Zele et al., 2004). During the learning process, students acquire new concepts and link them to prior knowledge. Existing concepts are thus enriched and accumulate additional meaning as they become connected to new ones.

The clusters that form can be viewed as concept maps, which are powerful tools for investigating students' knowledge structures within a social-constructivist perspective. In this thesis, students' expressed concept structures are used to identify the conceptual organisation of their ideas (Fisher, 2000).

Study II employs the recently developed framework of *conceptual demography*, which was specifically designed for this study. The framework focuses on the frequency, distribution, and relationship patterns of domain-specific concepts (Wahlberg & Gericke, 2018). Conceptual demography is used to discern the potential meaning-making capacity of textbooks' descriptions in selected chemistry and biology textbooks. By doing so, it captures the density and textual properties related to spatiotemporal relationships, that is relationships in space and time, of technical terms. In this thesis, conceptual demography serves to discern the meaning-making potential of protein synthesis descriptions and, consequently, the underlying ideas and subject specific contextual differences expressed through the concepts that describe it.

Study III (Wahlberg, Haglund & Gericke, 2025) extends the line of inquiry by applying the theoretical framework of *conceptual metaphor theory* (Lakoff & Johnson, 1980) to identify and characterise figurative language structures. This framework is grounded in the idea that metaphors are not just merely linguistic expressions but reflects underlying conceptual structures. Metaphors connect one domain to another by mapping elements from a source domain onto a target domain. These mappings are grounded in recurring embodied experiences that arise from how we move, perceive and interact with our environment. Such experiences, in turn, shape our cognitive processes, behaviours, and cultural practices. In this thesis, I employ this framework to discern underlying conceptual structures by identifying the metaphors used in textbooks in descriptions of protein synthesis and analysing how the connections between source and target domains are structured.

Study IV introduces the theoretical framework of *mechanistic reasoning* (Moreira et al., 2019) with a focus on the levels of sophistication identified in students' descriptions. This theoretical framework originates from Machamer et al. (2000), was expanded by Russ et al. (2008), and later refined by Moreira et al. (2019), who describes four relevant categories for analysing mechanistic reasoning: entities, activities, properties, and organisation. In this thesis, these categories were applied to protein synthesis while omitting the fifth category of representations. These categories are used to identify the sophistication of mechanistic reasoning. The framework distinguishes four levels: the descriptive level (Level I), the relationship level (Level II), the simple causal level (Level III), and the emergent mechanistic level (Level IV). In this thesis, the students' written responses were used to determine the levels of sophistication in mechanistic reasoning.

Although these four studies employ different conceptual and methodological perspectives, they converge on several central points. Firstly, domain-specific language in the molecular life sciences is treated as a fundamental cognitive tool that both shapes students' understanding and makes that understanding visible. Secondly, they share the view that representations, whether they are conceptual, linguistic or figurative, mediate learning by structuring how students encounter and understand disciplinary knowledge. Thirdly, they collectively demonstrate that students' understanding is contingent upon how concepts are structured and linguistically framed within instructional contexts, in this case, the textbook and students' own writing. Taken together, these studies illuminate students' challenges with learning protein synthesis. A central insight is that these challenges are not solely attributable to the inherent complexity of protein synthesis itself, but also how the associated concepts are introduced, organised, and framed in an educational setting. By examining this issue through complementary analytical and conceptual lenses, this thesis offers a multifaceted account of how conceptual structures and linguistic representation shape students' opportunities to develop a coherent understanding of protein synthesis.

Methodology

The educational context – Swedish upper secondary school

Swedish upper secondary school is a non-compulsory school for students aged 16-18. Students enrol in either academic programs that prepare them for university studies or vocational programs that lead to occupations such as healthcare, industrial, or media work. All programs include common core subjects such as Swedish, English, science studies, mathematics, and physical education and health (Swedish National Agency of Education, 2013). The specific science subject courses of chemistry, biology, and physics were included in the national Natural Science Programme. Since protein synthesis is used as an example of a complex molecular life science system, the studies focused on parts of the molecular life sciences components included in both chemistry and biology courses of the National Science Program (Swedish National Agency of Education, 2011a; 2011b).

Study I was undertaken during the Lpf 94 curriculum (Ministry of Education, 1994). where protein synthesis was studied in the Chemistry B (Swedish National Agency of Education, 1994b) and Biology A (Swedish National Agency of Education, 1994a) courses. Studies II-IV were carried out during the Gy11 curriculum where protein was studied in the Chemistry 2 and Biology 1 courses. In both curricula, chemistry and biology were organised into two sequential courses of approximately 80-100 hours' duration (Chemistry 1 and 2, or A and B; Biology 1 and 2, or A and B). The syllabi shared a common structure and began with an introduction, an articulation of the aim of the subject, specification core content, and knowledge requirements (Swedish National Agency of Education, 2011a; 2011b). Course 1 in Gy11, and Course A in Lpf 94 in each subject built on compulsory school education and was succeeded by course 2 in Gy11 or Course B in Lpf 94 in the same subject. More specialised courses in the science subjects, for instance in biotechnology or molecular biology, could also have been offered, depending on the school's profile and competences.

In the Lpf 94 curriculum, the syllabi for Chemistry B specified the goals that the student were expected to achieve upon completing the course (my translation):

- Be able to describe, in schematic terms, the structure and properties of some biologically important molecules, and describe the main biochemical features of the cell's metabolism and reproduction.

(Swedish National Agency of Education, 1994b)

The Biology A course in the Lpf 94 curriculum specified that the goals that the student must achieve after completion of the course were (my translation):

- Have knowledge of the structures of the genetic material and understand the relationships between these structures and the characteristics of the individual.
- Have knowledge of methods for genetic engineering, and be able to discuss the possibilities and risks of genetic engineering from an ethical perspective.

(Swedish National Agency of Education, 1994a)

The core content section of the Chemistry 2 syllabus specified that instruction should address the following core content:

- The genetic flow of information, including the main elements of the replication of biochemical processes, transcription and translation.
- The main facets of human metabolism at the molecular level.
- Structure and function of proteins, with special focus on enzymes.

(Swedish National Agency of Education, 2011b)

The core content section of the Biology 1 syllabus encompassed an even broader scope, specifying that teaching should address the following core content:

- Properties and functions of eukaryotes and prokaryotes.
- The structure of gene pools and the laws and mechanisms of heredity. Cell division, DNA replication and mutation.
- Gene expression. Protein synthesis, monogenic and polygenic characteristics, heredity and environment.
- Genetic applications. Opportunities, risks and ethical issues.

(Swedish National Agency of Education, 2011a)

These excerpts indicate that protein synthesis was addressed across both chemistry and biology syllabi, albeit with distinct disciplinary emphases. In the Chemistry B syllabus, the emphasis was placed on specific molecules and on cellular metabolism and reproduction. In contrast, the Chemistry 2 syllabus addressed the topic more implicitly by framing it as genetic flow of information and the processes of transcription and translation within a biochemistry context. The

Biology A syllabus instead highlighted knowledge of the structures of genetic material, and an understanding of how these structures relate to individual characteristics, as well as genetic engineering and ethical concerns associated with it. In comparison, the Biology 1 syllabus defined protein synthesis as core content within the life science context. Both Chemistry 2 and Biology 1 syllabi, emphasised the use of domain-specific concepts in the description of protein synthesis and its sub-processes. However, the biology syllabus additionally required the students to demonstrate an understanding of inheritance mechanisms also at the lowest performance levels.

Study I

Sample of respondents in Study I

The respondents in Study I, as presented in Table 1, were eleventh year students at an upper secondary school in a medium-sized town in the middle of Sweden. All of them were 16 years of age or older, and science majors. They had all taken the Chemistry B and Biology A courses. Syllabi of these courses included protein synthesis. A convenience sample was formed of 12 students from three classes, who volunteered to participate.

Data collection and analysis in Study I

The students were interviewed in groups since peer discussions reportedly enhance data collection and richness of the acquired material (Kvale, 1989). This has also been specifically corroborated for discussions concerning molecular life sciences (Smith et al., 2009). Two groups of three students and three groups of two students were enrolled in the study, formed by application of convenience criteria. The interviews were audiotaped. The interviews followed recommendations to divide each interview into defined phases (Kvale, 1997): introductory briefing phase, main phase, debriefing phase, and reflection phase. An additional 'acclimatisation phase' was added before data collection, during which students received some information on the topic of the interview.

Table 1

Overview of Studies I-IV

Study	Data	Analysis
I		
Clusters of concepts	Semi-structured group interviews with 12 Swedish upper secondary science students	Cluster concept mapping, adaptation of Fisher (2000)
II		
Conceptual demography	Swedish upper secondary chemistry (n=4) and biology textbooks (n=3)	Content analysis with the SEMMA-protocol, adaptation of Shmueli et al. (2010), analytic statistics of differences between biology and chemistry textbooks
III		
Metaphors	Swedish upper secondary chemistry (n=4) and biology textbooks (n=3)	Metaphor Identification Procedure (MIP; Pragglejazz group, 2009), conceptual metaphor analysis (Jahic Pettersson et al., 2020; Lakoff & Johnson, 1980)
IV		
Mechanistic reasoning	24 written tasks from Swedish upper secondary science students	Analysis of mechanistic reasoning, adaptation of Moreira et al. (2019)

In the introductory phase of the interview, the students were asked to explain what they knew on protein synthesis. They were then asked to relate sample terms, starting with DNA, gene and chromosome. When the group mentioned protein or polypeptide in any way, the main phase of the interview began. When any of the other technical terms in the sample were mentioned, the group was asked to explain and relate the new term to those they had already mentioned. During each interview, the group was asked to build a concept map (see for instance Novak & Gowin, 1984), while talking about the technical terms. After the interviews, the recordings were transcribed verbatim and the transcribed excerpts were categorised in an iterative process to identify patterns of the students' technical term usage (Novak & Gowin, 1984).

One level of analysis was the scientific accuracy in the students' explanations of the concepts. The students' explanations were analysed separately and compared to scientific explanations of each technical term, a process called *labelling* in Article I.

The analysis in Study I was done at three levels: concept, cluster, and relationship levels. At the concept level, the focus was on the students' ability to accurately explain the technical terms of the sample. The concept level analysis involved identification of technical terms in the transcripts, recording of the order in which they were mentioned, and evaluation of the scientific accuracy of the explanations. At the cluster level, the analysis focused on the students' tendency to group technical terms that they addressed together, i.e. in clusters (Fisher, 2000). The relationship level analysis focused on the students' ability to express scientifically accurate relationships within and between the identified clusters, by comparison of their statements with scientific explanations found in university level texts. The concept maps were regarded as facilitators of communication in the group interviews, rather than material for analysis in that study. Based on the interview transcripts, concept maps were constructed, and the relationships between clusters were derived from these maps.

Studies II and III

Studies II and III are derived from the same dataset (see Table 1) although Study III used updated editions. Certain sections of the chapter will therefore be presented together.

Sample of textbooks and texts in Studies II and III

To obtain a systematically grounded understanding of textbooks' presentations of protein synthesis, a set of textbooks that met the following criteria was selected for Study II: Each book had to be the latest edition commercially available in Sweden at the time of data collection in 2014. They also had to be expressly written to facilitate teaching and learning in the Biology 1, or Chemistry 2 courses, in accordance with the curriculum at the time of analysis, implemented in 2011. This resulted in inclusion of seven textbooks in the study: three chemistry textbooks (Andersson et al., 2013; Borén et al., 2012; Henriksson, 2012b) and four biology textbooks (Björndahl et al., 2011;

Brynhildsen et al., 2011; Henriksson, 2012a; Karlsson et al., 2011). This sample included all the textbooks that could target these courses and were available from the largest publishers in Sweden at the time of data collection.

Notably, several editions of each textbook and a few other textbooks from other publishers were available for teachers to choose from. These textbooks were however omitted to focus on a manageable set of the most widely used textbooks. There was also a substantial time gap between the two studies. Studies II and III used the same dataset, with the addition of newer editions of the books that had been included in Study II and were available at the time of data analysis for Study III in 2020: four textbooks in biology (Björndahl et al., 2017; Brynhildsen et al., 2011; Henriksson & Bosson, 2020; Karlsson et al., 2011) and three textbooks in chemistry (Andersson et al., 2013; Borén et al., 2020; Henriksson et al., 2019). This means that other textbooks that were published between the time for data collection for Study II and analysis for Study III were omitted from the sample to maintain a cohesive dataset across both studies. In Study II, data were collected from the selected textbooks by identifying sections that addressed protein synthesis. This was done by an examination of the table of contents in each textbook to locate chapters or sub-sections focused on protein synthesis or related sub-processes such as translation. Before the data analysis in Study III, the texts were compared with the original editions. Only minimal or no changes were found between the editions.

Data collection and analysis Study II

The collected data were analysed by applying a SEMMA (Sample, Explore, Modify, Model and Assess) approach to content analysis derived from data mining techniques, as proposed by Shmueli et al. (2010). The analysis identified how technical terms were used in selected chemistry and biology textbooks and revealed the texts' conceptual demography, i.e. the frequencies, distributions and relationships of the core and peripheral concepts in them. All the SEMMA steps outlined above were applied except for the final step. The process began by clarifying the purpose of the analysis and to identify the required dataset. The data were then explored, cleaned/reduced, and/or pre-processed after which the computational

task was defined. An appropriate analytical technique, interpreted in Study II as a purpose-built algorithm, was thereafter constructed and implemented.

The algorithm was constructed in Microsoft Excel® to count and pinpoint the selected terms in the sampled texts. The algorithm counted total frequencies of occurrences of each technical term in each text. Summary statistics such as mean values were calculated from its output. The data were extracted from spreadsheets generated by the algorithm. Because the texts differed in length, I used frequencies normalised to text length rather than absolute counts. The frequency of each sample term was extracted from word counts in the spreadsheets, and mean values were calculated. Mean normalised values for the indexed texts representing each context were then derived. A two-sided t-test with a significance threshold of 0.05, implemented in SPSS® (v. 22), was used to identify significant differences between frequencies of terms in the two contexts of chemistry and biology. The distribution of each technical term was characterised by counting the occurrences in each sentence, from the beginning to the end of each text. Relationships between technical terms within each text were defined as their coexistence in a sentence. Each such relationship was counted only once per sentence by the algorithm, regardless of how many times either technical term appeared.

Data collection and analysis Study III

In Study III, the Metaphor Identification Procedure (MIP; Pragglejaz Group, 2007) was used to identify metaphorical lexical units. The procedure began with establishing a general understanding of the text, followed by the determination of the lexical units. Next, each unit was examined to see if it had a more basic meaning in other contexts, typically more concrete, bodily related or more precise. For this step, each lexical unit was compared to the explanation provided by the National Swedish Dictionary (SAOL, Svenska Akademien, 2014) or National Swedish Thesaurus (SAOB, Svenska Akademien, n.d.). A lexical unit was coded as metaphorical if its contextual meaning differed from its basic meaning but could still be understood in relation to the that basic meaning.

Following the approach of Jahic Pettersson et al. (2020), the identified metaphorical lexical units were categorised through the lens of Conceptual Metaphor Theory (Lakoff & Johnson, 1980). In this process, the results from the MIP analysis were inductively paired with source domains and grouped into categories until the authors reached consensus. The categorisation generated a two-level hierarchical system that consists of main metaphor systems and corresponding sub-metaphor systems.

Study IV

Sample of respondents in Study IV

Study IV was conducted in western Sweden, where two upper secondary schools in different towns volunteered to participate. A criterion for inclusion was that the students had studied protein synthesis in at least one of the two subjects of chemistry or biology. This requirement substantially narrowed the pool of eligible schools due to timetabling constraints. Of the approximately 50 invited students, 25 consented to participate, and 24 submitted a written response (see Table 1).

Data collection and analysis for Study IV

Data were collected through a digital written task that I administered in the students' regular classrooms during scheduled lesson hours. An open-ended task was designed by me to elicit students' mechanistic understanding of protein synthesis. Students first completed background questions, including which science courses they had taken, and were then presented with the following prompt: "Picture the following scenario: A friend of yours is struggling to understand how proteins are formed in a living cell and asks you to write an explanation. What would you write?". The students recorded their responses in a textbox within the questionnaire and were given 45-60 minutes, depending on the lesson duration, to complete their written response.

For Study IV, data analysis followed the framework proposed by Moreira et al. (2019) with a focus on students' written descriptions of protein synthesis to identify entities, properties, activities, and organisation features. To establish each student's baseline use of

domain-specific concepts, the total number of words, all unique domain-specific concepts, and total number of domain-specific concepts used in each text were counted. The 16 most frequent domain-specific concepts identified as entities formed the sample for the analysis of mechanistic reasoning. Because domain-specific concepts may appear in multiple linguistic forms, an inclusive categorisation was applied. Properties were defined as descriptive and intrinsic attributes of entities, and activities were coded as actions expressed through active verb forms. Organisation was identified by the spatial orientation between entities.

Through an iterative process examining how concepts were paired with their properties, their activities, and organisational relationships, the unit of analysis was established as each line of reasoning. Sentences were analysed individually, followed by an examination of how domain-specific concepts were used across sentences to articulate the entities, their properties, activities and organisations. Bridging paragraphs were identified by tracing a student's line of reasoning until a conceptual halt or shift occurred, for instance when the description of mRNA transitioned into an explanation of protein composition, without linking the two. Such shifts were treated as separate lines of reasoning. In the final analytical cycle prior to validation, each text was reviewed holistically to ensure consistency and accuracy. Levels of mechanistic sophistication were then determined by iteratively comparing each line of reasoning with the criteria specific by Moreira et al. (2019).

Trustworthiness, reliability and transferability

In Study I, a mixed method approach was adopted (Robson, 2008). The interviews followed the same interview guide, were conducted by the same interviewer and took place at the same school to support the validity of the data. The interview guide was piloted beforehand, and minor adjustments were made prior to data collection. The concept maps that were generated during the interview served as tools for facilitating the respondents' discussion during each interview (Kvale, 1989). At the end of each interview, the students validated their reasoning by the construction of concept maps and were asked to revisit and articulate the constructed concept maps.

In Study II, a novel computer-generated algorithm was created specifically for the project. Rigorous validation was therefore essential to fine-tune it. This process was carried out in iterative rounds, testing multiple sets of texts with diverse content, such as instructions, brochures, and online texts. The output from the algorithm was manually reviewed in repeated cycles to identify and correct errors. As the algorithm had never been run before, spot-checks were also conducted, and the output was cross-checked against original texts to validate the process. Each text underwent successive cycles until no further errors were found. Because all the data are computer generated, the reliability of the study is excellent. It is important to note, however, that the volume of analysed data was lower than in most studies using data mining and a SEMMA-approach. Nevertheless, the protocol was found to be useful even for processing the relatively small datasets in Study II, as described and discussed in the methodology section.

In Study III, I conducted all the steps used in the MIP-procedure. To ensure validity (Robson & McCartan, 2016), approximately 10 % of the sample was jointly validated jointly by the three authors at each stage. I performed the initial analysis, after which the second author reviewed the selected subset and provided feedback. The third author then examined the revised subset and added further comments. The three of us discussed the results until reaching consensus. Regarding both Studies II and III, the textbooks were originally written in Swedish, and English translations were prepared only after the analyses had been completed. These translations were rendered as literally as possible to ensure that the analytical focus remained on metaphor use rather than on the specific nuances of the Swedish wording.

To validate the analysis for Study IV, two of the authors independently analysed several texts and resolved differences through discussions until consensus was reached. One researcher was an expert in genetics education and mechanistic reasoning, while the other, namely I, was trained in educational aspects of protein synthesis. Both of us were skilled in students' use of domain-specific concepts in a molecular life science. Although neither of us had previously applied the mechanistic framework (Moreira et al., 2019), the iterative consensus-building process strengthened the validity of the analysis.

In terms of transferability, the data has been collected and analysed over an extended period. The results are pointing in the same direction, that domain-specific language when communicating protein synthesis is still challenging. That would indicate that although the data comes from a limited cohort, it is reasonable to assume that students in other upper secondary schools would encounter similar challenges, and that textbook presentation would vary little over time.

Ethical considerations

All studies followed the ethical guidelines for data collection and analysis published by the Swedish Research Council (2002 to current edition), as well as the local ethical vetting procedure in place at Karlstad university at the time of data collection and analysis. Studies I and IV were conducted using human participants, whereas studies II and III were carried out using commercially available textbooks. Data for Studies I and IV were gathered from human participants who were all over 16 years of age. Participation was therefore decided by the students themselves. The two studies were conducted with a considerable timespan between them, which meant that different regulations, both locally and those issued by the Swedish Research Council, were in effect at the respective times of data collections. Before being interviewed or asked to fill out the written task, the participating students received clear written descriptions about their participation. In accordance with the local guidelines at Karlstad University, they were informed about the purpose of the study, their right to withdraw from the study at any time without giving any reason and without any penalty, data treatment, and, specifically for Study IV, information on GDPR. All participants consented to participation prior to data collection. Studies II and III also underwent ethical vetting prior to data collection and analysis, and the data was collected accordingly. When working with data from commercially available written textbooks, the requirement for assessing human participation is not regarded in that sense.

Results

Main results of Study I

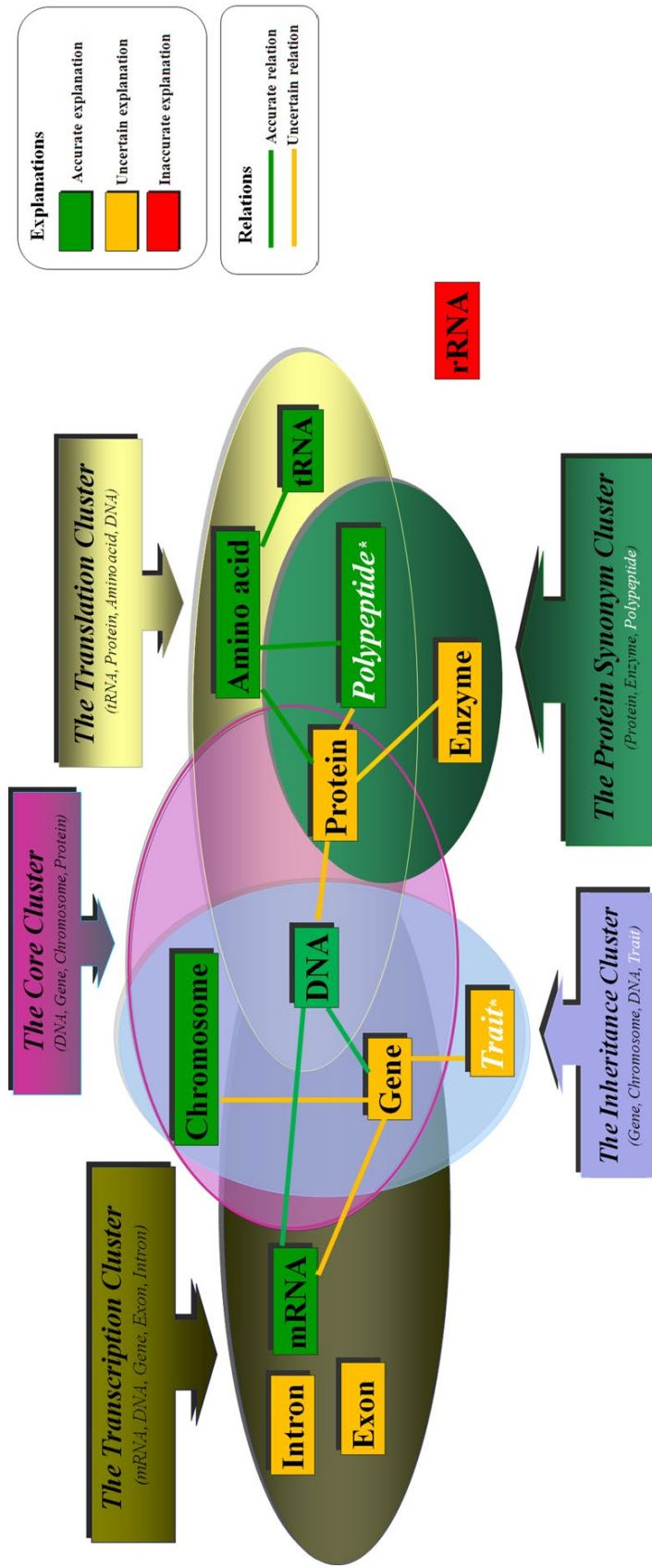
The main findings from Study I show that upper secondary science students struggle to comprehend the overarching conceptual structure of protein synthesis. Instead, students tend to compartmentalise their understanding by what can be described as ‘islands of technical terms’, here referred to as *clusters*. Technical terms that the students used together were grouped into clusters. Five such clusters of concepts were identified in the students’ recorded reasoning (incorporated technical terms are shown in brackets, see also Figure 3):

- a) Core Cluster (chromosome, DNA, gene, and protein)
- b) Transcription Cluster (DNA, exon, gene, intron, and mRNA)
- c) Translation Cluster (Amino acid, DNA, protein, and tRNA)
- d) Protein Synonym Cluster (Enzyme, polypeptide, and protein)
- e) Inheritance Cluster (Chromosome, DNA, gene, protein, and trait)

The core concepts of DNA and chromosome were generally labelled correctly by the students, but they showed vague comprehension of the concepts of gene and protein. The peripheral concepts mRNA, tRNA, and amino acids were correctly explained. The students showed weak understanding of intron and exon. The concept trait was not part of the original sample and was introduced spontaneously by the students. However, they displayed a vague comprehension of the concept and its relationships.

As can be seen as green lines in Figure 3, the students could relate the core concepts gene–DNA, and link core and peripheral concepts through the mRNA–DNA relationship. However, they struggled with the link between mRNA and gene as shown as yellow lines in Figure 3. The relationships between technical terms that are related to later stages of protein synthesis, tRNA–amino acid, and amino acid–protein, were addressed correctly, but the tRNA–protein relationship was not detected in the students’ recorded reasoning. Technical terms associated with the maturation of mRNA, i.e. intron and exon, were not related to any of the other concepts in the sample. Hence, they lack connections in Figure 3.

Figure 3
Labelled Concepts, Identified Clusters, and Relationships



Note. Figure from Article I. Concepts shown in black colour text are part of the original sample. Concepts in white colour text were introduced spontaneously by the students. Reprinted with permission from the journal. Original image: <https://doi.org/10.1080/00219266.2012.716785>.

The key bridging relationships were those that were linking the cluster of core technical terms (DNA–Protein) to three other clusters: the protein synonym cluster, the translation cluster, and inheritance cluster. The core cluster was also bridged to the transcription cluster via the DNA–mRNA relationship. The same relationship also served to connect the transcription cluster to the translation cluster. These five identified clusters reveal a compartmentalised and fragmented understanding of protein synthesis among upper secondary students. Although students frequently invoked core concepts, their reasoning tended to organise these concepts into discrete, poorly connected clusters, rather than into an integrated explanatory framework.

Overall, the way in which students cluster concepts point to a conceptual landscape where they possess many of the relevant technical terms but struggle to organise them into a coherent description of protein synthesis. Their reasoning reflects an accumulation of isolated vocabulary rather than an integrated understanding of the processes that connect DNA, gene structure, and protein formation. This conceptual fragmentation highlights the need for instructional approaches that explicitly support students in constructing a unified conceptual framework rather than merely acquiring isolated technical terms.

Main results of Study II

The findings from Study II demonstrated pronounced context-dependent differences in how chemistry and biology textbooks conceptualised and described protein synthesis. These differences were identified through an analysis of the conceptual demography characteristics of each disciplinary context. Chemistry textbooks devoted considerable attention to peripheral concepts, most notably the tRNA, and did not differentiate between the relative importance of core and peripheral concepts. Biology textbooks, in contrast, consistently emphasised core concepts. Although both contexts adhered to the canonical representation of protein synthesis, chemistry textbooks further compartmentalised the process through dense technical terminology, outlining transcription, splicing and translation as discrete sub-processes. This approach in the chemistry textbooks produced a disproportionate emphasis on the latter stages of protein

synthesis, where tRNA was positioned as the central concept. Biology textbooks also acknowledge the role of tRNA, but with lower frequency and fewer conceptual relationships. Technical terms associated with the starting point as presented in the canonical representation, DNA and gene, are treated differently in the two contexts. The term gene appears with greater evenness throughout the length of the protein synthesis description in biology textbooks, than across chemistry textbooks. While DNA exhibits similar distribution patterns in both contexts, chemistry textbooks place slightly greater emphasis on the term in the first half of their texts. The concept of mRNA emerged as central and was evenly distributed in both disciplinary contexts. However, biology textbooks employed the term more frequently in absolute numbers and relate it to a wider array of core concepts which indicated a more central position than in chemistry textbooks.

Technical terms associated with mRNA maturation, intron and exon, receive limited attention overall. When they do appear, they are confined to short, conceptually dense mid-text sections. Biology textbooks differ slightly by reintroducing these terms towards the end of the texts, a pattern not observed in chemistry textbooks. Finally, the product of protein synthesis is framed differently across the two contexts. Chemistry textbooks employ both peptide and protein, whereas biology textbooks exclusively use protein. This distinction could reflect a disciplinary distinction in terminology and conceptual framing. Such differences warrant careful consideration in educational settings as they have implications for how students construct and integrate their understanding of protein synthesis.

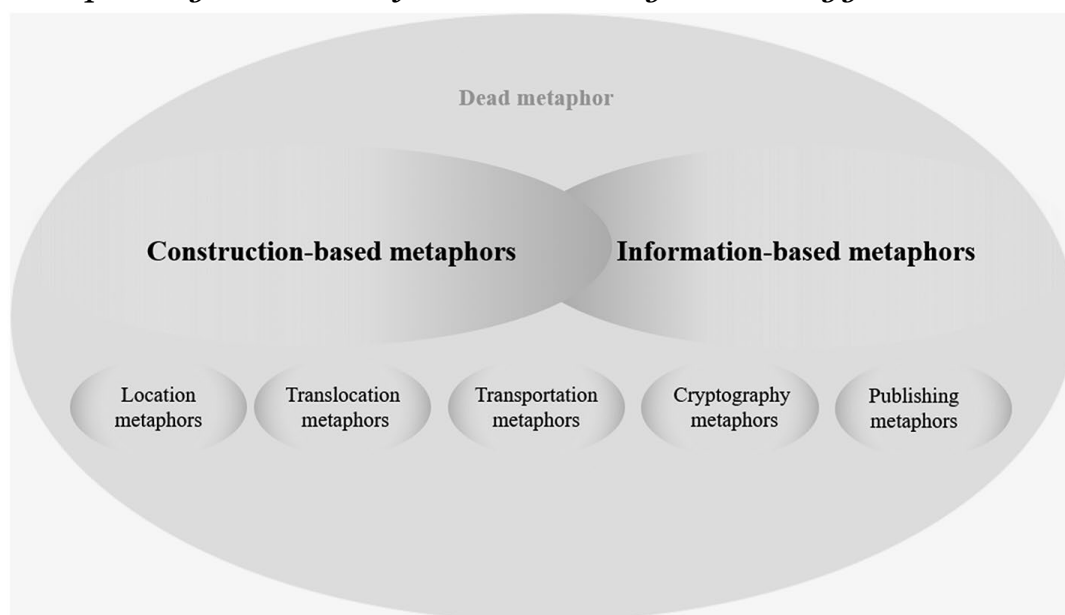
Main results of Study III

The main findings of Study III show that chemistry and biology textbooks employed an extensive range of metaphors drawn from diverse source domains to conceptualise protein synthesis. These metaphors primarily fell into two main overarching metaphor systems of *construction-based metaphors* and *information-based metaphors*, which together provide parallel explanatory frameworks. Construction-based metaphors drew on language from the building industry as their source domain. They typically frame concepts as structures with components where things can be built, made of parts,

or assembled. Textbooks use both nouns and verb forms. Information-based metaphors instead borrow from informatics where the transmission of information to a recipient is emphasised. This system included terms such as publishing, message, and cut and paste.

Under these systems, five sub-systems emerged (Figure 4): the *location sub-metaphor system* addressing spatial orientation; the *translocation* and *transportation sub-metaphor systems* regarding spatial movement; and the *cryptography* and *publishing sub-metaphor systems* regarding the handling of messages. These metaphor systems are however not applied exclusively. Instead, textbooks often shifted between source domains in their descriptions of the same target process, where the choice of metaphor varied according to context. As a result, the metaphors form an interconnected network that produced a more complex overall conceptual structure.

Figure 4
Metaphor Systems Identified in Chemistry and Biology Textbooks



Note. The main metaphor systems are the construction-based metaphors and the information-based metaphors. The sub-metaphor systems are the location, transportation, translocation, cryptography, and publishing sub-metaphor systems. Both the cryptography sub-metaphor system and the publishing sub-metaphor system are more related to the information main metaphor system than the other three sub-metaphor systems. They are therefore placed further to the right in the figure. Original image found at: <https://doi.org/10.1007/s11165-024-10197-y>

The location sub-metaphor system drew on spatial prepositions to describe orientation, by metaphors such as in, on, or under, with the in-metaphor particularly emphasised in the textbooks. These metaphors framed structures as confined spaces that could contain something and thereby functioned as conceptual containers. Examples were the cell or the ribosome.

The translocation metaphor system highlighted movement from one location to another with an emphasis on starting and endpoints as well as direction. Verbs such as walk, move, and flow signalled this metaphorical sub-system. For instance, this could be applied to the movement of the ribosome when it interacts with mRNA.

The transportation sub-metaphor system drew on meanings from the transportation sector to depict movement between two points facilitated by an active agent. This was typically realised through verbs such as haul or carry, or through nouns that denoted the active transporter, such as a carrier. In the textbooks, this metaphor was frequently applied to transfer RNA (tRNA) framed as a transporter of amino acids.

The cryptography sub-metaphor system drew on the specialised language of cryptography as a source domain. Terms such as codes and ciphers were used to describe how genetic information was transmitted and interpreted, which further emphasised the metaphorical framing of genetic processes as acts of encoding and decoding.

Finally, the publishing sub-metaphor system drew on vocabulary from writing and publishing industry as a source domain. Terms such as write, rewrite, template, imprint, copy, paste, and cut appear in this system. The textbooks employed metaphors from these domains when describing the process from DNA to immature mRNA, and through the subsequent maturation of mRNA. All textbooks used some version of a blueprint metaphor, such as template, imprint, or both, to explain the structure and function of DNA. The notion of a template referred to DNA or the gene itself but could also be described as the mRNA either before or after splicing.

Main results of Study IV

The main results from Study IV show that students were able to use domain-specific concepts throughout their explanation of protein synthesis, from genetic material to protein, but they compartmentalise their explanation into separate sub-processes with different level of sophistication within the explanation of each sub-process. The students' texts contained over 40 unique domain-specific concepts. The analysis focused on the 16 most frequent concepts, all of which were aligned with the central dogma. The concept of protein synthesis was not prioritised, while mRNA, protein and amino acids were the most common concepts. Students referred to DNA more often than the gene, highlighted nitrogenous bases over nucleotides when they describe functional impacts, and tended to focus on translation-related concepts, with ribosome-related actions being the only organisational elements that were consistently described.

The sophistication of the students' mechanistic reasoning varied both between individuals and within their explanations of different sub-processes of protein synthesis. This variation appeared to stem from how students compartmentalised the process of protein synthesis into smaller segments and sometimes fragmented further than the underlying sub-processes and biochemical steps. The results also show that the level of mechanistic reasoning is generally higher when students describe translation than when they describe other sub-processes in the protein synthesis. Although students were able to display rather sophisticated mechanistic reasoning, they tend to do so in isolated areas, mainly one sub-process at a time without relating the areas of text to another. Across the explanations, characteristic domain-specific concepts and structural elements marked each level of reasoning. About half of the students began at the least sophisticated levels and only the later stages of protein synthesis, particularly the translation process, tended to elicit higher-level reasoning.

Level I explanations of protein synthesis were characterised by a reliance on structural descriptions or simple restating of the concept of protein synthesis. These descriptions often focused on identifying structures or locations rather than processes. Level II explanations centred on simple relationships and demonstrated some recognition of

some relevant properties and activities. Most students who did not begin their explanation at a Level I, started their explanations Level II, meaning that their initial explanations typically involved basic descriptions or straightforward relational statements. Level III explanations, in contrast to Levels I and II, were characterised by causal reasoning. The few students who began their explanations at this higher level either focused solely on translation or maintained a high level of sophistication throughout their explanations. Finally, Level IV explanations represent emergent explanations of protein synthesis in which entities, organisation, activities, causality, and properties co-functioned. Only five of the 24 students displayed Level IV explanations, and only one student consistently provided Level IV reasoning across the entire process of protein synthesis.

Discussion

The aim of this thesis is to provide insights into how domain-specific language is structured and used in upper secondary school to communicate complex molecular systems, using protein synthesis as an example. In the discussion section, I focus on critical facets derived from the results that have implication for teaching and learning protein synthesis. These facets embrace and build on the facets that were identified in the licentiate thesis. In this thesis, a *facet* is understood as a factor that either facilitates or hinders teaching and learning of protein synthesis in relation to domain-specific vocabulary usage (Wahlberg, 2019). I discuss the following aspects: potential learning obstacles associated with the facet, associated meaning-making capacities, and the ways in which the chemistry and biology contexts influence the content.

In the licentiate thesis, I identified four facets derived from the categorisation of the results from Studies I and II: mechanistic or conceptual descriptions; compartmentalisation; mRNA as a core concept; and canonical representation. In this doctoral thesis, these have evolved into facets on **mechanistic reasoning** (incorporating the previous facet on mechanistic and conceptual descriptions); **compartmentalisation** (expanded to include findings from Studies III and IV); **mRNA as a central core concept** (integrating the earlier facet on mRNA as a core concept); **metaphors**, which has developed into a distinct facet in its own right; and **the canonical representation in molecular biology** (also expanded to include findings from Studies III and IV).

Facet on mechanistic reasoning

This facet is concerned with the level of mechanistic reasoning identified in the students' understanding. This facet addresses the importance of mechanistic reasoning for a coherent understanding of complex molecular systems. As an emergent phenomenon, protein synthesis involves a multitude of interactions and many entities at multiple organisational levels in the cell. In addition, there is a spatiotemporal aspect of the whereabouts of the entities, their interactions and their activities at any given time.

Students' difficulties across Studies I and IV point towards a consistent mechanistic challenge. The students could recall and use a suitable package of domain-specific vocabulary, yet they struggled to integrate the entities, activities and organisations into a coherent spatiotemporal explanation of the complex molecular system, here exemplified with protein synthesis. Their reasoning remained compartmentalised with accurate descriptions of isolated sub-processes but was limited in their ability to connect these systems into emergent system-level mechanisms. Furthermore, written descriptions in the textbooks, analysed in Studies II and III, followed the same pattern of compartmentalisation and lacked emergent mechanistic relationships. This pattern of leaving sub-processes and the links between them unidentified, unbridged, or invisible suggests that there are black boxes (Haskel-Ittah, 2023) that students fill with memorised terms (Machamer et al., 2000) rather than engaging in causal reasoning. Across the four Studies I-IV, a central insight is that there is a challenge in integrating domain-specific concepts into a coherent, emergent mechanistic explanation of the studied complex molecular phenomenon.

To ensure a consistent emergent mechanistic reasoning at Level IV, a student must not only identify the central participating entities, their properties, activities and organisation, but also do so for every single event that occurs in the cellular process. It was found in the results in Study IV that the students were unable to reach or consistently sustain high-level mechanistic reasoning throughout the protein synthesis explanation. A similar pattern was found in the textbooks in Study II, where relationships between sub-processes were often left implicit. This pattern aligns with the findings of Duncan and Reiser (2007), that students need explicit support to understand how biological processes are causally linked. Contrasting the findings in Study IV with the Gy11 syllabus in Biology 1 (Swedish National Agency for Education, 2011b), which explicitly requires students to express mechanistic reasoning, it appears unlikely that students could realistically meet these expectations based on the findings of this project. Consequently, students may be left unprepared to fulfil the stated requirements.

In contrast to detailed mechanistic descriptions found in chemistry textbooks texts on protein synthesis in Study II, conceptual reasoning that was identified in the biology textbooks has the advantage of providing explanations of overarching structures. Thus, chemistry and biology textbooks seem to have different aims in their presentations of protein synthesis. Students may take biology and chemistry courses at different points in their educational timetable, or they may study them in parallel. When courses are separated in time, educational challenges can arise. When students encounter material later in their education, they need to revisit fundamental concepts from early courses. At that stage, however, their time and cognitive resources are often already committed to other subjects which can be ineffective and potentially overwhelming. Conversely, separating the courses can risk weakening conceptual connections between chemistry and biology that are essential for developing a coherent understanding of life sciences. Studying chemistry and biology in parallel avoids the temporal gap but exposes students to disciplinary differences between the subjects which may create its own set of learning difficulties.

I would argue that a suitable order, at least for learning protein synthesis, is to take Biology 1 before Chemistry 2. In this order, students first encounter the overarching structures of protein synthesis in biology and later deepen their understanding in chemistry. This progression allows black boxes that were introduced in biology to be unpacked with higher level of sophistication of the mechanistic reasoning in the subsequent chemistry course. Haskel-Ittah and Yarden (2018) argued that introducing central underlying mechanisms early could support student learning. There is no inherent contradictory in doing so within a biology context. However, findings in this thesis suggest a potential risk that beginning with a detailed mechanistic reasoning rather than overarching ideas may lead to compartmentalisation in the students understanding. I conclude that students either struggle to construct an emergent mechanistic explanation, or that instruction material, here the written textbook, fails to provide a coherent mechanistic approach to complex molecular systems based on protein synthesis as the example.

Facet on compartmentalisation

This facet concerns the compartmentalisation that occurs throughout the explanation of protein synthesis. The constructed maps in Study I, the conceptual demography descriptions in Study II, the metaphor systems in Study III, and the isolated emphases in explanations in Study IV all demonstrated fragmented ideas. A characteristic pattern in students' conceptual organisation in Study I is the emergence of isolated 'concept islands'. Relationships were identified between a limited subset of domain-specific concepts while maintaining conceptual compartments, or clusters.

Areas between clusters remained unknown to the students and formed black boxes (Haskel-Ittah, 2023) in the students' reasoning. For instance, in student S1's Level I description in Article IV, a black box was identified between the statements "certain sequence of nitrogenous bases" and "which can form a specific kind of protein". In this case, the black box represents a rather large gap in the student's reasoning where the nitrogenous bases enter a mental black box in one end, and a protein shows up in the other. The steps between these points were missing in the student's reasoning and left the reader unable to see how the student arrived at the conclusion. Black boxes could indeed serve as temporary hallmarks for the students' conceptual development (Machamer et al., 2010), but Study IV suggests that even after completing relevant courses, these black boxes have not yet been unpacked by the students. According to Novak and Cañas (2006), students' ability to bridge concepts within and across conceptual compartments reflect the depth of their conceptual understanding. To develop a coherent understanding of cellular processes therefore requires that students recognise and integrate the relationships among central domain-specific concepts (Lewis et al., 2000).

Collectively, these findings suggest that more integrated representations of conceptual relationships, both within and between clusters of concepts, may support deeper learning in molecular life sciences if used to deepen knowledge structures or to unpack black boxes in the understanding. The findings in Study IV reinforced this point by showing that students' fragmented lower-level sophistication of mechanistic reasoning could be originated from a

compartmentalised way of thinking. I therefore conclude that protein synthesis is persistently and inherently presented in a compartmentalised manner in upper secondary education.

Facet on mRNA as a central core concept

This facet concerns how one domain-specific concept, mRNA, was found to be central in explaining protein synthesis. Results from Study I indicated that mRNA, initially regarded as a peripheral concept, should instead be understood as a core concept in descriptions of protein synthesis. This interpretation was corroborated across all four studies, in which mRNA consistently functioned as a central connector linking multiple sub-processes in protein synthesis. Despite this centrality, surprisingly little research has focused on teaching and learning related to mRNA.

Given its frequent use by both textbooks and students, together with its strong connections to many other domain-specific concepts, mRNA warrants explicit attention in research on students' understanding of protein synthesis. As Tibell and Rundgren (2010) emphasise, the essence of a technical term lies in its meaning-making capacity. The position, definition and relationships of mRNA with other technical terms in textbooks underscore its importance in learning protein synthesis. The extensive linking of mRNA to other technical terms contributed to a net-like conceptual structure, as defined by Kinchin et al. (2000), in both students' reasoning in Study I and textbooks' presentations in Study II. mRNA could thus be regarded as a bridging technical term, that interconnected several clusters in students' reasoning in Study I, and most of the technical terms examined in the textbooks in Study II. In Study IV, the students highly acknowledged mRNA, and in Study III, the metaphors used to describe mRNA shifted throughout the protein synthesis description. Because the meaning of mRNA changes throughout a description, an unguided or unclear change could risk fragmenting the students understanding. Based on the findings, I conclude that the technical term mRNA is crucial for a holistic understanding of protein synthesis.

Facet on metaphors

This facet is concerned with the use of metaphors in descriptions of abstract complex molecular systems. From the results of Study III, protein synthesis can be conceptualised through two distinct metaphorical frameworks: construction-based metaphors and information-based metaphors. Each framework represents a separate system that offers its own vocabulary and perspective for how protein synthesis works. As could be seen in Study III, these two main metaphor systems are not used in isolation. Instead, different metaphor systems can be applied in different situations that involve the same entities, for instance mRNA which can be explained as being constructed, transported or translated, or carrying information. An undeliberate use of the shifts between metaphor systems could obscure the student's understanding of what the metaphor is meant to convey at any given moment, especially if they are unable to track the shifts between systems. This means that they need guidance on where and how the shifts happen, as well as on the limits of that explanation.

Under the two main metaphor systems, there are sub-metaphor systems that are associated with either of these. As noted in the background section, some of the domain-specific language used to communicate protein synthesis is rooted in the field of cryptology and informatics (Kay, 2000). These metaphors remain tied to their original source domains, but their explanatory power may have been diminished because they no longer connect to familiar, everyday experiences. An example would be the reference to code keys in old spy novels, as addressed in Article III. This requires students to unpack the contextual meaning of the terms, explain what a code key is and how it functions, and then relate that function to the conventions of the spy novel genre. They must then understand the connection to the target domain to fully understand the metaphor. When old spy novels were common cultural references, this metaphor would have well. Given their probable decline in general relevance of old spy novels today, the metaphor loses resonance and the conceptual bridge between source and target domains may weaken or disappear. When that happens, the black boxes (Haskel-Ittah, 2023) that may appear in students' understanding cannot be filled because they become effective invisible and linguistically inaccessible.

As shown in the results reported here, dead metaphors (Rundgren, 2006) are present in the textbooks. That means that today, when for example splicing of ropes is merely a hobby or strictly associated with specific occupations rather than everyday life as it was some 70 years ago, the technical term splicing may not be as easily understood through its original metaphorical ground. It is therefore important to recognise how societal fluctuations influence language more broadly, Kay (2000) cautioned, and how these fluctuations specifically affect domain-specific vocabulary and the selection and use of metaphors. Based on these findings, I conclude that the selection and use of metaphors are key features in a successful mechanistic reasoning.

Facet on the canonical representation in molecular biology

This facet concerns how protein synthesis is represented, as outlined in the background chapter, either through the depiction of protein synthesis via the linear central dogma of molecular biology (Crick, 1958), or the canonical representation definition used nowadays. All four studies in this thesis rely on the central dogma of protein synthesis in some way, however with varying emphasis. One difference identified in Study II between chemistry and biology textbooks was that chemistry textbooks tended to emphasise the T-model, with a clear focus on the translation sub-process. In contrast, biology textbooks employed an overarching structure aligned with the TST-model. Similarly, the results in Study IV, which highlight students' varying emphases on the sub-processes of protein synthesis, mirror these findings: some students provided elaborate answers that covered the TST-model, while others reduced protein synthesis to the translation process alone, reflecting descriptions found in chemistry textbooks.

The Swedish upper secondary chemistry syllabus (Swedish National Agency of Education, 2011b) required a mechanistic presentation of protein synthesis, with emphasis on its sub-processes, particularly translation and transcription. In contrast, the biology syllabus (Swedish National Agency of Education, 2011a) placed greater emphasis on the overarching structure of protein synthesis. Although the curricula stated what content to teach rather than how it should be taught, it is reasonable to assume that the different curricular contexts of chemistry and biology influenced how protein synthesis was

presented and understood in the classroom. Notably however, neither of the Chemistry B or Chemistry 2 (Swedish National Agency for Education, 1994a; 2011a), Biology A or Biology 1 syllabi (Swedish National Agency for Education, 1994b; 2011b) required explanation framed explicitly in terms of the central dogma. Yet, the linear explanatory model persists in students' reasoning and continues to dominate textbooks' explanations today. I conclude that although there are no explicit curricular requirements to teach according to the central dogma of molecular biology, its underlying structure has nevertheless persisted over time.

Methodology discussion

A central strength of this thesis is the consistent focus on domain-specific language within the given setting of understanding and describing complex molecular systems in upper secondary education with protein synthesis as the example. This allows the same issue to be examined from multiple perspectives. A key strength across the methodologies in the four studies lies in the fact that they were carried out by using texts or speech in Swedish, but it would also have been possible to use texts and speech in another language such as for instance in English. The words that were involved in the analysis of the clustering of concepts, conceptual demography, metaphor identification, and mechanistic reasoning could have been adapted, modified, or translated to the desired language without altering the underlying analytical approach. However, the nuances of each language, for instance how prepositions are used in different languages, will inevitably influence how concepts are rendered. This could make translation of such words into a methodologically delicate process that is sensitive to interpretation.

A central limitation of the methodology is that the studies were conducted during different curricula, each with its own syllabus requirements. Although broadly comparable, subtle differences in course design and emphasis may have influenced how the content was presented to students. A more coherent project design with a larger and more continuous sample would have strengthened comparability across studies, though this doctoral thesis aims for qualitative exploration rather than statistical comparison. For this purpose, the

sample size in each study was sufficient, and data saturation was achieved. During the project, the commercial textbook market shifted as the curriculum changed from Lpf 94 to GY11. Several major textbooks that were released during the transition period had to be excluded to maintain a coherent dataset. Although only minor changes were made between some of the included textbooks it would, in retrospect, have been preferable to include all textbooks available since the initial data collection to more clearly underpin that statement.

Study I was conducted prior to the implementation of the curriculum for Swedish upper secondary school in 2011, under which Studies II-IV were carried out. Although no systematic comparison was made between teaching and learning of protein synthesis, and the content of the textbooks, it seems that the textbook treatment tends to change slowly over time, at least in the absence of a paradigm shift. Accordingly, it seems reasonable to assume that the selection of technical terms in Studies I and II, would show only little variations if the studies were to be repeated under present conditions. This was supported by the findings in Study IV that the students use the same sample of domain-specific concepts.

The technical term mRNA however, emerged in Study II as considerably more salient than previously acknowledged in earlier research. Because this insight was not available at the time of Study I was designed, it could not be incorporated into the methodological decisions that underpinned that study. Were Study I to be conducted again, particular attention would be devoted to the role of mRNA as a core concept in the selection of the technical terms. I conclude that the use of semi-structured interviews and written tasks for assessing students' understanding in Study I, was successful as it made students discuss with each other and provide data which revealed several patterns of understanding. Furthermore, the use of concept maps (Novak & Gowin, 1984) in Study I, and mechanistic reasoning framework in Study IV (Moreira et al., 2019), revealed structural features of students' reasoning that would have been difficult to identify through other methods.

In Study II, I employed content analysis supported by linguistic data mining techniques. Content analysis is a traditional approach for investigating frequencies of technical terms (see e.g. Krippendorff, 2004; Graneheim & Lundman, 2004). In this study, however, I expanded the approach by analysing not only frequencies, but also distributions and relationships between technical terms, in what I named ‘conceptual demography’. Adding this distribution and relationship dimension created a 3D-view that traditional content analysis cannot provide. Data mining typically relies on computational operations to detect patterns that would be difficult or impossible to identify manually (Feldman & Sanger, 2007). The computational operations often use enormous quantities of text and large semantic networks (see for instance, Shmueli et al., 2010). The dataset in Study II was far smaller, and the analytical complexity was correspondingly lower. Still, I needed a method that could reveal more complex structures beyond what ordinary content analysis could offer.

While commercially available software packages could have handled specific aspects of the conceptual demography (e.g., most text-editing programs provide straightforward word counting facilities), no program could analyse frequencies, distributions and relationship components of the conceptual demography. Therefore, a tailored algorithm was constructed to specifically target the relevant conceptual demography for Study II, yet easy to operate. I conclude that the developed algorithm for Study II successfully revealed the structural features of the textbooks use of domain-specific concepts. Although the dataset was considerably smaller than usual data mining sets, the SEMMA procedure (Sample, Explore, Modify, Model and Assess; Shmueli et al., 2010) proved to be methodologically productive. Furthermore, the metaphor identification protocol (Pragglejaz Group, 2007) functioned robustly in the textbook analyses, even though it is typically applied to much larger datasets.

In Study III, which examined metaphors in textbooks’ descriptions of protein synthesis, the same dataset as in Study II was analysed using the well-established Metaphor Identification Procedure (Pragglejaz Group, 2007). The analysis was conducted in Swedish by native Swedish speakers and thereafter translated into English. The

metaphoric identification could differ due to local variations in the language use that would not be present in English. After completion of the MIP analysis, the findings were interpreted by use of the conceptual metaphor framework by Lakoff and Johnson (1980) through comparison of source and target domains. Given the subtle distinctions and overlapping character of the source domains, the categorisation in the methodology section represents the most analytically defensible way to divide them into the sub-metaphor systems under the construction-based and information-based main metaphor systems. Inevitably, certain metaphors could plausibly have been assigned to more than one source domain. In such cases, classification was based on the option that was judged most salient or theoretically coherent. The overlaps between different sub-metaphor systems were nevertheless documented, as they illustrate the inherent complexity of metaphor use in texts.

In Study IV, the students' written task was designed specifically for the study and administrated as a digital questionnaire. The mechanistic framework proposed by Moreira et al. (2019), was based on a water-ethanol solution with only two entities, limiting the number of possible interactions. Reducing the number of entities that operate in an isolated system to only two makes the mechanistic reasoning framework rather easy to apply. In contrast, Study IV employed the analytical framework in a substantially more complex context by selecting protein synthesis as the focal phenomenon, thereby introducing a far greater number of entities, activities, and organisation levels. Consequently, it was found to be more challenging for students to engage in the higher levels of mechanistic reasoning. Nevertheless, the framework was found to be sufficiently robust to support the analysis. When reasoning in complex molecular systems, I conclude that the that the mechanistic framework proposed by Moreira et al. (2019) is applicable also to larger and more complex systems such as protein synthesis, than those previously examined.

Conclusions

The aim of this thesis was to provide insights into how domain-specific language is structured and used in upper secondary school to communicate complex molecular systems, using protein synthesis as an example. The conclusions of the thesis are structured as responses to the research questions.

Students' structure and use of domain-specific concepts

Research question 1 is: In what ways do upper secondary students structure and use domain-specific concepts to explain protein synthesis?

From the findings from Studies I and IV, I conclude that students' understanding of protein synthesis is organised into several relatively discrete clusters of domain-specific concepts with limited relationships between the clusters. Students reason more easily within a cluster than between clusters. mRNA emerged as a particularly important domain-specific concept in the reasoning as it served as a point of connection between several conceptual clusters. The clustering pattern can both support and constrain understanding. When clusters lack explicit causal links, mechanistic reasoning remains underdeveloped or fragmented. Taken together, the conclusion is that students' understanding may consist of small, internally coherent conceptual systems. Moreover, I conclude that students hold multiple, and sometimes contradictory conceptions simultaneously that may underscore the pluralistic nature of their knowledge. This is for instance shown in the fragmented presentation of protein synthesis in Studies I and IV, where black boxes (Haskell-Ittah, 2023) are apparent in the clustering of domain-specific concepts. Students' ability to express their understanding by their use of domain-specific concepts in a coherent explanation of a complex molecular system is thereby limited.

Students' sophistication of mechanistic reasoning

Research question 2 is: How do upper secondary students' use of domain-specific concepts reflect the sophistication of their mechanistic reasoning about protein synthesis?

From the findings of Studies I and IV, I conclude that students' mechanistic reasoning about protein synthesis appears fragmented as they typically explain isolated sub-processes without integrating them into a coherent description of protein synthesis. This pattern is consistent with earlier findings that highlight difficulties to connect sequential biological mechanisms (Haskel-Ittah, 2023). Although students employ domain-specific concepts in expected ways, such as pairing tRNA with amino acids, the heavy conceptual and linguistic load of molecular life sciences (Knippels, 2002; Tibell & Rundgren, 2010) could encourage memorisation strategies rather than a genuine understanding of the complexity of a system. Thus, a detailed explanation does not itself automatically guarantee that reasoning is mechanistic. In Study IV, students' explanations often included sub-mechanisms, but the sophistication of domain-specific concept use varied across the sub-processes. I also conclude that the translation process was identified as the most articulated sub-process, which yielded the highest level of sophistication. Students tended to demonstrate a higher level of sophistication in mechanistic reasoning in one area while offering lower levels in others. This pattern suggests compartmentalisation and may indicate that their understanding remains context-based rather than systematically connected.

Domain-specific language in textbooks

Research question 3 is: In what ways do upper secondary chemistry and biology textbooks employ domain-specific language to describe protein synthesis?

Based on the results of Studies II and III, the conclusion is that upper secondary chemistry and biology textbooks adopt a highly context-dependent use of domain-specific concepts on the topic of protein synthesis. Chemistry textbooks tend to use more domain-specific concepts with less emphasis on each, whereas biology textbooks tend to use fewer domain-specific concepts but with a higher emphasis when describing the example of protein synthesis. Chemistry textbooks tend to lean towards a more mechanistic description whereas biology textbooks tend to present a more conceptual description of the same topic, in this case the example of protein synthesis.

Study III shows that metaphoric language is highly used in two distinct tracks of construction-based metaphors and information-based metaphors. Between them, several sub-metaphor systems emerge that borrow their language from various source domains, such as in publication, transportation, or cryptography. I conclude that metaphors as structuring resources in protein synthesis descriptions can be viewed as key resources that shape how clusters of domain-specific concepts are formed and used. Metaphors have a structural role as they provide organising schemas, which can guide students to group concepts and interpret relationships between them. Some metaphors can promote a non-mechanistic or oversimplified view, such as implying intentionality or linearity, when in fact the underlying mechanism is more complex. Nevertheless, I conclude that metaphors can serve as a transitional tool that can help students move from an intuitive explanation towards more mechanistic accounts when the metaphors are critically examined and refined. Metaphors are both inputs to, and products of conceptual clusters and the structure of the conceptual demography, since they form clusters based on the different source domains and influence which domain-specific concept clusters become dominant.

Implications for teaching complex molecular systems

The implications for teaching and learning are drawn from the five facets that were presented earlier. From the **facet on mechanistic reasoning**, students rarely develop fully integrated mechanistic reasoning about protein synthesis. I therefore argue that teachers should avoid expecting a high level of sophistication and instead treat the black boxes in students' reasoning as productive starting points for building conceptual connections. One useful approach could be to begin with an inventory of the students' prior conceptions, and thereafter introduce the needed core concepts in teaching, such as protein, DNA, and mRNA in the context of protein synthesis. By doing so, the cognitive load can be reduced while still providing a holistic overview. As students gain more background knowledge, the identified black boxes can be gradually unpacked with concepts such as amino acid, gene, and tRNA. I would also recommend that teachers provide instructional material, for instance movie clips and textbook excerpts, that present mechanistic explanations at different levels of detail and sophistication, and having students compare and fill in gaps. Such approach could potentially scaffold the difference between an emergent mechanistic reasoning and a detailed lower-level reasoning in the students' understanding.

From the **facet on compartmentalisation**, I recommend that teaching should directly confront the strong compartmentalisation found in students' understanding of protein synthesis, as this limits their ability to recognise relationships across the topic. I argue that instruction should make these connections explicit to help students move beyond fragmented knowledge. By actively challenging compartmentalised domain-specific concept organisation, teaching can support the development of a more coherent and interconnected understanding of complex molecular life science systems. One teaching approach to address this is to give students core concepts written on individual pieces of paper, and ask them to explain and relate the concepts in concept maps, in ways similar to the method used in Study I.

From the **facet on mRNA as a central core concept**, I recommend that teachers use mRNA as a bridging concept throughout the teaching of protein synthesis. To support students' understanding, teachers should explicitly focus on the role of mRNA, for instance what it does and where it is located. This could be achieved by letting students follow mRNA on a journey from cell nucleus to ribosome, for instance through enactment or theatre in groups, or through a comparison between instructional material.

From the **facet on metaphors**, I recommend that teachers make metaphor use in teaching protein synthesis explicit and carefully structured. Teachers should identify which metaphor system is used and when, signal when they shift between systems, and explain how such shifts affect meaning, such as the meanings attached to mRNA. I recommend that one main metaphor system is used at a time. Teachers should also be attentive to metaphors rooted in outdated or unfamiliar source domains, as these no longer provide intuitive support and may create conceptual gaps. I argue that if teachers actively guide students through choices, uses and limits of each metaphor system, they can be helped to develop more coherent and flexible descriptions of molecular mechanisms. This could be achieved by letting students make an inventory of the different metaphors found in an array of instructional material, sort them and build a coherent description using the best metaphors to their knowledge.

From the **facet on the canonical representation in molecular biology**, I recommend that teachers could clearly guide students into the different ways to present protein synthesis to help them discern where each model is useful. Since both textbooks and students rely on the linear central dogma, teachers should point out the limitations and introduce more integrated representations without providing too much detail. If students are given a chance to compare models, it can strengthen their understanding. Teachers should also highlight how chemistry and biology emphasise different aspects of protein synthesis so students can be guided in these differences. One approach to this is to side by side compare how descriptions in chemistry and biology differ in their focus and what they include under the concept of the phenomenon in question, in this case, protein synthesis.

Further studies

From the results presented in this thesis, several directions for future research emerge. For example, little is known about how upper secondary students use metaphors when they describe protein synthesis. Gaining an understanding of students' self-generated analogies and metaphors use could provide valuable insights into how students structure both their verbal and written language. In addition, examining the conceptual demography of students' descriptions would add another layer to understanding how students structure their vocabulary and, consequently, their conceptual understanding.

This thesis has primarily focused on the student and textbook perspectives. To broaden this picture, it would be valuable to investigate teachers' views on domain-specific language use in the classroom in the context of teaching complex life science phenomena, for instance through interviews or classroom studies. Such studies would make it possible to compare students' conceptions of a disciplinary content with the teachers' instructional intentions and goals of the same area. By studying how teachers' aims relate to students' learning, future research could provide important implications for how teaching can be adapted to support both the intended learning outcomes, and students' conceptual understanding.

References

- Allchin, D. (2000). Mending Mendelism. *The American Biology Teacher*, 62(9), 632–640. [https://doi.org/10.1662/0002-7685\(2000\)062\[0632:MM\]2.0.CO;2](https://doi.org/10.1662/0002-7685(2000)062[0632:MM]2.0.CO;2)
- Amin, T. G. (2009). Conceptual metaphor meets conceptual change. *Human Development*, 52(3), 165–197. <https://doi.org/10.1159/000213891>
- Andersson, S., Ellervik, U., Rydén, L., Sonesson, A., Svahn, O., & Tullberg, A. (2013). *Gymnasiekemi 2* (6th ed.). Liber.
- Aubusson, P. J., Harrison, A. G., & Ritchie, S. M. (2006). Metaphor and analogy. In P. J. Aubusson, A. G. Harrison, & S. M. Ritchie (Eds), *Metaphor and analogy in science education* (pp. 1–9). Springer. https://doi.org/10.1007/1-4020-3830-5_1
- Bahar, M., Johnstone, A. H., & Hansell M. H., (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*. 33(2) 84–86. <https://doi.org/10.1080/00219266.1999.9655648>
- Bergqvist, A. (2017). *Teaching and learning of chemical bonding models: Aspects of textbooks, students' understanding and teachers' professional knowledge* [Doctoral dissertation, Karlstad University]. DiVA. <https://kau.diva-portal.org/smash/record.jsf?pid=diva2:1093420>
- Björndahl, G., Landgren, B., & Thyberg, M. (2011). *Spira 1*. Liber.
- Björndahl, G., Landgren, B., & Thyberg, M. (2017). *Spira 1* (2d ed.). Liber.
- Borén, H., Larsson, M., Lindh, M. Lundström, J., Ragnarsson, M., & Sundkvist, S.-Å. (2012). *Kemiboken 2* (5th ed.). Liber.
- Borén, H., Johansson, A., Lundström, J., Stenberg, C., & Wästeby, N. (2020). *Kemiboken 2* (6th ed.). Liber.
- Butler, S., Urrutia, K., Buenger, A., Gonzalez, N., Hunt, M., & Eisenhart, C. (2010). *A review of the current research on vocabulary instruction*. National Reading Technical Assistance Center, RMC Research Corporation.

- Brynhildsen, L., Brändén, H., & Ehinger, M. (2011). *Insikt Biologi 1* (2d ed.). Natur & Kultur.
- Carver, R. B., Castéra, J., Gericke, N., Evangelista, N. A. M., & El-Hani, C. N. (2017). Young adults' belief in genetic determinism, and knowledge and attitudes towards modern genetics and genomics: The PUGGS questionnaire. *PloS One*, *12*(1), e0169808. <https://doi.org/10.1371/journal.pone.0169808>
- Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. *PLoS Biology*, *15*(9), e2003243. <https://doi.org/10.1371/journal.pbio.2003243>
- Crick, F. (1958). The biological replication of macromolecules. *Symposia of the Society for Experimental Biology*, *12*, 138–163.
- Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561–563. <https://www.nature.com/articles/227561a0>
- Condit, C. M. (1999). *The meanings of the gene: Public debates about human heredity*. University of Wisconsin Press.
- Danielsson, K., & Selander, S. (2021). *Multimodal texts in disciplinary education: A comprehensive framework*. Springer.
- Driver, R., Leach, J., & Millar, R. (1996). *Young people's images of science*. Open University Press.
- Dogma. (2026). In *Oxford English Dictionary*. <https://www.oed.com>.
- Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understanding of molecular life sciences. *Journal of Research in Science Teaching*, *44*(7), 938–959. <https://doi.org/10.1002/tea.20186>
- Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education*, *75*(6), 649–672. <https://doi.org/10.1002/sce.3730750606>
- Edling, A. (2006). *Abstraction and authority in textbooks: The textual paths towards specialized language* [Doctoral dissertation, Uppsala University]. DiVA. <https://www.diva-portal.org/smash/record.jsf?pid=diva2:168583>

- Ekvall, U. (2001). Den styrande läroboken. In B. Melander, & B. Olsson (Eds.), *Verklighetens texter: sjutton fallstudier* (pp. 43–80). Studentlitteratur.
- Eriksson, S. (2025). *Analogy competence for science teachers: The development and validation of the ACT framework* [Licentiate thesis, Karlstad University]. DiVA. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A2006468&dswid=-540>
- Fang, Z. (2004). Scientific literacy: A systemic functional linguistics perspective. *Science Education*, 89(2), 335–347. <https://doi.org/10.1002/sce.20050>
- Felius, M., Beerling, M. L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., & Lenstra, J. A. (2014). On the history of cattle genetic resources. *Diversity*, 6(4), 705-750. <https://doi.org/10.3390/d6040705>
- Fisher, K. M. (2000). Overview of knowledge mapping. In K. M. Fisher, J. H. Wandersee, & D. E. Moody (Eds.), *Mapping biology knowledge* (pp. 5-23). Kluwer Academic Publishers.
- Focant, I., Rezsöhazy, R., & De Kesel, M. (2025). Secondary school students' misconceptions in genetics: Obstacles to learning and to address complexity. *Journal of Biological Education*, 1-19. <https://doi.org/10.1080/00219266.2025.2574579>
- Franklin, R. E., Klug, A., Finch, J. T., & Holmes, K. C. (1958). On the structure of some ribonucleoprotein particles. *Discussions of the Faraday Society*, 25, 197–198.
- Gagatsis, A. (2012). Compartmentalization in Learning. In N. M. Seel (Ed.) *Encyclopedia of the Sciences of Learning*. Springer. https://doi.org/10.1007/978-1-4419-1428-6_133
- Gericke, N. (2009). *Science versus school-science: Multiple models in genetics—The depiction of gene function in upper secondary textbooks and its influence on students' understanding* [Doctoral dissertation, Karlstad University]. DiVA. <https://www.diva-portal.org/smash/record.jsf?pid=diva2:128104>

- Gericke, N., Carver, R., Castéra, J., Evangelista, N. A. M., Marre, C. C., & El-Hani, C. N. (2017). Exploring relationships among belief in genetic determinism, genetics knowledge, and social factors. *Science & Education, 26*(10), 1223-1259. <https://doi.org/10.1007/s11191-017-9950-y>
- Gericke, N. M., & Hagberg, M. (2010). Conceptual variation in the depiction of gene function in upper secondary school textbooks. *Science & Education, 19*(10), 963-994. <https://doi.org/10.1007/s11191-010-9262-y>
- Gericke N. M., & Hagberg, M. (2007). Definition of historical models of gene function and their relation to students' understanding of life sciences. *Science & Education, 16*(7-8), 849–881. <https://doi.org/10.1007/s11191-006-9064-4>
- Gericke, N. M., Hagberg, M., & Jorde, D. (2013). Upper secondary students' understanding of the use of multiple models in biology textbooks: The importance of conceptual variation and incommensurability. *Research in Science Education, 43*(2), 755–780. <https://doi.org/10.1007/s11165-012-9288-z>
- Gericke, N., Hagberg, M., Santos, V. C., Joaquim, L. M., & El-Hani, C. N. (2014). Conceptual variation or incoherence? Textbook discourse on genes in six countries. *Science & Education, 23*(2), 381–416. <https://doi.org/10.1007/s11191-012-9499-8>
- Gericke, N., & Mc Ewen, B. (2023). Defining epigenetic literacy: How to integrate epigenetics into the biology curriculum. *Journal of Research in Science Teaching, 60*(10), 2216-2254. <https://doi.org/10.1002/tea.21856>
- Gilbert, J. K. (2006). On the nature of “context” in chemical education. *International Journal of Science Education, 28*(9), 957–976. <https://doi.org/10.1080/09500690600702470>
- Glaser, R. & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40*, 88–96.
- Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*(2), 105–112. <https://doi.org/10.1016/j.nedt.2003.10.001>

- Guzman, K., & Bartlett, J. (2012). Using simple manipulatives to improve student comprehension of a complex biological process: Protein synthesis. *Biochemistry and Molecular Biology Education*, 40(5), 320-327.
<https://doi.org/10.1002/bmb.20638>
- Haglund, J. (2012). *Analogical reasoning in science education: Connections to semantics and scientific modelling in thermodynamics* [Doctoral dissertation, Linköping University]. DiVA. <https://www.diva-portal.org/smash/get/diva2:571154/FULLTEXT02>
- Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Falmer.
- Hamilton, L. D., Barclay, R. K., Wilkins, M. H. F., Brown, G. L., Wilson, H. R., Marvin, D. A., ... & Simmons, N. S. (1959). Similarity of the structure of DNA from a variety of sources. *The Journal of Cell Biology*, 5(3), 397-404.
<https://doi.org/10.1083/jcb.5.3.397>
- Haskel-Ittah, M. (2023). Explanatory black boxes and mechanistic reasoning. *Journal of Research in Science Teaching*, 60(4), 915-933. <https://doi.org/10.1002/tea.21817>
- Haskel-Ittah, M. (2022). How can we help students reason about the mechanisms by which genes affect traits? In M. Haskel-Ittah, & A. Yarden (Eds.), *Genetics education: Current challenges and possible solutions* (pp. 71-86). Springer International Publishing.
- Haskel-Ittah, M., Yarden, A. (2017). Toward bridging the mechanistic gap between genes and traits by emphasising the role of proteins in a computational environment. *Science & Education*, 26(10), 1143-1160. <https://doi.org/10.1007/s11191-017-9927-x>
- Haskel-Ittah, M., & Yarden, A. (2018). Students' conception of genetic phenomena and its effect on their ability to understand the underlying mechanism. *CBE—Life Sciences Education*, 17(3), ar36. <https://doi.org/10.1187/cbe.18-01-0014>
- Henriksson, A. (2012a). *Iris Biologi 1*. Gleerups.
- Henriksson, A. (2012b). *Syntes Kemi 2* (2d ed.). Gleerups.

- Henriksson, A., & Bosson, C. (2020). *Iris Biologi 1* (2d ed.). Gleerups.
- Henriksson, A., Johansson, A., & Zetterberg, E. (2019). *Syntes Kemi 2* (3d ed.). Gleerups.
- Höst, G. E., Larsson, C., Olson, A., & Tibell, L. A. E. (2013). Student learning about biomolecular self-assembly using two different external representations. *CBE—Life Sciences Education*, 12(3), 471–482. <https://doi.org/10.1187/cbe.13-01-0011>
- Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics*, 33(Suppl 3), 245–254. <https://www.nature.com/articles/ng1089z>
- Jahic Pettersson, A., Danielsson, K., & Rundgren, C.-J. (2020). ‘Traveling nutrients’: How students use metaphorical language to describe digestion and nutritional uptake. *International Journal of Science Education*, 42(8), 1281–1301. <https://doi.org/10.1080/09500693.2020.1756514>
- Kay, L. E. (2000). *Who wrote the book of life? A history of the genetic code*. Stanford University Press.
- Karlsson, J., Kringsman, T., Molander, B.-O., & Wickman, P.-O. (2011). *Biologi 1*. Liber.
- Kinchin, I. M., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42(1), 43-57. <https://doi.org/10.1080/001318800363908>
- Kjosavik, F. (2007). From symbolism to information? Decoding the gene code. *Biology & Philosophy*, 22(3), 333–349. <https://doi.org/10.1007/S10539-006-9044-8>
- Knippels, M. C. P. J. (2002). *Coping with the abstract and complex nature of life sciences in biology education – The yo-yo learning and teaching strategy* [Doctoral dissertation, Utrecht University]. CD-β Press.

- Knudsen, S. (2005). Communicating novel and conventional scientific metaphors: A study of the development of the metaphor of genetic code. *Public Understanding of Science*, 14(4), 373–392. <https://doi.org/10.1177/0963662505056613>
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage.
- Koonin, E. V. (2012). Does the central dogma still stand? *Biology Direct*, 7(1), 27. <https://link.springer.com/article/10.1186/1745-6150-7-27>
- Kvale, S. (1989). *Issues of validity in qualitative research*. Studentlitteratur.
- Kvale, S. (1997). *Interviews: An introduction to qualitative research interviewing*. Sage Publications.
- Ladyman, J. (2002). *Understanding philosophy of science*. Routledge.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: Re-visiting students' understandings of life sciences. *International Journal of Science Education*, 26(2), 195–206. <https://doi.org/10.1080/0950069032000072782>
- Lewis, J., Leach, J., & Wood-Robinson, C. (2000). All in the genes? – Young people's understanding of the nature of genes. *Journal of Biological Education*. 34(2), 74-79. <https://doi.org/10.1080/00219266.2000.9655689>
- Lewis, J., & Wood-Robinson, C. (2000). Genes, chromosomes, cell division and inheritance – Do students see any relationship? *International Journal of Science Education*, 22(2), 177-195. <https://doi.org/10.1080/095006900289949>
- Livni Alcasid, G. A., & Haskel-Ittah, M. (2025). “Black boxes, full of them”: Biology teachers' perception of the role of explanatory black boxes in their classroom. *Research in Science Education*, 55(2), 341-358. <https://doi.org/10.1007/s11165-024-10191-4>

- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1-25.
<https://doi.org/10.1086/392759>
- Machová, M., & Ehler, E. (2023). Secondary school students' misconceptions in genetics: origins and solutions. *Journal of Biological Education*, 57(3), 633-646.
<https://doi.org/10.1080/00219266.2021.1933136>
- Marbach-Ad, G. (2001). Attempting to break the code in student comprehension of genetic concepts. *Journal of Biological Education*, 35(4), 183-189.
<https://doi.org/10.1080/00219266.2001.9655775>
- Marbach-Ad, G., & Stavy, R. (2000). Students' cellular and molecular explanations of genetic phenomena. *Journal of Biological Education*, 34(4), 200-205.
<https://doi.org/10.1080/00219266.2000.9655718>
- Martínez-Gracia, M. V., Gil-Quílez, M. J., & Osada, J. (2006). Analysis of molecular life sciences content in Spanish secondary school textbooks. *Journal of Biological Education*, 40(2), 53-60. <https://doi.org/10.1080/00219266.2006.9656014>
- Mikk, J. (2000). *Textbook: Research and writing* (Vol. 3). Peter Lang Publishing Inc.
- Ministry of Education (1994). *Curriculum for the non-compulsory School System: Lpf 94*.
<https://gupea.ub.gu.se/items/981b4e6b-e56b-4e2f-8517-5e722d8d67ff>
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2005). Assessing science understanding: A human constructivist view. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 1-13). Elsevier Academic Press.
- Moreira, P., Marzabal, A., & Talanquer, V. (2019). Using a mechanistic framework to characterise chemistry students' reasoning in written explanations. *Chemistry Education Research and Practice*, 20(1), 120-131.
<https://doi.org/10.1039/C8RP00159F>

- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment for the scientific research literature on reading and its implications for reading instruction.*
<https://www.nichd.nih.gov/sites/default/files/publications/publications/nrp/Documents/report.pdf>
- Nelson, J. (2006). Hur används läroboken av lärare och elever? *Nordic Studies in Science Education*, 2(2), 16–27.
<https://doi.org/10.5617/nordina.421>
- Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct and use them.* Florida Institute for Human and Machine Cognition.
<https://cmap.ihmc.us/publications/researchpapers/theoryunderlyingconceptmaps.pdf>
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* Cambridge University Press.
- Ogden, C. K., Richards, I. A., Malinowski, B., & Crookshank, F. G. (1923). *The meaning of meaning.* Kegan Paul.
- Olander, C., & Ingerman, Å. (2010). Students' language use when talking about the evolution of life-negotiating the meaning of key terms and their semantic relationships. *Nordic Studies in Science Education*, 6(1), 92-106.
<https://doi.org/10.5617/nordina.272>
- Orgill, M., & Bodner, G. (2007). Locks and keys. *Biochemistry and Molecular Biology Education*, 35(4), 244–254.
<https://doi.org/10.1002/bmb.66>
- Pearson, J. T., & Hughes, W. J. (1988). Problems with the use of terminology in genetics education: 1, A literature review and classification scheme. *Journal of Biological Education*, 22(3), 178–182. <https://doi.org/10.1080/00219266.1988.9654979>
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1–39. <https://doi.org/10.1080/10926480709336752>

- Pray, L. (2008). Discovery of DNA structure and function: Watson and Crick. *Nature Education*, 1(1), 100.
<https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/>
- Reichenberg, M. (2016). Explaining teachers' use of textbooks. *Journal of Educational Media, Memory and Society*, 8(2), 145–159. <https://doi.org/10.3167/jemms.2016.080208>
- Reinagel, A., & Speth, E. B. (2016). Beyond the central dogma: model-based learning of how genes determine phenotypes. *CBE–Life Sciences Education*, 15(1), art4.
<https://doi.org/10.1187/cbe.15-04-0105>
- Robson C. (2008). *Real world research*. Blackwell Publishing.
- Robson, C., & McCartan, K. (2016). *Real world research* (4th ed.). John Wiley & Sons, Inc.
- Rundgren, C.-J. (2006). Att börja tala 'biokemiska' – Betydelsen av metaforer och hjälpord för meningsskapande kring proteiner. *Nordic Studies in Science Education*, 2(3), 30–42.
<https://doi.org/10.5617/nordina.412>
- Rundgren, C.-J., Hirsch, R., & Tibell, L. A. (2009). Death of metaphors in life science? – A study of upper secondary and tertiary students' use of metaphors in their meaning-making of scientific content. *Asia-Pacific Forum on Science Learning and Teaching*, 10(1), 1–21.
https://www.eduhk.hk/apfslt/download/v10_issue1_files/rundgren.pdf
- Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 92(3), 499-525.
<https://doi.org/10.1002/sce.20264>
- Schrödinger, E. (1944). *What is life?* Cambridge University Press.
- Shapiro, J. A. (2009). Revisiting the central dogma in the 21st century. *Annals of the New York Academy of Sciences*, 1178(1), 6-28. <https://doi.org/10.1111/j.1749-6632.2009.04990.x>

- Shmueli, G., Patel, N. R., & Bruce, P. C. (2010). *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. John Wiley and Sons.
- Smith, M. U., & Gericke, N. M. (2015). Mendel in the modern classroom. *Science & Education*, 24(1), 151-172.
<https://doi.org/10.1007/s11191-013-9629-y>
- Smith, A. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T.T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910), 122–124. DOI:10.1126/science.1165919
- Smith-Walters, C., Bass, A. S., & Mangione, K. A. (2016). Science and language special issue: Challenges in preparing preservice teachers for teaching science as a second language. *Electronic Journal of Science Education*, 20(3), 59–71.
<https://ejrsme.icrsme.com/article/view/15869>
- Stadig Degerman, M., & Tibell, L. A. (2012). Learning goals and conceptual difficulties in cell metabolism: An explorative study of university lectures' views. *Chemistry Education Research and Practice*, 13(4), 447–461. DOI: 10.1039/C2RP20035J
- Svenska Akademien (2014). *SAOL: Svenska Akademiens ordlista* [The Swedish Academy Dictionary]. <https://svenska.se/saol/>
- Svenska Akademien (n.d.). *SAOB: Svenska Akademiens ordbok* [The Swedish Academy Thesaurus]. <https://svenska.se/saob/>
- Swedish National Agency of Education (1994a). *Biology*.
<https://syllabuswebb.skolverket.se/subject/BIO/2/pdf>
- Swedish National Agency of Education (1994b). *Chemistry*.
<https://syllabuswebb.skolverket.se/subject/KEM/2/pdf>
- Swedish National Agency for Education (2011a). *Biology*.
<https://www.skolverket.se/download/18.4fc05a3f164131a7418104a/1535372296309/Biology-swedish-school.pdf>
- Swedish National Agency for Education (2011b). *Chemistry*.
<https://www.skolverket.se/download/18.4fc05a3f164131a7418104e/1535372296640/Chemistry-swedish-school.pdf>

- Swedish National Agency of Education (2013). *Curriculum for the upper secondary school*.
<https://www.skolverket.se/publikationer?id=2975>
- Swedish Research Council (2002). *Ethical principles for research in the humanities and social sciences*. Swedish Research Council.
<https://www.vr.se/uppdrag/etik/god-forskningssed---ny-utgava.html>
- Taylor, C., & Dewsbury, B. M. (2018). On the problem and promise of metaphor use in science and science communication. *Journal of Microbiology & Biology Education*, 19(1),
<https://doi.org/10.1128/jmbe.v19i1.1538>
- Thörne, K., & Gericke, N. (2014). Teaching life sciences in secondary classrooms: A linguistic analysis of teachers' talk about proteins. *Research in Science Education*, 44(1), 81–108.
<https://doi.org/10.1007/s11165-013-9375-9>
- Thörne, K., Gericke, N. M., & Hagberg, M. (2013). Linguistic challenges in Mendelian life sciences: Teachers' talk in action. *Science Education*, 97(5), 695–722.
<https://doi.org/10.1002/sce.21075>
- Thörne, K., Gericke, N., & Mc Ewen, B. (2026). Introducing epigenetics into secondary school classrooms—an educational design study. *Science Education*, 110(2), 379-399.
<https://doi.org/10.1002/sce.70009>
- Tibell, L. A. E., & Rundgren, C.-J. (2010). Educational challenges of molecular life science: Characteristics and implications for education and research. *CBE—Life Sciences Education*, 9(1), 25–33. <https://doi.org/10.1187/cbe.08-09-0055>
- Van Zele, E., Lenaerts J. & Wieme, W. (2004). Improving the usefulness of concept maps as a research tool for science education. *International Journal of Science Education*, 26(9) 1043–1064. <https://doi.org/10.1080/1468181032000158336>
- Venville, G., & Donovan, J. (2006). Analogies for life: a subjective view of analogies and metaphors used to teach about genes and DNA. *Teaching Science: The Journal of the Australian Science Teachers Association*, 52(1), 18–22.

- Venville, G. J., & Treagust, D. F. (1998). Exploring conceptual change in life sciences using a multidimensional interpretive framework. *Journal of Research in Science Teaching*, 35(9), 1031–1055. [https://doi.org/10.1002/\(SICI\)1098-2736\(199811\)35:9<1031::AID-TEA5>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1098-2736(199811)35:9<1031::AID-TEA5>3.0.CO;2-E)
- Venville, G. J., & Treagust, D. F. (2002). Teaching about the gene in the genetic information age. *Australian Science Teachers Journal*, 48(2), 20–24. <https://www.proquest.com/openview/e829445f698795796095cb832a95d492/1?pq-origsite=gscholar&cbl=47307>
- Voet, D., & Voet, J. G. (2010). *Biochemistry*. John Wiley & Sons.
- Wahlberg, S. (2019). *Teaching and learning protein synthesis through domain-specific language in upper secondary education* [Licentiate thesis, Karlstad University]. DiVA. <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1276806>
- Watson, J. D., & Crick, F. H. (1953). The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 18, 123–131. doi:10.1101/SQB.1953.018.01.020
- Wood, E. J. (1990), Biochemistry is a difficult subject for both student and teacher. *Biochemical Education*, 18(4), 170–172. <https://eric.ed.gov/?id=EJ488610>
- Wood-Robinson, C., Lewis, J., & Leach, J. (2000). Young people's understanding of the nature of genetic information in the cells of an organism. *Journal of Biological Education*, 35(1), 29–36. <https://doi.org/10.1080/00219266.2000.9655732>
- Woody, W. D., Daniel, D. B., & Baker, C. A. (2010). E-books or textbooks: Students prefer textbooks. *Computers & Education*, 55(3), 945–948. <https://doi.org/10.1016/j.compedu.2010.04.005>
- Wright, L. K., Fisk, J. N., & Newman, D. L. (2014). DNA→RNA: What do students think the arrow means? *CBE—Life Sciences Education*, 13(2), 338–348. <https://doi.org/10.1187/cbe.cbe-13-09-0188>



Lost in Translation

Learning complex systems in the molecular life sciences is challenging for students, in large part because of the domain-specific language. This thesis provides insights into how domain-specific language is structured and used in upper secondary school to communicate complex molecular systems, using protein synthesis as an example. Across four studies, students' conceptual structures and mechanistic reasoning about protein synthesis, and how chemistry and biology textbooks describe it through metaphors and domain-specific concepts were investigated. The results show that students use domain-specific concepts in fragmented clusters rather than formulating descriptions showing an integrated understanding of protein synthesis. Students and textbooks follow the central dogma of a linear process from DNA to protein and compartmentalise concepts with mRNA as a key bridging idea. Few students reach an emergent mechanistic level of reasoning, and only in isolated sub-processes. Textbooks use two main metaphor systems: Information-based and construction-based metaphors. Teaching should emphasise links between clusters of concepts and make concepts and metaphors explicit to help students develop a coherent understanding of protein synthesis.

ISBN 978-91-7867-686-6 (print)

ISBN 978-91-7867-687-3 (pdf)

ISSN 1403-8099

DOCTORAL THESIS | Karlstad University Studies | 2026:20
