
# Teaching and Learning Protein Synthesis through Domain-Specific Language in Upper Secondary Education

Sara Wahlberg

Faculty of Health, Science and Technology

Chemistry

LICENTIATE THESIS | Karlstad University Studies | 2019:3

# Teaching and Learning Protein Synthesis through Domain-Specific Language in Upper Secondary Education

Sara Wahlberg

LICENTIATE THESIS | Karlstad University Studies | 2019:3

Teaching and Learning Protein Synthesis through Domain-Specific Language in Upper Secondary Education

Sara Wahlberg

LICENTIATE THESIS

Karlstad University Studies | 2019:3

urn:nbn:se:kau:diva-70684

ISSN 1403-8099

ISBN 978-91-7063-834-3 (print)

ISBN 978-91-7063-975-3 (pdf)



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 2019

WWW.KAU.SE

## Abstract

The aim of this licentiate thesis is to contribute to understanding of upper secondary teaching and learning of protein synthesis with a focus on domain-specific language. It is based on two studies, designated Studies I and II. Study I addressed upper secondary students' understanding of protein synthesis through their usage of domain-specific concepts. Data collected through semi-structured group interviews show that students can better reason about core concepts than peripheral concepts, and they compartmentalise the concepts into five clusters. Study II focused on chemistry and biology textbooks' presentation of protein synthesis through domain-specific concept usage and effects of context on these presentations. The textbooks were analysed using a content analysis approach involving data mining techniques implemented by a computer-generated algorithm. The results reveal that chemistry textbooks focus more on peripheral concepts and generally tend to identify fewer relationships among more concepts than biology textbooks, which emphasise core concepts and tend to highlight more relationships among fewer concepts. Jointly, Studies I and II reveal four facets of teaching and learning protein synthesis: 'mechanistic or conceptual descriptions', 'compartmentalisation', 'mRNA as a core concept' and 'canonical representation'. By acknowledging the results reported herein, teaching can improve the facilitation and reduce the hindrance in learning protein synthesis through the awareness of the domain-specific language usage.

## Populärvetenskaplig sammanfattning

Proteiner är livsnödvändiga. Inga proteiner – inget liv. Denna koppling kan vid första anblicken verka både enkel och rättfram men är allt annat än det. Varje sekund bildas tusentals och åter tusentals proteiner i varje cell i vad som kallas för *proteinsyntesen*. Från informationen som finns lagrat i organismens DNA byggs de specifika protein som behövs för olika uppgifter i cellerna som transport, struktur eller som hormon. I och med proteinsyntensens centrala roll för att förstå livets mekanismer, räknas ämnet idag som en av hörnstenarna inom de molekylära livsvetenskaperna, som är ett tvärvetenskapligt område inkluderande delar av biologi- och kemivetenskaperna. Detta område har – precis som alla andra ämnesspecifika områden – ett gemensamt språk för att kunna kommunicera sådant som är centralt.

Idag finns mycket lite forskning att tillgå kring undervisning och lärande kopplat till proteinsyntesen, och ännu mindre kan hittas som fokuserar på det ämnesspecifika språk som används för att kommunicera vad proteinsyntesen handlar om. Därför bidrar denna licentiatavhandling till förståelse för lärande och undervisning om proteinsyntesen i den svenska gymnasieskolan, med fokus på ämnesspecifikt språk.

Avhandlingen består av två studier. Studie I fokuserar på gymnasieelever från det naturvetenskapliga programmet och deras förståelse av proteinsyntesen genom att undersöka hur de hanterar ämnesspecifika begrepp. Specifikt undersöks hur dessa begrepp beskrivs och vilka mönster som framträder i hur begreppen relateras till varandra. Studie II fokuserar på hur läromedel – specifikt kursböcker i kemi och biologi för gymnasiets kurser Biologi 1 och Kemi 2 – presenterar proteinsyntesen genom att undersöka tekniska termers frekvens (antal), distribution (var i texten termen förekommer) och relationer (hur termerna relaterar till varandra) som tillsammans beskriver textens *konceptuella demografi*. Beskrivningarna mellan kemi- och biologiböckerna analyseras och jämförs.

I studie I samlades data in genom semi-strukturerade gruppintervjuer med 12 studenter. Resultaten från intervjuerna analyserades sedan via

framtagna begreppskartor på tre nivåer: begreppsnivå, relationsnivå och klusternivå. Resultaten i studie I visar att eleverna generellt sett har svårt att förstå proteinsyntesen som helhet men har lättare att förstå kärnbegrepp än perifera begrepp. Eleverna klustrar samman begreppen i isolerade ”begreppsöar”. Fem kluster kunde identifieras: arvskluster, kärnkluster, proteinsynonymkluster, transkriptions-kluster och translationskluster. Resultatet visar också att gymnasieelever har lättare att relatera begrepp inom ett kluster än mellan olika kluster.

I studie II samlades data in från fyra biologiböcker och tre kemiböcker. Data analyserades genom innehållsanalys som genomfördes med hjälp av tekniker för datautvinning. En algoritm togs fram specifikt för studien. Resultatet från studie II visar att det finns kontextbundna skillnader mellan kemi- och biologiböckers sätt att presentera proteinsyntesen. Generellt sett fokuserar kemiböcker på mindre centrala, så kallade perifera begrepp som tRNA och aminosyra, och gör ingen tydlig skillnad mellan betydelsen av dessa och mer centrala begrepp som DNA och protein. Biologiböcker å andra sidan, fokuserar mer på de centrala begreppen. Både kemi- och biologiböckers presentationer följer den kanoniska beskrivningen av proteinsyntesen inkluderandes translations- splicing och transkriptionsprocesserna. Kemiböcker beskriver proteinsyntesen genom att dela upp proteinsyntesen sekventiellt i olika textavsnitt med distinkt olik användning av tekniska termer, vilket skiljer sig från biologiböckernas övergripande presentation där alla tekniska termer kontinuerligt relateras till de mest centrala. I kemiböckerna görs färre relationer mellan ett större antal olika tekniska termer. Biologiböcker fokuserar i stället på ett färre antal olika tekniska termer och gör fler relationer mellan dem.

Från resultaten i studie I och II identifierades fyra fasetter för lärande och undervisning av proteinsyntesen. Dessa är: kanonisk beskrivning, mekanistisk eller konceptuell beskrivning, mRNA som kärnbegrepp och sekventiell uppdelning. Fasetterna kan ses som olika fokusområden som är viktigt för lärande och undervisning om proteinsyntesen. Sammantaget kan sägas att kemiböcker tenderar att beskriva proteinsyntesen på ett mekanistiskt sätt medan biologiböcker

tenderar att göra en mera konceptuell beskrivning. Kontexten spelar alltså roll för hur texten skrivs fram genom hur de tekniska termerna används. Ett exempel är begreppet mRNA som i studie II visats ges olika fokus i kontexterna, och som i studie I visat sig vara central för förståelsen av proteinsyntesen.

En slutsats som kan dras från fasetterna är att kursböcker kan bidra till de strukturer som finns i elevers uppfattningar kring proteinsyntesen sett utifrån deras begreppsanvändning. Hit räknas sättet som både elever och kursböcker klustrar samman begrepp, vilka relationer som görs och vilka begrepp som fokuseras. Både kemiböcker och biologiböcker följer den kanoniska beskrivningen kring proteinsyntesen, men med olika fokus. Ett konceptuellt perspektiv att beskriva proteinsyntesen har en fördel i att den kan ge förklaringar för övergripande strukturer, medan ett mekanistiskt perspektiv kan ge fördjupade och detaljerade delar av specifika passager av proteinsyntesen. Det gör att de två perspektiven kan komplettera varandra i undervisningen.

I studie I rekommenderar jag att undervisning om proteinsyntesen bör börja med att fokusera på de relationer som eleverna redan kan beskriva mellan olika kluster och att fokusera undervisningen runt de tekniska termer som finns beskrivna i två eller flera kluster. Exempel på sådana är de centrala begreppen DNA, gen och protein samt mRNA. Genom en medvetenhet kring användningen av det ämnesspecifika språket utifrån de resultat som presenterats i denna licentiatavhandling, kan undervisningen överbrygga svårigheter, och öka möjligheterna, för elevers lärande av proteinsyntesen.

## Preface and acknowledgements

It is no secret that this licentiate thesis has taken substantial time to finish. So, what happened? The short answer is that this thing called life happened. Or as Forrest Gump puts it: "*… and that's all I have to say about that*". I have come to realise over the years that this stretching of the process has given me a chance to grow into the role of researcher. The process may have been long, but this simply extended the opportunities and possibilities! However, this text did not write itself and would not have been possible to complete without the love and support of peers, colleagues, friends and family. Many have come, and many have left over the years between start and completion, and your impact shall never be forgotten. I remember you all and I would like to embrace you all with a big and warm **THANK YOU**. Nevertheless, I wish to directly acknowledge some persons and organizations below.

The licentiate was funded by **the Erna and Victor Hasselblad Foundation** through the research school. I wish to thank the foundation for the opportunity! I also want to thank the **Scientific Board** of the research school and my employer **Karlstad University** for providing the opportunities.

My supervisor, Prof. **Niklas M. Gericke**, who has stood beside me since my diploma work in the teacher education program in 2005 – thank you for spreading your knowledge, scaffolding and continuous belief that this licentiate thesis would actually be finished someday. This provided crucial help along the path – one step at a time. I could not have done this without you and hope for many years of collaboration to come! Thank you. I also thank my co-supervisor, Prof. **Lena Tibell**, who helped me a lot in early stages of the process and provided support along the way. My second co-supervisor early in the process, **Michal Drechsler,** also provided valuable insights in the beginning of process and later as a teaching colleague.

I thank the **upper secondary students** that participated in Study I.

I must mention a trio of professors who, more or less directly encouraged me to take the leap into the big unknown world of research: **Lars Blomberg, Onno de Jong and Gustav Helldén**. Thank you from the bottom of my heart for believing in me and encouraging me.

I must also acknowledge Dr **Donald F. Ross** at Computer Sciences at Karlstad University for invaluable input and discussions on computational strategies and language usage. I thank you.

Several readers of my work during the research years should be acknowledged. **Inger Edfors** – thank you for frank, constructive and helpful comments about material prepared for my 90% seminar. **Jesper Haglund** reader of a recent manuscript – thank you for valuable insights on triangles, words and general comments that considerably improved this thesis! **Birgitta McEwen** – thank you for reading of an early version of this manuscript. I cherish your efforts and comments! Also, a big thank you to colleagues in the department of English at Karlstad University and **Elisabeth Wennö** for fruitful comments on the language of the thesis. Also, **Inga-Lill Nilsson** and **Annelie Ekberg-Andersson** at Karlstad University Library are acknowledged for valuable input on referencing strategies. I also acknowledge **Reidar Lyng** for his frank and encouraging comments in the very beginning. Thank you.

**Birgitta Frändberg**, **Miranda Rocksén** and **Anne Solli**, – you made an irreplaceable contribution to this thesis as we began this journey together and your words and friendship have helped me along the way. I cherish them dearly.

**Anna B, Daniel, Nina, Susanne, Teresa and Torodd**. I only say 'thanks' as words cannot describe the feeling of being surrounded by you, even though I have not physically always been around.

**Friends and colleagues at the Department of Chemistry and Biomedical Sciences**, and **SMEER (Science, Mathematics, Engineering Education Research)** at Karlstad University. I can now say: Finally! Thank you for supporting, cheering, asking and caring along the way! For my former colleagues **Birgitta Svensson**,

“my partner in teaching crime” and **Inger Arenö** who took me along for the ride, a ‘thank you’ is not enough. I also wish to make a special acknowledgement to the administrators **Pia Eriksson** and **Lotta Klang** for invaluable assistance of all the practicalities surrounding the thesis *per se* and ‘the big day’. Also, my colleagues **Micke Andersen**, **Jörgen Samuelsson** and **Jan van Stam** - you made invaluable contributions by keeping my Java, lab and sanity intact and on track. To the biochemists **Thomas Nilsson, Maria Rova**, **Anna Smedja Bäcklund** – thank you for interesting discussions of the fabulous world of the cell! I also want to cherish the memory of our late colleague **Tommy Björk** who always encouraged me into someday take the leap into the unknown world of research. It turned out all right, just as you predicted.

**Karin** – I could not have done this without you. I will happily go by B7 any day of the week as long as it is with you! Thank you from the bottom of my heart.

**Nenne** and **Farbror Mats** – without our weekend breakfasts I would truly have lost my way in the world of words, concepts, life sciences and all the other things this thesis is about. Thank you for being there!

**My family** – my **mother**, **father**. Well. None of this would have been written if you hadn’t encouraged my curiosity and let me freely explore the world from an early age. I also truly cherish the love and encouragement of my **sister and her family**, who come now and then with cheering and encouraging words! Ett tack också till **svärmor** och **svärfar** för alla hejarop och hjälp längs vägen. Och **farmor**: Det är en lic. Du kan berätta för alla du känner och inte känner! Tack för alla uppmuntrande ord från både dig och från dig **Jörgen** – tack, tack, tack! **Morfar**: Nu är den klar. Tack för att trott på mig. Och **mormor** som inte finns med oss längre: Det finns ingen doktor i släkten än, men den är på god väg. Synd att du inte fick vara med till den början.

Last but always first: **My partner** and our **children**. None of this would have been possible without you and I love you more than I can express. This is for you!

## List of papers

**Paper I**

**Clusters of Concepts in Molecular Genetics: A Study of Swedish Upper Secondary Science Students' Understanding.**

Niklas M. Gericke and Sara J. Wahlberg

Published in 2013 in *Journal of Biological Education*, *47*(2), 73-83.
DOI: 10.1080/00219266.2012.716785

**Paper 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

Published in 2018 in *CBE—Life Sciences Education*, *17*(3), ar51, 1-14.
DOI: 10.1187/cbe.17-12-0274

## Authors' contributions

This thesis is based on the two papers listed above, which are referred to as 'Paper I' and 'Paper II'. In addition, for convenience, the studies reported in them are referred to as Studies I and II, respectively. These articles were produced in collaboration with my supervisor, Niklas M. Gericke, as outlined below. Both authors contributed to writing of all sections, then read and approved the manuscripts' contents before submission.

### Authors' contributions to Paper I

Contributions of the first author (Niklas M. Gericke) were:

- writing most of the article,
- providing guidance regarding the idea, planning of the project, data collection and writing,
- validating and monitoring the data analysis,
- executing the submission process including contact with the editor of the journal after discussion with the second author.

Contributions of the second author (Sara J. Wahlberg) were:

- providing the initial idea and plan of the study's design,
- providing the initial design of the data collection method,
- collecting all the data,
- main analysis of the data and graphical presentation of results,
- construction of initial drafts of the paper and contribution to writing of all parts of the manuscript.

### Authors' contributions to Paper II

Contributions of the first author (Sara J. Wahlberg):

- writing most of the article,
- developing the idea of 'conceptual demography',
- developing the project idea and planning the data analysis,
- providing the main idea for the data analysis software and its construction in collaboration with a programmer,
- validating the data analysis software,
- data collection and main analysis of the data,
- constructing visual presentations of the results,

- executing the submission and review process.

Contributions of the second author (Niklas M. Gericke) were:

- providing guidance regarding the idea, planning of the project, data collection, analysis and writing,
- supervising use of the methodology,
- co-developing the idea of 'conceptual demography',
- validating the data analysis.

# Table of contents

## Introduction

Proteins are crucial to life: no proteins – no life. At first glance, this connection seems both simple and straightforward, yet it is anything but. Every picosecond, thousands of proteins are constructed in each cell. These proteins are responsible for the construction and maintenance of cellular structures and functions, through involvement in myriads of life-supporting reactions. Production of the required proteins involves a series of reactions, starting from reading of the genetic material to the finished products – proteins. This production process is referred to as protein biosynthesis or, more briefly, *protein synthesis*.

Due to the importance of understanding mechanisms of life, protein synthesis is globally regarded as a cornerstone of molecular life sciences and education in the field (Reinagel and Speth, 2016). Sweden is no exception, and the topic of protein synthesis is a small, but vital, part of two courses in the upper secondary science curricula for science majors: Biology 1 and Chemistry 2 (Swedish National Agency for Education, 2011a; 2011b).

Today, we are aware of students' difficulties in understanding molecular life sciences, as illustrated by numerous papers published in the domain-specific journals *Biochemistry and Microbiology Education* and *CBE–Life Sciences Education*. A major general identified obstacle is the domain-specific language used by the molecular life science community when communicating information about cellular processes. The molecular life sciences – like any area of expertise – have their own collective language (Pearson and Hughs, 1988; Tibell and Rundgren, 2010), and are verbally overloaded (Fang, 2004). This, together with rapid expansion of molecular life sciences research and knowledge, makes key aspects difficult to comprehend. However, acquisition of such comprehension is crucial for any student seeking to understand cellular reactions and their connections to life.

In any learning process, the student faces several information sources. In science education, for instance chemistry education, written textbooks are reportedly among the most widely used (Edling, 2006).

As in every text, words used in textbooks are connected to provide a flow of information that communicates a central message (Halliday and Martin, 1993). In the case of texts associated with a specific research field or community, the choice of words is related to the tradition and joint vocabulary of the discipline, known as domain-specific language. Thus, to become scientifically competent, a student must be able to handle the specialised language used in science (Fang, 2004).

Within the domain-specific language, there are words that are that are crucial for communication of descriptions or labels of parts of the domain and associated phenomena. These words are referred to as technical terms in semantics. Almost 100 years ago, Ogden et al. (1923) established the idea of the semantic triangle, in which a word or technical term is a *symbol* of a *referent* in the real world (see Figure 1). The technical term refers to the referent, while at the same time expressing the abstract meaning of it (*thought* or *reference*), or what we could call a concept (the top of the triangle in Figure 1). Hence, when a technical term is assigned a meaning, this meaning can be referred to as a concept. These relationships are focal concerns of this thesis, particularly those of concepts explaining protein synthesis. For example, as a symbol 'polypeptide' is the label for a chain of amino acids, which is the referent. As a reference, the polypeptide is not only a chain of amino acids but attributed with a meaning and hence can be regarded as a concept.

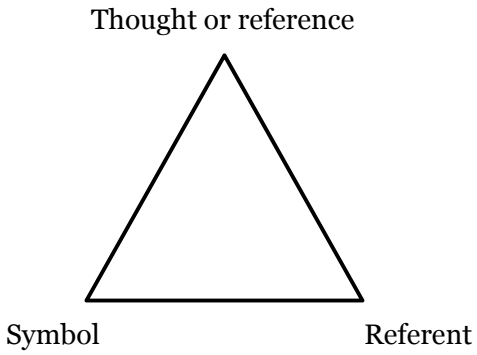


*Figure 1: The three parts of a semantic triangle as proposed by Ogden et al. (1923, p. 14): the symbol, the referent and the thought or reference.*

However, the conceptual element is merely one part of a word, as presented in the sematic triangle as presented in Figure 1. The symbol-level, internal representation as thought or reference, and the referent

itself are the three parts that constitutes the semantic triangle as proposed by Ogden et al. (1923).

Domain-specific concepts that are important for understanding a phenomenon can be regarded as *core concepts* (Driver et al., 1996). In addition, concepts that are important but less strongly emphasised in communication of the phenomenon, are designated *peripheral concepts* in this thesis.

In molecular life sciences, the domain-specific language and heavy usage of domain-specific concepts have been shown to cause difficulties for students of all ages (Knippels, 2002). Moreover, the insights (or lack thereof) provided by previous research on domain-specific vocabulary usage in the field raises questions regarding teaching and learning the specific topic of protein synthesis. Thus, the overall objective of the studies this thesis is based upon was to explore the domain-specific concept usage in students' reasoning and textbooks' presentations, thereby obtaining insights regarding the teaching and learning of protein synthesis from a linguistic perspective.

### Outline of the thesis

The thesis has the following layout: *introduction, background, aim and research questions, methodology, results, discussion* of both results and methodology, *implications* and *recommendations for further research*.

In the background section, I present an overview of the current state of research on molecular life sciences education generally, and teaching and learning protein synthesis specifically. The theoretical frameworks used in the studies and thesis are presented.

Significant developments of methodology were required to achieve aims of both Studies I and II, as well as identification of key facets of the focal topic. These developments are described in the methodology section and further considered in the 'Discussion of methodology' section. The methodology section also addresses protein synthesis education in the context of the Swedish school system by outlining the

relationship between the topic and syllabi of the two upper secondary courses Chemistry 2 and Biology 1.

Following the methodology section, there is an overview of the main results from Studies I and II, which provides foundations for the presentation and discussion of identified facets for teaching and learning protein synthesis. The methodology is discussed in this section. In the last part of the thesis, implications for teaching and learning protein synthesis in upper secondary school through domain-specific vocabulary usage are presented. Finally, I present closing remarks including recommendations for further research projects.

## Background

Research in chemistry and biology has become increasingly extensive and diverse. Due to this diversity, there are several overlapping fields between chemistry and biology, such as cellular processes (Orgill and Cooper, 2015), which are studied within the evolving interdisciplinary research field of life sciences.

### Life sciences

Life sciences is a collective name for interdisciplinary branches of science concerned with living organisms and associated molecules and systems (see *e.g*. Tibell and Rundgren, 2010). Life sciences can be studied at multiple levels, for example molecular life sciences focus particularly on molecular aspects of life, while ecology considers phenomena in entire ecosystems. Molecular life sciences are important components of fields of larger areas of science, such as chemistry and biology, including biochemistry, biotechnology or cell biology. Protein synthesis can be studied within any of these areas but, regardless of the area, the subject matter of molecular life sciences has been shown to cause problems for people wanting to master the subject (Jenkinson and McGill, 2012; Höst et al., 2013). The following sections address this issue.

To keep up with the rapid developments in life sciences, practicing biochemists and molecular/cell biologists need diverse and robust

foundations of knowledge in various domains. Trained biochemists and cell biologists are professionals that master the subject. They are confident in their knowhow of the field within which they are operating. For students in school, in contrast, this polymathic area has been shown to be associated with challenges that may become learning obstacles. In the following section, these difficulties are presented.

### Molecular life sciences – overwhelming, diverse and overloading?

From an educational perspective, biochemistry and cell biology are key areas of life sciences that encompass large and abstract bodies of knowledge for students to process (Orgill and Bodner, 2007). From the student's perspective, this complexity can be potentially overwhelming (Wood, 1990). Driver et al. (1996) proposed that core concepts, processes and principles are important for a good understanding of science. However, students' understanding of constituents of the core content in molecular life sciences connected to protein synthesis has been poorly investigated. Hence, there is a need to identify students' understandings of this core content, and optimal ways to communicate it to them.

The choice of terminology used in molecular life sciences has been shown to play a significant role in students' difficulties. For instance, many of the words used when communicating biochemistry have no everyday referents (Tibell and Rundgren, 2010). This may make the scientific language difficult to comprehend from the students' everyday life perspectives. It seems that the more confidential students are in moving between languages, colloquial and scientific, the more mature their understanding (ibid.).

It is therefore challenging to transform the science found in the research community into school science and to do so without sacrificing the essence of the language of science (Olander and Ingerman, 2010). Thus, an important task for research in life science education is to explore the meanings students make of scientifically related concepts.

### Domain-specific language

When a subject matter is communicated within a community, talking and writing are carried out in certain manners (Halliday and Martin, 1993). The molecular life sciences community is no exception. The language shared by a certain community is referred to as domain-specific language (Halliday and Martin, 1993), and the vocabulary is a vital part of this language. The domain-specific vocabulary is composed of words that describe (for instance) processes, items or phenomena – so-called technical terms – that are used to convey certain messages linked to a topic in the community. In this thesis, the focal community is comprised of all those involved in molecular life sciences education in two broader contexts (biology and chemistry) and the focal topic is the domain-specific language related to protein synthesis.

The domain-specific terminology in the field of molecular life sciences has been identified as a learning obstacle for several reasons. A putatively major one is the vast, daunting and increasing amount of domain-specific vocabulary (Knippels, 2002), and associated accumulation of domain-specific concepts (Reiss et al., 1999; Ananiadou et al., 2006). However, it has been shown that students' learning is facilitated if they can identify the central concepts in molecular life sciences (Stadig Degerman and Tibell, 2012). In order to help students' learning, the concepts must be described and explained in sufficient depth and detail (Tibell and Rundgren, 2010). An important part of learning life sciences is therefore learning to handle a domain-specific language where the conceptual foundations are key factors for students' learning.

Life sciences are taught within the two school subjects of biology and chemistry, which have different aims, with different perspectives and can therefore be regarded as different contexts. These contexts have their own domain-specific languages, which have substantial overlaps, but potentially different nuances and emphases. The students therefore often need the ability to apply concepts from different contextual domains, such as biology and chemistry, in new contexts within the molecular life sciences (Villafañe et al. 2011; Wood, 1990). An example of such a concept is energy, which may be fundamentally defined in the same way in each context but is related to different wordings to explain

domain-specific processes (Lancor, 2014). The differences in word usage between contexts may putatively reduce cognitive load by allowing students to consider individual meanings associated with concepts in isolation (Gilbert, 2006). However, such contextual use could foster the idea that the physical phenomena described by certain terms are not the same in different contexts. Thus, applying a concept in a new context may either hinder or facilitate learning.

### The central dogma and canonical representation

The focal topic in this thesis is currently referred to as protein synthesis. This has been a cornerstone of life sciences ever since discovery of the role of DNA in genetics, *inter alia* through contributions by Franklin, Wilkins, Crick and Watson (based on Levene and Chargaff's idea of genetic material) in the 1950s (Crick, 1958; Franklin et al., 1958; Hamilton et al., 1959; Pray, 2008; Watson and Crick, 1953). This triggered a major paradigm shift, leading to the pathway from information embedded in genetic material to mature proteins being presented in the so-called "central dogma[1] of molecular biology" (Crick, 1958), which is schematically represented in Figure 2.

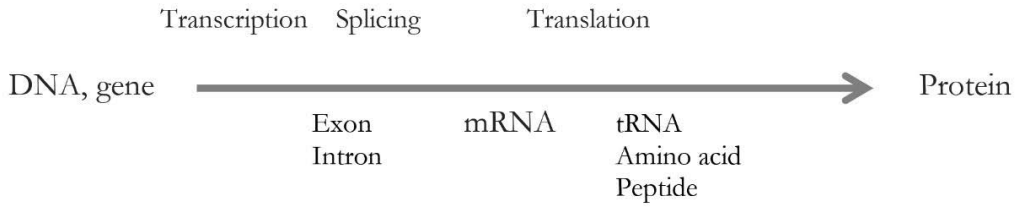


*Figure 2: The central dogma of molecular biology, as presented in Paper II. Reprinted with permission from CBE–LSE. Original image: https://doi.org/10.1187/cbe.17-12-0274.*

The original idea that laid the foundation for the 'central dogma of molecular life sciences' was first presented on the 19th of September 1957 in London, resulting in the well-known manuscript entitled "On

[1] A "dogma" can be defined as *"a principle or set of principles laid down by an authority as incontrovertibly true"* ("Dogma," 2018).

protein synthesis" (Crick, 1958). The main idea expressed in the manuscript is that the "main function of the genetic material is to control (not necessarily directly) the synthesis of proteins" (Crick, 1958, p. 138). The idea evolved into a dogmatic view of protein synthesis, which has subsequently been considered one of the cornerstones of life sciences (Pray, 2008; Cobb, 2017).

New linguistic structures are required to communicate new findings. The language used for communicating the new findings regarding protein synthesis was adopted from the then new field of informatics (Kay, 2000). Earlier ideas suggested that cells could learn from experience through some sort of 'cellular memory'. Thus, traits could be passed from generation to generation. The new insights in the 1950s provided an understanding (subsequently modified by discoveries of epigenetic phenomena, as described below) that this was not the case. Instead, DNA sequences provided sorts of templates, or messages, enabling the growth, development, functions and responses of any organism. These sequences became regarded as kinds of ciphers that needed a code key to be fully interpreted and understood (Kay, 2000).

This was a time of rapid technological development, and ideas from several rapidly developing fields, including cryptology and informatics, were used to understand and communicate the nature of the messages embodied in organisms' DNA (Kay, 2000). In the search for suitable language to describe protein synthesis, metaphors and analogies of informatics processes proved useful. For example, the conversion of genetic information in a sequence of RNA by ribosomes was regarded as analogous to translation of a text from one language to another. Thus, it was called translation, and the RNA used in the process was called 'messenger RNA' (mRNA). However, the new language describing protein synthesis could be seen as 'caged' by the conceptual apparatus now defining the process (ibid.). The way the domain-specific language for protein synthesis was developed based on informatics controlled the language, and hence the idea of protein synthesis in a possibly unnecessarily dogmatic fashion. Thus, the concepts borrowed from the field of informatics became facilitators for communicating and understanding the processes of protein synthesis, but at the same time contributed to the rigidity of the descriptions.

Hence, the scientific description of the protein synthesis became metaphoric when it was constituted (ibid.).

### Conceptualisation of protein synthesis through metaphors

Metaphors play important roles when communicating scientific phenomena (Lakoff and Johnson, 1980), for several reasons. For instance, they can be used to capture interest, make abstract ideas more concrete or contribute to an approximation of a scientific phenomenon (Rundgren et al., 2009).

In molecular life sciences generally, and protein synthesis specifically, there are many examples. One is the concept of *splicing*, which is derived from splicing ropes to understand how different parts of an overarching structure are taken apart and assembled. Another example is the process of translation, a concept derived from interpretation of languages. Rundgren et al. (2009, p. 2) claim that these metaphors can be considered dead, or as they put it: "science utilizes metaphors that have been drained of their 'poetic' potential to a degree that we can say that they have been conventionalized to death". This means that all or part of the meaning of a concept is changed by its metaphoric use.

Rundgren et al. (2009) also postulate that dead metaphors may come to life again in their communication, through the poetic content, and hence give rise to alternative conceptions in the learning of molecular life sciences. On the other hand, the conventionalisation may be a sound development of a concept, for instance in the concept of the *cell*. It is possible to conceptualise a cell within an organism without considering the origin of the metaphor – the cell of a prison or a monastery.

Using metaphors in communication of scientific phenomena and acknowledging the balances in their different usages, could provide important learning possibilities. On the other hand, neglecting the power of the usage of metaphors may result in alternative conceptions in students' knowledge acquisition processes.

### Canonical representation of protein synthesis

The original expression of the central dogma of molecular biology states that proteins are constructed in a specific one-way order (Crick, 1958). This means that information that has been transcribed and translated into proteins, cannot make its way back to the genetic material. Thus, information cannot be transferred from protein to DNA or between proteins (Crick, 1958). These findings were widely accepted in the research community at the time. However, Crick (1970) subsequently acknowledged that the phenomena are probably more complex than presented in his original model.

Today, almost 60 years later, it is known that the construction of proteins is not as linear and straightforward as the dogma indicated. For example, recent findings regarding epigenetic regulation show that information can also pass in the other direction, from proteins to the genetic material (Jaenisch and Bird, 2003). These new insights challenge the dogmatic idea. Teaching the central dogma in a way that retains the original idea of the central dogma (Crick, 1958), might therefore risk induction of scientifically wrong ideas about the essence of protein synthesis.

Further, presenting protein synthesis as a *'dogma'* is problematic as the dogma concept includes the idea of a truth. Truths are usually subjective (with exceptions in interpretations, for instance, in mathematics and logic) or validated by an irrefutable authority (Ladyman, 2002). For the central dogma, this would require Crick and colleagues to be regarded as deities, and preclude any questioning of its representation of protein synthesis. Thus, protein synthesis is seldom presented as a "dogma" nowadays, and often by the notion of *the canonical representation of molecular life sciences*, or specifically *canonical representation of protein synthesis* instead. This shift from defining the description of protein synthesis as a 'dogma' to regarding it as a 'canonical description', removes the problematic truth nuances connected to a dogmatic perspective.

I use both definitions in the thesis but for different purposes. The 'dogma' is mostly used for describing the authority and historical background when addressing protein synthesis. When referring to

more contemporary writings and discussions in relation to teaching and learning, I choose to use the term 'canonical description'.

### Protein synthesis and explanatory models thereof

During the last 40 years, there have been massive developments in biochemistry and molecular cell biology research. New insights have led to the development of new models and lines of reasoning. However, the original idea of considering the protein synthesis as starting from the genetic material and ending when complete proteins are released (from ribosomes) is still valid. Protein synthesis means production of proteins, and the information required to produce proteins is indeed present in the genetic material.

Protein synthesis is carried out in cells. This thesis focuses on the process in eukaryotic cells, which are also the dominant cell type in human beings. In contrast to prokaryotic cells, such as bacteria, eukaryotic cells have a confined space – the nucleus – that contains most of the genetic material (although some is also present in the mitochondria). This material is composed of DNA and can be regarded as the source of all information needed for the cells to function. As stated at the beginning of this thesis: proteins are crucial for life. Thus, I will focus on the function of the DNA and its role in the production of proteins in cells – protein synthesis.

Protein synthesis is the processes whereby proteins are constructed in cells from information in the genetic material. The protein synthesis includes several mechanisms, processes and structures. In the following section I briefly overview how proteins are formed in a eukaryotic cell, focusing on the process in human cells, as presented by Alberts et al. (2008) and Sadava et al. (2014). I deliberately omit detailed information on every step, such as roles of various enzymes and specific pathways, to provide a coherent description that is hopefully accessible for any reader. I assume that the reader knows the fundamental structure of DNA molecules and that these molecules are packaged in *chromosomes* only when cells are about to divide. As the processes of cell division – mitosis and meiosis – are beyond the scope of this thesis, chromosomes are not further considered. They were also

addressed to a minor degree in Study I, and not considered at all in Study II.

### *Overview of protein synthesis processes*

The description of protein synthesis below summarises the three major processes involved in production of a protein: *Transcription*, where a sequence of DNA is transcribed (or "copied") in the form of an immature mRNA molecule; *splicing*, where non-coding parts of the mRNA are removed, yielding a mature mRNA molecule; and *translation*, where the sequence of the mRNA is translated into a strand of amino acids. Thereafter, the amino acid strand will be subject to folding and then posttranslational modifications. The strand of amino acids subsequently folds into a functional conformation, often with the assistance of 'chaperones'. It may also be linked with other amino acid strands in the formation of larger structures, such as insulin molecules, which play a major role in homeostasis of human blood sugar levels. In addition, it may thereafter be subjected to various chemical posttranslational modifications, such as glycosylation or phosphorylation. However, folding and all posttranslational steps, reviewed for instance by Mann and Jensen (2013), are ignored here as they are beyond the scope of this thesis.

The proteins any organism needs must be synthesised by the organism itself. The information required to construct these proteins is present in the DNA. From this information, amino acids present in the organism's cells are assembled into myriads of functional proteins by cellular structures acting in highly regulated, complex processes. In the human body, these amino acids are released by digestive processes in the small intestine and delivered to all the cells in the body via the blood stream. Thus, the proteins we eat are not used directly. Instead, they are disassembled then re-assembled in the required forms by each cell's protein building or *protein synthesis* machinery.

Whenever there is a need for synthesis of a specific protein, a certain part of one strand of the DNA encoding all or part of the protein – a gene – is copied into a mRNA in a process called the transcription (Figure 3). This occurs in the nucleus of the cell, which is surrounded by a membrane. The nuclear membrane protects the DNA (which is

degraded if it enters the cytoplasm of the cell) and has other functions that are beyond the scope of this thesis.

At this stage, the newly produced mRNA is often referred to as a primary RNA transcript or 'immature mRNA' to clarify that it is not the final template for the protein. As can also be seen in Figure 3, non-coding parts of the immature mRNA that will not contribute to the finalised protein – *introns* – are removed and remaining parts – *exons* – are combined into the finalised mRNA in a process called *splicing*. The mRNA is now regarded as a mature form of mRNA and the final template for the protein to be synthesised. The transcription process is schematically illustrated in Figure 3.

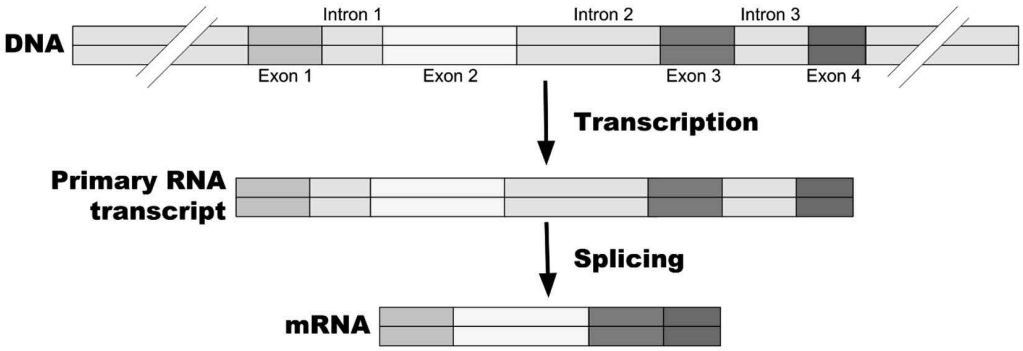


*Figure 3: Schematic representation of transcription and splicing, https://commons.wikimedia.org/w/index.php?curid=72095048. CC BY-SA 4.0*

The produced mRNA strands bind to structures in the cell called *ribosomes*, often referred to as 'cells' protein factories'. Their sequences are read by the ribosomes and translated into corresponding sequences of amino acids, which are assembled as the ribosome moves along from the start to the end of each mRNA, as visualised in Figure 4. The amino acids are each transported to the ribosome by specific molecules called tRNAs. The ribosomes then assemble the amino acids in the order described in the mRNA. When the reading – 'translation' – of each mRNA is finished, a translated counterpart of the information contained in the DNA sequence transcribed in the mRNA (a polypeptide) has been formed, as shown in Figure 4.

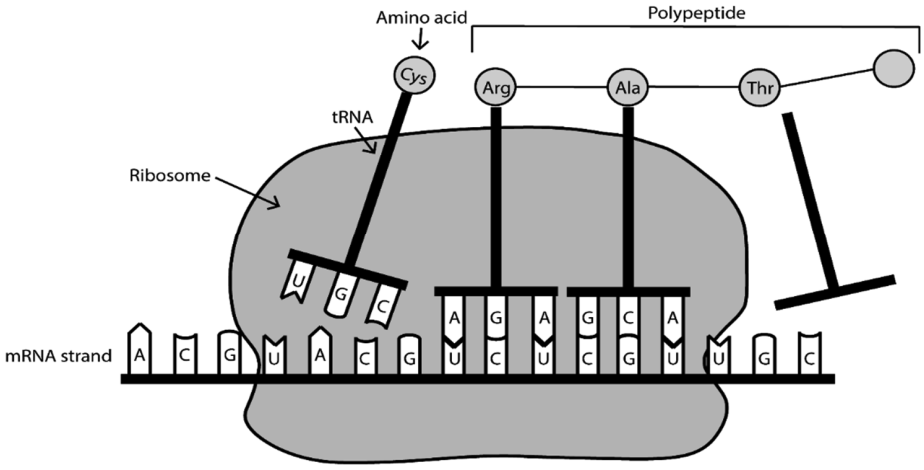


*Figure 4: Schematic diagram of the translation process. The letters found on the mRNA-strand refers to its corresponding nucleotide. The abbreviations found in the circles on the polypeptide chain indicates the corresponding amino acid, https://commons.wikimedia.org/wiki/File:Protein_Synthesis Translation.png. CC BY-SA 4.0.*

The translation product then undergoes post-translational modifications that change it to its functional 3D-structure. In the preceding paragraph it is called a 'polypeptide', but sometimes it is called a 'protein' or simply 'peptide'. Various authors (and students who participated in Study I) have used the number of amino acids in the chain to distinguish between polypeptides and proteins. A pair of amino acids is called a dipeptide, a chain with three amino acids a tripeptide. The difference between a polypeptide and a protein is however imprecise and academic where the common adoption is set to an upper limit of around 100 amino acids ("Peptide," 2018). All above is often regarded as proteins. Nevertheless, the terms peptide, polypeptide and protein are often used interchangeably in the literature, regardless of numbers of amino acids, as long as a chain is formed. Shortly after the reading is finished, mRNA will be degraded in the cytoplasm, and its constituent parts can be used again for building new mRNA-strands.

### Three models representing protein synthesis

In a brief content analysis of some widely used biochemistry and molecular biology textbooks at university level, I identified three educational representations of protein synthesis. In the thesis, they are named according to the molecular processes leading to synthesis of a

protein they include: the translation model ('t-model'), the transcription–translation model ('tt-model') and the transcription-splicing-translation ('tst-model'). These models are briefly presented in Paper II and described below.

### a) The translation model ('t-model')

In this model (the most condensed of the three), protein synthesis is defined as consisting solely of the translation process in the cell, so protein synthesis is equated to translation. As used, for instance, by Nelson and Cox (2013), and Tymoczko et al. (2013), the 't-model' describes how information embedded in mRNA is interpreted, by ribosome-mediated assembly of an amino acid chain. The product is described as either a polypeptide or a protein.

The crucial concepts for understanding protein synthesis according to the 't-model' are: 'amino acid', 'mRNA', 'polypeptide', 'protein', 'ribosome' and 'tRNA'.

### b) The transcription-translation model ('tt-model')

In the 'tt-model', protein synthesis is defined as consisting of the transcription process followed by the translation process, as explained for instance by Campbell et al. (2003). The transcription component of the model describes the process whereby an mRNA-strand corresponding to a sequence in the genetic material is created. The translation component is the same as the process described in the 't-model'.

The concepts that are important for understanding protein synthesis according to the 'tt-model' are: 'amino acid', 'DNA', 'gene', 'mRNA', 'polypeptide', 'protein', 'ribosome' and 'tRNA'.

### c) The transcription – splicing – translation model ('tst-model')

The 'tst-model', the most elaborate of the three, includes the processes transcription, translation and mRNA maturation (splicing). Thus, it adds splicing to components of the 'tt-model'. The splicing process is described as removal of parts of the mRNA, denoted introns, and joining of the remaining parts, exons, to form a mature mRNA. This mRNA is then translated as presented in the 'tt-model'. For examples of its use see Alberts et al. (2008) and Sadava et al. (2014).

The concepts that are important for understanding protein synthesis according to the 'tst-model' are: 'amino acid', 'DNA', 'exon', 'gene', 'intron', 'mRNA', 'polypeptide', 'protein', 'ribosome' and 'tRNA'.

As already mentioned, following completion of the translation process, the product may be subjected to several other processes that form the translation product into an active protein. Neither these so-called *posttranslational modifications*, nor epigenetic aspects of protein synthesis, are included in any of the presented models. These aspects of protein synthesis are not covered in this thesis either, because they are not included in any of the upper secondary level teaching material identified and considered in Studies I and II. Thus, these processes are apparently irrelevant to analysis of current practices of teaching protein synthesis at that level. Nevertheless, there may of course be other models that include more complex aspects (notably models that students may encounter on the Internet) and may be of interest for developing protein synthesis education beyond the scope of this thesis.

In this thesis, the default interpretation of protein synthesis is the 'tst-model' as it is the most elaborate of the three models and includes the aspects covered in current upper secondary protein synthesis education.

### Obstacles for learning protein synthesis

In the life sciences, subject matter is sometimes presented using analogies and metaphors to highlight or explain phenomena and processes, as respectively illustrated by Haglund and Jeppsson (2012) and Venville and Treagust (2002). Their use may be crucial for learners to understand scientific phenomena (Lakoff and Johnson, 1980), but it may also potentially cause problems for students, as described in the previous section. If students cannot interpret the essence of the analogies and metaphors, they may misinterpret the central message (Rundgren et al., 2009; Tibell and Rundgren, 2010; Venville and Treagust, 2002). Whenever this happens, or there is an alternative conception of the biochemical concepts, there is a risk of the student learning the wrong conception or turning away from learning the life sciences due to the complexity of the communication of the field (Tibell and Rundgren, 2010).

### Students' understanding of protein synthesis

Protein synthesis education has received little research attention to date, although some aspects have been addressed by authors including Rundgren (2006) and Martínez-Gracia et al. (2006). However, numerous studies have considered educational aspects of other areas of molecular life sciences, and associated students' learning difficulties, which have at least partial relevance to learning protein synthesis. These include studies on understanding of central concepts in life sciences, and mapping of associated learning obstacles (Allchin, 2000; Duncan and Reiser, 2007; Gericke and Hagberg, 2007; Gericke et al., 2013; 2014; 2017; Lewis and Kattman, 2004; Smith and Gericke, 2015; Thörne and Gericke, 2014; Thörne et al., 2013; Venville and Treagust, 1998; 2002). More specifically, considerable attention has been paid to students' understanding of the structure and function of genes (Duncan and Reiser, 2007; Lewis et al. 2000; Marbach-Ad, 2001; Marbach-Ad and Stavy, 2000; Thörne and Gericke, 2014; Wood-Robinson et al., 2000). All of these studies indicate that students have difficulties in understanding central concepts in communicating life science processes.

On the topic of protein synthesis, there are some reports of students' difficulties in understanding its canonical representation (Marbach-ad,

2001; Wright et al., 2014). Although some 60 years have passed since the first attempts to explain the formation of proteins, students still often display weak conceptual understanding of the different descriptions of how proteins are synthesised (Wright et al., 2014). One of few studies concerning the communication of protein synthesis is by Wright et al. (2014). They challenge representations of the central dogma, particularly the use of arrows, which they claim can lead to misconceptions, such as students thinking that DNA is converted into RNA and the RNA is then converted into protein. Thus, Wright et al. (2014, p. 338) suggest that "use of this representation during instruction can be counterproductive unless educators are explicit about the underlying meaning." This, and the general scarcity of research on how to present and teach protein synthesis, and how students comprehend the topic, indicates a clear need for further studies.

### Science textbooks

There are several sources of information that students come across in the learning of science. These are both oral, such as teachers, peers, TV and the Internet, or written, for instance textbooks, other books and texts from media such as the Internet. Students learn subject matter in both formal (largely school) settings and informal settings. In this thesis (and Study II), I have focused on textbooks because of their importance for formal learning of life sciences within the school system (Nelson, 2006).

Textbooks are rich sources of information (Nelson, 2006) that play a significant role in students' learning (Ekvall, 2001). Moreover, despite the rapid growth and accessibility of digital resources, students still seem to prefer textbooks over E-books (Woody et al., 2010). However, little is known about how science textbooks present specific topics of molecular life sciences and even less about how textbooks present the same content within different contexts.

The limited investigations to date include a contribution by Gericke et al. (2014), who found great similarities in usage of gene function models related to classical genetics among textbooks used in different countries with different curricula. In addition, Martínez-Gracia et al.

(2006) found that textbooks need to be updated to provide an understanding of basic technical terms by putting the terms into broader contexts. The impact of context on learning vocabulary is elaborated in the following section.

### Learning domain-specific language in different contexts

Following Duranti and Goodwin (1992), context is regarded here as the discourse embedded in a cultural setting, and the usage of a specific language as an important component of context. Thus, biology and chemistry textbooks are investigated as emanations from two similar and related, but different contexts, with potentially significant differences in domain-specific language.

The US National Reading Panel (National Institute of Child Health and Human Development, 2000) states that vocabulary should be presented in ways that enable learners to find it useful in different contexts, and content should be taught in rich contexts to enhance learning of vocabulary. It has also been argued that teaching concepts in various contexts will increase the likelihood of students understanding and learning the concepts (Butler et al., 2010; Gilbert, 2006; Stahl and Kapinus, 2001). Thus, teaching concepts in several contexts is advocated to enhance vocabulary learning.

Texts that are embedded in different contexts will have different linguistic features (Fang, 2004). The US National Reading Panel argues that students' understanding of domain-specific texts will be better if the vocabulary in them is derived from a domain-specific content learning material (National Institute of Child Health and Human Development, 2000). Textbooks are examples of such material. However, different chapters may contain technical terms that are not familiar to the student (Pittman, 1999), which have been shown to be potential learning obstacles in learning genetics from textbooks (Gericke and Hagberg, 2010a: b; Gericke et al. 2013).

In chemistry textbooks, protein synthesis is usually described in a biochemistry context, whereas biology textbooks usually describe it in molecular biology chapters.

### *Students' learning from texts*

Learning the vocabulary of relevant domain-specific language is one of the most important components of learning from a textbook. Traditionally, vocabulary teaching in foreign language education has focused on de-contextualisation of words or teaching them as 'isolated islands' (see, for instance, Criado and Sánchez, 2009). However, Holliday and Cain (2012) claim that unguided trial-and-error methods for reading science texts are inefficient, so there is a need to develop reading comprehension strategies in science education, i.e. techniques enabling students to gasp the meaning of texts they read. Such techniques may include inference of meanings from contexts, identification and summary of key points, use of semantic or graphic organisers, development of questioning strategies, and self-monitoring by students of their own understanding and difficulties (see for instance Butler et al., 2010; Gilbert, 2006; Holliday and Cain, 2012; Stahl and Kapinus, 2001)

Investigating students' understanding and the learning potential of textbooks' presentation of topics that are known to be difficult for students to understand is an important task for educational research to address in order to develop effective teaching and learning strategies. To assist efforts to address this task, this thesis highlights and discusses difficulties in understanding protein synthesis through domain-specific language.

### Structure of language

The thesis focuses on the domain-specific language of molecular life sciences, specifically the conventional language of biochemistry and molecular biology. Thus, it is important to acknowledge that language operates in contexts. According to Halliday et al. (2014, p. 32) this means that language "is always theorized, described and analyzed within an environment of meaning". Therefore, communication depends on the context in which it occurs.

Words, or more precisely concepts, were focal concerns in both studies underlying this thesis. In Study I, spoken words were treated as mental descriptions, i.e. conceptions, of students' understanding, while in Study II words in the form of technical terms in texts, and their roles in

the texts' meaning-making capacity, were considered. In Study II, sentences were also considered, as indicators of domains in which technical terms are used and related to other technical terms. As already suggested, and further discussed in the next section, there are many semantic concepts with similar meanings.

### *Language, word, term and concept*

To investigate the complexity of language construction and communication, it is important to define the constituent elements of a language. This is done here by addressing nuances of the central terms in this thesis – *language*, *word*, *term* and *concept*.

The Oxford English Dictionary defines language as "the method of human communication, either spoken or written, consisting of the use of words in a structured and conventional way" ("Language," 2018*)*. Fundamental elements of a language are *the words* that it consists of and the grammar used to construct texts or oral statements with the words. A word can be seen an entity that consists of a specific combination of letters (Hultman, 2003). It can be used to "describe a thing or express a concept" and when a word is seen as a specific expression, we can call it a *term* ("Term," 2018). However, a term is merely a model for describing part of the world, and thus may have several meanings, as mentioned in the background section, where the semantic triangle was introduced (Ogden et al., 1923).

A word may also be related to a certain knowledge domain, for instance 'molecular life sciences' or 'biochemistry', and can then be regarded as a *technical term* (Halliday and Martin, 1993). Here, technical explicitly expresses that it is related to a specific domain of knowledge. When assigning a term with a meaning, the term, or technical term, if it is related to a specific domain of knowledge, can be regarded as a *concept*.

*A concept* can be regarded as a mental description of the meaning of a word (Löbner, 2002). The use of concept instead of the term thereby includes a level of symbolic representation of the corresponding term. The meaning of a word may also be related to mental constructions in the *conceptions* of an individual. That is, there may be mental constructions that differ from the established or scientific meaning of

the term, and thus 'alternative conceptions' (Tibell and Rundgren, 2010). However, as noted above, a concept can be regarded as being expressed by a term ("Term," 2018). Thus, the term also acts as a label for the concept, regardless of any meaning-making process and other possible conceptions.

As can be seen from the preceding sections, usage of the words *word, term* and *concept* is complex. In the research literature, term and concept are often used synonymously even when addressing students' understanding. Hence, this usage adds a level of meaning-making to the discussion without separating the term from the corresponding concept, and vice versa, and taking the definitions into consideration. As Brown and Ryoo (2008) stress, concepts exist independently from the corresponding terms. The concepts (or conceptions) themselves can be regarded as representations of understandings of the terms. Differences in uses of terms and concepts in published studies complicate interpretation of results of analyses of 'students' understanding' of different contents, and highlight a need for caution when *term* and/or *concepts* are used. Therefore, comparing the more general aspect of a concept and students' subjective conceptions of the same matter is not straightforward.

Technical terms specifically used in a certain domain, in this case molecular life sciences, are regarded here as 'domain-specific vocabulary'. The full semantic properties of particular words were not addressed in either Study I or Study II, as students' understanding and meaning-making capacity were the primary concerns in the respective studies. Thus, I used 'concept' in slightly different ways in the two studies, partly because of development of my own knowledge of semantic models. Therefore, I used 'concept' in Study I and 'technical term' in Study II.

Based on the theoretical framework already outlined in this thesis, 'technical term' is used when referring to elements of domain-specific vocabulary, and 'concept' is only used when specifically referring to a mental description or meaning of a technical term. In addition, I use the words 'core concept' and 'peripheral concept', as they are well defined in Paper I, and core concerns of the paper, so changing these

central words would be potentially confusing. I also choose to use 'concept mapping' for mapping words in hierarchical structures, as proposed by Novak and Gowin (1984), because this phrase is deeply rooted and associated with the specific structure and method. Relabelling it as "technical term mapping" would not contribute substantially to the coherence of the thesis, and thus is not done.

### Students' compartmentalisation and clustering

An important aspect of learning molecular life sciences is the ability to relate different technical terms, which (as mentioned) are commonly referred to as concepts in the literature (*e.g.* Novak and Gowin, 1984), to one another. This can be visualised, for instance, in concept maps, which display the handling of concepts by someone, a community, texts or other material, and their relationships (Novak and Gowin, 1984).

One way that concepts can be related to each other is in 'isolated islands', by *compartmentalisation*. This refers to the recognition of relationships among a group of concepts that are isolated from other groups, in a *compartment* or *cluster*. In the thesis, compartment and cluster are used synonymously.

As shown by Marbach-Ad (2001), students tend to compartmentalise concepts of molecular life sciences. Several studies have shown that students can often reason about domain-specific concepts, but perceive very few relationships between groups of concepts that they can reason about (Bahar et al., 1999; Lewis and Kattmann, 2004). Lewis et al. (2000) claim that students must be able to understand relationships between central concepts used to communicate cellular processes in order to form a coherent conceptual understanding. This suggests that a more integrated presentation of the relationships between concepts and groups of concepts within life science education would enhance learning.

To show how concepts are related, the concepts can be sorted into concept maps. These can be used for several purposes, for instance to visualise students' understanding (Lewis et al., 2000; Glaser and Bassok, 1989; van Zele et al., 2004). The patterns created through concept maps may be translated into indicators of how well the

students understand a particular topic through the choice and use of terms for explaining the topic and the ability to relate them to each other (Mintzes et al., 2005; Novak and Gowin, 1984).

Concept maps may have various hierarchical levels of complexity, which Kinchin et al. (2000) refer to as 'spoke', 'chain' and 'net' structures (Figure 2). This hierarchy can be used to evaluate students' level of understanding of a topic.

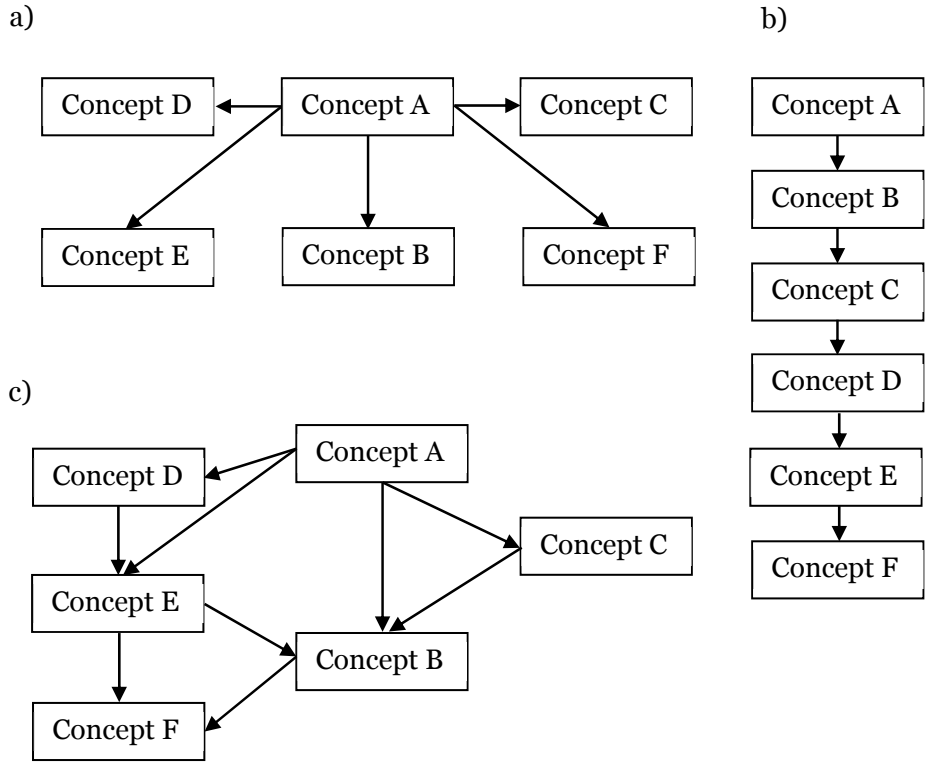


*Figure 2: Illustration of three concept map structures described by Kinchin et al. (2000, p. 47): a) 'spoke' b) 'chain' and c) 'net'.*

According to Kinchin et al. (2000), 'spoke' structures (Figure 2a), are the most basic of the three types, indicating that all of the included concepts are related to a core concept. The next level, 'chain' (Figure 2b) refers to a sequential presentation of the concepts, where each concept is linked to two others. The most complex of the three ways of relating concepts in concept mapping is the 'net' structure (Figure 2c), in which concepts are heavily interrelated. Thus, the complexity of concept maps required to represent students' understanding is positively correlated with the depth of their knowledge, and analysis of

the relationships and cross-linking in a concept map can shed light on their level of understanding (Novak and Cañas, 2008). In Study I, these insights were applied to probe overarching structures of students' understanding of central concepts in protein synthesis.

## Theoretical framework

The focal concern throughout the writing of this thesis, and the underlying studies, has been communication through domain-specific language, and it has been addressed using a constructivist perspective of learning. Study I focused on students' understanding through a social constructivist perspective (Brooks and Grennon, 1999; Fisher, 2000; Scott et al., 2007). Study II focused on the meaning-making capacities of texts in terms of possibilities for students to understand communicated messages, thereby constructing the student's understanding as in a constructivist framework.

### Reasoning in science: mechanistic vs conceptual approaches

A molecular or cellular process can be explained from either a mechanistic or conceptual perspective. According to Machamer et al. (2000) a mechanistic explanation of a cellular phenomenon includes the "how" and "what" of the associated cellular structures and activities. That is, a mechanistic explanation involves explanation of how a phenomenon is produced, through participating structures (so-called *entities*), that must be explained. What the entities do – the activities – and their relationships also need to be addressed (Craver and Darden, 2013). Learning mechanisms is no easy task for students (Duncan and Reiser, 2007; Haskel-Ittah and Yarden, 2017). However, Haskell-Ittah and Yarden (2018) suggest that students benefit from hints of mechanisms early in teaching sequences, even if the mechanisms will be considered later in the teaching. A conceptual explanation provides another description of the same content knowledge, such as protein synthesis, based on causal rather than mechanistic relationships between technical terms, or meanings rather than processes and activities (Scott et al., 2007). Thus, a conceptual explanation may include a more overarching presentation of a phenomenon. However, research in life science education has shown

that students also find conceptual aspects very challenging (Gericke and Smith, 2014).

## Aim and research questions

The aim of this licentiate thesis is to contribute to the understanding of teaching and learning protein synthesis through domain-specific language in upper secondary education.

Within the context of protein synthesis, the research questions are:

- What similarities and differences may be discerned between patterns of technical terms in student's reasoning and in chemistry and biology textbooks' presentation of protein synthesis?
- What impact does the biology and chemistry context have on the use of technical terms in chemistry and biology textbooks, respectively, and how does this relate to students' reasoning about protein synthesis?

## Methods

The methodology used to gather and analyse data is presented in more detail in this section. In Study I, data were collected through semi-structured group interviews and analysed by content analysis using a concept mapping approach. In Study II, textbook sections were selected and analysed by a content analysis approach using data mining techniques.

### Selections and sampling

Several samples were used in the studies, as described in the following section. Similar samples of concepts were used in both studies, but in Study I there was also a sample of respondents, while in Study II there were also samples of textbooks and texts in the textbooks.

### *Selection of technical terms*

In Study I, upper secondary students' textbooks were examined with a brief content analysis, to obtain an overview of the technical terms it was reasonable to assume that students of this level would have met in texts when learning the subject. These technical terms were then compared to those in descriptions of protein synthesis in several randomly chosen university-level biochemistry and molecular biology books, in efforts to identify those commonly used to present protein synthesis.

Most technical terms included in the resulting sample are related to the traditional way of expressing protein synthesis through the central dogma, as described in the background section. However, technical terms that are commonly used to present the mRNA maturation process were also included, as in the 'tst'-model, which includes the largest number of technical terms of the three models presented in the background section.

The technical terms that were chosen as most important for understanding protein synthesis were, in alphabetical order: amino acid, DNA, enzyme*, exon, gene, intron, mRNA, peptide, protein, rRNA* and tRNA. The asterisks indicate technical terms that were only addressed in Study I.

Some of the sample terms were labelled *core concepts* and others *peripheral concepts*. As previously explained, the use of 'concept' and 'technical term' varies between the two studies. In Study I, 'chromosome', 'DNA', 'gene' and 'protein' were labelled core concepts. These have previously been described as central concepts in life sciences education (Marbach-Ad, 2001). These four technical terms were also the most frequently mentioned technical terms in molecular life sciences in media at the time (Tseng et al., 2010). I therefore assumed that it was reasonable to think that the participating students would recognise these technical terms as they were used in several contexts in their everyday life, including formal teaching. The other concepts in the sample were regarded as peripheral.

Supporting structures such as ribosome and various enzymes were not included in either study, because they are emphasised less in upper secondary school education. As RNAs play a central role in protein synthesis teaching at secondary level it seemed reasonable to include 'rRNA' in the sample, although it can be classified as a constituent of ribosomes. As shown in Paper I, the participating students were unaware of rRNA, and it was omitted in Study II.

The technical term 'enzyme' was also omitted in Study II as it is related to the function of a specific group of proteins rather than the more general form of translation products, and students rarely mentioned the term in Study I.

In Study II, the 'core concepts' identified and addressed in Study I were reduced to three technical terms: 'DNA', 'gene' and 'protein' because 'chromosome' is not presented in relation to protein synthesis in the examined textbooks. In addition, the sample of peripheral concepts mentioned in Paper I was reduced by excluding 'rRNA' and 'enzyme', as previously described.

### *Sample of respondents in Study I*

The respondents in Study I were eleventh graders 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 Chemistry A and Biology B courses, which were equivalent to current Chemistry 1 and Biology 2 courses, respectively. Syllabi of these courses included protein synthesis. A convenience sample was formed of 12 students from three classes, who expressed willingness to participate.

### *Sample of textbooks in Study II*

Various textbooks used in Sweden are commercially available, and each school and teacher can choose what textbooks to use in their classes. Thus, several textbooks covering similar content may be used in the same school. To obtain a systematically grounded understanding of currently used textbooks' presentations of protein synthesis, a set of textbooks that met the following criteria was selected. Each book had to be the latest edition commercially available in Sweden at the time of

data collection, and expressly written to facilitate teaching and learning in the Biology 1 or Chemistry 2 courses, in accordance with the current curriculum, 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 targeting these courses that were available from the largest publishers in Sweden at the time.

Notably, several editions of each textbook and a few other textbooks from other publishers were available for the teachers to choose from. However, they were omitted to focus on a manageable set of the most widely used textbooks.

### The educational context – Swedish upper secondary school

Swedish upper secondary school is non-compulsory school for students between 16 and 18 years old, who enrol for either academic programs that provide preparation for university studies or vocational programs that provide preparation for occupations such as healthcare, industrial or media work. All programs include common core subjects such as languages (Swedish and English), general science, mathematics, and physical education and health (Swedish National Agency of Education, 2013).

The specific science subject courses of chemistry, biology and physics are only included in the national Natural Science Program and only taken by science majors, who also take courses in advanced mathematics. Studies I and II addressed parts of the molecular life sciences components of chemistry and biology courses of the National Science Program (Swedish National Agency of Education, 2011a; b).

Chemistry and biology are taught in four courses, each of approximately 80-100 hours duration (Chemistry 1 and 2; Biology 1 and 2). In each case, the second course follows the first course, but in other respects the order in which they are taken depends on the school's timetabling. The syllabi follow a common structure, beginning with an introduction and the 'aim of the subject' followed by a

description of the 'core content' and 'knowledge requirements' (Swedish National Agency of Education, 2011a; b). Course 1 in each subject builds on compulsory school education, or equivalent, and is succeeded by course 2 in the same subject. More specialised courses in the science subjects, for instance in biotechnology or molecular biology, may also be offered, depending on the school's profile and competence. Protein synthesis is currently taught in Chemistry 2 and Biology 1 courses.

The 'core content' section of the Chemistry 2 course syllabus states that teaching should cover the following core content (of biochemistry):

- 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 course syllabus states that teaching should cover the following core content (of life sciences):

- 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 show that protein synthesis is included in both chemistry and biology curricula, but with different emphases. In the Biology 1 syllabus, protein synthesis is explicitly classified as core content, as a key part of the "life sciences" content. In the Chemistry 2 syllabus, the 'protein synthesis' concept is implicitly rather than explicitly mentioned, in terms of *the genetic flow of information* and *the processes of transcription and translation* in the "biochemistry" content (Swedish National Agency of Education, 2011a).

### Data collection

Data were collected in Study I through semi-structured group interviews, and in Study II from the texts describing protein synthesis in the sample of chemistry and biology textbooks.

### *Data collection in Study I*

The students were interviewed in groups, as 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, such as all the students having the same timetable. All the students in each group were familiar with each other.

The interviews were structured in accordance with recommendations to divide each interview into defined phases (Kvale, 1989): an introductory briefing phase, main phase, debriefing phase and reflection. An 'acclimatisation phase' was also added as the students were provided some pre-information on the topic of the interview.

In the introductory phase, the students were asked to explain and relate the 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 were asked to explain and relate the new term to those they had already mentioned. During each interview, the group was asked to build a so-called 'concept map' (see for instance Novak and Gowin, 1984) while talking about the technical terms.

### *Data collection in Study II*

In Study II, data were collected from the selected textbooks by identifying sections covering protein synthesis. This was done by scrutinising the table of contents in each textbook to identify sections that address protein synthesis or processes that could be connected to it, such as sections focusing on translation. The sampling was the first step of a SEMMA (Sample, Explore, Modify, Model and Assess) analysis, which is further described in the data analysis section below.

## Data analysis

The data analysis procedures in each of the studies are summarised in this section. For a more thorough description please see the papers.

### *Data analysis in Study I*

In Study I, the interviews were transcribed verbatim. The transcribed excerpts from the interviews were categorised in an iterative process to reveal patterns of technical term usage (Novak and Gowin, 1984). One level of analysis was the scientific accuracy in the students' explanations of the different concepts. The students' explanations were analysed separately and compared to scientific explanations of each technical term, called *labelling* in Paper I.

The analysis was done at three levels: concept-, cluster-, and relationship levels. The students' ability to accurately explain the technical terms of the sample was analysed at the concept level. The concept level analysis involved identification of technical terms in the transcripts, then recording of the order in which they were mentioned and evaluation of the scientific accuracy of their explanations. At cluster level, the analysis focused on the students' tendency to form groups of the technical terms, which they were found to address together, i.e. in clusters. The relationship level analysis focused on the students' ability to express scientifically accurate relationships (through comparison with scientific explanations in university level texts) within and between the identified clusters.

The concept maps were regarded as facilitators of communication in the group interviews rather than material for analysis in that particular study. Maps showing the clusters were generated from the transcripts, and relationships were extracted from them.

### *Data analysis in Study II*

Uses of technical terms in the selected chemistry and biology textbooks were analysed to reveal the texts' 'conceptual demography', i.e. the frequencies, distributions and relationships of the core and peripheral concepts in them. The collected data were analysed by applying a SEMMA approach to content analysis derived from data mining techniques, as proposed by Shmueli et al. (2010). The protocol guiding the treatment of the data is presented in more detail in Paper II. The SEMMA approach was deployed throughout, with the exception of the last step. It is also important to acknowledge that the volume of analysed data was much lower than in most studies using data mining

and a SEMMA-approach. However, the protocol was found to be useful even for processing the relatively small amounts of data acquired in Study II, as described in more detail and discussed in the discussion of methodology section.

Briefly, the SEMMA protocol includes an initial routine that focuses on understanding the purpose of the analysis to pinpoint what needs to be analysed. Then the focus is on obtaining the dataset that will be analysed. The next stage includes exploration, cleaning/reducing, and/or preprocessing of the data followed by definition of the computational task. The last stages of the protocol involve choosing an appropriate data technique for the task (in Study II, this was interpreted as an appropriate algorithm, which had to be constructed) and its implementation.

The algorithm was constructed in Microsoft Office Excel® with the aim of counting and pinpointing the selected terms in the sampled texts. The algorithm counted total frequencies of occurrences of each technical term in each text, and summary statistics such as mean values were manually calculated from its output. The data were extracted from spreadsheets generated by the algorithm. As the texts varied in length, I chose to work with frequencies normalised with respect to text length rather than absolute frequencies.

The frequency of each sample term was extracted from word counts in the spreadsheets, mean values were calculated and mean normalised values for the indexed texts representing each context were calculated. A two-sided t-test with a 0.05 significance threshold, implemented in SPSS® (v. 22), was used to identify significant differences between frequencies of terms in the two contexts. The distribution of each technical term was characterised by the algorithm as the number of appearances in each sentence, from the beginning to the end of each text. Relationships between technical terms in each text were defined as their coexistence in a sentence. Each such relationship was only counted once by the algorithm, regardless of how many times each technical term was mentioned in a sentence.

### Validity and reliability

The methodology applied in Studies I and II were specifically designed for them. The validation of the methods and reliability of the studies are discussed in the following section.

In Study I, a mixed method approach was adopted (Robson, 2008). The interviews followed the same interview guide, with the same interviewer, at the same school in order to generate valid data. The interview guide was tested before the interviews and small changes were made before the data collection. The students validated their reasoning by constructing concept maps. For this, they were asked to repeat their reasoning and validate their comments at the end of the interview by going through the constructed concept maps, which were used as tools for facilitating the respondents' discussion during each interview (Kvale, 1989).

A computer-generated algorithm was programmed specifically in Study II. Thus, rigorous validation was crucial to fine-tune the novel algorithm. This was done in rounds, by testing it using multiple sets of texts with diverse types of content, such as instructions, brochure texts and online texts. The output from the algorithm was manually scanned in cycles to check for errors. Spot-checks of the texts was conducted, and the output from the program was cross-checked by scrutinising the original texts to validate the process, as the algorithm had never been run before the study. The cycles for each text were completed when no more errors were found, resulting in fewer and fewer cycles with increasing number of scanned texts. Because all the data are computer generated, the reliability of the study is excellent.

### Ethical considerations

Some of the data presented in Paper I were gathered from real people. When dealing with living beings, care must be taken to minimise any harm that could be potentially caused to participants. It is also crucial to carefully choose an appropriate research design that meets all relevant ethical criteria. Therefore, both studies followed ethical guidelines for data collection and analysis published by the Swedish Research Council (2002 to local). The design was also carefully

discussed with peers, according to guidelines at Karlstad University, prior to data collection and analysis.

The data in Study I were gathered from respondents who were all over 16 years of age. Therefore, the students themselves decided whether to participate or not. Careful instructions were provided for the students to read before they were interviewed. The purpose of the study was also presented orally directly before the start of the data collection. In addition, they were informed that they could withdraw from the study at any time without giving any reason and without any penalty.

## 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. They tend to compartmentalise their understanding by building 'islands of technical terms', here denoted *clusters*. Technical terms the students used together were placed in the same cluster. The following five concept clusters were identified in the students' recorded reasoning (incorporated technical terms in brackets, see also Figure 5):

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

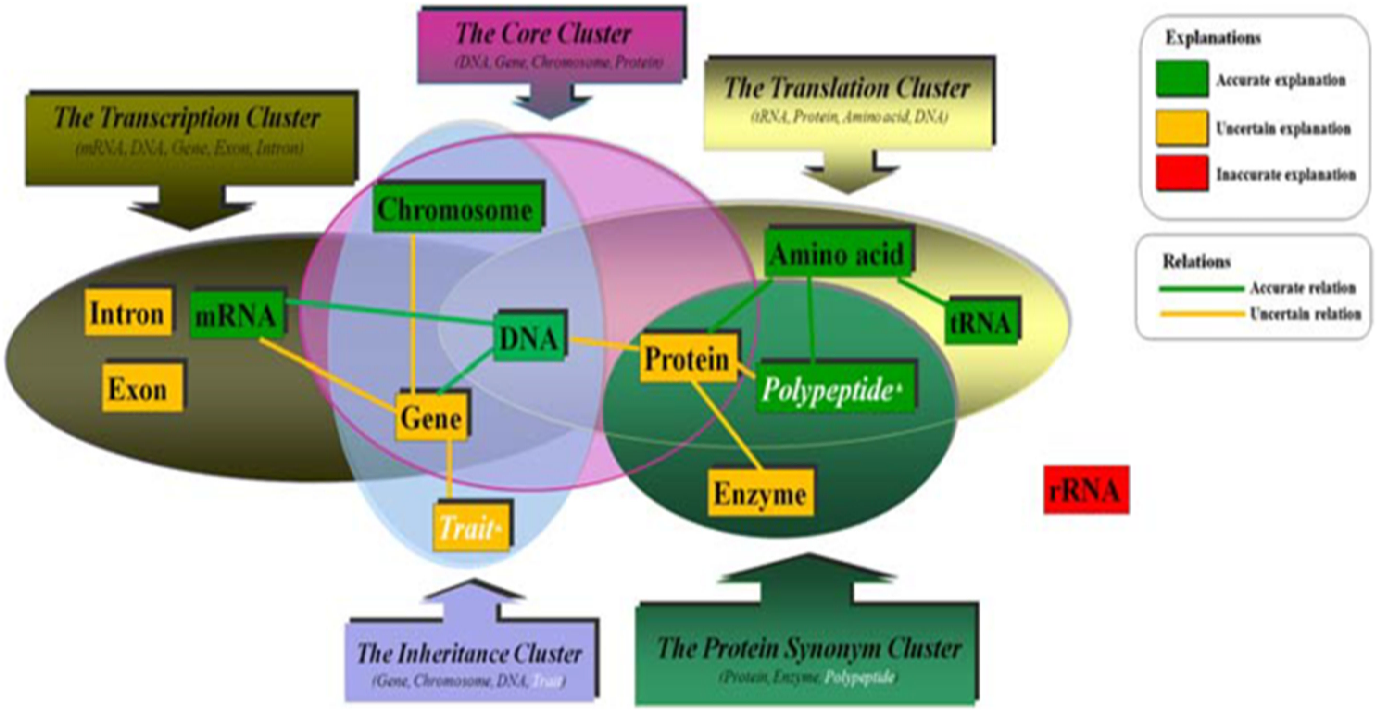


*Figure 5: Labelling of concepts, identified clusters and relationships between the concepts. Concepts in black colour text are part of the original sample. Concepts in white colour text were introduced spontaneously by the students. Reprinted with permission from Journal of Biological Education. Original image: https://doi.org/10.1080/00219266.2012.716785.*

The core concepts ‘DNA’ and ‘chromosome’ were generally labelled correctly by the students, but they showed vague comprehension 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 product of protein synthesis caused difficulties as the students were unable to differentiate between ‘protein’, ‘polypeptide’ and ‘enzyme’, but ‘polypeptide’ was the most scientifically accurate label of these three. The concept ‘trait’ was not part of the original sample and was introduced spontaneously of the students. However, they displayed a vague comprehension of the concept and its relationships.

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’. 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. The 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.

The key relationships bridging clusters were those linking the cluster of core technical terms (‘DNA’–‘Protein’) to three other clusters. The ‘core cluster’ was also bridged to the ‘transcription cluster’ via the ‘DNA’–‘mRNA’-relationship. In addition, the ‘transcription cluster’ was bridged to the ‘translation cluster’ by the same relationship.

### *Main results of Study II*

The main findings from study II show that there are context-linked differences in descriptions of protein synthesis between chemistry and biology textbooks. These differences were revealed through analysis of the conceptual demography of the textbooks associated with each context.

Overall, the chemistry textbooks focus on peripheral concepts, especially ‘tRNA’, and draw no distinction between the importance of

core and peripheral concepts. Biology textbooks, in contrast, put more emphasis on the core concepts.

The descriptions in both contexts follow the canonical presentation of protein synthesis. Chemistry textbooks compartmentalise it through technical term usage into its component processes of transcription, splicing and translation, strongly focusing on the latter part of protein synthesis, making 'tRNA' the central concept in protein synthesis. Biology textbooks also note the role of 'tRNA', but with less emphasis in terms of both frequency and relationships.

The technical terms associated with the starting point as presented in the canonical representation – 'DNA' and 'gene' – are treated differently in the two contexts. 'Gene' is more evenly distributed in biology textbooks than in chemistry textbooks throughout the entire text lengths. The technical term 'DNA' follows the same distribution pattern in both contexts, but chemistry textbooks tend to focus slightly more on the term in the first half of the texts than biology textbooks.

Notably, 'mRNA' is a central and evenly distributed concept in both contexts. However, biology textbooks use 'mRNA' more frequently in absolute numbers and relate it to larger numbers of other core concepts than chemistry textbooks, making 'mRNA' a more central technical term in biology textbooks.

The technical terms associated with maturation of 'mRNA' – 'intron' and 'exon' – are given less attention than the other considered terms. In cases where they are acknowledged, they are only mentioned a few times in short, dense sections of the textbooks spanning a few sentences in the middle of relevant texts. Notably, biology textbooks also revisit both terms at the end of the texts, but not chemistry textbooks. In the cases where they are acknowledged, they follow similar patterns: the frequency is low, the distribution of them is expressed in short, dense section of the textbook spanning over a few sentences in the middle of the text.

In terms of the product of protein synthesis, chemistry textbooks use both the 'peptide' and 'protein' technical terms, while biology textbooks only use 'protein'.

## Discussion

In this section, I mainly discuss critical facets derived from the results for teaching and learning protein synthesis. A facet is regarded in this thesis as a facilitator or hindrance for teaching and learning protein synthesis related to the domain-specific vocabulary usage.

I highlight 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'. I discuss the following aspects of each facet: learning obstacles that may be related to it, associated meaning-making capacities, and impacts of the biology and chemistry contexts on the content.

### Mechanistic or conceptual descriptions

As shown in Study II, chemistry textbooks tend to promote a rather mechanistic explanation of protein synthesis. In a mechanistic description, entities sequentially participate in spatiotemporally ordered sub-activities, which have well-defined starting points and endpoints (Machamer et al., 2000). In Study II, it became apparent that sections of chemistry textbooks describing protein synthesis typically use a couple of technical terms (i.e., entities) and relate them to each other, explaining their involvement in a particular mechanistic sub-process of protein synthesis. Hence, in each relevant text segment there is a strong focus predominantly on a specific phase of protein synthesis, so these segments are very detail-rich. The following excerpt from a chemistry textbook illustrates the mechanistic treatment of the translation phase in protein synthesis (Andersson, 2012, p. 336, my translation):

> The amino acid-tRNA-complex is attached to the mRNA chain by a sequence of three bases on the tRNA - a base triplet – which is bound to a complementary base triplet on the mRNA chain. [Original text in

> Swedish: “Aminosyra-tRNA-komplexet fästs till mRNA-kedjan genom att en följd av tre baser på tRNA – en bastriplett – binds till en komplementär bastriplett på mRNA-kedjan”.]

Thus, the chemistry textbook texts adopt a rather narrow, mechanistic approach in descriptions of the synthesis of peptides and proteins, focusing largely on their assembly *per se*.

Biology textbooks tend to promote more conceptual reasoning, involving presentation of protein synthesis in overarching conceptual structures. They focus more clearly on gene function, presenting a wider perspective of protein synthesis from genetic material to product in a conceptual manner, paying less attention to sub-processes. This is illustrated by the following passage from Karlsson et al. (2012, p. 170, my translation):

> Every gene codes for a certain protein. Which protein a gene gives rise to is decided by the DNA sequence (the order of the nitrogen bases) in the part of DNA which constitutes the gene. [Original text in Swedish: “Varje gen kodar för ett visst protein. Vilket protein en gen ger upphov till bestäms av DNA-sekvensen (kvävebaseras ordning) den DNA-sträcka som utgör genen”.]

Conceptual reasoning has the advantage of providing explanations of overarching structures, but chemistry textbooks may provide more detailed presentations of specific parts of protein synthesis, as can be seen in the examples above. Thus, chemistry and biology textbooks seem to have different aims in their presentations of protein synthesis. In addition, it seems reasonable to assume that students take the biology and chemistry courses in different orders. The optimal order, at least for learning protein synthesis, may be to take biology course I before chemistry course II. Learners will then encounter the overarching structures of protein synthesis in the biology course and details in the chemistry course later.

As several researchers have shown (e.g. Duncan and Reiser, 2007; Haskell-Ittah and Yarden, 2017; 2018), mechanistic reasoning is difficult for students to adopt. However, understanding of mechanisms may facilitate understanding of overarching structures, and vice versa. Together with Haskell-Ittah and Yarden (2018), I agree that addressing

the underlying central mechanisms in early stages of protein synthesis education could help student learning.

### Compartmentalisation

This facet is concerned with compartmentalisation of technical terms in both students' understanding and textbook texts on protein synthesis. Students' compartmentalisation in understanding of protein synthesis has similarities to both chemistry and biology textbooks' ways of presenting the process. Similar compartmentalisation has also been described by Marbach-Ad (2001). As Novak and Cañas (2006) argue, students' bridging ability (i.e. ability to link concepts within and between clusters) in their reasoning about technical terms reflects their understanding. Thus, the ways textbooks associated with the different contexts relate technical terms are connected to the meaning-making capacity of the texts, according to Novak and Cañas (2006). By focusing on groups of technical terms together in textbook texts, students might – sometimes falsely – get the impression that these technical terms only need to be addressed together within the same cluster and are not related to technical terms of other clusters.

As shown in Study II, chemistry and biology textbooks may contribute to the compartmentalisation of technical terms in students' understanding, as concentrations of sets of technical terms (notably 'intron' and 'exon') in small parts of texts were detected in textbooks of both contexts. Failure to revisit some terms in other parts of texts where other technical terms are addressed may also contribute to compartmentalisation. This can be seen in chemistry textbook texts, for instance, regarding the technical terms 'DNA' and 'gene'. Use of the terms 'DNA' and 'gene' is confined to early parts of the texts whereas 'tRNA' and 'amino acid' are used in later parts, and no relationship between these two sets of technical terms is highlighted.

Focusing only on a specific term may lead to a misunderstanding that a technical term is more central to understanding protein synthesis than it actually is. In addition, if teaching focuses on details, students may miss overarching relationships that are important for a coherent understanding of a topic, as proposed by Novak and Cañas (2006).

### mRNA as a core concept

Results of both studies indicate that 'mRNA' – which was initially regarded as a peripheral concept – needs to be regarded as a core concept in protein synthesis descriptions. However, very little research to date has focused on teaching and learning connected to 'mRNA'. Thus, 'mRNA' needs to be addressed explicitly and given greater attention in research on students' understanding of protein synthesis, as it is frequently used by both textbooks and students, and linked to so many of the other technical terms. As Tibell and Rundgren (2010) stress, the essence of a technical term needs to be highlighted through its meaning-making capacity. Therefore, the position, definition and relationships of 'mRNA' with other technical terms in textbooks highlights its importance in learning protein synthesis.

The extensive linking of 'mRNA' to other technical terms contributes to a net-like conceptual structure, as defined by Kinchin et al. (2000), in both students' reasoning and textbooks' presentations. It can be regarded as a bridging technical term, because 'mRNA' was used as an interconnection between several clusters in students' reasoning recorded in Study I and most of the considered technical terms in books examined in Study II. This central position probably aids the meaning-making capacity of the texts and students' knowledge construction. Based on these findings, I conclude that the technical term 'mRNA' is a critical aspect for successful learning of protein synthesis.

### Canonical representation

In the background section, three models of protein synthesis are presented. Treatments of the process in the chemistry textbooks resemble the 't-model', with a clear focus on the latter part of protein synthesis. They also include rich details, indicating relationships between many peripheral concepts associated with specific parts of the canonical representation. By doing so, chemistry textbooks may become overwhelming for students, as cautioned for instance by Tibell and Rundgren (2010). The different contexts, biochemistry and molecular biology, focus on different parts of protein synthesis, as presented in the overarching structure of protein synthesis in the canonical presentation.

The biology textbooks use an overarching structure that resembles the 'tst-model' as presented in the background section. Focusing on this model may help students to understand the process as a whole. However, their learning of the overarching structure may be at the expense of in-depth learning of individual technical terms and relationships between technical terms. From the results reported here, it seems reasonable to assume that students would benefit from a more conceptual presentation of the overarching structures of protein synthesis at upper secondary school level. However, this needs to be supplemented with in-depth studies of individual technical terms for a coherent understanding

The Swedish upper secondary chemistry curriculum (Swedish National Agency of Education, 2011a; 2011b) includes a largely mechanistic presentation of protein synthesis, focused on its sub-processes, particularly translation and transcription. Analogously with the findings reported here, the biology curriculum advocates more emphasis on the overarching structure, for instance by placing protein synthesis within life sciences and specifically addressing protein synthesis. Although the curricula state the content, but not the fashion in which this content should be taught, it seems reasonable to assume that the curricular contexts of chemistry and biology will affect the teaching and learning of protein synthesis.

As addressed in the background section, the domain-specific language used to communicate protein synthesis is rooted in the field of cryptology and informatics (Kay, 2000). As can be seen in the results reported here, dead metaphors (Rundgren, 2006) are present in the textbooks. That means that today, when (for instance) splicing of ropes is merely a hobby or strictly associated with specific occupations rather than everyday life, as it was some 60 years ago, the understanding of the technical term 'splicing' may not be as easily grasped through the metaphor today. In addition, the increasing use of increasingly complex office and household computers which may strongly influence language and understanding, together with everyday interactions through "social media", will affect how we address items, entities or structures. Hence, it is important to acknowledge and understand the impact of societal fluctuations on any language, as cautioned by Kay

(2000), but on domain-specific language structures and the choice and usage of metaphors in particular.

## Discussion of methodology

In this section, I reflect on the methodology applied in the two studies. Each study had specific features, such as the data collection methods, and shared features, such as most of the domain-specific vocabulary. However, the differences in methodology were substantial, so I address the two studies separately.

Some years ago, I started Study I: the first study I had ever conducted. Although I prepared for every step of the way very carefully, a novel researcher has far less skill than an experienced researcher should have acquired. Interviewing is an artform in itself, which requires experience. However, a researcher in the making must start somewhere. There is no other way to build experience than to practice, practice and practice again. That means that if the study was carried out today, it would be different.

It is important to acknowledge that Study I was carried out before implementation of the current curriculum for Swedish upper secondary school in 2011. So, the textbooks were also slightly different. No attempt was made to compare the teaching and learning of protein synthesis to the content of the textbooks, but the latter seems to change slowly over time (at least in the absence of a paradigm shift). Therefore, it seems reasonable to assume that the selection of technical terms for Study I (and Study II) would change very little if the studies were repeated today. However, the technical term of 'mRNA' seems more important than was generally acknowledged in previous research (Study II). As this was not known at the time of Study I, it was impossible to consider this insight when designing that study, but if I repeated the study again, I would carefully consider the role of 'mRNA' in the selection of technical terms.

In Study II, the methodology involved content analysis based on use of linguistic data mining techniques. Content analysis is a traditional approach for investigating frequencies of technical terms (see *e.g.* Krippendorff, 2004; Graneheim and Lundman, 2004). However, in

this study, further dimensions of content analysis were added by jointly analysing frequencies, distributions and relations of technical terms – 'conceptual demography'. By not only addressing frequencies and distributions of technical terms, but also considering a third dimension of relationships, this provides a 3D-presentation of texts in ways that would be impossible with traditional content analysis.

A fundamental aspect of data mining is use of computational operations for detecting patterns that would be difficult or impossible to detect manually (Feldman and Sanger, 2007). The computational operations often focus on enormous quantities of text and construction of large semantic networks (see, for instance, Shmueli et al., 2010). Thus, computer software used in data mining is often designed for statistical and structural analysis of huge datasets, such as whole dictionaries or massive internet data sources (e.g., Twitter flows and Facebook interactions) and presentation of the outputs.

In Study II, the quantity of data acquired and analysed was nowhere near these amounts. Hence, the complexity of the analysis was far lower. However, there was a need in Study II to analyse texts in a novel way that could reveal more complex structures than ordinary content analysis, and no computer programs were available for such analysis. Because my analysis focused on small, but important, parts of textbooks, there was a need to construct a computer program that could specifically target the relevant conceptual demography. Several commercially available software packages could have been suitable for targeting specific parts of the conceptual demography (e.g., most text-editing programs provide straightforward word counting facilities). However, no program capable of analysing frequencies, distributions and relationship components of the conceptual demography was available. Therefore, I identified a need to construct computer software that was easy to operate but capable of performing these tasks in Study II.

## Conclusions and implications

The aim of this licentiate thesis is to contribute to understanding of teaching and learning protein synthesis through domain-specific language in upper secondary education. The patterns revealed in the studies show that students compartmentalise technical terms that are important for understanding protein synthesis into five clusters with various connections within and between clusters. Similar patterns can be detected in textbooks. Chemistry textbooks tend to adopt more mechanistic descriptions whereas biology textbooks tend to provide more conceptual descriptions. Both students' reasoning and textbooks' presentations follow the canonical representation of protein synthesis. These findings reveal four important facets of learning protein synthesis: 'mechanistic or conceptual descriptions', 'compartmentalisation', 'mRNA as a core concept' and 'canonical representation'.

It is clear from the current literature that students struggle to understand central topics in molecular life sciences. I would like to change the perspective and turn these learning difficulties into learning opportunities. As results of the studies show, students are faced with a hard task to cope with learning protein synthesis through the domain-specific language. The understanding of basic technical terms is important for the development of students' conceptual frameworks in molecular life sciences, as argued by Lewis et al. (2000) and corroborated by findings presented here.

As pointed out in Paper I, I recommend that teaching of protein synthesis should begin by focusing on students' existing connections between technical terms of different clusters. Then technical terms that appear in two or more clusters such as 'DNA', 'gene', and 'protein' should be highlighted. In addition, 'mRNA' needs to be regarded as a core concept.

I encourage teachers and textbook authors to use the domain-specific language as a facilitator for planning, teaching and evaluating life science education generally, and protein synthesis education in particular.

## Future research

The results of Studies I and II indicate a need to broaden investigation of upper secondary education in protein synthesis. To extend the findings reported here, the next steps could be to further explore the meaning-making capacity of protein synthesis descriptions by applying linguistic theories and methods more deeply in analyses of textbooks. For example, systemic functional linguistics approaches could provide valuable insights into the language used in presenting protein synthesis by investigating several dimensions of the texts outside the domain-specific language in ways proposed by Halliday et al. (2014). Students' texts and talks, for instance in discussions, could also be analysed using the same framework.

It would also be interesting to investigate the textbook authors' and editors' views on presentations of protein synthesis in textbooks through domain-specific language. Such a study could enhance understanding of the ideas and considerations behind the texts' structure. To build on these ideas, the ways that students interpret textbook descriptions of protein synthesis, and the role of the context in these interpretations, could add valuable insights into the relationship between students' reasoning and textbooks' presentations. In addition, a Delphi-study on biochemistry and molecular biology experts' views of how to handle the domain-specific language could provide important insights into the structure of what to incorporate and highlight when teaching protein synthesis.

Finally, it would be interesting to investigate the usage of analogies and metaphors in upper secondary education of protein synthesis, for instance using a similar approach to Haglund and Jeppson (2012). The investigation of analogies and metaphors in textbooks' texts on protein synthesis, as well as in students' and teachers' talk and texts on the topic, could provide valuable insights into effects of their uses in the teaching of protein synthesis on students' learning.

## References


Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular biology of the cell*, 5th ed. New York, NY: Garland Science.

Allchin, D. (2000). Mending Mendelism. *The American Biology Teacher*, *62*(9), 632–640.

Ananiadou, S., Kell, D. B., & Tsujii, J. I. (2006). Text mining and its potential applications in systems biology. *Trends in Biotechnology*, *24*(12), 571–579.

Andersson, S., Ellervik, U., Rydén, L., Sonesson, A., Svahn, O., & Tullberg, A. (2013). *Gymnasiekemi 2*, 6th ed. Stockholm, Sweden: Liber.

Bahar, M., Johnstone, A. H., & Hansell M. H., (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*. *33*(2) 84–86.

Björndahl, G., Landgren, B. & Thyberg, M. (2011). *Spira 1*. Stockholm, Sweden: Liber.

Borén, H., Larsson, M., Lindh, M. Lundström, J., Ragnarsson, M. & Sundkvist, S-Å. (2012). *Kemiboken 2*, 5th ed. Stockholm, Sweden: Liber.

Brooks M. G., & Grennon, J. (1999). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association for Supervision & Curriculum Development.

Brown, B. A., & Ryoo, K. (2008). Teaching science as a language: A "content-first" approach to science teaching. *Journal of Research in Science Teaching*, *45*(5), 529–553.

Brynhildsen, L., Brändén, H., & Ehinger, M. (2011). *Insikt Biologi 1*, 2d ed. Stockholm, Sweden: Natur & Kultur.

Butler, S., Urrutia, K., Buenger, A., Gonzalez, N., Hunt, M., & Eisenhart, C. (2010). *A review of the current research on vocabulary instruction*. Retrived from: https://www2.ed.gov/programs/readingfirst/support/rmcfinal1.pdf

Campbell, N. A., Mitchell, L. G., Reece, J. B., & Bishop, J. (2003). *Biology: Concepts & connections*, 4th ed. San Francisco, CA: Benjamin Cummings.

Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. *PLoS Biology*, 15(9), e2003243.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. Chicago, IL: University of Chicago Press.

Criado, R., & Sánchez, A. (2009). Vocabulary in EFL textbooks. A contrastive analysis against three corpus-based word ranges. In A. Sánchez & P. Cantos (Eds.), *A Survey on corpus-based research / Panorama de investigaciones basadas en corpus* (pp. 862-875). Murcia, Spain: Editum.

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.

Driver, R., Leach, J., & Millar, R. (1996). *Young people's images of science*. Bristol, PA: Open University Press.

Dogma. (2018). In *Oxford English Dictionary*. Retrieved from: https://en.oxforddictionaries.com/definition/dogma

Dos Santos, V. C., Joaquim, L. M., & El-Hani, C. N. (2012). Hybrid deterministic views about genes in biology textbooks: A key problem in genetics teaching. *Science & Education*, *21*(4), 543–578.

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.

Duranti, A., & Goodwin, C. (2009). Rethinking context: An introduction. In A. Duranti, & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon*, vol. 11 (pp. 1-42). Cambridge, UK: Cambridge University Press.

Edling, A. (2006). *Abstraction and authority in textbooks: the textual paths towards specialized language*. (Doctoral dissertation, Uppsala Universitet). Uppsala, Sweden: Acta Universitatis Upsaliensis.

Ekvall, U. (2001). Den styrande läroboken. In: B. Melander, & B. Olsson (Eds.), *Verklighetens texter: sjutton fallstudier* (pp. 43–80). Lund, Sweden: Studentlitteratur.

Fang, Z. (2004). Scientific literacy: A systemic functional linguistics perspective. *Science Education*, *89*(2), 335–347.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. New York, NY: Cambridge University Press.

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). Dordrecht, Netherlands: Kluwer Academic Publishers.

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.

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.

Gericke N. M., & Hagberg, M. (2010a). Conceptual incoherence as a result of the use of multiple historical models in school textbooks. *Research in Science Education, 40*(4), 605–623.

Gericke N. M., & Hagberg, M. (2010b). Conceptual variation in the depiction of gene function in upper secondary school textbooks. *Science & Education*. *19*(10), 963–994.

Gericke, N. M., & Smith, M. U. (2014). Twenty-first-century life sciences and genomics: contributions of HPS – Informed research and pedagogy. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching,* vol. I (pp. 423–467). Dordrecht, Netherlands: Springer.

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.

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.

Gilbert, J. K. (2006). On the nature of "context" in chemical education. *International Journal of Science Education*, *28*(9), 957–976.

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.

Haglund, J., & Jeppsson, F. (2012). Using self-generated analogies in teaching of thermodynamics. *Journal of Research in Science Teaching*, *49*(7), 898–921.

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. London, UK: Falmer.

Halliday, M., Matthiessen, C. M., & Matthiessen, C. (2014). *An introduction to functional grammar*. Cornwall, UK: Routledge.

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.

Haskel-Ittah, M., & Yarden, A. (2017). Toward bridging the mechanistic gap between genes and traits by emphasizing the role of proteins in a computational environment. *Science & Education*, *26*(19), 1143–1160.

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.

Henriksson, A. (2012a). *Iris Biologi 1*, 1st ed. Malmö, Sweden: Gleerups Utbildning.

Henrikson, A. (2012b). *Syntes Kemi 2*, 2d ed. Malmö, Sweden: Gleerups Utbildning.

Holliday, W. G., & Cain, S. D. (2012). Teaching science reading comprehension: A realistic, research-based approach. In B. Fraser, K. Tobin, & C. McRobbie (Eds.), *Second international handbook of science education* (pp. 1405–1417). Berlin, Germany: Springer.

Hultman, T. G. (2003). *Svenska akademiens språklära*. Stockholm, Sweden: Svenska akademien.

Höst, G. E., Larsson, C., Olson, A., & Tibell, L. A. (2013). Student learning about biomolecular self-assembly using two different external representations. *CBE—Life Sciences Education*, *12*(3), 471–482.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33*, 245–254.

Jenkinson, J., & McGill, G. (2012). Visualizing protein interactions and dynamics: Evolving a visual language for molecular animation. *CBE—Life Sciences Education*, *11*(1), 103-110.

Kay, L. E. (2000). *Who wrote the book of life?: A history of the genetic code*. Redwood City, CA: Stanford University Press.

Karlsson, J., Krigsman, T., Molander, B-O., & Wickman, P-O. (2011). *Biologi 1*. Stockholm, Sweden: 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.

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). Utrecht, Netherlands: CD-β Press.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Kvale, S. (1989). *Issues of validity in qualitative research*. Lund, Sweden: Studentlitteratur.

Ladyman, J. (2002). *Understanding philosophy of science*. Abington, UK: Routledge.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, Il: University of Chicago Press.

Lancor, R. (2014). Using metaphor theory to examine conceptions of energy in biology, chemistry, and physics. *Science & Education*, *23*(6), 1245–1267.

Language. (2018). In *Oxford English Dictionary*. Retrieved from: https://en.oxforddictionaries.com/definition/language

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.

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.

Löbner, S. (2002). *Understanding semantics*. London, UK: Arnold.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. Philosophy of comprehension of genetic concepts. *Journal of Biological Education*, *35*(4), 183–189.

Mann, M., & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, *21*(3), 255-261.

Marbach-Ad, G. (2001). Attempting to break the code in student comprehension of genetic concepts. *Journal of Biological Education*, *35*(4), 183–189.

Marbach-Ad, G., & Stavy, R. (2000). Students' cellular and molecular explanations of genetic phenomena. *Journal of Biological Education*, *34*(4), 200-205.

Martínez-Gracia, M. V., Gil-Quilez, M. J., and Osada, J. (2006). Analysis of molecular life sciences content in Spanish secondary school textbooks. *Journal of Biological Education*, *40*(2), 53–60.

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). Burlington. MA: Elsevier Academic Press.

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*. Washington, DC: National Institute of Health.

Nelson D. L. & Cox, M. M. (2013). *Lehninger - Principles of Biochemistry*, 6th ed. New York: Worth publishing.

Nelson, J. (2006). Hur används läroboken av lärare och elever? *NorDiNa, 4*, 16–27.

Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct and use them*. Pensacola, FL: Florida Institute for Human and Machine Cognition.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge, UK: Cambridge University Press.

Ogden, C. K., Richards, I. A., Malinowski, B., & Crookshank, F. G. (1923). *The meaning of meaning*. London, UK: 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.

Orgill, M., & Bodner, G. (2007). Locks and keys. *Biochemistry and Molecular Biology Education*, 35(4), 244–254.

Orgill, M., & Cooper, M. M. (2015). Teaching and learning about the interface between chemistry and biology. *Chemistry Education Research and Practice*, *16*(4), 711–713.

Pearson, J. T., & Hughes, W. J. (1988). Problems with the use of terminology in life sciences education: A literature review and classification scheme. *Journal of Biological Education*, *22*(3), 178–182.

Peptide. 2018. In *Encyclopaedia Britannica*. Retrieved from https://www.britannica.com/science/peptide

Pittman, K. M. (1999). Student generated analogies: another way of knowing? *Journal in Research in Science Teaching*, *36*(1), 1–22.

Pray, L. (2008). Discovery of DNA structure and function: Watson and Crick. *Nature Education*, *1*(1), 100.

Reinagel, A., & Speth, E. B. (2016). Beyond the central dogma: model-based learning of how genes determine phenotypes. *CBE–Life Sciences Education*, *15*(1).

Reiss, M. J., Millar, R., & Osborne, J. (1999). Beyond 2000: Science/biology education for the future. *Journal of Biological Education*, *33*(2), 68–70.

Robson C. (2008). *Real world research*. Madden, MA: Blackwell Publishing.

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.

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.

Sadava, D. E., Hillis, D. M., Heller, H. C., & Berenbaum M. (2014). *Life – the science of biology*, 10th ed. Sunderland, MA: Sinauer Associates.

Scott, P., Asoko, H., & Leach, J. (2007). Student conceptions and conceptual learning in science. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 31–56). Mahwah, NJ: Erlbaum.

Shmueli, G., Patel, N. R., & Bruce, P. C. (2010). *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. Hoboken, NJ: John Wiley and Sons.

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.

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.

Stahl, S., & Kapinus, B. (2001). *Word power: What every educator needs to know about teaching vocabulary*. Washington, DC: NEA Professional Library.

Swedish National Agency of Education (2013). *Curriculum for the upper secondary school*. Retrieved from: https://www.skolverket.se/publikationer?id=2975

Swedish National Agency for Education (2011a). *Biology*. Retrieved from: https://www.skolverket.se/download/18.4fc05a3f164131a7418104a/1535372296309/Biology-swedish-school.pdf

Swedish National Agency for Education (2011b). *Chemistry*. Retrieved from: https://www.skolverket.se/download/18.4fc05a3f164131a7418104e/1535372296640/Chemistry-swedish-school.pdf

Swedish Research Council (2002). *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning*. Retrieved from: http://www.codex.vr.se/texts/HSFR.pdf

Term. (2018). In *Oxford English Dictionary*. Retrieved from: https://en.oxforddictionaries.com/definition/term

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.

Thörne, K., Gericke, N. M., & Hagberg, M. (2013). Linguistic challenges in Mendelian life sciences: Teachers' talk in action. *Science Education*, *97*(5), 695–722.

Tibell, L. A., & 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.

Tseng, Y-H., Chang, C-Y., Chang Rundgren, S-N., & Rundgren C-J. (2010). Mining concept maps from news stories for measuring civic scientific literacy in media. *Computers & Education*, *55*(1), 165–177.

Tymoczko, J. L., Berg, J. M. & Stryer, L. (2013). *Biochemistry*. New York, NY: W.H. Freeman & Co.

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.

Watson, J. D., & Crick, F. H. (1953). The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, *18*, 123–131.

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), 1031–55.

Venville, G. J., & Treagust, D. F. (2002). Teaching about the gene in the genetic information age. *Australian Science Teachers Journal*, *48*(2), 20–24.

Villafañe, S. M., Bailey, C. P., Loertscher, J., Minderhout, V., & Lewis, J. E. (2011). Development and analysis of an instrument to assess student understanding of foundational concepts before biochemistry coursework. *Biochemistry and Molecular Biology Education*, 39(2), 102–109.

Wood, E. J. (1990), Biochemistry is a difficult subject for both student and teacher. *Biochemical Education*, 18(4), 170–172.

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.

Woody, W. D., Daniel, D. B., & Baker, C. A. (2010). E-books or textbooks: Students prefer textbooks. *Computers & Education, 55*(3), 945–948.

Word. (2018). In *Oxford English Dictionary*. Retrieved from: https://en.oxforddictionaries.com/definition/word

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.

# Teaching and Learning Protein Synthesis through Domain-Specific Language in Upper Secondary Education

Proteins are crucial to life: no proteins – no life. Every picosecond, thousands of proteins are constructed in each cell in what is referred to as protein synthesis. Due to its importance for understanding the mechanism for life, protein synthesis is globally regarded as a cornerstone of molecular life sciences and education in the field. Like any area of expertise, the molecular life sciences share a domain-specific language. However, research on teaching and learning protein synthesis through this domain-specific language is scarce. The aim of this licentiate thesis is thereby to contribute to understanding of upper secondary teaching and learning of protein synthesis with a focus on domain-specific language. The thesis is based on two studies: Study I addressed students' understanding where the results show that students can better reason about core concepts than peripheral concepts, and they compartmentalise the concepts into five clusters. Study II focused on chemistry and biology textbooks' presentation and effects of context on these presentations. The results reveal that chemistry textbooks focus more on peripheral concepts and generally tend to identify fewer relationships among more concepts than biology textbooks, which emphasise core concepts and tend to highlight more relationships among fewer concepts. Jointly, Studies I and II reveal four facets of teaching and learning protein synthesis. By acknowledging the results reported herein, teaching can improve the facilitation and reduce the hindrance in learning protein synthesis through the awareness of the domain-specific language usage.

ISBN 978-91-7063-834-3 (print)

ISBN 978-91-7063-975-3 (pdf)

ISSN 1403-8099

LICENTIATE THESIS | Karlstad University Studies | 2019:3