Conceptual Demography in Upper Secondary Chemistry and Biology Textbooks' Descriptions of Protein Synthesis: A Matter of Context?

Sara J. Wahlberg^{†*} and Niklas M. Gericke[‡]

¹Department of Engineering and Chemical Sciences and ¹Department of Environmental and Life Sciences, Karlstad University, 65188 Karlstad, Sweden

ABSTRACT

This study investigates how the domain-specific language of molecular life science is mediated by the comparative contexts of chemistry and biology education. We study upper secondary chemistry and biology textbook sections on protein synthesis to reveal the conceptual demography of concepts central to the communication of this subject. The term "conceptual demography" refers to the frequency, distribution, and internal relationships between technical terms mediating a potential conceptual meaning of a phenomenon. Data were collected through a content analysis approach inspired by text summarization and text mining techniques. Chemistry textbooks were found to present protein synthesis using a mechanistic approach, whereas biology textbooks use a conceptual approach. The chemistry texts make no clear distinction between core terms and peripheral terms but use them equally frequently and give equal attention to all relationships, whereas biology textbooks focus on core terms and mention and relate them to each other more frequently than peripheral terms. Moreover, chemistry textbooks typically segment the text, focusing on a couple of technical terms at a time, whereas biology textbooks focus on overarching structures of the protein synthesis. We argue that it might be fruitful for students to learn protein synthesis from both contexts to build a meaningful understanding.

INTRODUCTION

Protein synthesis is the process by which information encoded in genetic material is interpreted and used to produce specific proteins. As such, it is one of the most fundamental processes in living organisms. Because of its importance for understanding the mechanisms of life and the molecular aspects of inheritance, protein synthesis is a cornerstone of the molecular life sciences (Reinagel and Speth, 2016). However, there have been few studies on its teaching and learning at any educational level.

In upper secondary school teaching, protein synthesis is typically included in both chemistry and biology curricula. However, students struggle to comprehend the canonical representation of protein formation through the central dogma of protein synthesis (Wright *et al.*, 2014). They also struggle greatly to understand the function of genes (Gericke *et al.*, 2013) and their relationship to proteins (Duncan and Reiser, 2007) and have difficulties explaining and relating concepts that are central to the communication of protein synthesis (Gericke and Wahlberg, 2013). The reasons for these difficulties are poorly understood and investigated, but the domain-specific vocabulary used when communicating cellular structures and mechanisms has been identified as a major obstacle (Knippels, 2002). For instance, the life sciences use their own domain-specific language (Pearson and Hughes, 1988), with a daunting number of concepts to be learned (Tibell and Rundgren, 2010).

Life science education is developing rapidly, and students today are faced with a vast range of learning opportunities, not least through social media and the Internet.

Adele Wolfson, Monitoring Editor

Submitted Dec 19, 2017; Revised Jun 21, 2018; Accepted Jun 29, 2018

CBE Life Sci Educ September 1, 2018 17:ar51 DOI:10.1187/cbe.17-12-0274

*Address correspondence to: Sara J. Wahlberg (Sara.Wahlberg@kau.se).

© 2018 S. J. Wahlberg and N. M. Gericke. CBE—Life Sciences Education © 2018 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/ licenses/by-nc-sa/3.0).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology. However, they seem to prefer printed textbooks to electronic resources (Woody *et al.*, 2010), so textbooks remain a dominant source of information for learning. Written material from textbooks is one of the most prominent reading materials in Swedish schools (Edling, 2006). Textbooks are thus key mediators of the complex language of life science to students, making it very important to understand how their use and presentation of this domain-specific language facilitates or possibly hinders learning.

In Swedish upper secondary schools, protein synthesis is most commonly taught within the contexts of biochemistry in chemistry and molecular biology in biology (Swedish National Agency for Education, 2011a,b). This means that students study the same subject matter in two similar but different contexts. It has been suggested that this way of "compartmentalizing" content knowledge in time and space across separate curricula is an important barrier to learning in the life sciences (Knippels, 2002), but there is very little empirical research on how such different contexts influence the presentation of topics in the life sciences in general, or protein synthesis in particular.

For these reasons, the aim of this study is to investigate how protein synthesis is described in upper secondary chemistry textbooks and biology textbooks and to compare the two contexts. We investigate how domain-specific language is mediated by the different contexts through the texts' *conceptual demography*. Conceptual demography is defined by how the technical terms (i.e., concepts) of a topic, in this case protein synthesis, are outlined in a written text by studying the frequency with which they occur, their distribution in the text, and the relationships that the text presents between different technical terms. Implications for teaching and learning based on the study's findings are also presented.

BACKGROUND

Protein Synthesis

The protein biosynthesis pathway is typically presented using the canonical representation of the central dogma (Crick, 1958, 1970). In his 1958 article, Crick argued that the "main function of the genetic material is to control (not necessarily directly) the synthesis of proteins" (Crick, 1958, p. 138). Today, we know that protein synthesis is far more complex than suggested by this simplified original model, and more elaborate models have been developed. However, students learning how proteins are synthesized often display a weak conceptual understanding of the different descriptions of protein synthesis (Wright *et al.*, 2014).





Broadly, there seem to be three ways of defining protein synthesis in current educational literature. The first treats protein synthesis essentially as a synonym for the translation process, that is, the process whereby the information carried by the mRNA is interpreted and used to guide the assembly of amino acids into polypeptides by the ribosome (cf. Nelson and Cox, 2013; Tymoczko et al., 2013). We consider this to be the most limited definition of the three. The second definition encompasses both translation and transcription (cf. Jouper-Jaan et al., 2004; Ehinger and Ekenstierna, 2008), that is, the process of generating mRNA using information encoded in DNA and of generating polypeptides using information encoded in mRNA. The third definition is the broadest, including transcription and translation as well as a number of additional processes and details such as mRNA maturation (cf. Alberts et al., 2008; Sadava et al., 2014). We use the third definition as our point of departure in this work because it is the most inclusive model and so minimizes the risk of overlooking something important in our textbook analysis.

The understanding of the mechanisms of protein synthesis is rapidly developing, but there are some key concepts that are typically included when communicating the subject, especially at the upper secondary level. Previous analyses of upper secondary textbooks, teaching, and curricula have shown that three concepts-DNA, gene, and protein-are typically given a central role; we refer to them as the core concepts of protein synthesis, because they are the main components of the central conceptual framework used at the upper secondary level, i.e., the central dogma (Marbach-Ad, 2001; Gericke and Wahlberg, 2013). Some additional peripheral concepts are also typically introduced when teaching this topic at the upper secondary level (Gericke and Wahlberg, 2013). The peripheral concepts considered in this article are amino acid, exon, intron, mRNA, peptide, and tRNA (Gericke and Wahlberg, 2013). The three major processes of protein synthesis (i.e., translation, splicing, and transcription) and the associated core and peripheral concepts are shown in Figure 1.

Several other important concepts are used when communicating protein synthesis at higher educational levels, such as various structures and enzymes. These were not considered in this work, because previous studies showed that they are given relatively little attention in upper secondary education (Marbach-Ad, 2001; Gericke and Wahlberg, 2013).

Students' Understanding of Protein Synthesis

Descriptions of protein synthesis can be described as either conceptual or mechanistic. Mechanistic explanations are important

> for understanding any life process (Craver and Darden, 2013) and play central roles in understanding many subtopics within the life sciences, particularly in biochemistry and molecular biology. Machamer *et al.* (2000) define mechanistic explanations as ontic (i.e., real) descriptions of the cellular entities and activities involved in a life science phenomenon and how these entities and activities are arranged and organized. Describing the mechanism of a phenomenon involves explaining how the phenomenon is produced in a regular way. Regularity is shown in the way that the mechanisms

typically work from beginning to end and in the way that the entities and activities are appropriately structured and ordered spatially and temporally (Machamer *et al.*, 2000). In most biological domains, mechanisms are described at the cellular or molecular levels (van Mil *et al.*, 2013). Using the broadest definition (see preceding section), the phenomenon of protein synthesis can be described within the framework of the central dogma as the mechanistic process whereby information encoded in the genes (DNA) is translated to produce fully developed protein structures. A mechanistic explanation of protein synthesis should thus include descriptions of the properties of the participating entities as well as their activities and interrelationships (Craver and Darden, 2013).

Conceptual explanations provide another way of describing the same content knowledge (Scott *et al.*, 2007). Learning about and understanding protein synthesis involves learning and understanding the concepts involved and their relationships, that is, understanding the central and peripheral concepts discussed in the preceding section, and the relationships between them (Gericke and Wahlberg, 2013). Moreover, in a conceptual explanation, the relationships between concepts are shown in a causal rather than a mechanistic way, as sketched in Figure 1. Much research on life science education has focused on the conceptual aspects of learning and has shown that students find these aspects very challenging (Gericke and Smith, 2014).

Most earlier educational studies on life science education focused on gene function, that is, the overarching gene-trait relationship. This relationship is a core concept in genetics education, but has long been described as educationally challenging and difficult for students to understand (cf. Venville and Treagust, 1998, 2002; Allchin, 2000; Lewis and Kattman, 2004; Duncan and Reiser, 2007; Gericke and Hagberg, 2007; Gericke et al., 2013; Gericke and Wahlberg, 2013; Thörne et al., 2013; Thörne and Gericke, 2014). It has been shown that students tend to think of genes at different levels of organization at the same time, as entities that convey information relating to traits (e.g., eye color) and the code for producing proteins (Duncan and Reiser, 2007). Students thus tend to see genes as having two distinct functions, one relating to the production of proteins and another that causes traits to appear. It seems difficult for students to grasp the idea that a gene can encode both structure and function. Understanding this duality requires a developed explanation of how different concepts interact with one another on one level of organization to produce patterns and effects at a different level of organization (Lewis and Kattmann, 2004; Duncan and Reiser, 2007). Duncan and Tseng (2011) found that students lack a fundamental understanding of proteins' functions as components of complex systems in genetic phenomena. They also struggle to decipher the exact role of a protein (Haskel-Ittah and Yarden, 2017). Therefore, for many students, the causal relationship between traits and genes is something of a "black box" (Reinagel and Speth, 2016). Haskel-Ittah and Yarden (2017) argue that students better understand the mechanisms underlying the gene-trait relationship if they are presented with examples of proteins as entities. However, little attention has been paid to understanding the learning and teaching of the mechanisms, processes, and concepts associated with protein synthesis.

We investigated upper secondary students' understanding of protein synthesis (Gericke and Wahlberg, 2013). In that

study, it was found that the participating students had difficulties relating the starting and end concepts of protein synthesis, that is, the core concepts of DNA and protein. This could partly be explained by the fact that the students had difficulties relating the mRNA concept to the protein concept, that is, describing the translation process. Another main finding of that study was that the students compartmentalized the concepts into clusters: the core cluster, the transcription cluster, the translation cluster, the protein synonym cluster, and the inheritance cluster (Gericke and Wahlberg, 2013). The students could relate the concepts within each cluster but had difficulties relating concepts between clusters. Consequently, the tRNA and mRNA concepts were totally isolated from each other in the students' minds, breaking any link between the transcription and translation processes. The tRNA concept was only meaningfully related to the concept of amino acids, but as Fisher (1992) found several decades ago, students often believe that amino acids are synthesized in the translation process. Wright et al. (2014) found that 36% of the fourth-year university students enrolled in their study believed that transcription is a chemical transformation of DNA into RNA, or that mRNA existed before the process of transcription took place.

Further, we found that upper secondary students understood the maturation process of mRNA only vaguely, and could not describe the meaning of the concepts of exons or introns, nor their relationships to mRNA (Gericke and Wahlberg, 2013). We also found that upper secondary students typically used the concepts of protein and polypeptide synonymously for the translation product. None of the students in the study could clearly differentiate between the concepts of enzyme, protein, and polypeptide.

To sum up, students lack an integrative understanding of protein synthesis; instead, they compartmentalize its subprocesses and the related concepts. Moreover, they generally do not distinguish between the core and peripheral concepts.

The Importance of Language for Teaching and Learning Molecular Life Sciences

A "word" is in this article regarded as an entity that consists of a specific combination of letters (Hultman, 2003), and when this specific word is related to a knowledge domain, we regard it as a "technical term" or, in short, "term" (Halliday and Martin, 1993). When a term is assigned a meaning, it will be regarded as a "concept," as it can be seen as a mental description of the meaning of a word (Löbner, 2002). Technical terms are thus symbolic representations of the corresponding concepts. The concept has often been used in the literature when discussing student's word usage without separating the term from the underlying concept. However, scientific concepts exist outside and independent of the corresponding terms, but the terms themselves are representations and resources for the higher understanding of concepts (Brown and Ryoo, 2008). In this article, "domain-specific vocabulary" or "technical term" is used when discussing the textbook text without addressing its meaning.

Every discipline such as molecular genetics and biochemistry uses a set of terms that are commonly known and form part of the discipline's shared body of knowledge—the lexicon of the discipline (Fromkin and Rodman, 1998). This lexicon is formed by the discipline's tradition and any new influence affecting that tradition. Within a specific discipline, the lexicon consists of terms that have a commonly understood meaning among disciplinary participants. The lexicons of molecular biology and biochemistry may both include terms such as "DNA," "gene," or "protein." When addressing specific parts of a phenomenon, different terms from the lexicon are used. In a communicative activity, terms are linked to each other in comprehensible flows, and the separate parts, which include more than technical terms to make the language comprehensible, work together to communicate a message (Shore and Kempe, 1999; van den Broek, 2010).

There is a vast amount of research on students' struggles, difficulties, and learning obstacles when encountering molecular life sciences. The domain-specific language of molecular life sciences encompasses a very large body of technical terms for students to handle and they risk getting overwhelmed by the amount of information to process (Wood, 1990). Further, the field has its own domain-specific vocabulary that is important to learn and understand (Pearson and Hughes, 1988) but that can at the same time also be a central obstacle to learning (Knippels, 2002). To understand a topic or phenomenon, it is important to know the meaning of technical terms and the ways individual terms are used, combined, and related to each other (Shore and Kempe, 1999). The students must be able to draw connections between large numbers of concepts to understand what is being communicated (Orgill and Bodner, 2007). Central principles, mechanisms, and core concepts must be identified and related during teaching to facilitate learning (Driver et al., 1996; Tibell and Rundgren, 2010).

We still know very little about how science texts regarding life science are constructed (van den Broek, 2010), and even less about texts addressing protein synthesis. For this reason, it is important to investigate the structure and usage of the domain-specific language when communicating molecular life science content in general and protein synthesis in particular. This study contributes to the body of knowledge in this area by providing new information on the potential of textbooks to facilitate learning within these knowledge domains.

Conceptual Demography—The Frequency, Distribution, and Relationships of Technical Terms

Science can be communicated through different media, and written texts are one of the most frequently used formats. These texts are embedded in a specific context, and domain-specific vocabulary is the most important mediator of the content's contextual meaning (Gilbert, 2006). Chemistry and biology students are faced with texts that were intended to communicate content knowledge as effectively as possible. The challenge for students is to decode their meaning. Science texts need to be presented in a way that will facilitate students' learning and increase the likelihood of learning taking place (van den Broek, 2010). A central message in comprehending science texts and learning from them is to be able to identify relationships between concepts in the texts and being able to relate this to prior knowledge (van den Broek, 2010).

In this study, we will investigate the potential meaning-making capacity of selected textbooks by investigating the conceptual demography of the technical terms used in their descriptions of protein synthesis. We will do this by investigating three components of the texts' conceptual demography: the frequency, distribution, and relationships of their domain-specific vocabulary. By combining these components into a whole, we will delineate what we call the texts' conceptual demography. Conceptual demography describes how the technical terms in a specific text describe a phenomenon in terms of how frequently the terms are used, their distribution, and the way they are related throughout the text. In that way, the conceptual demography catches those properties of a text that relate to spatial and temporal relationships of technical terms. Hence, we can say something about the meaning-making potential of the phenomenon, and thus the underlying ideas, that is, the concepts describing it. In the following sections, we describe the components of conceptual demography.

Frequency

To delineate the conceptual demography of the studied texts, we determine the frequency with which each domain-specific core and peripheral term relating to protein synthesis is used in the chosen term samples. This makes it possible to identify over- and underrepresented sample terms in the text. Over the length of a text, certain terms may be repeated. The frequency of their recurrence has been referred to as the "vocabulary frequency," which quantifies the number of times a learner comes into contact with specific terms (Godev, 2009). It is suggested that increasing the number of times a learner encounters a term will deepen his or her learning of that term (Urzúa et al., 2006, in Godev, 2009). According to Tzeng et al. (2005), the reactivation of a term through a text is important for establishing conceptual connections to other terms in the mind of the reader. The National Reading Panel (National Institute of Child Health and Human Development, 2000) states that multiple exposures and repetition of vocabulary are important in the learning process. The frequency of a technical term being used can thus serve as an indicator of its importance, making term frequency a vital aspect of a text's conceptual demography.

Distribution

Another important aspect of a text's conceptual demography is the order in which technical terms are introduced and addressed throughout the text, that is, the spatiotemporal aspect of the text. The distribution is defined by the locations of the sample terms in the text, starting at its beginning. Analysis of the distribution can clarify the structural arrangements presented in the text and reveal which technical terms student encounter first, last, and in between as they interact with the text. The ordering of the technical terms strongly affects both the text's meaningmaking capacity and its readability (Perfetti, 2007). In particular, the duration of a term's use (i.e., the extent to which it occurs regularly in the text rather than being confined to specific sections) affects the length of time over which students encounter that term, which has been suggested to enhance the likelihood of learning a concept (Godev, 2009). Linderholm and colleagues (2004) point out that the spatiotemporal aspect of a text is important for coherence-based retrieval of the meaning of a text. An important aspect, according to Linderholm et al. (2004), is in what way the reader can recall technical terms from previous reading segments and link them conceptually to the currently read sentences. A longer distance between technical terms makes comprehensibility more difficult. Hence, the

distribution aspect of the text is important for its learning potential.

Relationships

The final aspect of conceptual demography investigated in this work is the way in which sample terms are related to one another throughout the text. As presented by van den Broek (2010), the relationships made in a text represent one of the central features in student learning from texts. Here, a relationship is defined in terms of the co-occurrence of sample terms within individual sentences. Lemke (1990) claims that the technical terms of a text create a "thematic pattern" that provides the meaning of the content knowledge through the relationships between the terms. According to both Baker et al. (1998) and van den Broek (2010), it is important for a text to relate its central concepts to one another to communicate a deeper understanding to the reader and support learning. Therefore, a communicative text should contain many passages in which technical terms are related. This will expose the students to many term interactions and create opportunities for them to develop a thematic pattern that enables understanding of the scientific phenomenon as a whole.

Technical Terms in the Contexts of Chemistry and Biology

One widely recognized challenge within molecular life science education is the overload resulting from the ever-accelerating accumulation of scientific concepts (Millar and Osborne, 1998; Ananiadou et al., 2006). The different meanings given to concepts by different contexts could create learning barriers. However, contextual differences have also been suggested to simplify this content load by allowing students to consider individual meanings associated with concepts in isolation (Gilbert, 2006). This study investigates the role of context in the presentation of protein synthesis in science textbooks, with the contexts in question being the subdisciplines of chemistry and biology education. Because of its importance in the life sciences, the topic of protein synthesis is included in both chemistry and biology curricula worldwide. In chemistry, it is typically presented in the biochemistry sections of upper secondary curricula and textbooks, while in biology it is typically presented in sections dealing with molecular biology.

In this study, we define a context as a discourse embedded in a cultural setting where the use of a specific language is an important element of understanding (Duranti and Goodwin, 1992). In school, students are required to learn in different contexts and to be involved cognitively by applying previous knowledge in new contexts (Shin *et al.*, 2009). Vocabulary therefore must not be learned in isolation, but should always be presented in a meaningful context (Shin, 2006), that is, a context that gives meaning to phrases, sentences, and terms.

Gilbert (2006) claims that there are many different contexts in science education and that the meanings of concepts are dependent on their applications, which in turn depend on the context. One example is the concept of energy (see Cooper and Klymkowsky, 2013), which is a core idea in science but is explained in different ways (which are sometimes inconsistent with one another) in different scientific subdisciplines. For instance, the role of the energy concept in understanding changes in matter is presented in different ways in chemistry and biology. In chemistry education, the energy concept is often

used in relation to thermodynamics and processes at the sub-microlevel to describe chemical reactions, whereas in biology education it is often used in a more colloquial way to describe simplified macrolevel processes and the way energy is distributed as new compounds are formed or degraded in ecosystems (Cooper and Klymkowsky, 2013). If these differences are not addressed in teaching, the different meanings of concepts in different contexts can be difficult for students to discern, creating learning barriers. However, if these differences are explicitly addressed and compared, different contexts can provide complementary descriptions of concepts that may enhance students' learning by providing multiple perspectives (Gilbert, 2006). Therefore, when teaching science, it is important to identify the way technical terms are given different conceptual meanings in various contexts. A way to identify such differences can be to analyze the conceptual demography of texts portraying the same content knowledge while representing different contexts.

The Textbook and the Importance of Language as a Knowledge Mediator

The textbook is a rich source of information (Nelson, 2006) and is part of the process whereby scientific knowledge is transformed into teachable school knowledge (Moody, 1996; Mikk, 2000). As part of the learning process, the textbook plays a significant role as reading material for the student (Ekvall, 2001). Woody *et al.* (2010) found that students typically choose printed textbooks in preference to electronic alternatives.

The textbook is important for secondary and upper secondary science teachers (Bergqvist, 2012). However, science textbooks are often dense in facts (Nelson, 2006), and the average science textbook may have significantly more new vocabulary terms than is recommended (Groves, 1995). For example, an elementary textbook may have as many as 30 new terms per chapter (Smith-Walters *et al.*, 2016). This can make the textbook's content challenging for students to grasp (Edling, 2006).

The National Reading Panel (National Institute of Child Health and Human Development, 2000) states that to be effective, vocabulary learning should occur in rich contexts, and the vocabulary should be chosen and presented to ensure that the learner will find it useful in many contexts. It has also been argued that teaching concepts in various contexts will increase the likelihood that these concepts will be understood and learned by the students (Stahl and Kapinus, 2001; Gilbert, 2006; Butler *et al.*, 2010). This suggests that addressing a concept from two or more viewpoints enhances student learning. The National Reading Panel (National Institute of Child Health and Human Development, 2000) also states that students will be better equipped to deal with specific content-area texts if the vocabulary in those texts derives from content learning material such as the textbook.

There have been only a few textbook studies focusing on the life sciences, most of which have dealt with genetics education. Martínez-Gracia *et al.* (2006) reported that Spanish secondary biology textbooks describe procedural ideas but do not facilitate the learning of the main ideas of genetics. Gericke *et al.* (2014) reported that upper secondary chemistry and biology textbooks struggle to clearly present the relationship between genes and traits. Descriptions of gene function are presented in different ways in textbooks for different subdisciplines, and

books from all disciplines tend to use simplistic explanations that avoid biochemical explanations (Gericke and Hagberg, 2010). Finally, students' understanding of the gene–trait relationship was found to be hindered by the different presentations of this relationship in the contexts of textbooks from different subdisciplines (Gericke *et al.*, 2013).

To sum up, textbooks are among the most important teaching materials and knowledge mediators in life science education. However, upper secondary students often have problems understanding the life science content in their chemistry and biology textbooks because of the use of scientific language in different contexts. In this study, we investigate the conceptual demography of chemistry and biology textbooks to determine whether and how these different contexts are reflected in the conceptual demography of protein synthesis descriptions of the textbooks.

Aim and Research Questions

The aim of this study is to explore the importance of context as a differentiator between the meaning conveyed of the same life science phenomenon in biology and chemistry texts. Specifically, we compare the conceptual demography of written protein synthesis descriptions in Swedish upper secondary biology and chemistry textbooks to reveal implications for teaching and learning protein synthesis.

Within the comparative contexts of chemistry and biology textbooks, relating to the sample words "amino acid," "DNA," "exon," "gene," "intron," "mRNA," "peptide," "protein," and "tRNA," the research questions are

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

METHODOLOGY

The textbooks were investigated using a content analysis approach inspired by text summarization and text mining techniques. Text summarization is a way of reducing texts to visualize what is being communicated without the need of reading the whole text (Reeve *et al.*, 2006). The focus of text miningresearch design is on extracting useful data from document collections to reveal patterns of information (Feldman and Sanger, 2007). Text mining uses computational techniques such as algorithms for discovering patterns, preprocessing routines, and creating inputs for visualization tools that can be used to depict the patterns within texts (Feldman and Sanger, 2007).

In this study, we based the methodology on routines inspired by "the nine steps in data mining" as proposed by Shmueli *et al.* (2010). These steps encompass the standard SEMMA (sample, explore, modify, model, and assess) protocol, and include

- 1. Development of an understanding of the purpose of the analysis
- 2. Obtaining the data set to be used for the data analysis
- 3. Exploring, cleaning, and preprocessing the data
- 4. If necessary, reducing the data
- 5. Determining the data task
- 6. Choosing the data technique
- 7. Performing the task using an algorithm

- 8. Interpretation of the results of the algorithm
- 9. Deploying the model by integration, for example, in operational systems

The aim of this study is not to integrate any operational systems, so step 9 was omitted. Steps 1–8 were conducted as follows:

Analytical Steps 1 and 2: Selecting the Data Set

The first step was determined by the research questions. The goal was to reveal the conceptual demography of the chosen texts and its components (frequency, distribution, and relationships). Microsoft Office Excel was used to record, analyze and visualize the data. Separate spreadsheets were created for each component of the conceptual demography, that is, frequency, distribution, and relationships.

The Textbook Sample

The textbook sample consisted of every commercially available (in March of 2015) Swedish upper secondary textbook in chemistry (n = 3; Borén *et al.*, 2012; Henriksson, 2012b; Andersson *et al.*, 2013) and biology (n = 4; Björndahl *et al.*, 2011; Brynhildsen *et al.*, 2011; Karlsson *et al.*, 2011; Henriksson, 2012a). The chosen textbooks are all published by the three largest publishing companies in the country. All the books claim to follow the most recent curriculum implemented in Sweden in 2011 (Swedish National Agency for Education, 2011a,b). Older versions and other textbooks were excluded.

Textbook Section Sample

The table of contents of each book was read through, and contiguous chapters and paragraphs concerning the molecular processes of constructing proteins and the function of the gene were selected, typically leaving one or two chapters of interest. Some of the books contained periodic, short, isolated sections discussing some aspect of cellular activity that was not directly associated with the surrounding text. These passages were carefully scrutinized to clarify their relationship to the topics of interest and were excluded from the analysis where appropriate. Because we were interested in the textbooks' vocabulary, the images, diagrams, and other graphics and captions were all excluded.

Chemistry textbooks use 785 \pm 384 words in a whole section, where 100 \pm 27 words represent one of the sample terms (95% confidence interval). Biology textbooks use 771 \pm 302 words in a whole section, where 100 \pm 45 represent one of the sample terms (95% confidence interval). The final samples from chemistry and biology textbooks contained on average 64 sentences and 58 sentences, respectively. No difference in the number of sentences could be discerned statistically between chemistry textbooks and biology textbooks in the chosen sections (p = 0.25). Therefore, the contexts are comparable, as the number of sample terms per sentence is similar in both contexts. For a more detailed presentation of the textbooks, see Table 1.

The Technical Term Sample

The term sample consists of a selection of terms from the domain-specific vocabulary of biochemistry and molecular genetics that in previous studies been found to be commonly used when describing protein synthesis in upper secondary education (Marbach-Ad, 2001; Gericke and Wahlberg, 2013). Specifically, it

TABLE 1.	The raw data of the tot	al word counts and th	he calculated perce	entage of sample terr	ns in each text sample
----------	-------------------------	-----------------------	---------------------	-----------------------	------------------------

Subject	References	A Total WC	B S	C Total ST	D % ST/WC	E WC/S	F Total ST/S
Chemistry	Andersson et al. (2013)	566	53	93	16.4	10.7	1.75
Chemistry	Borén <i>et al.</i> (2012)	1176	91	132	11.2	12.9	1.45
Chemistry	Henriksson (2012b)	614	48	76	12.4	12.8	1.58
Biology	Björndahl et al. (2011)	512	41	69	13.5	12.5	1.68
Biology	Brynhildsen et al. (2011)	765	47	82	10.7	16.3	1.74
Biology	Henriksson (2012a)	612	47	81	13.2	13	1.72
Biology	Karlsson et al. (2011)	1196	97	166	13.9	12.3	1.71

A, the total word count (WC) of the chosen sections; B, the total number of sentences (S) in the text sample; C, the total amount of sample terms (ST) in the text sample; D, the proportion of sample terms in the text (ST/WC); E, the average number of total words per sentence (WC/S); and F, the density expressed in percent of sample terms per sentence (ST/S).

consisted of the core and peripheral terms (concepts) presented in the introductory section: "amino acid," "DNA," "exon," "gene," "intron," "mRNA," "peptide," "protein," and "tRNA."

Analytical Steps 3 and 4: Processing the Data Set

The selected text was digitalized and compared with the original text in two rounds by two persons to minimize errors. Particular attention was paid to the punctuation in the text, because the analytical software used in the study employs punctuation as an end marker. The algorithm was programmed to search for a specific combination of letters—a *string*—that was the technical term in the sample. Each selected term in the sample made up one string. For instance, the string "protein" will target all words including this specific combination of letters in the exact order as described to the algorithm.

Terms that could cause trouble (such as those containing the letter sequence "RNA") were manually removed, and a "-" was inserted in the search term to exclude hits not corresponding to real matches. For example, the Swedish term "gärna" (Eng. "gladly") would produce a false positive hit without this measure. All texts were manually scanned from beginning to end after each removal of unwanted terms to reveal further errors; this process was performed iteratively until no further errors were found.

Analytical Steps 5 and 6: Algorithm Construction and Validation

The data mining task was set up to analyze the frequency, distribution, and relationship of the terms in the term sample. The sentences in the processed texts were numbered according to the order in which they appeared in the corresponding textbook, and each analysis began at the first sentence of the chosen text. The frequency was computed by counting the occurrences of each sample term. The distribution was defined by recording the number of each sentence in which the relevant sample term appeared. Relationships between two sample terms were characterized by counting the number of sentences in which both of those two terms appeared. The count was increased by only one for each sentence containing both terms, even if one or both terms in question occurred multiple times in the sentence. Therefore, if one sentence contained "DNA" twice and "gene" once, the program would record one instance of the "DNA"-"gene" relationship. For example, in the sentence: "The gene is a part of the <u>DNA</u> where a specific part of the <u>DNA</u> is targeted," gene is mentioned once and DNA twice, but only one DNA– gene relationship was recorded.

Analytical Step 7: The Algorithm Task

The algorithm was implemented in the commercially available program Microsoft Office Excel. To determine what operations should be performed and ensure that these operations were performed correctly, we implemented a training and data-validation process. This involved iterative manual searching for errors. Potential errors included positive hits in sentences that do not contain any of the sample terms, or double hits for sentences containing only one instance of a given sample term, among others. Because of these risks, it was essential to carefully fine-tune the program file using multiple texts before analyzing the results in detail. An analysis of the protein synthesis section from one textbook was defined as a round of analysis, so there were seven rounds of analysis in total.

Analytical Step 8: Calculating and Representing the Data

Frequency, distribution patterns, and relationship structures in chemistry textbooks and biology textbooks were extracted from the spreadsheets that had been generated in Microsoft Excel. Calculations that were not automatically performed by the spreadsheets, such as computations of mean values, were performed manually (see Table 1 and Supplemental Table S2). Because our aim was to compare a common depiction of the conceptual demography of the protein synthesis of all biology textbooks descriptions with a common depiction of all the chemistry textbooks, rather than individual textbooks, we decided to work with indexes instead of absolute numbers. In the process of indexation, the results were normalized against the texts' lengths.

Statistical analysis using SPSS v. 22 was used to conduct a two-sided t test (95% confidence interval) to determine differences between term frequencies between contexts.

Data-processing steps performed after the main analysis also included the construction of visualizations to facilitate communication of the results, as seen in the figures presented in the *Results* section.

RESULTS

Our aim when conducting this study was to investigate the conceptual demography of protein synthesis descriptions in



FIGURE 2. The frequency of each sample term expressed in percent of the whole word sample in each context referred to as an indexed value.

Swedish chemistry and biology textbooks, and to compare the two contexts. We will first address the frequency structure, followed by the results of the distribution study and the relationship structures between technical terms.

Frequency of Sample Terms

The results from the frequency study show that the most commonly used sample terms in both chemistry and biology textbooks are the core terms "DNA," "gene," and "protein" and the peripheral terms "mRNA," "tRNA," and "amino acid." The results can be seen in Figure 2 as the percentage of the sample terms per total amount of words, that is, the density of each sample term, and also as absolute numbers in the Supplemental Material. The results show that there are no statistical differences (two-sided *t* test and confidence interval) in frequencies between the two contexts of chemistry and biology, except for the peripheral terms "tRNA," for which p = 0.01, and "peptide," for which p = 0.006. The terms "tRNA" and "peptide" are used much more frequently in chemistry textbooks.

Figure 2 also shows that very little attention is paid to the peripheral terms "intron" and "exon" in either context. These terms are associated with the important process of mRNA maturation.

Distribution of the Sample Terms

Figure 3, A and B, ¹presents the distribution of sample terms in the texts in terms of their average number of occurrences in each consecutive sentence of the processed text, from first to last. The figures list the sample terms in an order corresponding to the sequence of the central dogma, starting with "DNA" and concluding with "protein."

If we first address the core terms, we can see that the distributions of the terms of "DNA" and "protein" follow similar patterns in both contexts. The term "DNA" is mostly used in the first halves of the texts and then only used a few times in the remainder, whereas the term "protein" is more evenly distributed in both chemistry and biology texts, except for a gap of several sentences toward the end of the first third of the chemistry texts (between units 10 and 17 at the axis showing the relative distribution in Figure 3, A and B). However, the usage of the core term "gene" differs between the biology and chemistry textbooks—it is relatively evenly distributed in the former case, but much more heavily concentrated at the beginning of the text in chemistry books.

The peripheral term "tRNA" is associated with the later part of the protein synthesis section in both chemistry and biology textbooks, but the chemistry texts tend to introduce this term somewhat earlier. The distribution of the term "amino acid" differs slightly between the contexts: it is more evenly distributed in



FIGURE 3. Distributions of the sample terms in chemistry (A) and biology (B) textbooks. The axis marked "Relative distribution" represents the average number of processed sentences from the beginning to the end of protein synthesis content in the text samples of the textbooks from each context. The axis marked "Average frequency" shows the average number of appearances of the sample term in each sentence. The technical terms of the sample can be seen on the axis marked "Technical term."

¹The presentation of a multidimensional image for the conceptual demography in landscape diagrams is inspired by van den Broek (2010).

biology texts, whereas in chemistry texts, it predominantly occurs in the second half.

The peripheral terms "intron" and "exon," which are central in communicating the process of mRNA maturation, are sparsely used in both contexts. However, their usage does differ slightly between the contexts. In cases in which the terms are used, chemistry texts generally feature them only in a couple of sentences in the middle of the text, whereas they are more evenly distributed (despite only occurring a couple of times per book) in the biology texts.

Finally, chemistry textbooks use "peptide" only in the last third of the text when discussing the translation product. The biology textbooks do not use the term "peptide" at all. As previously noted, the term "protein" is evenly distributed in both contexts.

Relationship Structures of Sample Terms

We have defined "a relationship" in terms of the presence of two sample terms in the same sentence. The results of the relationship analysis are presented in Supplemental Material Table S2 and Figure 4, A and B.

Biology textbooks prioritize some relationships over others, while chemistry textbooks do not. Therefore, the variation in the occurrence of relationships between pairs of sample terms is greater among the biology texts. This is clearly shown in Supplemental Material Table S2, which presents sample-term relationships in the texts. Chemistry textbooks allocate circa 55% of all found relationships in the text to the top eight relationships ("amino acid"–"mRNA"; "amino acid"–"tRNA" "DNA"–"gene"; "DNA"–"mRNA"; "DNA"–"protein"; "gene"–"protein"; "mRNA"–"protein"; "mRNA"–"protein"; "mRNA"–"protein"; "dots allocate 64% of all relationships to these top eight. The remaining relationships can be seen in Supplemental Material Table S2 and Figure 4, A and B.

Looking more closely at the relationships between the core terms ("DNA"-"gene," "protein"-"gene," and "DNA"-"protein"), we can see that all three of these relationships occur more frequently in biology texts than in chemistry texts. The most common relationships in the chemistry textbooks are (in consecutive order): "amino acid"-"tRNA"; "mRNA"-"tRNA"; and "mRNA"-"protein." Conversely, the most common relationships in the biology texts are (in consecutive order): "amino acid"-"tRNA"; "DNA"-"gene"; and "gene"-"protein." As can be seen, the biology texts emphasize the linking of the core terms, while the chemistry texts place a much stronger emphasis on relating the peripheral terms.

The most frequent relationship in both chemistry and biology texts is the relationship between "amino acid" and "tRNA." The relationship between "mRNA" and "tRNA" also occurs frequently. Notably, the term "mRNA" occurs in more relationships than any other sample term in both chemistry and biology texts. In general, the chemistry textbooks contain more relationships between terms associated with the later parts of protein synthesis (i.e., those relating to translation), with the term "peptide" being an exception.

Finally, the less frequent sample terms relating to RNA maturation—"intron" and "exon"—are weakly related to one another and the other sample terms in both contexts. However, "intron" and "exon" are involved in more relationships with other sample terms in biology textbooks than in chemistry textbooks.

The Conceptual Demography of Protein Synthesis

The conceptual demography of the treatment of protein synthesis in chemistry and biology textbooks can be described by simultaneously considering the frequency, distribution, and interrelationships of the sample terms. Figure 3, A and B, shows their relative distribution, while Figure 4, A and B, shows their absolute frequencies (*y*-axis) and relationships (indicated by the thickness of the lines linking the term boxes). The chemistry and biology textbooks use the core terms at quite similar relative frequencies (Figure 2), although the frequencies of the terms "DNA" and "protein" are somewhat higher in the biology texts. This trend is more readily apparent when considering the absolute numbers of occurrences for each core term (see the positions of the core terms on the *y*-axis in Figure 3, A and B).

The distributions and relationships of the sample terms in the two contexts seem quite different. To begin with, the use of the core terms in the chemistry textbooks is more compartmentalized. Notably, the term "gene" predominantly occurs in the beginning of the chemistry texts, and the same trend exists for the term "DNA." Additionally, there is a clear interruption in the occurrence of the term "protein" in the chemistry texts. The relationship data indicate that the core terms are much more central in the biology texts, because they are more strongly related to each other and the peripheral terms than is the case in the chemistry texts (see Figure 4, A and B).

The use of the peripheral terms exhibits greater variation. The terms "amino acid," "mRNA," and "tRNA" are the most frequently used peripheral terms in both contexts and have the greatest number of relationships; conversely, the terms "exon," "intron," and "peptide" are less emphasized in both contexts, and the latter is not used at all in the biology texts. The chemistry texts focus heavily on peripheral terms associated with the final stage of protein synthesis, that is, translation: the terms "amino acid," "peptide," and "tRNA" are used and related more often in the chemistry textbooks. The relationship most strongly emphasized in the chemistry texts is that involving "tRNA" and "amino acid," which are both primarily found in the later parts of the texts.

The sample term that differs most strongly in the conceptual demographies of chemistry and biology is "tRNA." Chemistry textbooks use this term much more frequently than biology textbooks, although it is primarily used in the same parts of the text in both contexts. Therefore, students reading chemistry textbooks encounter the term "tRNA" much more densely than those reading biology texts, mostly as a result of a substantial contribution of one of the three chemistry textbooks. Despite this, there is an overrepresentation of the usage of the technical term "tRNA" in chemistry textbooks compared with biology textbooks. In both contexts, "tRNA" is most often presented together with the sample terms "amino acid," "protein," "peptide," and "mRNA." However, the relationship between "tRNA" and "protein" is weak in both contexts, whereas the relationship between "tRNA" and "amino acid" is strongly emphasized in both contexts.

The sample term most strongly related to other sample terms in both chemistry and biology textbooks is "mRNA." Moreover, it is evenly and repeatedly distributed in the texts. Its conceptual demography differs only slightly in the two contexts: biology texts use it uniformly along their length, and at a



FIGURE 4. The relationships between selected sample terms and the average frequencies with which they occur in chemistry (A) and biology (B) textbooks. The numbers in parentheses indicate the average frequencies of the corresponding terms in absolute numbers, and the thickness of the lines linking pairs of term boxes reflect the strength of the relationship as a percentage of the total number of relationships found in the texts. The core term boxes are highlighted in red, and the peripheral term boxes are highlighted in blue. See also Supplemental Material Table S2, a and b.

somewhat greater overall density, whereas its use in chemistry texts decreases slightly toward the end.

The sample terms associated with the "mRNA" maturation system, "intron" and "exon," are presented in specific text sections in both contexts, but are more strongly related to each other and to other sample terms in biology textbooks than in chemistry textbooks. Their usage follows the same pattern in both contexts: they exist in isolated, dense, coexisting islands, and their overall emphasis is weak.

DISCUSSION

Our results reveal clear differences between the descriptions of protein synthesis presented in textbooks for chemistry (where it is presented in the context of biochemistry) and biology (where it is discussed in the context of molecular biology). From our results, we denote the approach used to present protein synthesis in the chemistry textbooks as a "mechanistic" approach, whereas the biology textbooks use what we denote as a "conceptual" approach.

As defined by Machamer *et al.* (2000), a mechanistic description in life sciences comprises a regularity in which the entities sequentially participate in spatiotemporally ordered subactivities with well-defined starting points and endpoints. In our

analysis of the conceptual demography, we can see that any one section of a chemistry text typically uses a couple of technical terms (i.e., entities) and relates them to each other, explaining their involvement in a particular submechanistic process of protein synthesis-for example, how amino acids are linked to tRNA. The entities represented by these core and peripheral terms and their relationships and activities with other entities (terms) are outlined and described in defined segments of the texts. This can be seen in the distribution patterns, which clearly show the repeated use of particular terms in specific sections of the text where the density of these technical terms is very high. Then, on moving to the next section of the textbook, another group of terms is more frequently mentioned and related. This approach maintains a focus on the mechanistic molecular processes of protein synthesis throughout the chemistry texts clearly corresponding to the processes of transcription, splicing, and translation, which also was verified after rereading the textbooks after analysis. A consequence of this approach is that no priorities are given to the core terms, so no sample terms other than "mRNA" are used as meaning makers throughout the text. Instead, the terms and their conceptual meanings are isolated and compartmentalized within the text segments. Each technical term (except to some degree the term "gene")

represents ontic entities with properties that at a molecular level realistically explain subactivities of protein synthesis. Hence, to explain each mechanistic activity, only a couple of technical terms are needed, and no references are made to overarching terms, such as the core terms, that provide a conceptual meaning of the phenomenon, that is, protein synthesis, as a whole. We refer to this way of presenting protein synthesis and molecular life sciences as a "mechanistic approach."

The biology textbooks also to some degree follow this fundamental pattern of regular segmentation of the technical terms used. Still, the biology texts focus instead on the core terms of "DNA," "gene," and "protein," together with the term "mRNA." These technical terms are used consistently throughout the texts and are related to the other technical terms (the peripheral terms) that appear compartmentalized in "islands" within the text, as shown in the distribution and relationship analysis. Moreover, the density of the technical terms is generally lower, indicating that the mechanistic process of protein synthesis is not presented at the same level of detail as in the chemistry texts. Instead, the results show that the biology texts convey a more conceptual understanding of protein synthesis, never losing sight of where it starts (DNA, gene) and ends (protein), and how the different subprocesses (transcription, splicing, and translation) relate to these points of reference. The term "gene" was particularly used for this purpose, being a non-ontic entity not involved in mechanistic explanations describing how properties of a molecule lead to a change in the activity (Machamer et al., 2000). The results show a focus on the overarching structure of protein synthesis in the biology textbooks, which was confirmed in the postanalysis rereading of the texts. Therefore, we describe this explanatory process as a "conceptual approach."

The chemistry textbooks' emphasis is primarily on sample terms relating to the later stages of protein synthesis. This focus is broadly consistent with the most condensed definition of protein synthesis (Nelson and Cox, 2013; Tymoczko et al., 2013), which focuses on the translation process. The biology textbooks, on the other hand, seem to adopt a wider perspective that in a higher degree includes the transcription and maturation processes. The descriptions of protein synthesis in the two contexts thus seem to have different aims. The biology textbooks that present protein synthesis in a molecular biology context are trying to present the process within the broader setting of its importance for gene function, whereas the biochemical context adopted in the chemistry textbooks necessitates a narrower focus on mechanistic explanations of how peptides and proteins are assembled. This is also shown in the way the biology texts do not use the term "peptide," which is closely linked to mechanistic molecular explanations, but stick to the term "protein," which has a wider conceptual meaning. It thus seems that the biology books more clearly address gene function, which has been identified as something students struggle to understand (Venville and Treagust, 1998; Lewis and Kattman, 2004; Duncan and Reiser, 2007; Gericke and Hagberg, 2007). It is important for biology and chemistry teachers to be aware of the existence of these two approaches when using these textbooks, which is discussed in the next section.

Implications for Teaching and Learning Protein Synthesis

Depending on the aim of a teaching setting, it may be fruitful to use the chemistry textbook's mechanistic approach as a rich source of detailed information on specific aspects of protein synthesis. However, as Tibell and Rundgren (2010) stress, there is a risk of drowning the students in detail at the expense of aiding their understanding of the protein synthesis process as a whole. On the other hand, addressing the overarching structures, as in the conceptual approach, may also impose difficulties on students, because they may miss the benefits of the mechanistic explanations relating to the molecular structures. Therefore, they may struggle to understand why specific reactions take place. We see here a need for future studies investigating the effects on learning through mechanistic and conceptual text approaches.

Technical Term Usage in Textbooks

We found that chemistry books discuss the term "gene" only at the very beginning of their sections on protein synthesis, whereas in the biology texts it occurs repeatedly and is evenly distributed throughout the text. These are two distinctly different ways of addressing a term. As discussed by Halliday *et al.* (2014), terms used to build a text must be arranged in a relevant flow. The biology textbooks never risk losing the important connection to the origin of the genetic material (the starting point of protein synthesis) as they move toward the construction of the protein. They may thus provide a more coherent level of understanding than the chemistry books.

Because chemistry books focus more on mechanistic details, this context may contribute to a deepened knowledge of specific stages of protein synthesis. Several terms are used many times throughout the chemistry texts and are also addressed numerous times within more confined sections, one example being "tRNA." We know that repetition makes students aware of a term's importance (Bybee, 1995), so if a term is repeated many times in a short section, the student may come to consider that term to be more important than others, which may be misleading, because peripheral terms were mentioned as often as core terms in many cases. Consequently, the student may risk missing central points about the overarching structures of protein synthesis. On the other hand, repetition fortifies the acquisition of learning.

The way students compartmentalize their understanding of protein synthesis may be strongly influenced by the textbooks' choice and usage of terms, which could be concluded based on other studies of text comprehension (van den Broek, 2010). However, the previous findings that students did not link the concepts "mRNA" and "tRNA" (Gericke and Wahlberg, 2013) somewhat contradict the results of this study, which shows the importance of the relationship between the terms "mRNA" and "tRNA" in the studied texts, indicating that current textbooks do not support student learning. As can be seen in Supplemental Material Table S2 and Figure 4, A and B; the term "tRNA" is not related to any other sample terms in the biology texts (except "mRNA"), and only to the term "DNA" on some occasions in the chemistry texts. This lack of relationships to other technical terms might lead to the textbooks' failure to support what Tzeng et al. (2005) denote as coactivation in the mind of the readers, that is, that the term in question is connected in a cluster with the other concepts concurrently activated in the text. This hypothesis might explain why this relationship is difficult for students to grasp, even though it is highlighted in the textbooks and therefore warrants further studies. Given this

discrepancy, students risk missing an important link that connects the processes of transcription and translation within a broader understanding of protein synthesis.

Chemistry textbooks pay more attention to the mechanistic explanations that sequentially relate a couple of peripheral terms at a time. Therefore, it would be of interest in future text comprehension studies to investigate whether this way of portraying a phenomenon really leads to coactivation (Tzeng *et al.*, 2005), and thereby coherent conceptual understanding for the reader. Based on our findings, it could be questioned whether the reader can recall concepts from previous text segments and link them to the currently read sentences in mechanistically arranged texts, because a longer distance between the technical terms makes comprehensibility more difficult (Linderholm *et al.*, 2004).

There is a need to scrutinize the labeling of specific structures, such as the somewhat ambiguous definition of the translation product in the textbooks, namely, the use of the technical terms "protein" and "peptide." "Peptide" is not used at all in the biology books, which implies that the product formed immediately after completion of the translation process is referred to as a "protein," that is, the focus is on a conceptual rather than a mechanistic level. Our findings suggest that textbooks may contribute to students' struggles in understanding the central role of "protein" if there is no clear guidance about when to use the terms "protein" or "peptide." This is consistent with the findings of Haskel-Ittah and Yarden (2017), who have shown that students are unable to clearly state the role and function of a protein. Our findings are also in line with the findings of Thörne and Gericke (2014), who stress the relationship between students' difficulties and ambiguous use of the "protein" concept in biology teaching. This warrants further investigation of the associated meaning.

Suggestions for Teaching

On the basis of our findings, we agree with the claim that teaching strategies that present the meaning of the context are important for students' conceptual learning (Gilbert, 2006). We therefore suggest, like Shore and Kempe (1999) and Gilbert (2006), that students should learn protein synthesis in rich contexts including conceptual and mechanistic approaches in combination. Gilbert (2006) claims that the teaching of concepts in one context will increase the likelihood that they will be understood in others. In a teaching situation, a teacher might highlight the differences between the two ways of presenting protein synthesis. For example, in a biology course, the teacher could use a chemistry textbook to highlight a specific subprocess of protein synthesis and enhance students' mechanistic understanding. Conversely, in a chemistry course, a teacher could refer to a biology textbook to place protein synthesis in the broader setting of its biological role.

Gilbert (2006) argues that the meaning of contexts can be used to reduce or simplify a content load. We agree with Gilbert's (2006) recommendation that the best way to facilitate learning is to identify the most important concepts within a subject and focus the learning effort on those concepts. As proposed by Haskel-Ittah and Yarden (2017), students are more likely to comprehend the underlying mechanisms if concepts are presented as a central entity, as in the case of the protein concept in the gene–trait relationship. We conclude that teachers and textbook authors and editors would benefit from recognizing the conceptual demography of the subjects they are teaching and writing about and the impact of the context as a scaffold that can facilitate students' learning. This could help students to identify and bridge gaps in their understanding of protein synthesis.

Limitations and Future Research

We acknowledge that students are not only offered information solely through the written material provided by a textbook, but through a multitude of sources. Visual representations and artifacts are also important for teaching and learning life sciences (Treagust and Tsui, 2013), and studies of those teaching materials could provide additional insights into the meaning-making capacity of the text as a whole.

Part of this study focuses on relationships between pairs of technical terms, because this is the most common term relationship in texts. However, in the future, it would be of interest to investigate possible co-occurrence of higher-order relationships, including three or more technical terms, that is, to study their contribution to the meaning-making capacity of a domain-specific text.

This study does not address language as a source of meaning in terms of the functions of grammar in creating and expressing meaning (including nontechnical terms), that is, the way semantic relations give functional meaning to the technical terms. An example of such a functional view of language is the theoretical framework of systemic functional linguistics (SFL) (Halliday *et al.*, 2014). An interesting further step would be to use SFL to investigate the meaning-making capacity of textbook sections on protein synthesis, which would make it possible to determine whether the differences regarding the different use of technical terms in biology and chemistry textbooks identified in this work affect or correlate with the texts' overall meaning-making capacity. Yet another interesting analysis would be a text comprehension analysis, as proposed by van den Broek (2010), of biology and chemistry texts portraying protein synthesis.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the Erna and Victor Hasselblad Research Foundation. We also thank B. E. Markus Blidh for helpful assistance with Microsoft Office Excel.

REFERENCES

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). Molecular biology of the cell (5th ed.). New York: Garland Science.
- Allchin, D. (2000). Mending Mendelism. 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 förlag.
- Baker, S. K., Simmons, D. C., & Kame'enui, E. J. (1998). Vocabulary acquisition: Research bases. In Simmons, D. C., & Kaméenui, E. J. (Eds.), What reading research tells us about children with diverse learning needs (pp. 183–218). Mahwah, NJ: Erlbaum.
- Bergqvist, A. (2012). Models of chemical bonding: Representations used in school textbooks and by teachers and their relation to students[°] difficulties in understanding (Licentiate thesis). Karlstad, Sweden: Karlstad University Press.
- Björndahl, G., Landgren, B., & Thyberg, M. (2011). *Spira* 1. Stockholm, Sweden: Liber Förlag.

- Borén, H., Larsson, M., Lindh, M., Lundström, J., Ragnarsson, M., & Sundkvist, S-Å (2012). Kemiboken 2 (5th ed.). Stockholm, Sweden: Liber Förlag.
- 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 (2nd 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 (pp. 1–20). Washington, DC: National Reading Technical Assistance Center, RMC Research Corporation.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.
- Cooper, M. M., & Klymkowsky, M. W. (2013). The trouble with chemical energy: Why understanding bond energies requires an interdisciplinary systems approach. CBE–Life Sciences Education, 12(2), 306–312.
- Craver, C. F., & Darden, L. (2013). In search of mechanisms: Discoveries across the life sciences. Chicago: University of Chicago Press.
- Crick, F. (1958). On protein synthesis. Manuscript published after a lecture given at Society for Experimental Biology symposium on the Biological Replication of Macromolecules, 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. Buckingham, UK: Open University Press.
- Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understanding of molecular genetics. *Journal of Research in Science Teaching*, 44(7), 938–959.
- Duncan, R. G., & Tseng, K. A. (2011). Designing project-based instruction to foster generative and mechanistic understandings in genetics. *Science Education*, 95(1), 21–56.
- Duranti, A. & Goodwin, C., (Eds.) (1992). *Rethinking context: Language as an interactive phenomenon* (Vol. 11). Cambridge, UK: Cambridge University Press.
- Edling, A. (2006). Abstraction and authority in textbooks: The textual paths towards specialized language (Doctoral dissertation). Uppsala, Sweden: Acta Universitatis Upsaliensis.
- Ehinger, M., & Ekenstierna, L. (2008). Bioteknik. Lund, Sweden: Studentlitteratur.
- Ekvall, U. (2001). Den styrande läroboken. In Melander, B., & Olsson, B. (Eds.), Verklighetens texter: Sjutton fallstudier (pp. 43–80). Lund, Sweden: Studentlitteratur.
- Feldman, R., & Sanger, J. (2007). The text mining handbook: Advanced approaches in analyzing unstructured data. Cambridge, UK: Cambridge University Press.
- Fisher, K. M. (1992). Improving high school genetics instruction. In Smith, M. U., & Simmons, P. E. (Eds.), *Teaching genetics: Recommendations and research: Proceedings of a national conference* (pp. 24–28). National Science Foundation.
- Fromkin, V., & Rodman, R. (1998). An introduction to language. Orlando, FL: Harcourt Brace College.
- Gericke, N. M., & Hagberg, M. (2007). Definition of historical models of gene function and their relation to students' understanding of genetics. *Sci*ence and Education, 16(7–8), 849–881.
- Gericke, N. M., & Hagberg, M. (2010). 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., & 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, 381–416.
- Gericke, N., & Smith, M. U. (2014). Twenty-first-century genetics and genomics: Contributions of HPS—informed research and pedagogy. In Matthews, M. R. (Ed.), International handbook of research in history, philosophy and science teaching (Vol. 1, pp. 423–467). Dordrecht, Netherlands: Springer.

- Gericke, N. M., & Wahlberg, S. J. (2013). Clusters of concepts in molecular genetics: A study of Swedish upper secondary science students' understanding. *Journal of Biological Education*, 47(2), 73–83.
- Gilbert, J. K. (2006). On the nature of "context" in chemical education. International Journal of Science Education, 28(9), 957–976.
- Godev, C. B. (2009). Word-frequency and vocabulary acquisition: An analysis of elementary Spanish college textbooks in the USA. *Revista de Lingüísti*ca Teórica y Aplicada, 47(2), 51–68.
- Groves, F. H. (1995). Science vocabulary load of selected secondary science textbooks. School Science and Mathematics, 95(5), 231–235.
- Halliday, M., Matthiessen, C. M., & Matthiessen, C. (2014). An introduction to functional grammar. Cornwall, UK: Routledge.
- Halliday, M. A. K., & Martin, J. R. (1993). Writing science: Literacy and discursive power. London: Falmer.
- 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, 10, 1–18.
- Henriksson, A. (2012a). *Iris Biologi 1* (1st ed.). Malmö, Sweden: Gleerups Utbildning AB.
- Henriksson, A. (2012b). *Syntes Kemi 2* (2nd ed.). Malmö, Sweden: Gleerups Utbildning AB.
- Hultman, T. G. (2003). *Svenska akademiens språklära*. Stockholm, Sweden: Svenska Akademien.
- Jouper-Jaan, Å., Lidesten, B.-M., & Strömberg, E. (2004). *Helix: I bioteknikens tjänst*. Lund, Sweden: Studentlitteratur.
- Karlsson, J., Krigsman, T., Molander, B.-O., & Wickman, P.-O. (2011). *Biologi 1*. Stockholm, Sweden: Liber Förlag.
- Knippels, M. C. P. J. (2002). Coping with the abstract and complex nature of genetics in biology education—The yo-yo learning and teaching strategy. Utrecht, Netherlands: CD-β Press.
- Lemke, J. L. (1990). *Talking science: Language, learning, and values*. London, UK: Ablex Publishing.
- Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: Re-visiting students' understandings of genetics. *International Journal of Science Education*, 26(2), 195–206.
- Linderholm, T., Virtue, S., Tzeng, Y., & van den Broek, P. W. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes*, 37(2), 165– 186.
- Löbner, S. (2002). Understanding semantics. London: Arnold.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25.
- Marbach-Ad, G. (2001). Attempting to break the code in student comprehension of genetic concepts. *Journal of Biological Education*, 35(4), 183–189.
- Martínez-Gracia, M. V., Gil-Quilez, M. J., & Osada, J. (2006). Analysis of molecular genetics content in Spanish secondary school textbooks. *Journal of Biological Education*, 40(2), 53–60.
- Mikk, J. (2000). *Textbook: Research and writing*. (Baltic Studies for Education and Social Sciences, Vol. 3). New York: Peter Lang.
- Millar, R. & Osborne, J. F., (Eds.), (1998). Beyond 2000: Science education for the future. London: King's College London.
- Moody, D. E. (1996). Evolution and the textbook structure. *Science Education*, 80(4), 395–418.
- 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 (NIH Publication No. 00-4769). Washington, DC: Government Printing Office.
- Nelson, D. L., & Cox, M. M. (2013). *Lehninger: Principles of biochemistry* (6th ed.). New York: Worth.

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

- Orgill, M., & Bodner, G. (2007). Locks and keys. Biochemistry and Molecular Biology Education, 35(4), 244–254.
- Pearson, J. T., & Hughes, W. J. (1988). Problems with the use of terminology in genetics education: 1. A literature review and classification scheme. *Journal of Biological Education*, 22(3), 178–182.

- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383.
- Reeve, L. H., Han, H., Nagori, S. V., Yang, J. C., Schwimmer, T. A., & Brooks, A. D. (2006). Concept frequency distribution in biomedical text summarization. In: *Proceedings of the 15th ACM international conference on information and knowledge management* (pp. 604–611).
- Reinagel, A., & Speth, E. B. (2016). Beyond the central dogma: Model-based learning of how genes determine phenotypes. *CBE–Life Sciences Education*, *15*(1), ar4. doi: 10.1187/cbe.15-04-0105
- Sadava, D. E., Hillis, D. M., Heller, H. C., & Berenbaum, M. (2014). *Life—The science of biology* (10th ed.). Sunderland, MA: Sinauer.
- Scott, P., Asoko, H., & Leach, J. (2007). Student conceptions and conceptual learning in science. In Abell, S. K., & Lederman, N. G. (Eds.), *The handbook* of research on science education (pp. 31–56). Mahwah, NJ: Erlbaum.
- Shin, F., Rueda, R., Simpkins, C., & Lim, H. (2009). Effective instructional strategies: Developing literacy in science for English language learners through content area instruction. In Coppola, J. (Ed.), One classroom, many learners. Newark, DE: International Reading Association.
- Shin, J. K. (2006). Ten helpful ideas for teaching English to young learners. *English Teaching Forum*, 4(2), 2.
- 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: Wiley.
- Shore, W. J., & Kempe, V. (1999). The role of sentence context in accessing partial knowledge of word meanings. *Journal of Psycholinguistic Research*, 28(2), 145–163.
- Smith-Walters, C., Bass, A. S., & Mangione, K. A. (2016). Science and language special issue: Challenges in preparing preservice teachers for teaching science as a second language. *Electronic Journal of Science Education*, 20(3), 59–71.
- Stahl, S., & Kapinus, B. (2001). Word power: What every educator needs to know about teaching vocabulary. (NEA Success in Reading Series). Washington, DC: NEA Professional Library.
- Swedish National Agency for Education. (2011a). *Biology*. Retrieved April 20, 2017, from www.skolverket.se/polopoly_fs/1.194789!/Menu/article/ attachment/Biology.pdf
- Swedish National Agency for Education. (2011b). *Chemistry*. Retrieved April 20, 2017, from www.skolverket.se/polopoly_fs/1.194837!/Menu/article/ attachment/Chemistry.pdf

- Thörne, K., & Gericke, N. (2014). Teaching genetics 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 genetics: 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.
- Treagust, D. F., & Tsui, C.-Y. (2013). *Multiple representations in biological education*. Dordrecht, Netherlands: Springer.
- Tymoczko, J. L., Berg, J. M., & Stryer, L. (2013). Biochemistry. New York: Freeman.
- Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the landscape model: Modeling inferential processes and memory representations of text comprehension. *Behavior Research Methods*, 37(2), 277–286.
- Urzúa, P., Sáez, K., & Echeverría, M. S. (2006). Disponibilidad léxica matemática: Análisis cuantitativo y cualitativo. *Revista de Lingüística Teórica y Apli*cada, 44(2), 59–76.
- Van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, 328(5977), 453–456.
- van Mil, M., Boerwinkel, D., & Waarlo, A. (2013). Modelling molecular mechanisms: A framework of scientific reasoning to construct molecular-level explanations for cellular behaviour. *Science & Education*, 22(1), 93–118.
- Venville, G. J., & Treagust, D. F. (1998). Exploring conceptual change in genetics using a multidimensional interpretive framework. *Journal of Research in Science Teaching*, 35(9), 1031–1055.
- Venville, G. J., & Treagust, D. F. (2002). Teaching about the gene in the genetic information age. Australian Science Teachers Journal, 48(2), 20–24.
- Wood, E. J. (1990). Biochemistry is a difficult subject for both student and teacher, *Biochemical Education*, *18*, 170–172.
- Woody, W. D., Daniel, D. B., & Baker, C. A. (2010). E-books or textbooks: Students prefer textbooks. Computers & Education, 55(3), 945–948.
- Wright, L. K., Fisk, J. N., & Newman, D. L. (2014). DNA \rightarrow RNA: What do students think the arrow means? *CBE*-*Life Sciences Education*, *13*(2), 338-348.