Instructional Science 17: 57–76 (1988)

Reasoning and the structure of knowledge in Biochemistry

DON PLOGER
Graduate School of Education, Tolman Hall, EMST, University of California, Berkeley, CA 94720, U.S.A.

Abstract. This study examines three aspects of reasoning in biological science: (1) how biochemists solve a problem involving a disorder of metabolism, (2) how they explain the solution and (3) how these experts differ from students. A methodology for explicitly describing biochemical knowledge is developed and applied to transcripts of reasoning to infer reasoning strategy from the way subjects use that knowledge. Two strategies are described: the normal-function strategy, which is based upon normal causal mechanisms, and the known-pathology strategy, which relies upon knowledge of pathology as well as normal causal mechanisms. The results suggest that in problem solving, experts use a greater variety of strategies than novices, and in the explanation, experts use a more general form of the normal-function strategy than novices. Three implications for science education are discussed: the organization of knowledge, problem-solving strategy, and explaining a solution to a problem.

Introduction: reasoning and the structure of knowledge in Biochemistry

What is the relationship between the structure of knowledge of a domain and reasoning strategy? How are the basic principles of a domain used to organize facts? How are those principles used to select relevant facts during reasoning? I will investigate these questions in the domain of metabolism, an important area of biochemistry which deals with the organization of the molecular and cellular processes necessary to sustain life. This study will present an organization of the knowledge of metabolism and determine how experts and students in biochemistry use knowledge to solve problems and give explanations.

In a review of cognition and science education, diSessa and Ploger (1987) noted that research on reasoning in physics is relatively well developed, clearly focused on an important area, Newtonian mechanics, and oriented toward education. In contrast, the research on reasoning in biological areas has focused on clinical medicine, which is not clearly related to educational practice.

Knowledge in Newtonian mechanics is well organized around a relatively small set of mathematical principles. Furthermore, this knowledge is used in systematic ways during reasoning. Larkin, McDermott, Simon and Simon (1980) found that experts differ from novices in their use of knowledge during problem solving. Experts reason forward from what is given in the problem toward the unknown. Novices, on the other hand, start with the unknown and work backward, trying to find a principle that involves current unknowns. Because of this difference in strategy, experts are faster than novices in solving problems.

Unlike Newtonian mechanics, knowledge in metabolism: (1) is non-mathematical, (2) contains relatively more facts in comparison to principles and explicit methods, (3) involves uncertainty, and (4) is concerned with the relation of normal and abnormal function. Knowledge in clinical medicine is more closely related to knowledge in metabolism. In fact, the MYCIN research project demonstrated that it was possible for a computer system to reason accurately with a large information base and without mathematical certainty (see Buchanan and Shortliffe, 1984, for a review).

Clancey (1983) attempted to modify MYCIN for teaching purposes and found that it was necessary to reorganize MYCIN's structuring of the knowledge. The new system, NEOMYCIN, clearly separated knowledge of the medical domain from strategic knowledge. Clancey noted that much of the medical knowledge was not based upon underlying causal mechanisms, but upon the empirical association of symptom and disease.

Studies of human reasoning have demonstrated that physicians do, in fact, reason using empirical associations. For instance, Kuipers and Kassirer (1984) found that while experts were able to correctly diagnose disease, "they gave very weak explanations of how [the phenomenon] are caused" (Kuipers and Kassirer, 1984, p. 369). In a second example of reasoning without reference to causal mechanism, Patel and Groen (1986) presented expert physicians with a case description of an unemployed 27-year-old male, who reported being scratched on the arm by a cat. The physicians who successfully solved the problem realized that the patient had lied about being scratched by a cat, that the scars were the result of intravenous drug use, and that the patient had bacterial endocarditis, an infection of the heart. These successful subjects relied upon knowledge of symptoms of the disease and a suspicion that this particular patient would take intravenous drugs and lie about it. They did not consider how the infection entered the body, or why the heart was affected. They reasoned with knowledge of empirical associations rather than of basic causal mechanisms of the disease.

In summary, the structure of knowledge in previously studied domains differs in important ways from the structure of knowledge in metabolism. In contrast to Newtonian mechanics, metabolism is not mathematically organized and involves more facts. In contrast to clinical medicine, metabolism is organized around normal causal mechanisms.

Although there is no direct evidence that the different structure of knowledge in metabolism leads to different types of reasoning, there is indirect evidence from reports on teaching problem-solving in biochemistry. Blanchaer (1982) observed that students were most interested in the relation of the clinical phenomena to basic biochemical mechanisms. This is remarkable because the problems were clinically oriented, involving descriptions of symptoms that a physician would typically see. Blanchaer did not, however, advance a systematic way to describe the relevant metabolic knowledge or how to detect it during reasoning.

The structure of knowledge in Metabolism

No previous cognitive study has presented a sufficiently detailed description of knowledge in any area of biochemistry. A major contribution of this study is to advance such a description, and to apply it to the analysis of reasoning. This description of the structure of knowledge of metabolism will draw heavily on important concepts that appear in standard biochemistry textbooks.

The concept of *levels of biological organization* is very useful in structuring knowledge in metabolism. The general concept of "levels" has proved useful in many areas of science. For instance, Newell (1981) presented a formulation of the concept of levels of organization in his description of the function of computer systems, defining levels by two characteristics: each level can be described, to a certain degree, independently of any other level, and each level can be described in terms of the next lower level. The concept of levels also applies to biological systems, where the levels most relevant to biochemistry are the cellular and the molecular levels. Adapting Newell's formulation: a cell can be described, to a certain degree, without reference to detailed molecular events, and cellular function can be described in terms of molecular interaction.

Knowledge in metabolism is also organized by principles relating normal and abnormal function. In a biochemically-oriented text on metabolic diseases (Stanbury, Wyngaarden, Fredrickson, Goldstein and Brown, 1982), the major focus is on the description of normal function. Abnormal function is viewed as a deviation of normal function, and it is understood by first understanding the normal. Although convenient, it is not necessary to describe abnormal function in this way. In fact, biochemists are aware of many specific pathologies and could potentially make use of this knowledge in problem solving. Unlike practicing physicians, however, the biochemist would be likely to use the knowledge of the specific pathology to suggest a causal mechanism.

In order to illustrate the relationship between the structure of knowledge in metabolism and reasoning, it is necessary to introduce some biochemical terminology. The problem which was used in the study: "How does pyruvate kinase (PK) deficiency lead to hemolytic anemia?" involves the relationship between an abnormal molecular condition and an abnormal cellular condition. The molecular condition, PK deficiency, refers to a deficiency in the activity of the molecule PK. The cellular condition is hemolytic anemia, which is an alteration of the red blood cell (RBC). Therefore, two aspects of the organization of biochemical knowledge are relevant to this problem: the relationship between molecular and cellular events, and the relationship between normal and abnormal function.

The relationship between cells and molecules has been described by a set of principles, such as Lehninger's (1975) "axioms". Of the ten axioms, two are relevant to the problem. (1) Energy-yielding processes are necessary for energy-requiring processes. The cell's energy currency is adenosine tri-phosphate (ATP). (2) Cells regulate their own metabolic activity. Figure 1 portrays these two axioms. At the top of the figure are two boxes: energy-yielding processes and energy-requiring processes. Energy-yielding processes provide energy, as ATP, for energy-requiring processes, as stated in Axiom 1. Furthermore, because this is a very important metabolic process, it is regulated, according to Axiom 2. Figure 1 also shows the two major types of energy-requiring processes: those that build components and those that remove components.

The relation of cell function to the more detailed molecular function is shown in Figure 2. The top portion of Figure 2, which is very similar to Figure 1, involves the cellular level. Figure 2 also includes molecular details, such as PK which appears at the bottom of the figure, and an intermediate level, metabolic pathways. Three metabolic pathways are shown under the energy production box: glycolysis, TCA (which refers to the tri-carboxylic acid cycle), and beta-oxidation, which are the three inter-related energy-producing pathways available to cells. PK is shown below the glycolysis box because PK is an enzyme in glycolysis.

The concept of metabolic pathway is very important in biochemistry because it provides a way to organize detailed knowledge of molecular events. A pathway also provides a way to relate knowledge at the molecular level to cellular function. Figure 2 shows how a molecular entity, such as PK, is related to broader cellular events, such as energy production, and ultimately to energy-requiring processes.

In order to solve the problem, it is not sufficient simply to relate the molecular entity, PK, to the cellular entity, RBC. It is also necessary to introduce abnormal function and to show how a deficiency of PK causes a particular disorder of the RBC, in this case hemolytic anemia. This can be done after considering normal function and first recognizing that normal PK is one of the molecular components that are necessary for energy production within the cell. If there is a defect in PK, there could be severe consequences for a cell due to inadequate energy production.

A second way to introduce abnormal function is to consider known pathologies that are related to the condition in the problem. This approach relies upon empirical association which is so prominent in clinical medicine. However, in metabolism, it is not sufficient to recognize an association; it is also necessary to provide a causal account.

The analysis of this problem indicates that the knowledge of metabolism differs in important ways from knowledge in fields previously studied in cognitive science. I hypothesize that because of these differences in the structure of knowledge, reasoning will differ in important ways. I will advance two reasoning strategies for solving problems in biochemistry, each based on a different conception of abnormal function. The *normal-function strategy* involves reasoning about normal function, before making reference to abnormal function. The *known-pathology strategy*, on the other hand, involves first introducing a known pathology and then determining whether this is relevant to the problem.

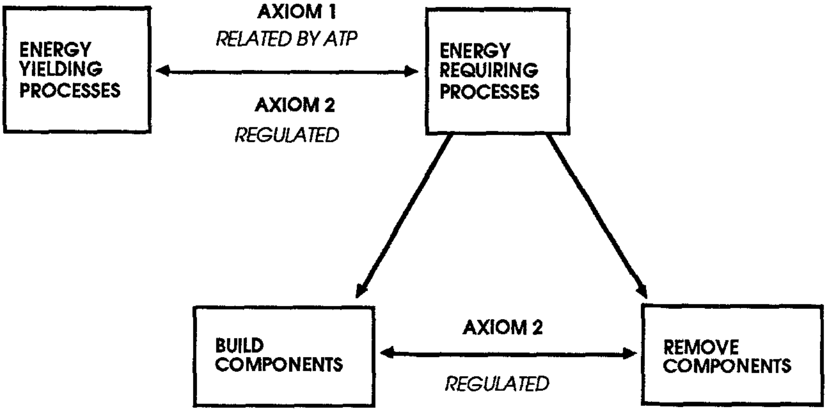

Figure 1. Axioms of cell function

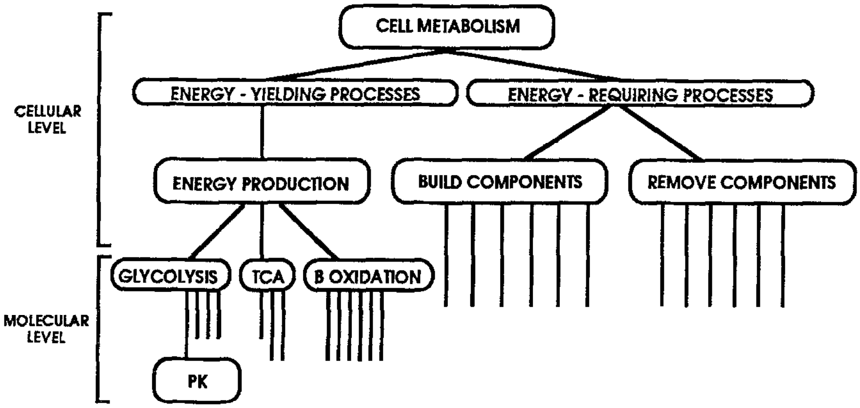

Figure 2. Organization of biochemical knowledge

In addition to problem solving, this study will also consider the generation of scientific explanations of solved problems. Although this issue has received considerable attention in artificial intelligence (Buchanan and Shortliffe, 1984; Clancey, 1983), there have been few psychological studies.

In this paper, I will develop a methodology for analyzing protocols of reasoning in biochemistry, and use that methodology to investigate the strategies of experts and novices as they solve and explain a problem in metabolism. Then I will describe how this line of research can bring to the study of reasoning in biology a coherence and relevance to education that is now present in the research on physics reasoning and education.

Method of data collection

Two experts and two novices served as subjects in the study. Both experts had a Ph.D. degree in a biological science and were involved in teaching and research in biochemistry. Both novices were first-year medical students who were tested after the relevant material had been covered in their biochemistry course, but before the specific problem had been discussed.

Subjects were interviewed individually. After a brief introduction to the purpose of the experiment, they were asked to practice the process of thinking aloud. When they were comfortable with the procedure, subjects were told that they would be given a problem and asked to solve it. They were told their verbalizations would be tape recorded. They were allowed to write down anything that they wished during the problem solving session. They were told that they could ask questions, and if the answer to a question was on the list of facts previously identified as relevant to the problem, the experimenter would answer the question. They were asked to summarize the instructions before the actual problem solving began.

After they gave a satisfactory summary of the instructions, they were given the card with the problem "How does a genetic deficiency of pyruvate kinase (PK) lead to hemolytic anemia", and asked to solve it. After they had finished solving the problem, they were asked to recall what they said during problem solving. Finally, they were asked to give an explanation of the problem as though they were speaking to someone with some basic background in biochemistry, but no knowledge of the specific problem. They were told that their explanation did not have to follow the original problem-solving activity.

The tape recorded sessions were transcribed, and the resulting transcripts are referred to as *verbal protocols*. For each of the four subjects, a problem-solving protocol and an explanation protocol was obtained. The recall protocols are not considered in this report. Copies of these protocols are available from the author upon request.

Overview of protocol analysis

This section describes the organization of biochemical knowledge, examines the problem, outlines two different reasoning strategies, and specifies a procedure for determining reasoning strategy from a protocol. It will illustrate the procedure with a detailed example of each strategy from an actual protocol.

This methodology depends upon the detection of certain *terms*, which are words or phrases referring to biochemical components or processes. A term can be at either the molecular level or the cellular level and can refer to either the normal state or an abnormal state of a component or process. This analysis will identify important terms used in problem-solving with respect to their relationship to the structure of biochemical knowledge and their function in solving problems of this sort.

All problems involving metabolic diseases require that a molecular defect be related to a cellular condition. In this particular case, PK deficiency must be related to hemolytic anemia. Both of these terms are abnormal conditions related to normal components. PK is the normal form corresponding to PK-deficiency, and is referred to as the *molecular problem-statement term.* RBC is the normal form corresponding to hemolytic anemia, and is referred to as the *cellular problem-statement term.*

An acceptable answer to the problem is "PK-deficiency causes decreased energy production and that decreased energy production impairs the RBC sufficiently to cause hemolytic anemia". It is acceptable to replace the term "decreased energy production" with either "decreased ATP production" or "impaired glycolysis". The *answer-relevant terms* for the PK problem are energy, ATP, and glycolysis. These terms are very important because at least one of them must be resent in the correct answer.

A solution must relate an answer-relevant term to both the molecular problem-statement term (in this case, PK), and the cellular problem-statement term (in this case, either hemolytic anemia or RBC). This relationship, which does not have to be a single relation, is defined as a *relation path.* For instance the statement "PK is an enzyme in glycolysis" is a relation path, consisting of one relation, between "PK" (the molecular problem-statement term) and "glycolysis" (an answer-relevant term). The set of statements "The RBC is a kind of cell"; and "Cells require energy" is a relation path, consisting of two relations, between "RBC" (the cellular problem-statement term) and "Energy" (an answer-relevant term).

The problem involves a molecular problem-statement term and a cellular problem-statement term. The correct answer includes both those terms, as well as an answer-relevant term. Therefore, if a subject correctly solves the problem he must relate the molecular problem-statement term to the answer-relevant term and relate the cellular problem-statement term to the answer-relevant term.

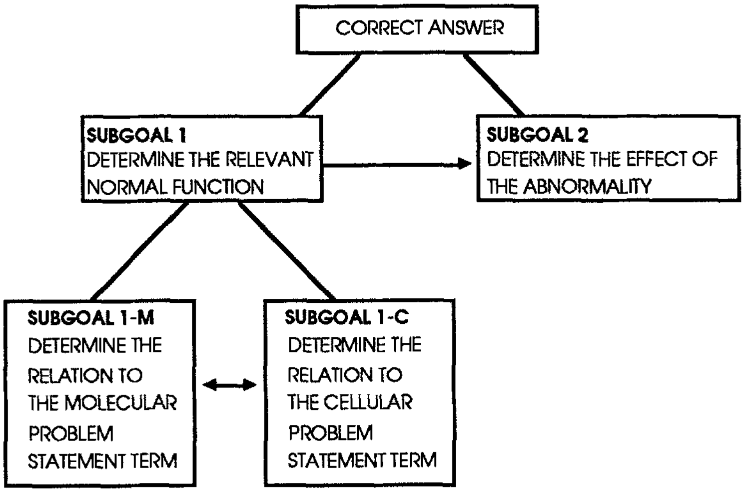

Figure 3. The normal-function strategy

The analytic procedure lists the answer-relevant terms and then determines three important locations in a successful protocol: (1) the molecular relation-path, where for the first time an answer-relevant term is related to the molecular problem-statement term, (2) the cellular relation-path, where an answer-relevant term is related to the cellular problem-statement term for the first time, and (3) the correct answer. This analysis is sufficiently precise to discriminate the strategies I have proposed, and yet feasible to conduct without an automated procedure.

I have advanced two reasoning strategies. In the normal-function strategy (Figure 3), normal function is primary, and the abnormal is understood as a deviation from the normal. There are two major subgoals which must be achieved in order, as indicated by the directional arrow. Subgoal 1, which is achieved by determining the relevant normal function, has two further subgoals. Subgoal 1-M requires a relation of some aspect of metabolism to the molecular problem-statement term, and Subgoal 1-C requires a relation of some aspect of metabolism to the cellular problem-statement term. These two subgoals can be achieved in either order as indicated by the two-headed arrow in the figure. Subgoal 2 is achieved by determining the effect of the abnormal condition on the normal function.

The known-pathology strategy (Figure 4) involves three subgoals, which must be achieved in order. Subgoal 1 is different from the normal-function strategy, because it involves a known pathology. However, once this pathology is introduced, normal causal mechanisms are considered. Subgoal 2 and 3 are essentially similar to the normal-function strategy.

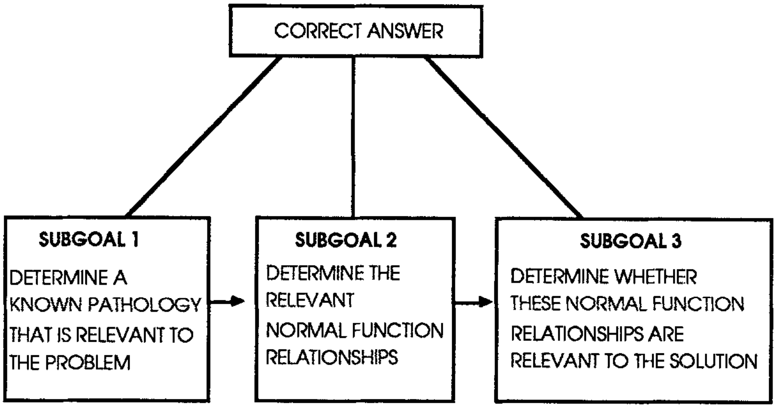

Figure 4. The known-pathology strategy

Example of the Normal Function strategy

Table 1 presents three locations in a protocol which were determined by the procedure. Note that the following answer-relevant terms are underlined: ATP in Statement 34, Energy in Statement 69, and ATP in Statement 77. They were found by listing the answer relevant terms and determining the first occurrence that was related to the molecular problem statement term, the first occurrence that was related to the cellular problem statement term, and the correct answer.

Statement 34 is the first occurrence of an answer-relevant term (ATP). In that statement, ATP is related to the terms phosphoenolpyruvate (PEP) and pyruvate. Those two terms had previously been related to PK in Statements 2 and 3. Therefore, there is a relation path between the molecular problem-statement term, PK, and the answer-relevant term, ATP. The procedure also determined that all terms during this phase of the protocol were at the molecular level and involved only normal function.

Statement 69 show the relation of the answer-relevant term (energy) to the cellular problem-statement term (erythrocyte, which means RBC). Therefore, the cellular relation-path is a single relation between energy and the RBC (erythrocyte) in line 69. Note that there has been no reference to abnormal function.

Statements 76-77 states the correct answer. At this point, the subject refers to abnormal function: “We would not be making as much ATP”.

Table 1. Problem solving protocol.

NORMAL FUNCTION SEGMENT

Molecular Relation Path

2 ALL RIGHT THE FIRST THING IS THAT **PYRUVATE KINASE** CATALYZES A REACTION

3 WHICH IS THE REACTION, I BELIEVE, OF PHOSPHOENOLPYRUVATE TO PYRUVATE

……

34 THE OTHER THING THAT'S HAPPENING IN THE PHOSPHOENOLPYRUVATE TO PYRUVATE REACTION IS THAT WE ARE MAKING A MOLECULE OF ATP

Cellular Relation Path

69 WELL I SUPPOSE ONE PART OF THE ANSWER COULD HAVE TO DO WITH THE NEED OF THE **ERYTHROCYTE** FOR ENERGY TO KEEP ITS MEMBRANE IN GOOD REPAIR.

Correct Answer

76 BUT IF WE DIDN'T HAVE **PYRUVATE KINASE**

77 WE WOULD NOT BE MAKING AS MUCH ATP

78 I WOULD SUSPECT THAT THAT WOULD BE A MAJOR MECHANISM HERE

From this sequence of relations, it is concluded the subject is following the normal function strategy. Statements 2, 3, and 34 satisfy Subgoal 1-M because they relate the molecular problem-statement term to the answer-relevant term. Statement 69 satisfies Subgoal 1-C, because it relates the cellular problem-statement term to the answer-relevant term. Only normal function has been referred to at this point. Then, Statements 76-77, which introduce normal function and provide the correct answer, satisfy Subgoal 2.

Example of the known pathology strategy

Table 2 presents a protocol portion that is an example of the known pathology strategy. This protocol requires a brief explanation. "Carbohydrate metabolism" refers to a number of processes including glycolysis.

In Statement 96, the subject refers to defects of the process hexose monophosphate pathway (HMP), which are a class of known pathologies. The detection of the occurrence of the term, HMP, is directly made by the analysis. The remainder of this section discusses the strategy more informally.

Table 2. Problem solving protocol.

KNOWN PATHOLOGY SEGMENT

95 WELL THERE ARE OTHER THINGS IN CARBOHYDRATE METABOLISM THAT ARE INVOLVED
96 THERE ARE DEFECTS IN THE HEXOSE MONO-PHOSPHATE PATHWAY CAN PRODUCE HEMOLYTIC ANEMIAS
97 IN FACT THERE ..... [This is a sentence fragment]
98 6-PHOSPHOGLUCOSE DEHYDROGENASE IS ONE OF THE MORE COMMON CAUSES OF A HEMOLYTIC ANEMIA
99 BUT I DON'T SEE WHERE THAT WOULD COME INTO PLAY HERE
100 IF ANYTHING, JUST THE MASS ACTION WOULD BACK UP THE PATHWAY MORE TO GLUCOSE-PHOSPHATE
101 AND IF ANYTHING SHUNT GLUCOSE INTO ALTERNATIVE PATHWAYS
102 SO I DON'T THINK THAT THAT'S A PROBLEM

The subject must determine whether or not there is a decrease in flow though HMP. In PK deficiency, there is an increase in flow, so this possible causal account should be rejected. The subject's reasoning proceeds as follows: In Statement 100, she states that PK deficiency will cause an increase in glucose-phosphate. In Statement 101, she indicates that there will be an increase in flow through all alternative pathways, one of which is HMP. Finally, in Statement 102, she correctly concludes, "I don't think that that's a problem".

Because the known pathology was introduced in Statement 96 and a possible causal role was rejected in Statement 102, the interval from 96 to 102 is assigned to the known-pathology strategy regarding HMP.

Summary

The analysis initially focuses on the introduction of information relevant to the correct answer. In order to determine reasoning strategies, the analysis detects three occurrences of the answer-relevant terms: the molecular-relation-path, the cellular-relation-path, and the correct-answer assertion. Then the location of any known pathology is noted. From the occurrence of this knowledge in a protocol, the subject's problem-solving strategy is inferred.

Results

The following section reviews the results of the expert discussed in the previous section, provides a way to display the use of knowledge in reasoning, and then analyzes the other subjects. Throughout this section the focus is on how each subject introduced information relevant to the correct answer in problem solving and how they used that information in giving an explanation.

Experts

Portions of the problem-solving protocol of Expert 1 are included in Tables 1 and 2; a portion of the explanation protocol of Expert 1 is included in Table 3. In addition, the relation paths in the two protocols are summarized in Figure 5. The top of the figure shows the relation path between the answer-relevant term and the molecular problem-statement term for the problem-solving protocol that occurs in Statement 34. The relation path consists of the following relations: PK is related to reaction in Statement 2; reaction is related to PEP and pyruvate in Statement 3; all three terms are related to ATP in Statement 34. Then in Statement 69, which is the cellular-relation-path, energy is related to the RBC-membrane (erythrocyte is synonymous with RBC). This involves a level shift, indicated by the crossing of the line that separates the molecular level from the cellular level in the figure.

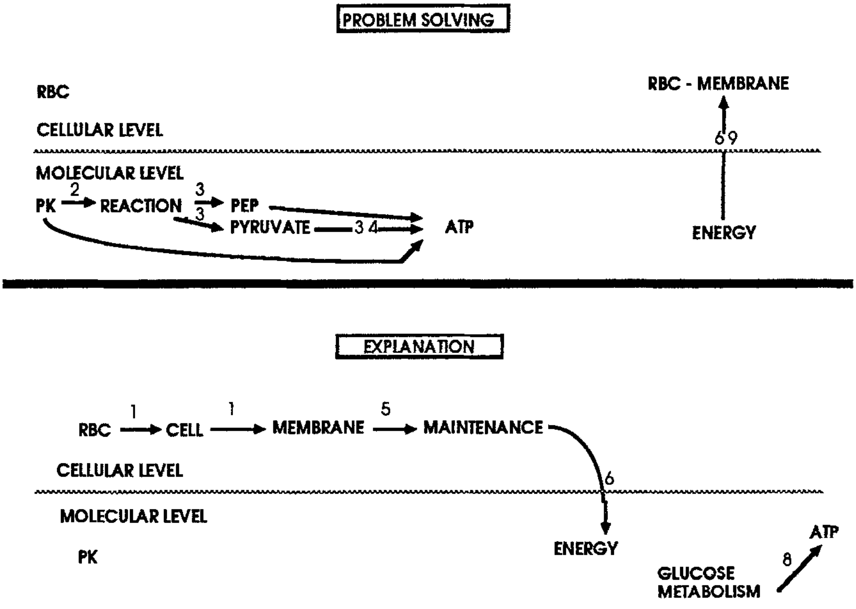

Figure 5. Results For expert 1

Table 3. Explanation protocol.

NORMAL FUNCTION SEGMENT

1 THE ERYTHROCYTE IS A CELL BOUNDED BY A MEMBRANE

2 THE CELL MEMBRANE OF ALL CELLS PLAYS A VITAL ROLE OF MAINTAINING THE INTEGRITY OF THE CELL

3 AND KEEPING THE CONTENTS OF THE CELL WITHIN [THE CELL]

4 AND SERVING AS A STRUCTURE WHICH REGULATES THE ENTRY AND EXIT OF MOLECULES INTO AND OUT OF THE CELL

5 THE MEMBRANE INTEGRITY HAS TO BE MAINTAINED THROUGH THE LIFETIME OF THAT CELL

6 AND THIS IS AN ENERGY-REQUIRING PROCESS

7 AS A RESULT, EVENTS WHICH INTERFERE WITH ENERGY PRODUCTION WITHIN THE CELL COULD BE EXPECTED TO HAVE AN EFFECT ON MEMBRANE FUNCTION AND INTEGRITY

8 IN THIS PARTICULAR SITUATION WE HAVE A DEFECT IN AN ENZYME THAT IS REQUIRED FOR METABOLISM OF GLUCOSE TO GENERATE ATP

9 AND SO WE HAVE IMPAIRED ATP PRODUCTION

10 ATP BEING THE ENERGY CURRENCY WITHIN THE CELL

11 AND WE COULD THEREFORE EXPECT THAT THIS....THAT SUCH A CELL THAT CANNOT MAKE NORMAL AMOUNTS OF ATP

12 CANNOT CARRY OUT MANY ENERGY-REQUIRING PROCESSES

The explanation is shown at the bottom of Figure 5. The cellular-relation-path relates the answer-relevant term to the cellular problem-statement term in Statement 6 (via the entities cell, membrane, and maintenance). This involves a level shift. Then the answer-relevant term is related to the molecular problem-statement term in Statement 8. PK is not referenced explicitly, but rather is referred to as an enzyme in glucose metabolism.

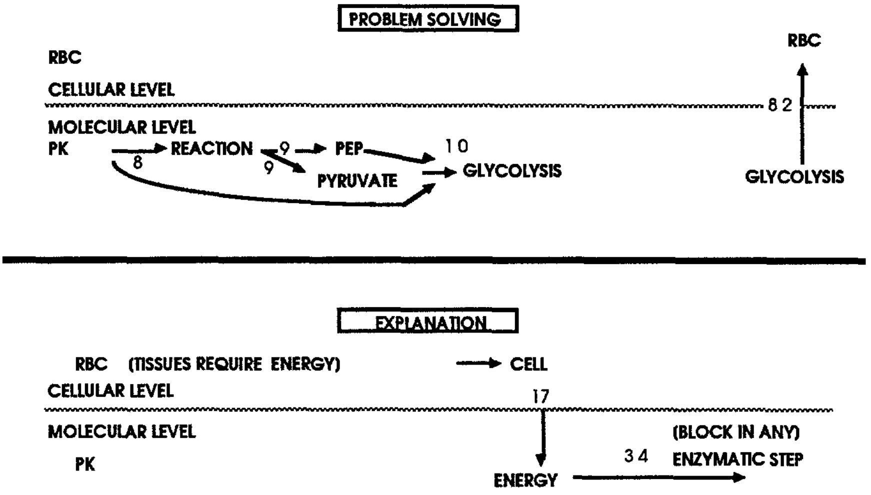

Figure 6. Results for expert 2

Figure 6 presents the corresponding comparisons for Expert 2. For problem solving, the molecular-relation-path occurs first. In this case, the answer-relevant term is glycolysis. The cellular-relation-path, which involves the RBC, occurs in statement 82. This is a level shift, indicated by crossing the line between levels.

In the explanation for Expert 2 (at the bottom of Figure 6), the first relation path is to the cellular problem statement term. There is a reference to the general cell, and its need for energy. In Statement 17, the subject states that the RBC, like all cells, needs energy. In Statement 34, the answer-relevant term is related to the molecular problem statement term. PK is not referred to explicitly, but rather as an enzymatic step in energy production.

Expert 1 and Expert 2 are similar in several ways. Both subjects used the specialized form of the normal-function strategy in problem solving, and the general form in the explanation. The explanation involved a re-ordering of the knowledge of problem solving; the first relation was to the cellular problem statement term. There was no explicit reference to any reaction mechanism; not even to the problem statement term, PK. Instead, both experts used relations involving the general cell and the general form of the relevant axiom.

Novices

The problem solving protocol for Novice 1 is presented at the top of Figure 7. The molecular-relation-path involves reaction mechanisms and the metabolic pathway, glycolysis, and occurs before the cellular-relation-path.

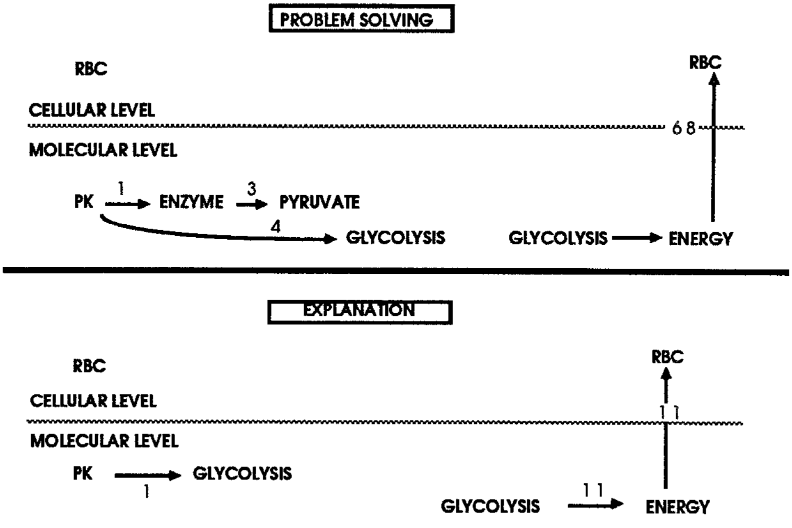

Figure 7. Results of novice 1

Although both of these relation paths involve normal function, Novice 1 introduced pathology (a decrease in energy) in Statement 50. Therefore, this is not a pure form of the normal-function strategy.

The explanation of Novice 1 (at the bottom of Figure 7) begins at the molecular level: PK is related to glycolysis in Statement 1. Then there is a relation to the cellular term in Statement 11, which is a reference to the specialized cell. This is not a pure form of the normal-function strategy, because the subject again refers to a pathology.

The problem-solving protocol of Novice 2 is shown in the top of Figure 8. The first relation-path is to the molecular problem-solving term in Statement 3. The first relation path to the cellular problem-statement term occurs in 53: "A lack of ATP could kill the cell". By cell, this subject is referring to the RBC, not to the general cell. This reference is to a pathology, and therefore is a deviation from the normal-function strategy.

The explanation of Novice 2 is shown at the bottom of Figure 8. The first relation path is to the cellular problem-statement term. However, the relations are specialized with respect to the particular cell: "In the RBC, you only have anaerobic glycolysis" (Statement 7). The first relation to the molecular problem-statement term occurs in Statement 12, where the subject explicitly refers to the PK reaction in detail.

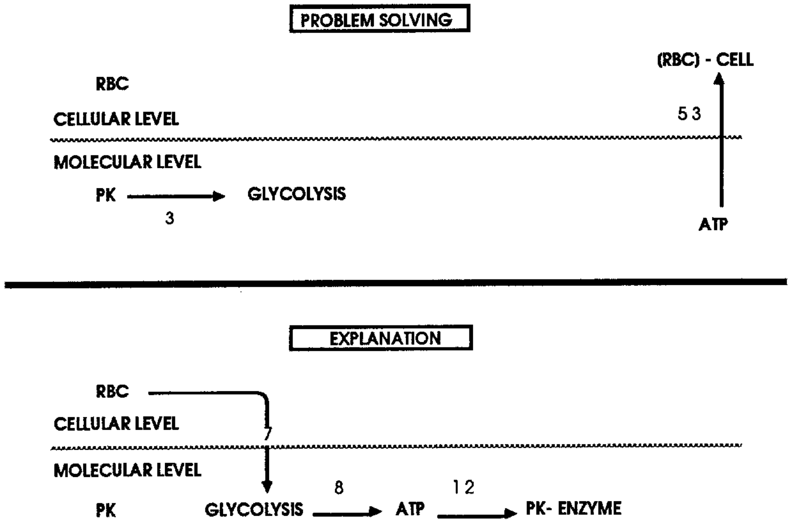

Figure 8. Results of novice 2

Comparison of Experts and Novices

In the problem-solving protocols of all subjects, the molecular-relation-path occurred before the cellular-relation-path. In each case, reaction mechanisms were involved, and there was explicit reference to the molecular problem-statement term, PK. For both experts and novices, the cellular-relation-path involved the specialized cell. Therefore, the problem-solving protocols of both experts and novices were similar in their search for the correct answer. All subjects began by introducing terms at the molecular level, without stating the relevance of those molecular terms to the cellular level and, when later making a level shift, all subjects then referred to the specialized cell.

In the explanation, both experts began with the cellular problem-statement term. They did not refer to properties of the specialized cell, but rather to properties of cells in general. The experts made no explicit reference to reaction mechanisms, or even to the problem-statement term, PK. The novices, on the other hand, made explicit reference to the PK reaction and to the specialized cell. Therefore in the explanation, both experts used the general form of the normal-function strategy; the novices did not.

In problem solving, the experts found the correct answer by a specialized form of the normal-function strategy. They considered the normal relations before considering an abnormal condition involving the answer-relevant term. In contrast, the novices made reference to abnormal conditions involving the answer-relevant term before fully establishing the normal relation.

Use of the Known-pathology strategy

As discussed in the previous section, Expert 1 used the known-pathology strategy. Expert 2 also used this strategy. He referred to a known pathology in Statement 13 and 14 of the problem solving protocol: "Certain types of anemia that involve NADPH" and "glutathione". This general anemia is related to HMP: the final product of HMP is used to form glutathione. The subject introduces the disorder in Statements 13 to 16, and asks for clarification in 17 through 32. From 33 to 64, he refers to the normal mechanisms related to the known pathology. Then he concludes: "That's really the opposite of what this is doing" (Statements 65-66). This reasoning is therefore a good approximation of the known-pathology strategy.

Neither novice made any reference to a known pathology.

Deciding to Terminate Reasoning

After finding the correct answer, Expert 1 continued to reason about this problem. The correct answer was found in Statement 77, but the protocol continued until 129. Similarly, Expert 2 gave the correct answer in 126, but continued to reason until 193. Novices, on the other hand, were much more willing to stop upon giving an answer which they considered to be correct. Novice 1 terminated reasoning after giving the correct answer in Statement 111. Novice 2 stopped problem solving before giving the correct answer, which he found during the recall.

Both experts considered other possible causal mechanisms during the interval from the correct answer to the conclusion of problem solving. Expert 1 determined the consistency of the correct answer with other knowledge of the domain, and then considered whether any other effects were related to the answer. Expert 2 considered "experiments which I would consider to test this hypothesis".

Summary

This study demonstrates the following important differences between experts and novices:

In problem solving, novices did not use the pure form of the the normal-function strategy to find the correct answer. Instead, they referred to abnormal function before completely understanding the normal function. On the other hand experts made no reference to the abnormal function of the answer-relevant term before the normal relation was stated.

Experts and novices also differed in their use of the other strategy, the known-pathology strategy. While both experts used it at least once in problem solving, neither novice used it at all. One reason for this difference is that experts know more pathologies. However, this difference is plausibly related to a different appreciation for the complexity of the domain. Both novices terminated reasoning when they found the correct answer. Experts, on the other hand, did not stop, but continued to explore at least one way of confirming their answers.

A clear difference between experts and novices is evident in their transition from the problem-solving protocol to the explanation. In problem solving, the experts employed a specialized form of the normal-function strategy, which involved reaction mechanisms and references to the specialized cell. In the explanation, they used the general form of that strategy. Instead of referring to reaction mechanisms, they related the problem to a basic principle. Novices, on the other hand, did not use the general form of the the normal-function strategy in their explanations. They continued to refer to reaction mechanisms and particular properties of the specialized cell, and did not refer to the general axiom.

General discussion

The results indicate that reasoning processes follow the basic properties of the organization of knowledge of a domain. In metabolism, new and interesting reasoning strategies that have no analog in areas previously studied were uncovered.

In particular, expert biochemists employed both the normal-function strategy, in which the relevant normal function is considered before abnormal function is introduced, and the known-pathology strategy, in which a known pathology is introduced in order to suggest a causal mechanism that is relevant to the problem.

Furthermore, experts used these strategies to explore a variety of possible causal mechanisms, followed by reasoning to eliminate less likely possibilities. Experts introduce possibilities which novices do not even consider, and therefore do not necessarily reach the correct solution more rapidly than novices, because experts realize that biochemistry is not logically complete, and that one likely causal mechanism does not rule out other, and perhaps more significant, causal mechanisms.

These findings are in sharp contrast to the results of studies of problem solving in Newtonian mechanics and clinical medicine. Because correct answers in Newtonian mechanics can be obtained by the legal application of a well-defined set of mathematical equations, experts achieve the correct answer more quickly than novices because they better understand these operations and how they are organized. Problems in clinical medicine typically have a generally accepted answer, and experts reach this more quickly because they have more knowledge and a better understanding of how that knowledge is classified. Although a correct answer exists in biochemistry problems, there is no way to confirm it without laboratory experiment. Consequently, it is important to consider other plausible alternatives before settling upon a final answer.

This study also obtained explanations following problem solving. These data provide an indication of how a subject altered knowledge that was involved in the solution of the problem. Experts used a general form of the normal-function strategy, which relates to an axiom of normal function, in a way consistent with the organization of domain knowledge. Novices used a more specialized form of the normal-function strategy, which relates to properties of a specific cell. Novices were concerned more with summarizing the answer to the particular problem, whereas experts were much less concerned with the particular problem and much more concerned with the way that it exemplified basic principles. The use of general principles by experts is consistent with prior work on scientific problem solving. However, this study extends those findings to show how fundamental principles in a domain are used as an explicit guide to the decomposition of a problem.

These findings indicate that reasoning in biological science differs fundamentally from reasoning in areas that have previously been studied in cognitive science. Expert biochemists have a different concept of a solution than their counterparts in Newtonian mechanics and clinical medicine. The biochemist does not attempt to find a mathematically verifiable solution as in Newtonian mechanics, nor the accepted solution as in clinical medicine, but rather to discover the most plausible and convincing solution only after examining and evaluating a wide range of possibilities.

Educational Implications

The structure of a domain is fundamental and needs to be described explicitly, and in a form that displays features that are related to reasoning.

The methodology of the present study is sufficiently detailed to guide the development of an automated information-retrieval system for educational purposes. Such a system would differ from Blanchaer's computer-assisted instruction programs by permitting students to access a wide range of facts rather than only the facts considered by the instructor to be important for a particular problem. Furthermore, those facts would be presented in a structured manner, clearly noting the relationship of details to basic principles. Such an information-retrieval system would also differ from programs such as MYCIN by having a clear relation to normal function.

In addition to information retrieval, another important part of problem solving in biochemistry is the evaluation of a solution. The results of this study indicate that experts use more strategies and are less likely to accept a plausible solution as necessarily the complete answer. This study described two strategies and detected their use during problem solving.

Experts have a different view of the solution to a problem: they are not as interested in a particular problem as they are in how that problem illustrates basic principles. Students can be introduced to different ways of explaining the solution to a problem. It is likely that this more general explanation leads to an improvement in the ability to solve related problems and also an increased appreciation of the generality and importance of the basic ideas. It is fortunate that Blanchaer's instructional programs involve the same general area, metabolic diseases. Therefore, much of the groundwork has been established for testing whether giving a general explanation to a particular problem leads to greater ease in solving other problems.

The educational implications of this study are not limited to the instruction of future biochemists. All people would benefit from a greater understanding of modern biology. However, the amount of information can discourage even an eager layperson. The description of the organization and use of biochemical knowledge can be understood by persons with far less familiarity with the domain than is required in existing work. This study, therefore, can make the knowledge of discoveries in biological science available to a much wider audience.

As an extension and realization of this work, we are constructing a biology information-support system in Boxer. Boxer is a general purpose computer language designed for pre-college science education (diSessa and Abelson, 1986). Boxer's hierarchical structure is well-suited to accommodate the representation of metabolic knowledge. In the not too distant future, we plan to have a structure for storing knowledge in biochemistry and physiology and a means for accessing this knowledge during reasoning consistent with the principles uncovered in this study.

Acknowledgements

This research was supported by National Science Foundation Grant NSF IST81-19889 awarded to Charles Schmidt. I thank my dissertation advisor, Charles Schmidt, and the members of my committee, Richard Harvey, Len Hamilton, Carl Helm, and Eliot Noma, for their constructive criticism and their interest in the work, and Loretta Kasper for her suggestions and support throughout this project.

References

Blanchaer, M. C. (1982). Microcomputer-based learning in a medical biochemistry course. *Biochemical Education*, **10**, 107–109.

Buchanan, B. G. and Shortliffe, E. H. (1984). *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project.* Reading, MA.: Addison-Wesley.

Clancey, W. J. (1983). The epistemology of a rule-based system : A framework for explanation. *Artificial Intelligence*, **20**, 215–251.

diSessa, A. A. & Ploger, D. (1987). Cognition and Science Education. The 1987 Forum for School Science. Students and School Science: Papers from the 1987 National Forum for School Science. Washington, D. C.: American Association for the Advancement of Science.

diSessa, A. and Abelson, H. (1986). Boxer: A reconstructable computational medium. *Communications of the ACM*, **29**, 859 – 868.

Kuipers, B. and Kassirer, J. P. (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science*, **8**, 363–385.

Larkin, J. H., McDermott, J., Simon, D. P. and Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science*, **4**, 317–345.

Lehninger, A. L. (1975). *Principles of Biochemistry.* New York: Worth.

Newell, A. (1981). The knowledge level. *AI Magazine.*, **2**, 1–20.

Patel, V. L. and Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, **10**, 91–116.

Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. and Brown, M. S. (1982). *The metabolic basis of inherited disease.* New York: McGraw-Hill.