The PepPro™ Peptide Synthesis Expert System

Philip R. Martz, Matt Heffron, Suresh Kalbag, Douglas F. Dyckes, Paul Voelker

Beckman Instruments, Inc.
2500 N. Harbor Blvd.
Fullerton, Ca. 92634

Abstract

Peptide synthesis is an important research tool. However, successful syntheses require considerable effort from the scientist. We have produced an expert system, the PepPro™ Peptide Synthesis Expert System, that helps the scientist improve peptide syntheses. To use PepPro the scientist enters the peptide to be synthesized. PepPro then applies its synthesis rules to analyze the peptide, to predict coupling problems, and to recommend solutions. PepPro produces a synthesis report that summarizes the analysis and recommendations. The program includes a capability that allows the scientist to write new synthesis rules and add them to the PepPro knowledge base. PepPro was developed on Xerox 11xx series workstations using Beckman's proprietary development environment (MP). We then compiled PepPro to run on the IBM PC. PepPro has limitations that derive from unpredictable events during a synthesis. Despite the limitations, PepPro provides several important benefits. The major one is that it makes peptide syntheses easier, less time-consuming, and more efficient.

Introduction

The ability to synthesize peptides is an important research tool in the biological sciences. A peptide is a sequence of amino acids. In the most common method, solid-phase peptide synthesis (Merrifield 1963), the scientist begins by anchoring the first amino acid of the sequence to a solid substrate. Then, using manual or automated (peptide synthesizer) chemistry, successive amino acids are coupled, one at a time, to the previous amino acid in the sequence until the peptide is complete.

Unfortunately, a synthesis can be difficult and time-consuming for the scientist. The synthesis may not go according to plan. The most likely problem is that an amino acid that should have coupled, has coupled poorly or not at all. If the scientist handles this and other challenges well, the synthesis will be successful. If the scientist does not handle these challenges well, the yield of the desired peptide will be reduced and the yield of *undesired* peptides will be increased. The purification of the peptide, which is a major task in itself, may then become even more difficult. If the yield of the peptide is too low the synthesis may have to be repeated. Finally, there will be economic losses in terms of time and materials.

We have developed a program, the PepPro™ Peptide Synthesis Expert System, that helps the scientist recognize and respond to the challenges of peptide synthesis. PepPro incorporates the knowledge of peptide synthesis experts. It uses these synthesis rules to analyze the peptide, to predict coupling problems, and to recommend appropriate procedures. PepPro produces a multi-page synthesis report that summarizes the analysis and recommendations. The scientist can use this report to conduct the synthesis. One of PepPro's most powerful capabilities allows the scientist to write synthesis rules and add them to PepPro's knowledge base.

PepPro is a commercial product of Beckman Instruments, Inc. It is a 1.5 megabyte program for IBM XT, AT, and PS/2 personal computers and their compatibles. It includes a manual of approximately 110 pages.

There are programs that address the planning of chemical syntheses. However, these are general purpose programs for organic syntheses and do not specifically address the requirements of peptide synthesis. Two peptide synthesis references that discuss solid-phase synthesis are Barany & Merrifield (1980) and Stewart & Young (1984).

Two of PepPro's functions are the subject of this paper. The **Synthesis** function analyzes the peptide and provides the synthesis advice. The **User**

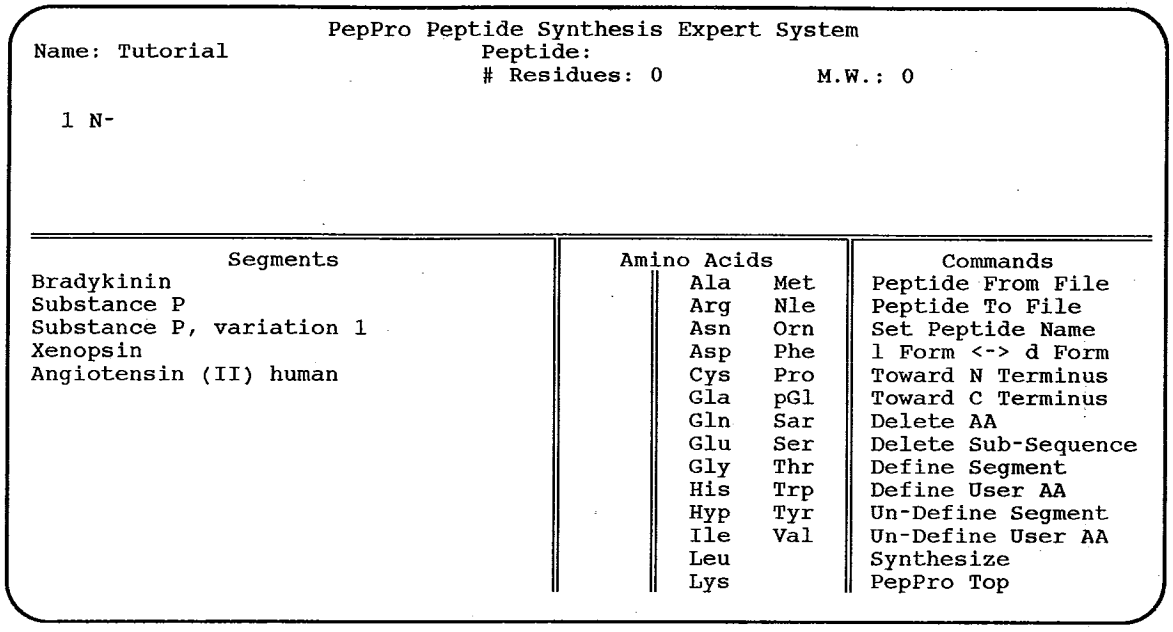

Figure 1: The peptide input screen.

Rules function allows scientists to write synthesis rules based on their knowledge or experience and to add these rules to PepPro's knowledge base. Two other functions, **Information** and the **User Profile Manager**, provide important support roles to the expert system but will not be discussed.

The Synthesis Function

The scientist enters almost all inputs by selecting commands with a mouse. To run the Synthesis function the scientist points at the word *Synthesis* on the PepPro Top Level screen and clicks the left mouse button.

The first screen of the Synthesis function is the peptide input screen (Figure 1). This screen provides the scientist with a refined peptide editing interface. This screen simplifies the process of entering the peptide and provides commands to edit amino acid sequences, to save peptides by name, to define new amino acids, and to enter peptides from a file. Selecting any of these commands with the right-hand mouse button invokes the on-line help; PepPro will display a statement of the command's function and operation.

Once the peptide is entered, the user selects *Synthesize.* PepPro then begins a question and answer dialogue in which the scientist specifies the desired characteristics of the peptide and the synthesis. PepPro analyzes the peptide using a backward chaining inference engine (MP) written by us.

PepPro's synthesis knowledge is encoded as *If-then* rules. A rule is shown in Figure 2. The rule is used each time PepPro checks whether to acetylate an amino acid (a process that can reduce the formation of impurities) after it is coupled to the peptide fragment. The rule causes PepPro to determine if BULKY.BULKY = Yes. BULKY.BULKY is a parameter that indicates whether the coupling of one amino acid to another is likely to be slow. PepPro determines the value of BULKY.BULKY in another rule (not shown here) that examines both the sequence and an amino acids database. If the coupling is likely to be slow, the rule checks two conditions. If the CURRENT.CHARGE of the peptide fragment is not equal to the PEP.CHARGE, the advice is to acetylate. The rule also succeeds if the number of amino acids in the fragment and the peptide differ by more than 20%.

PepPro's rules do not apply to the peptide as a whole. Instead, the rules apply to amino acids. PepPro uses its rules to analyze each amino acid separately in terms of its context of surrounding amino acids. It then determines the recommendation, and makes an entry in the reports. PepPro repeats this cycle (evaluate the context, apply rules, write entry in report) for each amino acid in the sequence until all amino acids have been checked.

PepPro produces the following reports:

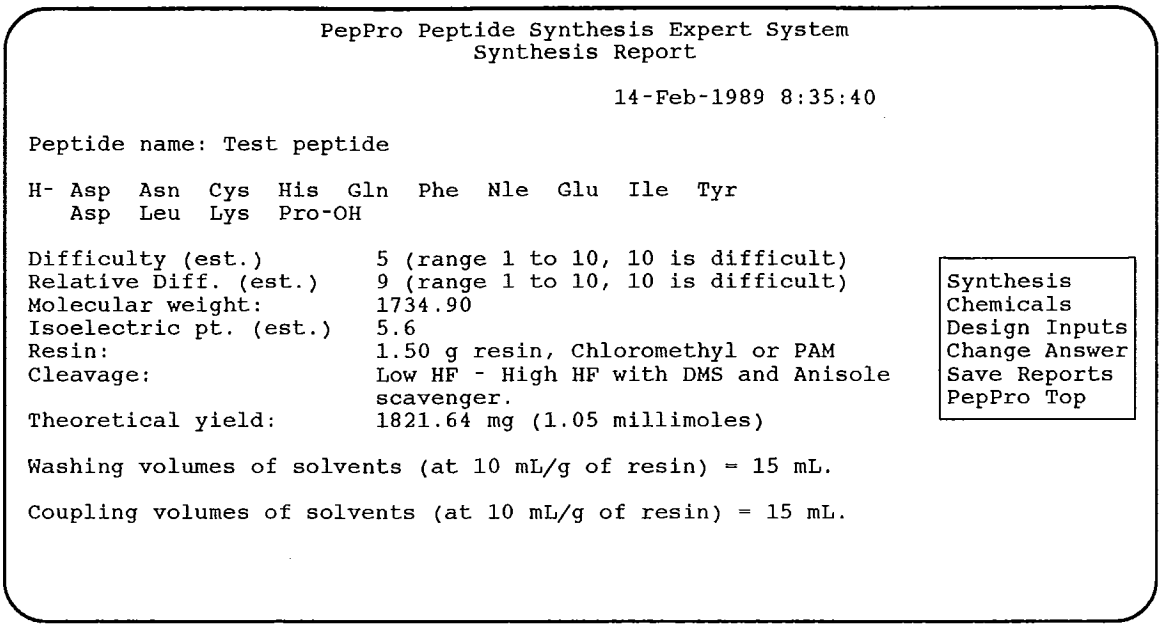

Figure 3: The General Recommendations of the Synthesis Report.

```
Rule 1103: Rulegroup Peptide.Plan.Rules

If: 1. BULKY.BULKY = Yes, and
    2. Either:
     A. (PEP.CHARGE - CURRENT.CHARGE)=1
     or
     B. (PEP.NUMBER / FRAG.NUMBER)>.2
Then: Conclude ACTION is ACETYLATE
```

Figure 2: One of PepPro's rules for acetylation.

- Synthesis Report—This is PepPro's analysis and recommendations of how to conduct the synthesis. The user can request that it be produced in a concise or an expanded format. The report has three parts:
  - General Recommendations
  - Fragments List
  - Amino Acid Recommendations
- Chemicals Report—A list of the amino acids, reagents, solvents, and scavengers required to do the synthesis. This report is tied closely to the recommendations made in the Synthesis Report so that it will reflect only those chemicals required for the synthesis in question.
- Design Inputs—A summary of the inputs during the question and answer dialogue.

The reports include a menu of commands along the right-hand side of the screen that allow for switching between reports (Figure 3). There is a command to save the reports to a file. The **Change answer** command allows the user to change the requirements and have PepPro reevaluate the synthesis. Commands along the last line of the screen allow for scrolling and paging through the reports.

General Recommendations

This part of the Synthesis Report outlines the major characteristics of the synthesis (Figure 3). The significance of the General Recommendations is that it establishes the primary considerations for conducting the synthesis: how to start the synthesis, how difficult it will be, how to end the synthesis, and what techniques will be required.

A more detailed description follows. PepPro gives two difficulty ratings for the synthesis. These ratings are estimates, on a scale of 1 to 10, with 10 indicating the most difficult. They are based on the peptide length plus rules that estimate how difficult it is to couple the amino acids in the peptide. Using these estimates, the scientist can plan an appropriate synthesis strategy by knowing how difficult a synthesis is likely to be. PepPro estimates the peptide's isoelectric point, information that is useful

	Charge	Residues	Source
Peptide:	-1	7	
Fragment:	-1	5	PepPro
Fragment:	0	4	PepPro
Fragment:	-1	3	PepPro
Fragment:	-1	2	User

Figure 4: PepPro's list of peptide fragments.

when the peptide is purified. It determines the appropriate resin linkage from the desired C-terminus form. The report shows the method for cleaving the peptide from the resin. It includes advice on appropriate chemical additives (scavengers). The report gives the theoretical yield based on 100% coupling efficiency and an estimate of actual yield that is determined by estimating the coupling efficiencies of the amino acids. Finally, the report gives the solvent volume and reagent (coupling) volume that the scientist should use.

Fragments List

PepPro's account of peptide fragments (impurities) that will be created during the synthesis is a significant capability (Figure 4).

The report shows the number of residues, the fragment charge, and the source (either PepPro or User) of the rule that is responsible for producing the fragment. The fragments list helps the scientist monitor small, but important, details. The list can alert the scientist that, given the current synthesis plan, a fragment very similar to the desired peptide will be produced. The list can be used to determine a purification strategy. By examining the list the scientist can decide whether purification based on size, charge, or a combination of the two will be required to separate the fragments from the peptide.

Amino Acid Recommendations

The coupling of the amino acids is a crucial step in the synthesis and the synthesis activity most responsive to sound recommendations. PepPro suggests a coupling strategy for each amino acid in the peptide. An example of an amino acid entry is shown in Figure 5.

The recommendation gives PepPro's concerns, methods, and justifications for the coupling step. The first line of the entry shows that this advice concerns the coupling of *Asp* (Aspartic acid) to *Arg* (Arginine) preceded by *Val* (Valine). PepPro has selected a chemical group, (OBzl), as a way to protect

#1	Boc-Asp(OBzl) → Arg-Val...Resin
Method:	Double-couple. *Merrifield coupling *15 mins.
Reason:	Double-coupling is advised due to the side groups of Asp and Arg.
Advise:	HOBt, 30 min
PepPro:	Acetylating will produce a fragment with the same charge as the peptide.
Lab:	Paul uses Merrifield for 15 mins.

Figure 5: An amino acid entry.

the amino acid throughout the synthesis. The advice indicates to: *Double-couple with Merrifield coupling* for 15 minutes. PepPro provides a justification **(Reason:)** and any additional notes or suggestions **(PepPro:)**.

When determining a recommendation, PepPro checks its rules as well as any rules that the user has written. Thus, the entry is a combination of recommendations coming from PepPro's rules and from the user's rules. In the figure, PepPro has marked *Merrifield coupling* and *15 minutes* with asterisks indicating that these recommendations came from a rule written by the user. PepPro has also written an **Advise** entry which indicates that the recommendation PepPro would have made conflicts with the recommendation in the user's rules. The **Advise** entry shows that PepPro would have recommended *HOBt coupling for 30 minutes.* The entry concludes with any notes **(Lab:)** that the scientist, by way of User Rules, has written for this step.

The User Rules Function

The User Rules function is one of PepPro's most significant features because it gives PepPro unusual flexibility. We showed the central outcome of this flexibility in the previous section; the scientist has a way to control, by way of user written rules, the recommendations that appear in the Synthesis Report.

We were motivated to provide User Rules by three factors. First, peptide synthesis is a growing science; its rules are not hard and fast. Our experts agreed that they could not represent the range of experience that exists in the field. Second, there was a diversity of opinion on how syntheses should be conducted and which methods should be used. PepPro represents a significant, but not the only, approach. User Rules allows for the diversity of opinion. Third, rapid increases in peptide synthesis knowledge could make PepPro obsolete. All three of

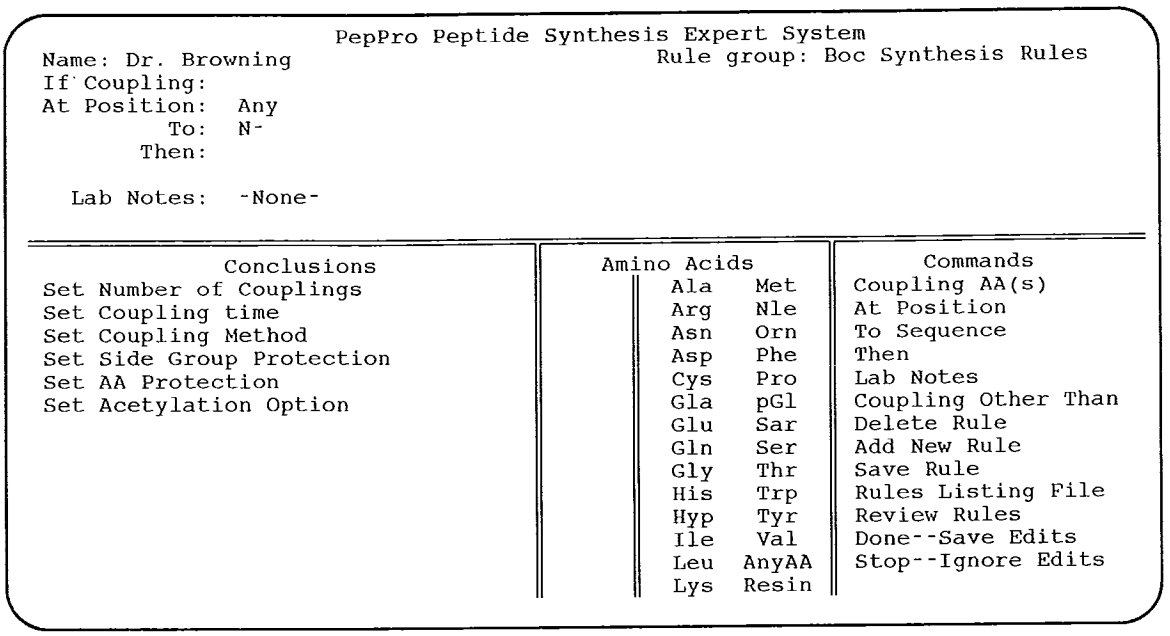

Figure 6: The rule writing screen.

these reasons indicated that PepPro required a flexibility that would allow the scientist to highly personalize PepPro in terms of opinion and knowledge.

From the scientist's view the reasons for writing rules are more immediate:

- to incorporate new knowledge, experience, or insights;
- to be more or less conservative than PepPro;
- to change one of PepPro's recommendations that is perceived to be wrong;
- to add one's comments to a synthesis report.

Rule Writing Mechanics

To run the User Rules function the scientist selects *User Rules* from the PepPro Top Level screen. The scientist can write rules for one of three rule groups depending on the type of synthesis. PepPro requests the name of the rule group that should be accessed and then displays the rule writing screen (Figure 6). This screen provides the commands necessary to write, edit, delete, and save rules. The rule format is shown in Figure 7. Using this format the scientist can concisely express synthesis knowledge. On one hand the scientist can write rules that apply only when very specific criteria are met. One example is shown in the first rule of Figure 8 where the **At Position** and the **To:** clauses establish a highly specific rule.

If Coupling:	*some AAs*
At Position:	*a range or list of positions*
To:	*a sequence of AAs*
Then:	*do the following ...*
Lab Notes:	*for these reasons ...*

Figure 7: The rule format for user written rules.

This rule will apply only when Asp is coupled to the sequence *Glu-Cys- Asp.* At the other extreme, the scientist can write completely general rules (second rule in Figure 8). This rule will always apply whenever any amino acid is coupled to any amino acid. The third rule demonstrates the use of rules as a sequence dependent note keeping device. The scientist is requesting that whenever PepPro encounters the pattern *Gln-AnyAA-Pro* at positions 2 through 8 it should enter the Lab Notes into the report. The fourth rule shows the use of the *Coupling Other Than* command. This command allows the user to state the amino acids to which the rule doesn't apply. This rule also shows that one can write that something not be done (*Don't Acetylate*). The flexibility of PepPro's rule writing scheme is noted by an interesting fact: The first rule in Figure 8 is more specific and the second rule is more general than any rule in PepPro's

If Coupling:	Asp
At Position:	8, 9, and 15
To:	N-Glu-Cys-Asp
Then:	Double-couple active ester 30 minutes Acetylate
If Coupling:	Any AA
At Position:	Any
To:	N-Any AA
Then:	Single-couple HOBt 45 minutes Acetylate
If Coupling:	Gln
At Position:	2 through 8
To:	N-AnyAA-Pro
Lab Notes:	Gln-AnyAA-Pro cyclize here. Pause for coupling test.
If Coupling:	Other than Ile, Val, and Leu
At Position:	Any
To:	N-AnyAA
Then:	Symmetric anhydride 20 mins. Don't Acetylate

Figure 8: Examples of user written rules.

knowledge base.

User written rules are not simply an adjunct to PepPro's rules. If desired, the scientist can write rules to establish a knowledge base that significantly enhances PepPro's knowledge base. The scientist can even replace PepPro's knowledge base by writing the most general rule premise, *If Coupling AnyAA to AnyAA* (second rule of Figure 8). Since the user's rules are checked first and since this general rule always applies, PepPro's rules will never be checked. With this rule in place, the scientist can write other rules for more specific situations.

Integration with PepPro's Rules

A truism of rule based systems is that each rule represents a unit of knowledge independent of other rules. In fact, this is rarely the case. Rules interact in unexpected ways. Adding one new rule can easily disrupt an expert system. This raised the question: *How to organize and use the scientist's rules and still maintain the integrity of PepPro's knowledge base?*

We addressed this question in three ways. First, we keep PepPro's rules separate from the user's rules. Second, PepPro checks the user's rules before it checks its rules. If a user's rule applies, PepPro will use that rule and ignore its rule. Third, PepPro does not treat the user's rules equally. It ranks them in order from the most to the least specific by calculating a score for every rule. For example, the more specific *If coupling Leu to Asp* is given priority over the less specific *If coupling Leu to Any AA*. When analyzing a peptide, PepPro checks the rules in order from the most to the least specific. This is a natural way to organize rules; the specific cases take precedence over the general case. Fourth, we do not prevent the scientist from writing unreasonable rules, primarily because we cannot determine confidently whether a rule is unreasonable. However, PepPro does use its rules to analyze the recommendations arising from the user's rules. PepPro's response to a rule that might be unreasonable is to place an *Advise* entry in the Synthesis report (Figure 5).

Development

We started the PepPro project in late 1986. The development team consisted of one primary expert, three consulting experts, two knowledge engineers, and support personnel. PepPro's development has included several versions of the software and a beta test. PepPro has been used internally for about one year. It is scheduled to become a commercial product in March 1989.

PepPro was developed on Xerox 11xx series Artificial Intelligence workstations and then compiled to run on the PC. On the Xerox workstation we used an expert system development environment (MP) written by us. MP was originally developed to support an earlier product, the SpinPro™ Ultracentrifugation Expert System (Martz, Heffron, & Griffith 1986). MP exists as Common Lisp source code with macros that conditionally expand based on whether the inference engine is to run on the PC or the workstation. It provides the developer with several capabilities that are not available in other PC based shells:

- It allows the developer to build the expert system on a workstation (Xerox 11xx series), thereby taking advantage of the workstation's power and multi-window bit mapped interface. MP runs the expert system in a window that simulates a PC screen.
- It produces the completed system for the IBM PC. This is significant because Beckman's marketing studies indicated that the IBM PC was the delivery vehicle of choice.
- It gave us control of the user interface and any specialized functions required by PepPro.

- It includes a sophisticated report writer that allows PepPro to write complex, multi-page reports.
- It provides a collection of knowledge base development utilities including editing rules and parameters, writing reports, noting inconsistencies, and finding redundant rules.

The initial development consists of building, reviewing, and editing the application in Xerox Common Lisp on the workstation using MP's tools and simulated PC environment. A "rule compiler" then converts the rules into functions. Finally, the knowledge base is moved to the PC where it is compiled using Golden Common Lisp™.

Discussion

During the knowledge acquisition phase of PepPro the experts were required to formalize answers to previously unasked questions: *How does one predict when to double-couple? How does one know when to acetylate?* The answers to questions such as these evolved over time. One example is PepPro's acetylation strategy. Acetylation is a reaction that can be done immediately after a coupling step. If done thoughtfully, acetylation can produce capped sequences that will improve the yield of the synthesis and make purification easier. Early on, several strategies were proposed and discussed. None of these comprised a complete strategy. However, the eventual result was a complete and consistent strategy that evolved from the early strategies. Interestingly, this strategy (the knowledge) did not exist until the experts were required to formalize an approach.

PepPro has limitations. First, some limitations derive from the scientific community's incomplete understanding of the process of peptide synthesis. This incomplete understanding translates directly into limitations in PepPro. For example, the scientist can predict some coupling problems and can address these prior to the synthesis. However, other coupling problems will arise that were not predicted. The scientist is forced to respond with a countermeasure after the problem is detected. Second, the User Rules function, although providing the structure for most of the rules a scientist might want to write, does not allow the scientist to write all the rules that apply to peptides. A third limitation is part criticism and part compliment. Some peptide chemists have noted that PepPro is good at finding the coupling problems they already know about, but not good at finding problems they didn't know about.

Despite these limitations, PepPro provides important benefits. The most important is the potential to improve the scientist's syntheses. PepPro should make these syntheses easier, less time consuming, and more efficient. A second major benefit of PepPro, and most expert systems, is the collection and formalization of the available knowledge on the subject. This formalization process improves the experts' understanding of their field and, consequently, it improves the expert system that results from their knowledge. A third major benefit of PepPro and other expert systems is that the collected information becomes knowledge. Information reported in the literature, although it may be the source for the rules in an expert system, is passive. It places the burden on the scientist to find, understand, and apply it. PepPro, however, actively responds to the problems posed to it. Finally, the User Rules function enhances all the benefits mentioned above: it provides a growth path to improved syntheses, it gives scientists a way to formalize and record their knowledge, and it turns their synthesis notes into active knowledge.

Acknowledgements

For their contributions the authors thank Phyllis Browning, Phil DeSouza, Ed Faulk, Ed Fong, Manny Gordon, Jon Harbaugh, Michael Mokotoff, Jim Osborne, Ted Shapin, John Stewart, Bruce Wintrode, and Janis Young.

References

Barany, G. and Merrifield, R.B. (1984). Solid-phase Peptide Synthesis. In Gross, E. and Meienhofer, J. (Eds.), **2**, 1-284. *The Peptides*. Academic Press, New York,

Martz, P.R., Heffron, M., and Griffith, O.M. (1986). An Expert System for Optimizing Ultracentrifugation Runs. In T.H. Pierce and B.A. Hohne (Eds.), 297-311. *Artificial Intelligence Applications in Chemistry*, American Chemical Society, Washington, D.C.

Merrifield, R.B. (1963). Solid-phase Synthesis: The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.*, **86**: 2149-2154.

Stewart, J.M. and Young, J.D. (1984). *Solid Phase Peptide Synthesis*, 2nd ed. Pierce Chemical Co., Rockford, Ill.