Expert System Simulations as Active Learning Environments

Douglas L. Brutlag
Department of Biochemistry
Stanford University School of Medicine
Stanford, CA 94305, USA

I discuss here a knowledge base capable of predicting the activity of DNA polymerase I under a wide variety of conditions that I developed using classical production rules together with a system for truth maintenance (Figure 1). These rule-based methods have the intrinsic property of being able to explain conclusions both in the form of English sentences and in the form of graphic representation of the flow of inference (Figure 2). Variables such as the presence of salts, nucleotides, temperature, ionic strength and pH are all represented graphically as active images of thermometers, gauges and switches (Figure 3). Properties of the physiological conditions and substrates can be changed by using a mouse-pointing device. In response to changes of any parameter, a series of rules defining the specificity and reactivity of enzymes is automatically invoked and specific conclusions are made (Figure 4). The methods for specifying the actions of enzymes are well defined, allowing ready simulation of other enzymes. These simulations can be coupled together to generate a discrete event simulation of multiple steps in a metabolic pathway.

The power of rule-based systems are multiple. We can represent the knowledge of the metabolism at many levels simultaneously. In some cases, the actual mechanism is known and intermediates in the reaction can be represented. In other cases only the substrates and products are known and these can be related by appropriate rules. The speed of rule-based inferences is much faster than that of mathematical models, making interactive simulation possible. Changes and additional knowledge about an enzyme can be accommodated readily since rules can be added or removed without regard to their order.

```
(ACTIVITY-RULE (IF (THE DIVALENT-CATIONS OF CONDITIONS IS Mg)
                   (THE REACTIVITY OF DNA-POLYMERASE-I IS BINDING)
                    THEN DEDUCE (THE REACTIVITY OF DNA-POLYMERASE-I IS ACTIVITY)))
(BINDING-RULE (IF (THE TEMPERATURE OF CONDITIONS IS ?T)
                  (LISP (< 0.0 ?T 45.0))
                  (THE IONIC-STRENGTH OF CONDITIONS IS ?I)
                  (LISP (< .001 ?I .3))
                  (THE pH OF CONDITIONS IS ?P)
                  (LISP (< 6.0 ?P 9.5))
                   THEN DO (THE REACTIVITY OF DNA-POLYMERASE-I IS BINDING)))
(MN-ACT-RULE (IF (THE DIVALENT-CATIONS OF CONDITIONS IS Mn)
                 (THE REACTIVITY OF DNA-POLYMERASE-I IS BINDING)
                  THEN DEDUCE (THE REACTIVITY OF DNA-POLYMERASE-I IS ACTIVITY)))
(DNA-SYN-RULE (IF (THE REACTIVITY OF DNA-POLYMERASE-I IS ACTIVITY)
                  (THE NUCLEOTIDES OF CONDITIONS IS dATP)
                  (THE NUCLEOTIDES OF CONDITIONS IS dTTP)
                  (THE NUCLEOTIDES OF CONDITIONS IS dGTP)
                  (THE NUCLEOTIDES OF CONDITIONS IS dCTP)
                  (CANT.FIND (THE NUCLEOTIDES OF CONDITIONS IS ddTTP))
                  (CANT.FIND (THE NUCLEOTIDES OF CONDITIONS IS ddATP))
                  (CANT.FIND (THE NUCLEOTIDES OF CONDITIONS IS ddCTP))
                  (CANT.FIND (THE NUCLEOTIDES OF CONDITIONS IS ddGTP))
                  (CANT.FIND (THE NUCLEOTIDES OF CONDITIONS IS ddTTP))
                   THEN DO (THE REACTIVITY OF DNA-POLYMERASE-I IS SYNTHESIS)))
(3-EXO-ACT-RULE (IF (THE REACTIVITY OF DNA-POLYMERASE-I IS ACTIVITY)
                     THEN DEDUCE (THE REACTIVITY OF DNA-POLYMERASE-I IS 3'-EXONUCLEASE)))
(5-EXO-ACT-RULE (IF (THE REACTIVITY OF DNA-POLYMERASE-I IS ACTIVITY)
                     THEN DEDUCE (THE REACTIVITY OF DNA-POLYMERASE-I IS 5'-EXONUCLEASE)))
(LIM-SYN-RULE (IF (THE REACTIVITY OF DNA-POLYMERASE-I IS ACTIVITY)
                  (OR (THE NUCLEOTIDES OF CONDITIONS IS dATP)
                      (THE NUCLEOTIDES OF CONDITIONS IS dTTP)
                      (THE NUCLEOTIDES OF CONDITIONS IS dGTP)
                      (THE NUCLEOTIDES OF CONDITIONS IS dCTP))
                   THEN DO (THE REACTIVITY OF DNA-POLYMERASE-I IS LIMITED-SYNTHESIS)))
(RIBO-INCORP-RULE (IF (THE DIVALENT-CATIONS OF CONDITIONS IS Mn)
                      (THE REACTIVITY OF DNA-POLYMERASE-I IS BINDING)
```

Figure 1. A few of the rules used to predict the activities of DNA polymerase I under a variety of environmental conditions. These rules can be used either to infer the activity of the enzyme as a function of the conditions (forward chaining of the rules) or to decide what conditions are necessary to obtain a certain activity (backward chaining from a conclusion to the required premises).

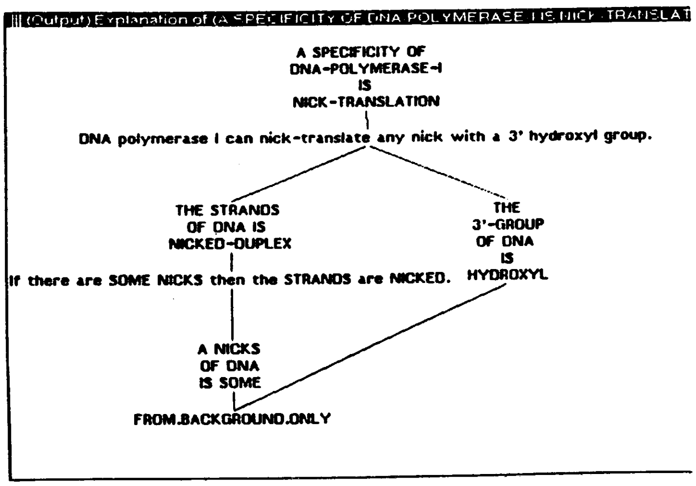

Figure 2. A graphic explanation of the conclusion that a particular DNA terminus is an effective primer terminus, based on the facts that the terminus is base paired and is a 3'-hydroxyl group. Terminus pairing was concluded from the user's statement that the DNA was nicked and the rule that nicked DNAs are usually assumed to contain paired termini.

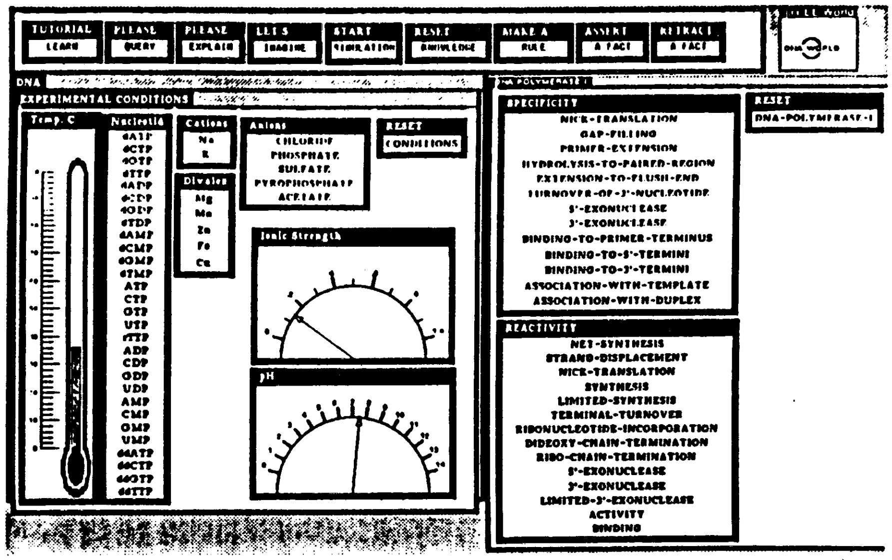

Figure 3. An image panel that reports the status of several environmental conditions to the user. These images are also active and the various values presented can be altered by pointing at the thermometer, dials and switches. Such graphic screens are the primary means of specifying facts to the knowledge system.

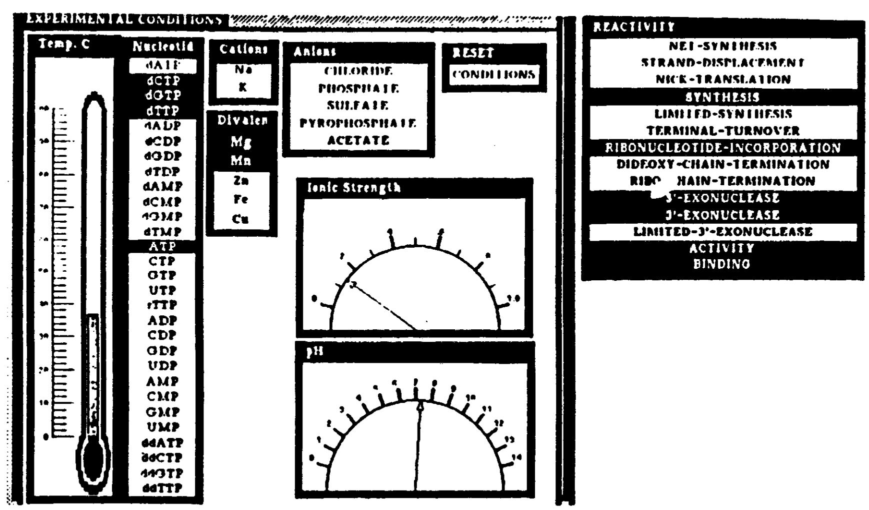

Figure 4. Conclusions about the specificity and reactivity of DNA polymerase I deduced by the knowledge system. Facts about the DNA structure and environmental conditions have been specified and the rules have automatically deduced the conditions shown with a black background.

There are several potential uses of such symbolic simulations. First, they can serve as an educational tool with which to build an interactive textbook. Unlike most other types of interactive texts, those based on symbolic simulations can explain their chain of reasoning. Descriptions of properties of enzymes can be graphically displayed for the reader. The simulation can serve as a model laboratory for students to carry out experiments with the enzyme under many experimental or physiological situations. Finally, the simulation can be used to predict or test the results to be obtained by actual experiments. The advantages of simulating an experiment prior to executing it are that unforeseen reactions and interactions can be brought to a scientist's attention.

Acknowledgements

This work was carried out using the KEE knowledge system designed by IntelliCorp, Inc., and made available under their University Grants Program. I would also like to thank both IntelliCorp and IntelliGenetics, Inc., for supporting my sabbatical during which this work was performed.