ELSEVIER Artificial Intelligence in Medicine 17 (1999) 303–321

Artificial Intelligence in Medicine

www.elsevier.com/locate/artmed

A qualitative process system for modeling NF-κB and AP-1 gene regulation in immune cell biology research

Robert B. Trelease [a,*], Richard A. Henderson [b], Jack B. Park [c]

[a] *Department of Pathology and Laboratory Medicine, UCLA School of Medicine, Los Angeles, CA 90095, USA*

[b] *School of Medicine, Wright State University, Dayton, OH 45323, USA*

[c] *ThinkAlong Software, Box 359, Brownsville, CA 95919, USA*

Received 1 July 1998; received in revised form 20 December 1998; accepted 28 February 1999

Abstract

An experiment-oriented integrated model of the regulation of the biologically ubiquitous NF-κB and AP-1 gene transcription promoters was built by extending a previously developed qualitative process system for simulating cell behavior in the immune system. The core knowledge base (KB) implemented a deep biological ontology including molecular, ultrastructural, cytological, histological, and organismic definitions. KB states, relationships, predicates, and heuristics also represented process interactions between reactive oxygen species, growth factors, and a variety of kinases phosphorylating intermediate molecules in the NF-κB and AP-1 regulatory signaling pathways. The system successfully simulated the molecular process steps underlying outcomes of eight different molecular genetics laboratory experiments, including those dealing with NF-κB and AP-1 regulation in immunodeficiency virus infection and tumor necrosis factor responses. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Qualitative modeling; Gene regulation; Immunology research

* Corresponding author. Tel.: +1-310-2069971; fax: +1-310-7947465.
E-mail address: trelease@ucla.edu (R.B. Trelease)

PII: S0933-3657(99)00021-4

1. Introduction

During the past two decades, there has been a vast expansion in biomedical research on and knowledge of the functions of regulatory genes, ranging from their involvement in processes of intercellular signal transduction to control of tissue regeneration and body form development. Of particular interest has been the study of ubiquitous protooncogenes that are involved in a range of fundamental cell regulatory processes, and that when mutated can effect the dysregulation of cellular proliferation characterizing various kinds of cancers [8]. Two different classes of such genes, for activation protein 1 (AP-1) and nuclear factor κ-B (NF-κB) transcription factors, are evolutionarily conserved and involved in a wide variety of cellular activation processes in many different classes of animals. In 1996, Sen and Packer [24] presented an integrative conceptual model of NF-κB and AP-1 transcriptional regulation that encompassed a wide array of findings from a number of different research studies. Their model identified a number of different process stages, from enzyme activation through promoter binding, at which reactive oxygen species (ROS) might interact with NF-κB and AP-1 activation mechanisms to modify the transcription of a variety of other cellular genes.

As reported, Sen and Packer's model is qualitative, with quantitative relationships unknown, undefined, or incompletely understood in many of the biological processes described. To produce a computer-based implementation of their conceptual model, it is thus most appropriate to employ qualitative reasoning (QR) methods. Although QR methods have been successfully applied to the solution of a variety of problems in physical sciences [27] and biomedical research domains [28,6], relatively few systems have modeled processes of gene regulation. Perhaps the most prominent have been Karp's GENSIM/HYPGENE discovery systems [14,15] which have been used to model the discovery of the tryptophan operon, and the EcoCYC project representing genetic mechanisms in the *E. coli* bacterium [16].

The purpose of this project was to implement an experiment-simulating qualitative modeling system that could represent AP-1 and NF-κB processes involved in cell regulatory mechanisms. The principal task involved developing monotonic extensions to incorporate into the knowledge base of a previously described system for modeling immune cell behaviors and infection-related cytokine functions [26]. The resulting integrated system was intended to simulate the reported specific molecular-level findings of published experiments on AP-1 and NF-κB functions in various aspects of cellular and humoral immunity. This project was conceived as the first component of a computer-based discovery system for working with molecular and cell biology research data.

2. Modeling system overview

The modeling system was developed and run on Apple Power Macintosh computers (Motorola Power PC Reduced Instruction Set processors running at 133–233 MHz) with 80–128 megabytes of random access memory. The related

knowledge bases, cell behavior rules, and experimental conditions were developed using The Scholar's Companion (TSC; ThinkAlong Software, Brownsville, CA) a symbolic programming environment composed of nested, extensible, multiple language interpreters/compilers, run-time utilities, and an envisionment builder.

Like other knowledge-based systems, TSC supports the definition of hierarchical sets of object frames representing actors (e.g. different types of cells), taxonomic, functional, and ontologic relationships, physical and physiological processes, and predicated interactions and symbolic behavioral process rules. Concepts can be qualitatively represented at multiple levels of abstraction (or 'granularity') [5] ranging from the molecular through cellular, tissue, organism, and populations levels. TSC also compiles functional experimental designs with statements of initial process conditions (states and relationships) for simulation trials to be run using defined actors (e.g. cells and constituent molecules). TSC's forward-chaining envisionment builder uses breadth-first knowledge base (KB) search and rule constraint satisfaction to simulate the process state changes and cell behavior interactions as they evolve from initial experimental conditions.

When a TSC KB is compiled, and prior to engaging the envisionment building process, rules and actor taxonomic relationships can be interactively queried via user input entered in a 'natural language' Conversation window. Rule taxonomies can be visualized through the 'Frame Browser' window graphics display.

When the envisionment builder is engaged after compilation, it begins evaluating the initial experimental conditions in terms of all of the process rules in the KB. Using a justify-assume-interpret cycle [18], it proceeds from initial through subsequent system conditions. Process rules evaluating to true ('firing') result in one or more sets of consequent system conditions (instantiated characteristic states of actors, predicates, and relationships). Each set of conditions, called an 'Episode' in TSC, is identified numerically and compared for identity with previous Episodes. When an identical Episode is encountered, expansion of the envisionment along that branch terminates as a way to reduce the intractable branching (combinatorial explosion) problem encountered in many QR environments [19]. The envisionment building cycle continues until there are no further rules that can fire given system conditions or until special stopping rules fire.

Each TSC experimental simulation produces three different types of envisionment output depicting the evolution of modeled processes state changes. Consistent with other qualitative simulations in physical sciences [27], all possible process outcomes are represented in the full envisionment for each experiment. Envisionment outputs include the Log file, Frame Browser graphics, and run-time Conversation window reports. An illustration of these envisionment elements is shown during a gene regulation experiment in Fig. 1, along with other TSC program function windows.

The Log file provides a permanent sequential text-based record of all conditions, process rule firings, and state transitions occurring while the envisionment builder operates during a given simulation experiment. Each specific set of experimental conditions resulting from process rule firing is listed as a numbered Episode, with the antecedent and consequent conditions and predicates shown. For example,

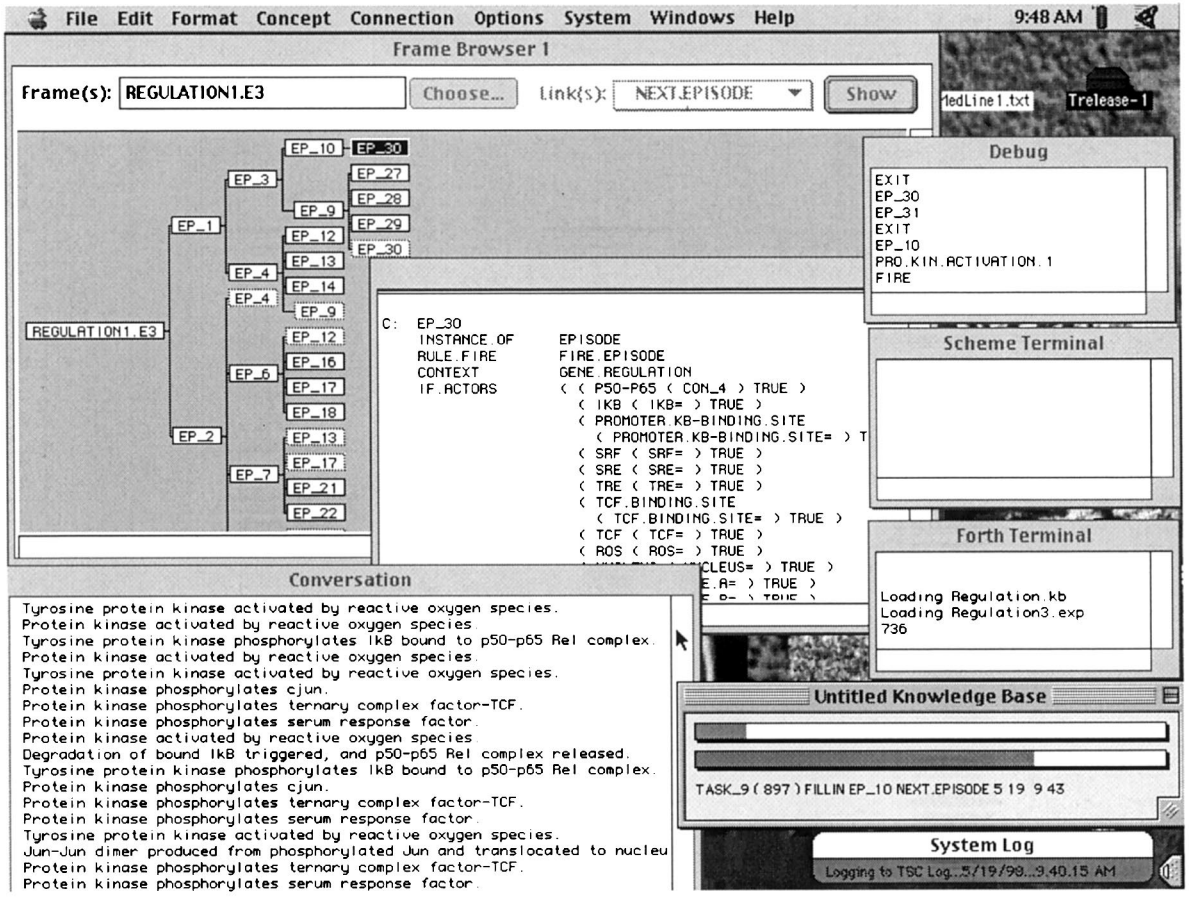

Fig. 1. Experiment runtime view of gene modelling system showing TSC conversation window with rule-firing reports (lower left), frame browser with graphic depiction of envisionment episode relationships (upper left), text expansion window with antecedent and consequent relationships for episode 30 (center), system debug, Scheme and Forth interface windows (upper right side), and envisionment memory/episode and logfile indicator windows (lower right).

Episode 1 may occur when the presence of ROS in initial conditions induces activation of a phosphorylating enzyme. Enzyme activation leads to Episodes 2 and 3, with phosphorylation of two different intermediate proteins.

The Frame Browser visualization graphically depicts the evolution of an envisionment's process states as a decision tree with specific Episodes shown as the nodes (branch points) of the tree structure. For a given experiment, the Browser thus generates a characteristic visual qualitative state diagram (see Figs. 4–8). During experiment runtime, individual Episodes shown in the Browser window can be expanded into a text window showing Episode antecedents and consequents (see Fig. 1, center).

The conversation (natural language) window produces a run-time display of cell process rules firing statements as the envisionment builder evaluates each Episode's conditions (Fig. 1, lower left).

2.1. Qualitative modeling system architecture

The general principles followed in constructing this modeling system were based on published principles of qualitative process (QP) theory [9] and qualitative simulation (QSIM) [18]. The formal modeling system KB was composed of four conceptual parts: (1) A core biological KB defining actors, taxonomic relationships, process definitions and predicates; (2) sets of behavioral rules for biologically active substances, cells, pathogens, and processes; (3) sets of behavioral rules for molecules, reactions, and components of NF-κB and AP-1 gene regulatory process pathways, and (4) sets of rules and conditions for the execution of experimental simulations. All rules and concepts (actors, predicates, etc.) were defined within TSC as frames in a Scheme dialect of the LISP programming language. Fig. 2 shows an example of process rule syntax.

In order to support incremental expansion of the KB and problem domains, the core model components were divided into a group of loadable source files. The base biological KB was composed of four incrementally compiled files. BIOPRIMITIVE.T included definitions of underlying physical, chemical, and biological actors (e.g. atoms, molecules, proteins, cells, organisms, endocytosis), organism taxonomy and inheritance (e.g. class, order, family, genus, species), and some basic predicate relationships between various actors. CELL.T included fundamental definitions of biological structures (e.g. nuclei, membranes, mitochondria, etc.) substances and predicates central to cell biology. IMMUNE.T was comprised of definitions of immune system cells, receptors, cytokines and other actors, substances, organs, and predicates. The specific heuristics for immune system cell processes, interactions, and behaviors were contained in IMMUNE.RBT. All concepts and process rules were derived from current theories and objective scientific information published in the current general biological [22], systematic [10], histological and cell biological [11,13], and specialized immunological [1,12] literatures.

The original regulatory substances taxonomy and ontology were greatly expanded by adding new actors, states, relations, and predicates to the pre-existing bioprimitive and immune system KB files as described above. Furthermore, process

rule sets were created for representing ROS-related interactions with cell enzymes and other regulatory process steps involved in AP-1 and NF-κB mediated transcription of several generic promoter/enhancer genes. The new gene regulation process rules were contained in an additional source file, REGULATION.T, which was compiled after the BIOPRIMITIVE.T, CELL.T, and IMMUNE.T modules prior to execution.

For each simulated experiment, an individual EXPERIMENT code file (e.g. REGULATION3.EXP) defined initial experimental conditions and setup an envisionment building cycle. To run a simulation, a specific.EXP file was loaded (compiled) by TSC, and the envisionment builder was engaged. Of the group of simulations reported here, initial experiments were designed to demonstrate the fundamental state transitions of prototypical NF-κB and AP-1 gene regulatory processes in the presence and absence of different factors (ROS and Rel regulatory proteins). Other experiments were designed to model processes demonstrated in a group of specific published research studies of these gene promoters in specific cell

```
c:    MAP.kinase.phosphoryl
      level            basic
      sub.of           phys.process
      instance.of      process.rule
      my.creator       rbt
      contextGene.transcription.regulation
      if.actors        ( ( IkB-p50-p65 ( *IkB-p50-p65 )  true )
                         ( IkB ( *IkB  ) true )
                         ( MAP.kinase ( *MAP.kinase ) true ) )
      if.states        ( ( activated ( *MAP.kinase) true ) )
      if.not.states    ( ( phosphorylated ( *IkB ) true ) )
      then.states      ( ( phosphorylated ( *IkB ) true ) )
      then.relates     ( ( phosphorylates ( *MAP.kinase *IkB) true ) )
      then.say         " MAP kinase phosphorylates IkB bound to p50-p65 Rel
                       complex."

c:    prot.kin.C.IkB.phosphoryl
      level            basic
      sub.of           phys.process
      instance.of      process.rule
      my.creator       rbt
      contextGene.transcription.regulation
      if.actors        ( ( IkB-p50-p65 ( *IkB-p50-p65 ) true )
                         ( IkB ( *IkB ) true )
                         ( p50-p65 ( *p50-p65 ) true )
                         ( protein.kinase.C ( *protein.kinase.C ) true )
                         ( T-cell ( *T-cell ) true )
                         ( HIV ( *HIV ) true ) )
      if.states        ( ( activated ( *protein.kinase.C ) true ) )
      if.relates       ( ( infects ( *HIV  *T-cell ) true ) )
      if.not.states    ( ( phosphorylated ( *IkB ) true ) )
      then.states      ( ( phosphorylated ( *IkB ) true ) )
      then.relates     ( ( phosphorylates ( *protein.kinase.C *IkB) true ) )
      then.say         " Protein kinase C phosphorylates IkB bound to p50-p65 Rel
                       complex. "
```

Fig. 2. Examples from TSC gene regulation KB illustrating process rule syntax.

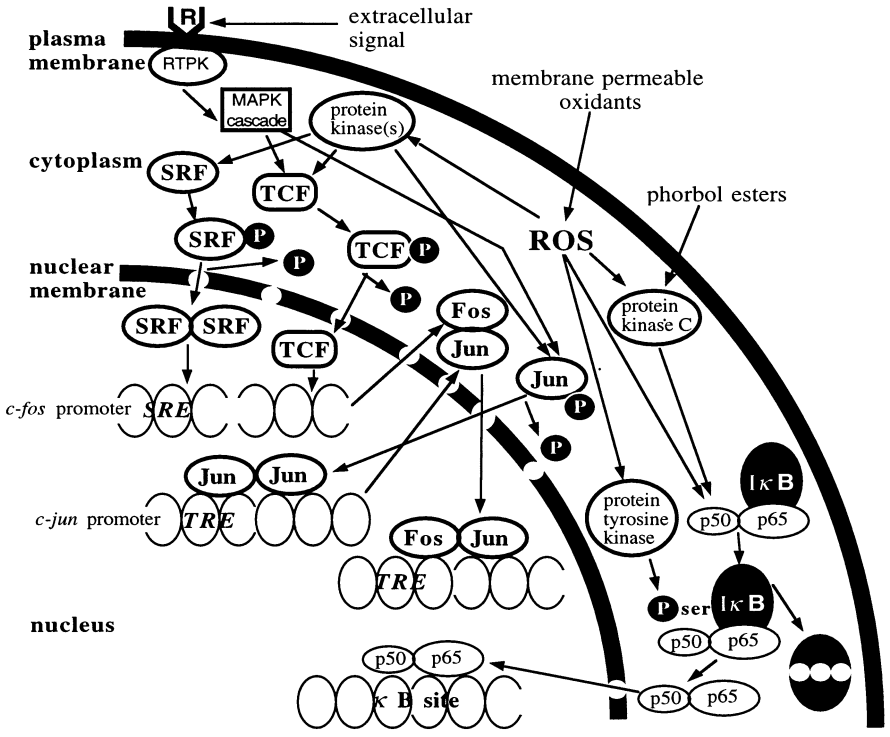

Fig. 3. Graphic depiction of the hypothetical scheme for steps in cellular NF-κB and AP-1 regulatory gene (protooncogene) activation that may be influenced by oxidants (ROS), antioxidants, and intracellular signaling molecules (e.g. growth factors) (modified and redrawn after Sen and Packer, [24]).

types under conditions of viral infection, cytokine signaling, and growth factor stimulation. These experiments are summarized in Section 3 below.

2.2. Conceptual model of NF-κB and AP-1 gene regulation

NF-κB and AP-1 represent two different classes of well-defined transcription factors that are implicated in the inducible expression of a variety of genes involved in intercellular signaling and oxidative stress mechanisms. NF-κB and AP-1 activation processes included in Sen and Packer's model [24] are depicted schematically in Fig. 3.

NF-κB is a member of the Rel family of proteins that are normally present in cell cytosol in inactive forms. There are two groups in the family, classified on the basis of their structures, functions, and mode of synthesis. Group I includes p50 and p52 proteins. Group II includes p65, Rel or c-Rel, RelB, and *Drosophila* differentiation proteins, dorsal and Dif. NF-κB/Rel proteins form hetero- or homodimers that are functionally required for DNA binding and transcriptional activation of certain genes. Inactive Rel protein complexes are found in two different forms in the

cytosol: (1) a Rel hetero- or homodimer bound to a member of the IκB inhibitory proteins, and (2) a complex between a Rel Group II protein and a Group I precursor protein. NF-κB/Rel activation is brought about by a variety of stimulus mechanisms (e.g. ROS oxidation or viral infection) inducing enzymatic phosphorylation of the IκB protein, with subsequent cleavage of IκB. Active NF-kB (Rel) dimers are then rapidly translocated to the nucleus, where they can bind to κB transcription sites in the promoter and enhancer regions of a variety of gene families, including those encoding cytokines, cytokine receptors, cell adhesion molecules, growth factors, and latent viral genes. NF-κB is thus crucial to many aspects of immunity and inflammation [17].

AP-1 represents a class of transcriptional activator proteins closely associated with tumor promotion. It includes Fos (cFos) and Jun (cJun) families, proteins that are translational products of the c-Fos and c-Jun protooncogenes. Fos and Jun proteins form hetero- and homodimers that are active in DNA binding and initiation of gene activation. AP-1 dimers exert their effects by binding to 12-*O*-tetradecanoylphorbol 13-acetate (TPA) response element (TRE) portions of cellular or viral genes. Oxidative or other stimuli initiate activation processes by inducing enzymatic phosphorylation of available Jun, with subsequent homodimerization and binding of JunJun to the TRE site on the c-Jun promoter gene. This leads to the production of additional Jun protein. Fos is produced by a more complex mechanism involving ROS-induced enzymatic phosphorylation of serum response factor (SRF) and ternary complex factor (TCF) proteins. SRF binds to a serum response element (SRE) located in the c-Fos promoter gene. TCF binds to a second site adjacent to SRE, and both SRF and TCF binding are needed for promotion of c-Fos with subsequent production of the cFos monomer. FosJun heterodimer binds to the TRE site in a variety of cellular genes and initiates gene transcription. AP-1 gene regulation may thus involve promotion of c-Fos and c-Jun as part of the process activating a given target gene.

Phosphorylation of NF-κB and AP-1 transcription factor components can be catalyzed by a variety of cellular enzymes, including protein tyrosine kinase (or 'tyrosine kinase'), protein kinase C, and receptor-associated protein kinases involved with extracellular signal transduction. Receptor associated protein kinases may initiate 'kinase cascade' reactions, with several intermediate enzymes phosphorylating and activating subsequent kinases (such as mitogen-activated protein kinase; MAPK). Such a chain reaction may lead to the eventual phosphorylation of IκB, Jun, and Fos proteins. Initial kinase activation can be mediated by ROS oxidation, phorbol esters, and receptor binding of regulatory substances such as growth factors and cytokines. The fundamental Sen and Packer conceptual model was expanded by adding explicit process rules to represent known MAPK chain reaction steps coupling growth factor receptor binding to promoter activation [4,7]. Activated MAPK can thus play the same role as protein kinase or tyrosine kinase in initiating AP-1 and/or NF-κB activating processes.

3. Simulation experiments

We designed a variety of different stimulation experiments for testing the KB and its gene regulation process models. In accordance with specific published laboratory research designs, individual sets of initial conditions were defined, stored, and individually compiled in separate TSC EXPERIMENT files. For example, virus-infected T-cells, hydrogen peroxide, enzymes, and cJun were defined as actors in the initial conditions of one simulation of proviral gene activation (experiment 6). The most important elements of all the experiments are summarized in Table 1 and further described below.

Experiments 1–4 were intended to demonstrate the prototypical (general rule class) process features of the Sen and Packer model [24]. Experiments 1–2 concerned themselves with activation of c-Fos, c-Jun, and prototype gene B via AP-1 mechanisms in response to ROS-induced activation of the protein kinase enzyme.

Table 1
Experiments simulated with the current TSC immunology/gene regulation knowledge base

Reg exp	Prime actors	Genes activated	Episodes	Scientific refs
E1	ROS, SRF, TCF protein kinase	c-fos	20	Sen and Packer (1996) [24]
E2	Ros, cjun, SRF, TCF, protein kinase	c-jun, c-fos, gene B	139	Sen and Packer (1996) [24]
E3	ROS, cjun, NF-κB, SRF, TCF, tyrosine and protein kinase	c-jun, c-fos, gene A, gene B	1392	Sen and Packer (1996) [24]
E4	ROS, NF-κB	Gene A	6	Sen and Packer (1996) [24]
E6	T-cell, HIV, ROS, H_2O_2	HIV proviral gene	6	Sappey et al. (1995) [23]
E6a	T-cell, HIV, DFO, low H_2O_2	None, protein, kinase C inhibited	2	Sappey et al. (1995) [23]
E6b	T-cell, HIV, no DFO	Proviral	7	Sappey et al. (1995) [23]
E7	Chondrocytes, TNF-α	c-fos	21	Lo et al. (1995) [21]
E8	Growth hormone, Shc, tyrosine, kinase, cRas, cRaf, MAPK, MAPK kinase	–	6	Cooper (1995) [8]Cobb and Goldsmith (1995) [7]Blenis (1993) [4]
E9	Same as E8, cjun, c-fos	c-fos	130	Cooper (1995) [8]Cobb and Goldsmith (1995) [7] Blenis (1993) [4]
E10	Endothelial cell, TF, TNF-α, NF-κB, AP-1 (× 2)	Tissue factor (TF)	2854	Bierhaus et al. (1995) [3]
E11	IL-1, IL-β, TNF-α	TNF-α	24	Lieb et al. (1996) [20]

Experiment 4 was designed to demonstrate the transcription promotion of prototype gene A via an NF-κB mechanism evoked by ROS activation of the enzyme tyrosine kinase. Experiment 3 was intended to simulate the combined features of activation of both NF-κB and AP-1 gene promoter mechanisms via ROS activation of both of the aforementioned kinase enzymes.

Induction of specific immune responses in human immunodeficiency virus (HIV) infection stimulates expression of proviral genes incorporated in the genome of affected cells. This mechanism is consistent with the ROS/oncogene regulation model of Sen and Packer. Sappey et al. [23] performed experiments in which hydrogen peroxide (H_2O_2; evolved during acute anti-infectious immune responses) induced proviral gene expression in HIV infected cultured T-cells via NF-κB protooncogene promotion. In one experimental group, the free radical scavenger deferoxamine (DFO) prevented HIV gene expression by eliminating the activating ROS (hydroxide radical) typically produced by H_2O_2. Experiments E6–E6B simulated the basic mechanisms of this proviral gene expression/suppression and the effects of H_2O_2 in the presence or absence of DFO.

Stimulation of AP-1 and NF-κB gene promotion may be evoked by different MAPK-dependent mechanisms in transduction of extracellular signals important to general immune responses, cell regulation by growth factors and cytokines, expression of CAMs, and neural transmission [4,7,8]. Again, these mechanisms are highly evolutionarily conserved, and specific MAPK-using signaling pathways are involved in many different types of cellular processes in organisms ranging from the unicellular through humans. Based on reaction processes summarized by Blenis, Cobb and Goldsmith [4,7], experiment 8 was designed to demonstrate the fundamental MAPK 'kinase cascade' reactions evoked by growth hormone (GH). Experiment 9 coupled these GH-induced MAPK processes to the production of AP-1 promoter.

A group of simulations focused on different aspects of the functionality of tumor necrosis factor α (TNF-α), a cytokine involved in a range of responses to a variety of pathogens, cancer cells, and inflammation. Lo et al. [21] demonstrated that TNF-α evokes ROS production in chondrocytes (cartilage cells) with subsequent AP-1 upregulation in inflammation (related to arthritis). This was simulated in experiment E7. The cytokine IL-1 has been shown to evoke expression of the TNF-α gene via an AP-1 mechanism during immune responses to infection [20]. This was simulated in experiment E11. TNF-α has also been shown to induce the release of the clotting mediator Tissue Factor (TF) from endothelial (blood vessel lining) cells via a mechanism involving a single NF-κB and dual AP-1 enhancer sites on the TF promoter gene [2]. In contrast to other mechanisms simulated in this project, TF regulation thus involved the functioning of three interacting gene promoter pathways, including separate, selective JunJun and JunFos AP-1 sites. The basic TNF-α-TF-promoter mechanism was represented in experiment E10.

4. Experimental observations and results

The full series of envisionments demonstrated that the modeling system KB was

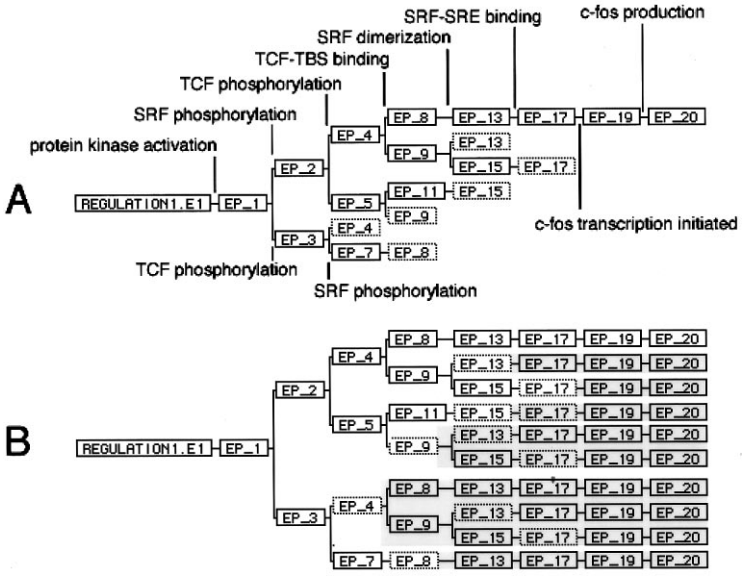

Fig. 4. (A) Annotated frame browser graphic depiction of episode relationships in the envisionment of experiment E1: promotion of transcription of the c-Fos gene via activation of protein kinase by ROS. Dotted boxes indicate episodes identical to those previously encountered in envisionment building, i.e. having the same consequent states and subsequent episodes. All branches thus converge toward episode 20, initiation of cFos production. (B) Modified graphic depiction of the envisionment of experiment E1, showing the effect of not terminating expansion on identical (repeated) episodes. Note that all terminal episodes are identical to episode 20 and there are 54 episodes depicted.

capable of qualitatively representing a range of different characteristic gene regulation processes and state transitions dependent on defined experimental initial conditions.

The general class behavior experiment envisionments (E1–E4) demonstrated the fundamental transcriptional promoter processes known to occur with ROS activation of phosphorylation of intermediate molecules (e.g. TCF and SRF) by kinase enzymes. Depending on initial conditions, such as the absence of a Rel promoter precursor molecule or the full complement of necessary agents, the model would produce envisionments showing partial activation (Figs. 4A and 5) or activation of a full aggregate of the appropriate gene promoter systems (Fig. 6). E1 and E2 demonstrated the ROS-evoked induction of fundamental AP-1 processes, including promotion of c-Jun and c-Fos gene transcription. E4 showed the relatively simple set of process state transitions characteristic of NF-κB transcription promotion. E3, in particular, generated a very large envisionment ($\sim$ 1395 episodes) showing the combinatorial explosion of different states encountered when considering all the processes evoked by two activating enzymes, AP-1 and NF-κB and their four related gene systems. E1–E4 thus fully implemented the general rule class behavioral features of the fundamental conceptual model of Sen and Packer [24] at their chosen level of abstraction.

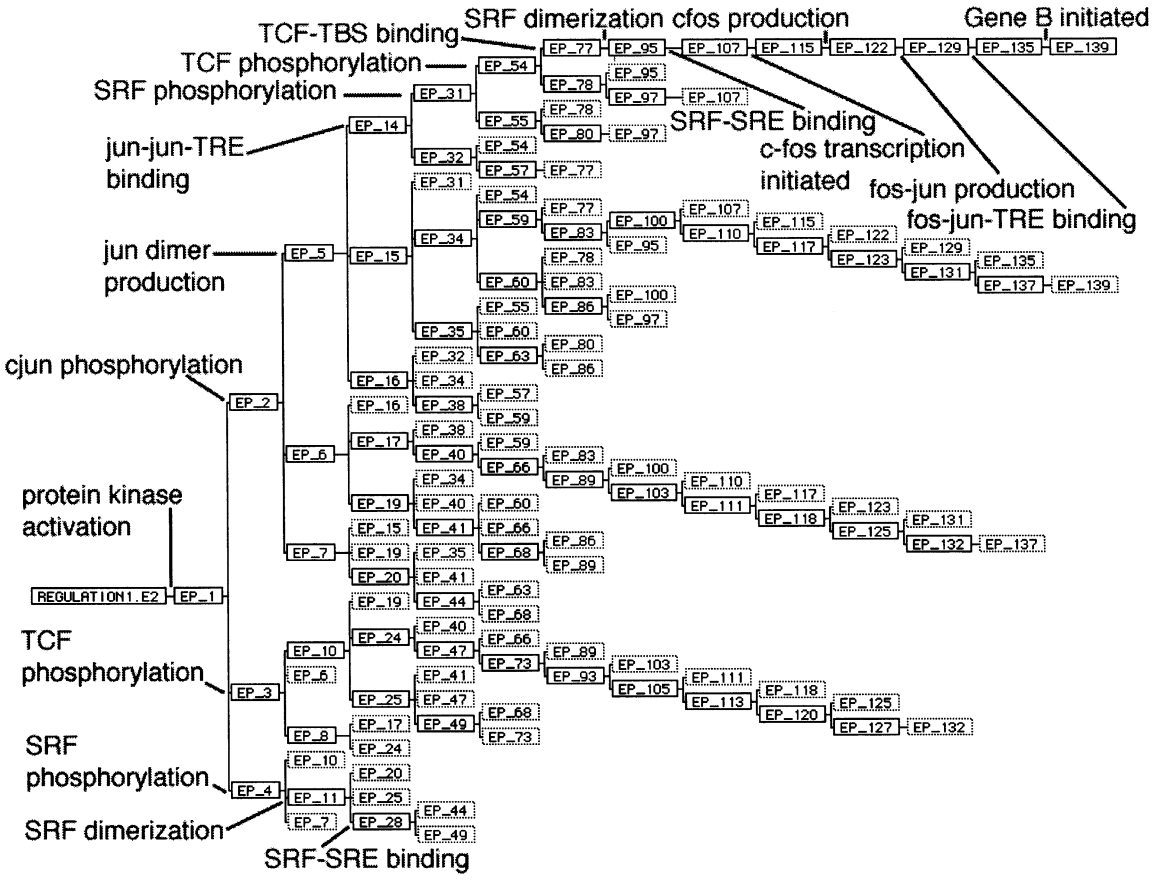

Fig. 5. Annotated frame browser graphic depiction of the envisionment of experiment E2: presence of ROS initiates protein kinase phosphorylation of available cjun promoter protein, which leads to AP-1 mediated promotion of transcription of the prototype gene B.

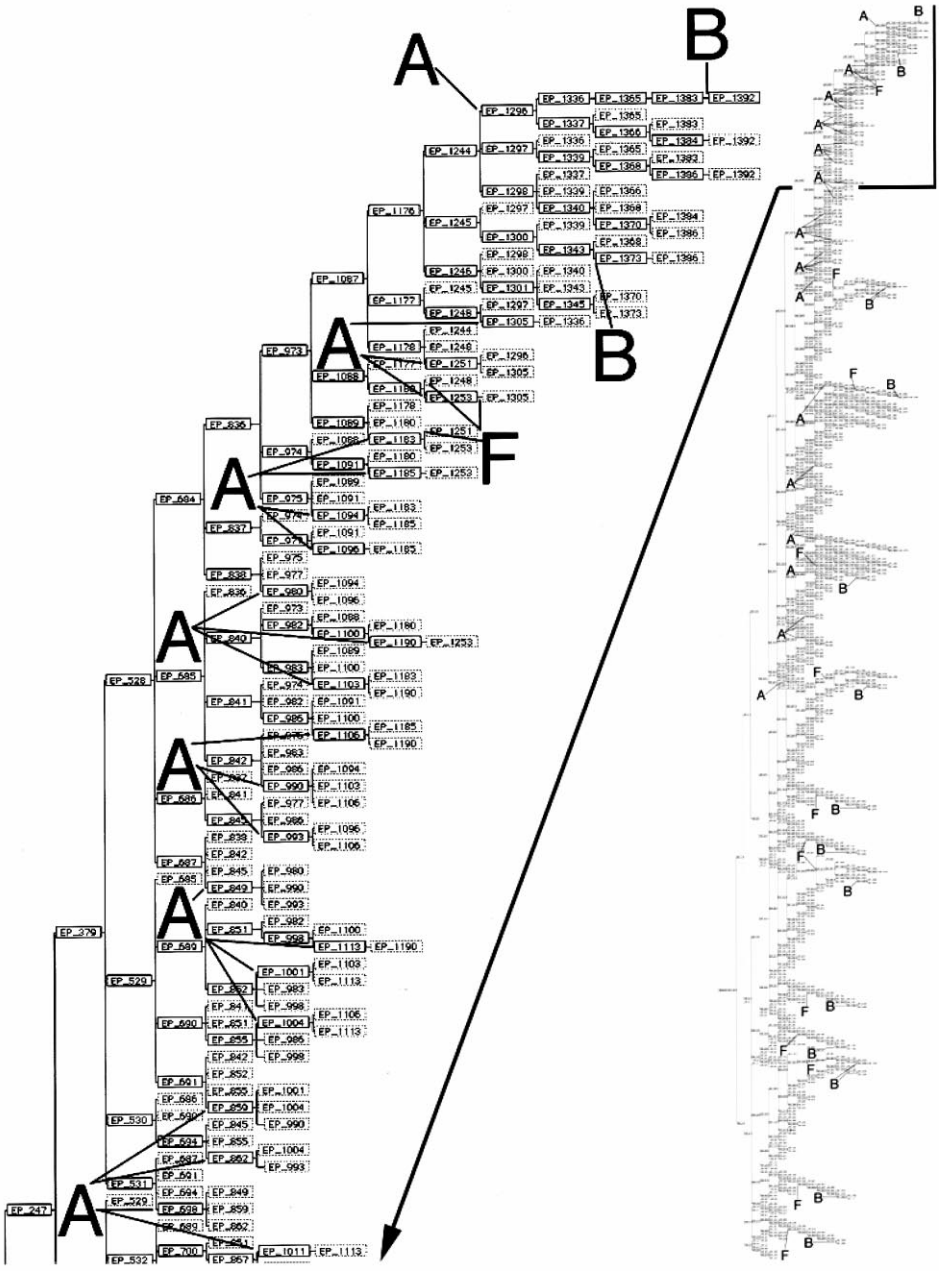

Fig. 6. Annotated frame browser graphic depictions of the envisionment of experiment E3: promotion of transcription of c-Fos, genes A, and B following ROS activation of protein kinase and tyrosine protein kinase enzymes. Right image, reduced view of entire 1392 episode envisionment 'tree'. Left image, terminal portion of envisionment expansion leading to gene B transcription, with lines at right indicating the portion represented. A, B, and F indicate episodes with gene A, B or c-Fos transcription promotion.

The first two simple envisionment graphs showed quite plainly the effect of the 'dotted Episode' method for redundant branch-point reduction as implemented by TSC. Assuming that identical antecedent conditions of given Episodes would produce identical expansions (Episode 'branching') of consequent conditions, termination of expansion from repeated identical ('dotted') Episodes greatly simplified the complexity of envisionments. This is graphically illustrated for E1 in Fig. 4A, B.

The E6 series of envisionments simulated the experimentally demonstrated effects of immune stimulation on expression of HIV proviral genes. In general, ROS elicited proviral gene promotion via a tyrosine kinase mediated production of NF-κB transcription promoter (Fig. 7B; E6). Administration of H_2O_2 would produce hydroxyl radical (a reactive oxygen species) that stimulated tyrosine kinase to initiate the NF-κB production mechanism leading to expression of HIV proviral genes (Fig. 7D; E6B). Presence of the DFO prevented the production of hydroxyl radical, and thus inhibited the activation of tyrosine kinase that would have evoked NF-κB-promoted proviral gene activation (yielding the abbreviated envisionment, E6A, shown in Fig. 7C).

In experiment E7, presence of TNF-α caused chondrocyte production of H_2O_2, inducing activation of protein kinase and leading to the production of cFos via AP-1-mediated activation of the c-Fos promoter gene (Fig. 7E). This represented a specific instantiation of the general AP-1 class behavior typified by E1 (Fig. 4).

In experiment E8, the envisionment demonstrated the main processes coupling membrane receptor activation with a multiple kinase enzyme reaction chain leading to the phosphorylation of MAPK (Fig. 7F). The stimulatory molecule was GH, representing a class of growth factor known to bind to receptor-associated enzyme initiating the kinase chain reaction. Experiment E9 further showed the coupling of GH-induced MAPK activation with the promotion of the c-Fos gene and the production of AP-1 transcription promoter (Fig. 8).

Process rules for experiment E10 successfully simulated the experimentally identified state transitions involved in TNF-α-stimulated upregulation of complex NF-κB/dual-AP-1 regions on the TF promoter gene. Like E3, E10 generated a very large envisionment (Table 1) reflecting the numerous different process states and transitions possible with the transcription of several different genes initiated by MAPK mechanisms. Due to the size of the envisionment, it was technically unfeasible to represent E10 in a Browser graphic Figure for this report.

Experiment E11 successully simulated AP-1 promoted transcription of the TNF-α gene evoked by IL-1-mediated activation of protein kinase c. Like E7, this represented another specific instantiation of the general AP-1 class behavior typified by E1 (Fig. 7G).

In all experiments, each envisionment converged toward a single final state (e.g. c-Fos, c-Jun, and AP-1 activation in E2). No process (rule firing) cycles were observed.

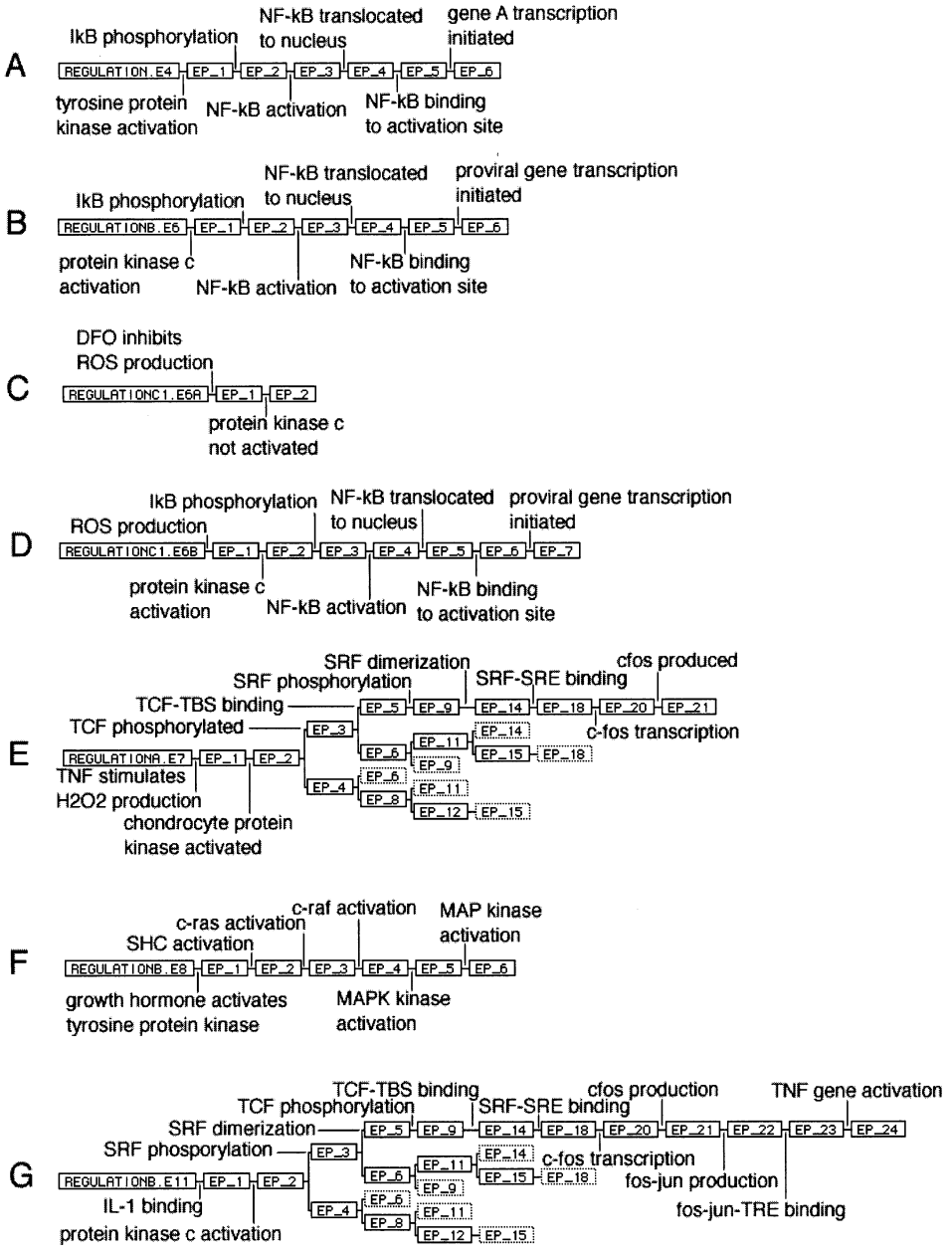

Fig. 7. Graphic depictions for seven gene regulation experiments yielding envisionments with relatively few episodes. (A) Browser graphic depiction of the envisionment of experiment E4. (B) Experiment E6: Induction of HIV proviral gene expresion by ROS species stimulation of protooncogene systems. (C) E6A; prevention of HIV proviral expression by DFO—a free radical scavenger that eliminates conversion of H_2O_2. (D) E6B: activation of AP-1 production pathway by H_2O_2 evolution of hydroxyl radical, leading to promotion of HIV proviral gene expression. (E) Experiment E7: TNF-α gene upregulation. (F) Experiment E8: initiation of MAPK cascade. (G) E11: IL-1-β receptor binding leads to AP-1 activation of the TNF-α promoter gene.

5. Discussion

This modeling system was able to simulate the most important features of fundamental NF-κB and AP-1 gene transcription promoter processes known to result from oxidant (ROS) activation of cellular phosphorylation enzymes. Depending on initial conditions, such as the presence of a promoter protein or a full complement of necessary agents, the model would produce envisionments showing partial or full activation of a set of interacting genes. The essential conceptual model of Sen and Packer was, thus, successfully implemented on a computer at a level of abstraction appropriate for exploring generalized hypotheses about cellular regulatory processes and biological signal transduction.

Beyond this, the process model system integrated additional intermediate reaction details of the MAPK kinase cascades coupling extracellular hormonal signaling and AP-1-mediated gene promotion [4,7,8]. Specific simulations also successfully

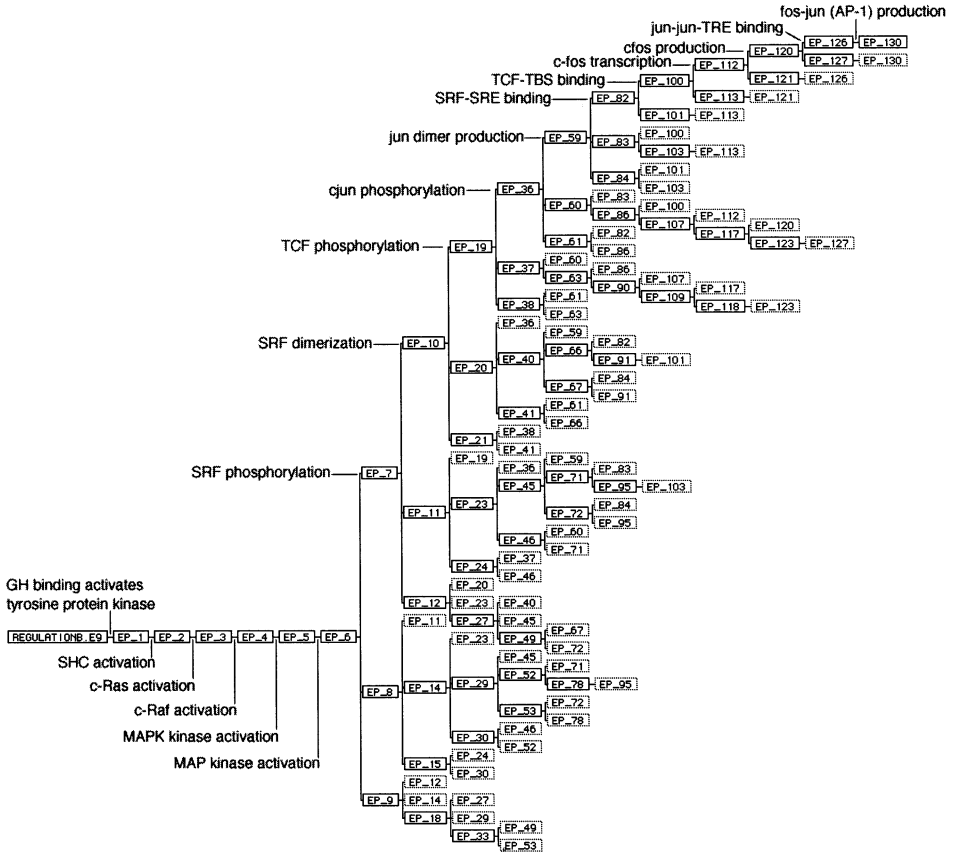

Fig. 8. Graphic depiction of the envisionment of experiment E9; initiation of MAPK kinase cascade by growth hormone, leading to AP-1 promoter production.

represented the underlying processes reported in the results of a group of published immunogenetics research experiments. In particular, they characterized the AP-1 and NF-κB gene regulatory processes mediating TNF-α-related cellular responses, including TNF-α receptor stimulation, ROS- and IL-1-stimulated upregulation of the TNF-α promoter gene. These latter simulation experiments took advantage of pre-existing (IMMUNE.T) immunobiology content in the KB, demonstrating the extensibility of the original qualitative modeling system design. This QR system thus incorporated new process content representing a deeper level of biological abstraction, in order to evolve from simple primary simulations of cell behavior to those encompassing underlying biochemical reactions involved in gene regulation.

For biomedical research, this work is significant because it demonstrates the practicality of using QR methods for modeling complex biological processes that are virtually intractable to representation by conventional computer-based mathematical approaches. These qualitative simulations can represent experimental research outcomes and underlying processes studied in contemporary molecular genetics research. This type of formalized genetic process modeling has proven to be unique, and it may become a useful experimental tool for understanding gene function and regulation during an epoch of biomedical science dominated by the Human Genome Project and widespread molecular genetics research.

There are, however, reasonable limitations to the simulations created by this system. Experiments involving multiple parallel processes can generate very large envisionments, representing combinatorial explosions of all the possible qualitative state transitions. Envisionments larger than 5000 Episodes may exceed the processing and storage capabilities of current computer systems. The largest simulations can take weeks to complete, even with the current generation of high-speed RISC processors (>200 Mhz) and large amounts of program random access memory (>128 megabytes). Analysis of large, complex envisionments can also take a good deal of time. Although such a time course may still be much shorter than that of a series of real biological laboratory experiments, practical use of such simulations for applied research will require judicious application of working scientific principles and knowledge engineering methods.

Beyond providing a tool and methodology for formal representation and testing of crucial hypotheses in contemporary molecular genetics and biological sciences, this sets the stage for development of more advanced computational tools for advancing scientific discovery through research experiments. In particular, given the global approach taken to creating process heuristics, it should be possible to build intelligent systems that study data from laboratory experiments, apply model-based reasoning to these data, and induce new model process rules. In fact, such a system based on TSC has been successfully created for studying polymer curing in materials sciences [2].

This last aspect of scientific discovery process automation has important implications for the more generalized application of systems for basic experimental biology research. Typically, scientists formulate hypotheses based on cognitive (mental) models of systems under study, then they devise methods and experimental procedures to test those conjectures. If comprehensive qualitative reasoning environ-

ments are used for formalized modeling of those systems and hypotheses, experimental designs can be objectively tested and refined independently of, but in conjunction with the performance of physical experiments [25]. One consequence of such use could be experiment optimization, a valuable asset in an era of limited research funding. If the chosen experimental methodology is amenable to computer-based process control, it may be possible to 'close the loop' in creating an automated discovery system that can iteratively evaluate hypotheses against experimental data, create new models/hypotheses, and implement new evaluative experiments. Ultimately, such computer-based discovery systems may be used to aid in the development of new molecular methods for treating resistant infections, enhancing wound healing, and inhibiting neoplastic cell proliferation.

Current work with this modeling system and future plans are proceeding along two lines of development. The first effort involves extending the core biological simulation system with encoder functions and heuristics to allow loading of laboratory experiment spreadsheet data for purposes of inference based on KB models. This is another step in the evolution of an experimental biology discovery system. The second endeavor entails transporting the generalized biological qualitative modeling system to a more widely accessible software environment, with potential for running simulations via the World Wide Web. Initial gene regulation experiments have been successfully run with a converted biology KB implemented using CLIPS (C Language Integrated Production System) and its Java-based subset, JESS (Java Expert System Shell). Ultimately, it is hoped that this biological simulation technology may be freely shared with other researchers by way of the Internet.

Acknowledgements

This work was supported in part by contract F49620-93-C0063 from the Air Force Office of Scientific Research. Research assistance was provided by Eve Isham. We thank Dr Steve LeClair for his support and helpful suggestions.

References

[1] Abbas AK, Lichtman AH, Pober JS. Cellular and Molecular Immunology. Philadelphia: W.B. Saunders, 1991.
[2] Abrams FL. Process discovery: automated process development for the contol of polymer curing, Ph.D. Dissertation, School of Engineering. Dayton, OH: University of Dayton, 1995
[3] Bierhaus A, Zhang Y, Deng Y, et al. Mechanism of the tumor necrosis factor α-mediated induction of endothelial tissue factor. J Biol Chem 1995;270:26419–32.
[4] Blenis J. Signal transduction via the MAP kinases: proceed at your own RSK. Proc Nat Acad Sci USA 1993;90:5889–92.
[5] Bredeweg B. Introducing meta-levels to qualitative reasoning. Appl Artif Intel 1989;3:85–100.
[6] Bylander T, Smith JW, Svirbely JJ. Qualitative representation of behavior in the medical domain. Comput Biomed Res 1988;21:367–80.
[7] Cobb MH, Goldsmith EJ. How MAP kinases are regulated. J Biol Chem 1995;270:14842–6.

[8] Cooper GM. Oncogenes, 2nd Edition. Boston: Jones and Bartlett, 1995.
[9] Forbus KD. Qualitative process theory. Artif Intel 1984;24(1984):85–168.
[10] Fortuner R, editor. 1993. Advances in Computer Methods for Systematic Biology. Baltimore: The Johns Hopkins University. Press.
[11] Goodman SR. Medical Cell Biology. Philadelphia: J.B. Lippincott, 1994.
[12] Janeway CA, Travers P. Immunobiology: the Immune System in Health and Disease, 2nd Edition. New York: Garland, 1996.
[13] Junqueira LC, Carneiro J, Kelly RO. Basic Histology, 7th Edition. Norwalk: Appleton and Lange, 1992.
[14] Karp PD. Frame representation and relational databases: alternative information-management technologies for systematic biology. In: Fortuner R, editor. Advances in Computer Methods for Systematic Biology. Baltimore: Johns Hopkins University Press, 1993:275–85.
[15] Karp PD. Hypothesis formation as design. In: Shrager J, Langley P, editors. Computational Models of Scientific Discovery and Theory Formation. San Mateo: Morgan Kaufman, 1990:276–317.
[16] Karp PD, Riley M, Paley SM, Pellegrini-Toole A, Krummenacker M. EcoCYC: encylopedia of *Eschericia coli* genes and metabolism. Nucleic Acids Res 1998;26:50–3.
[17] Kopp EB, Ghosh S. NF-κB and Rel proteins in innate immunity. Adv Immunol 1995;58:1–27.
[18] Kuipers B. Qualitative simulation. Artif Intel 1986;29:289–388.
[19] Kuipers B, Chiu C. Taming intractable branching in qualitative simulation, Proc. IJCAI-87 1987: 1079–1085.
[20] Lieb K, Kaltschmidt C, Kaltschmidt B, et al. Interleukin-1β uses common and distinct pathways for induction of the Interleukin-6 and tumor necrosis factor α genes in the human astrocytoma cell line U373. J Neurochem 1996;66:1496–503.
[21] Lo YYC, Cruz TF. Involvement of reactive oxygen species in cytokine and growth factor induction of c-fos expression in chondrocytes. J Biol Chem 1995;270:11727–30.
[22] Purves WK, Orians GH, Heller HC. Life: the Science of Biology. Sunderland: Sinauer, 1995.
[23] Sappey C, Boelaert JR, Legrand-Poels S, Forceille C, Favier A, Piette J. Iron chelation decreases NF-κB and HIV Type 1 activation due to oxidative stress. AIDS Res Human Retrovirus 1995;11:1049–61.
[24] Sen CK, Packer L. Antioxidant and redox regulation of gene transcription. FASEB J 1996;10:709–20.
[25] Shrager J, Langley P. Computational approaches to scientific discovery. In: Shrager J, Langley P, editors. Computational Models of Scientific Discovery and Theory Formation. San Mateo: Morgan Kaufman, 1990:1–25.
[26] Trelease RB, Park J. Qualitative process modeling of cell–cell-pathogen interactions in the immune system. Comput Methods Programs Biomed 1996;51:171–81.
[27] Weld DS, De Kleer J, editors. Readings in Qualitative Reasoning About Physical Systems. San Mateo: Morgan Kaufmann, 1992.
[28] Uckun S. Model-based reasoning in biomedicine. Crit Rev Biomed Eng 1992;19:261–92.