SENEX; CLOS/CLIM IN BIOMEDICAL EDUCATION

Sheldon S. Ball
Dept. of Pathology
University of Mississippi
2500 North State Street
Jackson, MS 39216
ssb@fiona.umsmed.edu

Vei H. Mah
Dept. of Neurology
Thomas Jefferson University
130 South 9th Street, Suite 400
Philadelphia, PA 19107

ABSTRACT

SENEX is a computer application in its fifth year of development focused on representation of molecular information, presentation of data, and reasoning with molecular information. It is written in a portable programming environment supported by Common Lisp, the Common Lisp Object System (CLOS), and the Common Lisp Interface Manager (CLIM). SENEX contains information about molecules, molecular events and disease processes, and provides tools for reasoning with and displaying this information in useful ways. This paper describes the architecture of SENEX and how that architecture facilitates development of a set a computer tools for education and research in molecular pathology. Several examples of how an individual might currently use SENEX to explore information related to the molecular basis of Alzheimer's disease are illustrated.

INTRODUCTION

Pathology is the study of disease. The study of disease begins with understanding of normal processes of biochemistry and physiology, for it is aberrations in these processes that manifest as disease. Molecular neuropathology deals with molecular aspects of disease. It is a discipline characterized by structures of variable complexity, events constrained by a variable number of factors, and incompletely understood phenomena. SENEX is a computer application in the domain of molecular pathology.

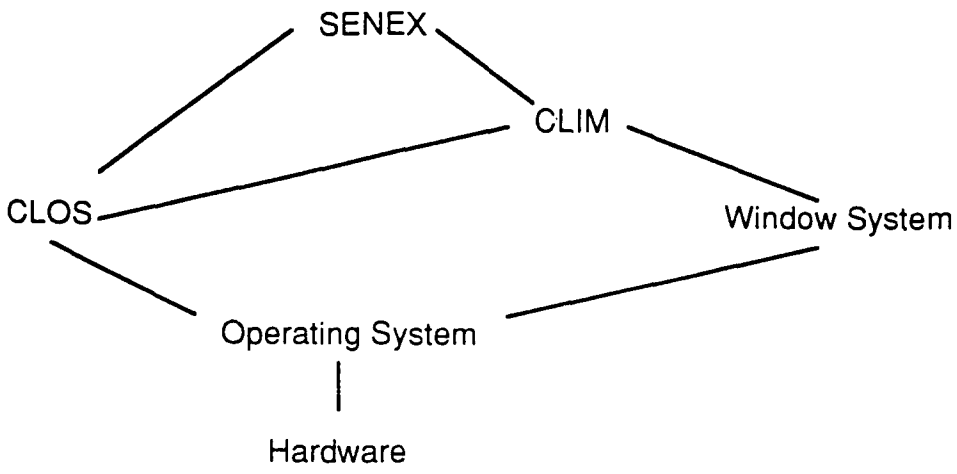
One of the representational challenges in building SENEX has been defining what information is desired for a complete, accurate and detailed description of the phenomena involved in the domain of molecular pathology, and how collection and organization of incomplete or partial information can function to place pieces of the puzzle in their proper perspective. Thus the design of SENEX has been oriented towards integrating data obtained from disparate sources by representation of partial information and remodeling that information as new data becomes available. This approach has been relatively successful due in large part to the powerful programming tools provided by CLOS/CLIM [1,2].

SENEX has grown in size and complexity since its conception, and adapted to new developments in program design as well as those in molecular pathology. These adaptations are not complete nor will they ever be lest the disciplines of program design and molecular pathology become static. Three major developments in programming design have shaped SENEX since its conception: 1) The ANSI standardization of the Common Lisp Object System (CLOS); 2) Introduction of the CLOS metaobject protocol; 3) Development of the Common Lisp Interface Manager. Developments in molecular pathology which have shaped the design of SENEX are too numerous to mention. However several domain associated concepts have

provided the greatest influence on system design: 1) Exploration of molecular information in a graphical context; 2) Visualizing molecular information at increasing levels of detail; 3) Pursuit of tangential issues; 4) Molecular predictions for the purpose of designing laboratory experiments.

CLOS and CLIM confer application portablility to SENEX.

A schematic diagram of how Senex is layered over the host system is shown below.



The application uses tools provided by CLIM and CLOS. CLIM uses tools provided by CLOS and the host window system. CLOS and the host window system use tools provided by the operating system, and the operating system uses the available hardware. SENEX currently runs on Macintosh and Sun SparcStations.

Molecules and molecular processes are represented in SENEX by means of Common Lisp objects.

A rough breakdown of the different classes of objects in the SENEX classification structure is shown below.

THE SENEX CLASSIFICATION STRUCTURE

	# of Classes
Entities	6358
Organisms	1019
Anatomic structures	266
Cells	148
Compartments	121
Molecules	4254
Proteins	2608
Genes	224
Motifs	295
Diseases	474
Events	125

An object in SENEX is an instance of an object class [2]. For example, Homo sapiens is a class of organism and Sheldon S. Ball (first author) is an instance of the class Homo sapiens. Similarly, protein phosphorylation is a class of event and there are many specific instances of protein phosphorylation in SENEX. To date, there are 6483 object classes and 11,484 unique objects in SENEX.

Objects have slots which provide a means of describing an object in detail.

Slots are specialized descriptors of objects defined with the most generalized class to have that attribute or property. Slots may assume default values specified with class definitions and most slot values themselves are instantiated as objects. Slots and their default values are inherited through the classification structure. The basis of the SENEX classification structure is the MEDICAL SUBJECT HEADINGS (MeSH) tree structures. Thus the SENEX classification structure is a biological classification structure. However, there are significant differences between the SENEX classification structure and MeSH, and a mapping between the two is implemented. There is a mapping of synonyms to the canonical forms used as class names and SENEX also has a word completion facility.

Slot default values provide a means of programming biological knowledge into SENEX.

Molecules are classified in chemistry and biology largely on the basis of their properties. Thus classes of molecules share particular properties which may be represented as slot values. All members of a class of molecules may inherit properties as default slot values. Classes of molecules may have multiple supertypes, so that the properties of a class of molecule may be determined by inheritance of slot default values from multiple molecular supertypes.

Inheritance of slot values is specialized depending upon the slot.

Proteins contain structural elements which give rise to the function(s) of the molecule. These structural elements are known as motifs. Most proteins consisting of a single polypeptide can be represented as an ordered set of motifs connected by peptide regions (see figure 10). Different classes of molecules contribute different motifs to their subclasses through slot default values. Thus when a class of molecules is defined with multiple supertypes, motifs are inherited from all supertypes. Inheritance of motifs is said to be inheritance by UNION as distinguished from inheritance by SHADOWING, the default method of CLOS inheritance. Motifs in addition to those motifs inherited from class supertypes may be specified as slot default values with the definition for the protein class.

Senex uses reflective techniques in context of the CLOS metaobject protocol.

Senex uses two enabling technologies: reflective techniques [3,4] which make it possible to expose the implementation of a language (in this case LISP), and object-oriented techniques which allow the implementation of the language to be locally and incrementally adjusted. The basic elements of CLOS - classes, methods, and generic functions - are accessible as metaobjects (objects that represent fragments of a program). A protocol operating on these metaobjects defines the behavior of CLOS. SENEX uses introspective protocols to access slot values of these metaobjects and intercessory protocols to change the behavior of CLOS in specializing the inheritance of motifs.

Ordinary functions, macros, and methods provide a means of programming biological knowledge into SENEX.

The inheritance of motifs by union necessitates processing of motif defaults to eliminate duplicates and identify specializations of more generalized motifs. In addition, certain rules for ordering motifs within a protein can be employed, and details regarding the structure of motifs may be dependent upon the state of the molecule. This type of molecular information is programmed into SENEX using ordinary functions, macros, and methods defined on specific classes of molecules.

SENEX uses CLIM presentations for display of data.

A CLIM presentation is a visual representation of an object linked directly to its semantics, thus facilitating the separation of the internal representation of objects from the presentation of data to users. For example, figure 1 shows a screen image obtained from browsing through the SENEX classification structure to find the Alzheimer amyloid precursor protein (APP). APP is the precursor of the 4 kD A4 peptide that forms the core of senile plaques, insoluble deposits of amyloid found in the brain of Alzheimer's disease patients and to a lesser extent normal elderly individuals. In the lower right hand corner of the screen, there is a window entitled Query that represents a menu of a sort. The Query window contains a set of commands for which the user simply completes the details, places the cursor, and presses <enter>. All subsequent actions are through selection of mouse-sensitive objects in the windows that appear in response to the initially entered command.

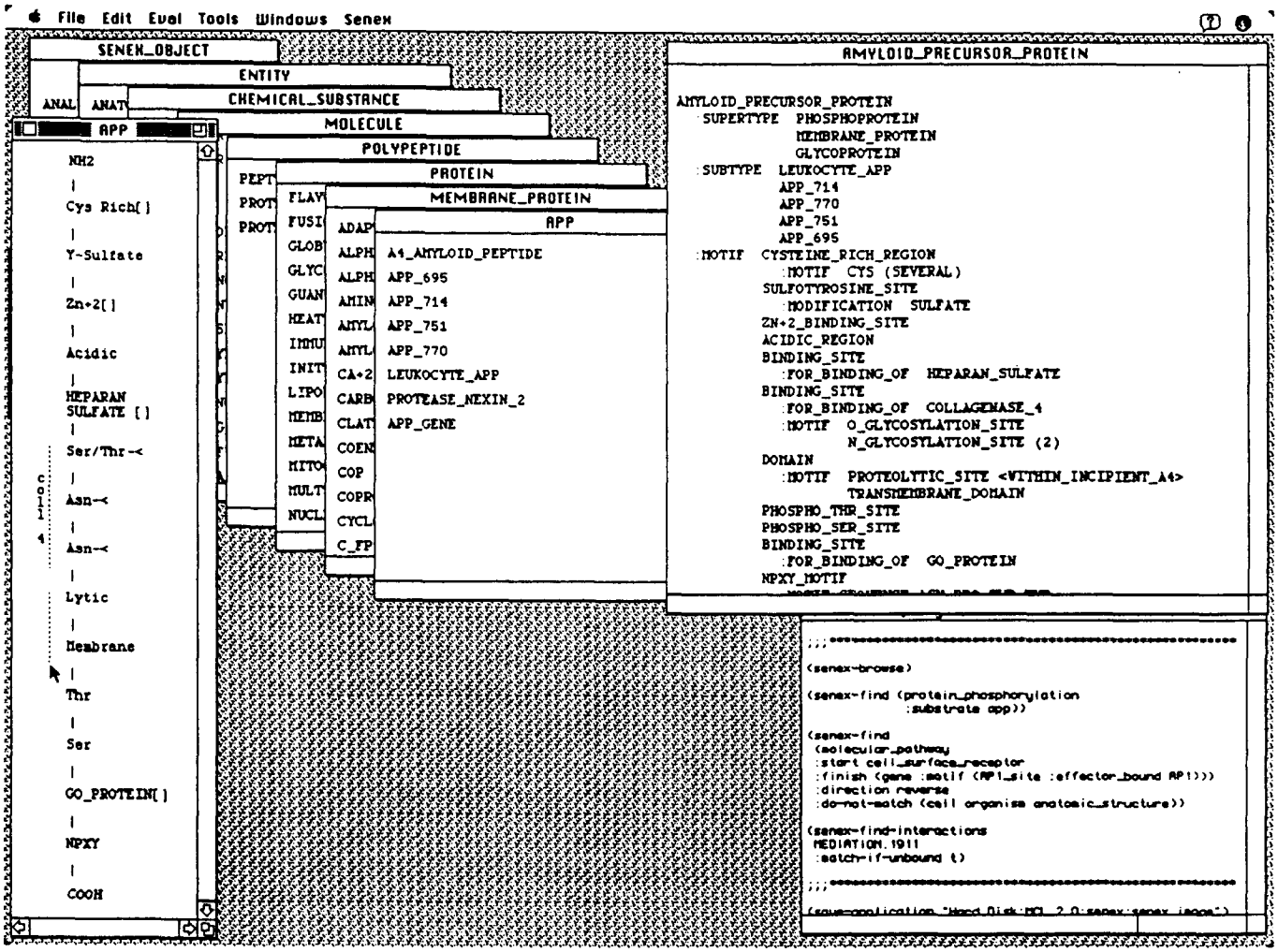


Figure 1

The command used in this instance was the command (senex-browse). A window in the upper left hand corner of the screen entitled SENEX_OBJECT appeared (now partially hidden by a set of offset descending windows). A selection through a mouse gesture in this window produced the adjacent window entitled ENTITY (offset to the right & down). All mouse-sensitive objects in a window highlight when the mouse pointer is passed over the object. Choices from these windows are (for the most part) members of

the SENEX classification structure, with each selection producing a new window showing subclasses of the selected class.[1] If no subclass exists, no subclass window is produced and a general description of the class appears in a window in the upper right hand corner of the screen. Even if a class has subclasses, a description of the class may appear if specified conditions for the class are met. The description of the class is the symbolic representation for that class with associated slot default values. An English description and reference may accompany the symbolic representation.

Senex computes graphical presentations from symbolic representations.

From this symbolic representation, SENEX computes a graphical presentation of the object representing the class and displays this presentation in the lower left hand corner of the screen. The line drawing of APP indicates that the protein consists of an ordered set of motifs connected by peptide regions. These motifs, like the protein containing the motifs, are represented in SENEX as objects. Classes of motifs constitute an important branch of the SENEX classification structure. The motifs shown in the line drawing of APP (lower left) are mouse-sensitive, that is, details of the motif will appear in a new window if the user selects the motif with a mouse gesture.

Generic functions select methods appropriate for specialized classes.

The line drawing of APP shows additional elements of complexity. Dashed lines to the left of the motifs represent domains containing the motifs. The more N-terminal (upper) domain is a binding site for binding and inhibition of collagenase-4 (hydrolyzes collagen of basement membrane). The more C-terminal domain containing the membrane region and proteolytic site represents the region of APP which gives rise to the amyloidogenic A4 peptide. Selecting this latter domain with the mouse brings up a new window (figure 1a) showing the peptide sequence of this protein region. This is accomplished through use of a specialized method defined on a generic function *draw*. The line drawing of this domain shows the sequence which gives rise to the amyloidogenic A4 peptide, the normal cleavage site of the APP secretase, the region of the amyloid precursor protein traversing the plasma membrane, and several mutations in APP found in various hereditary disorders.

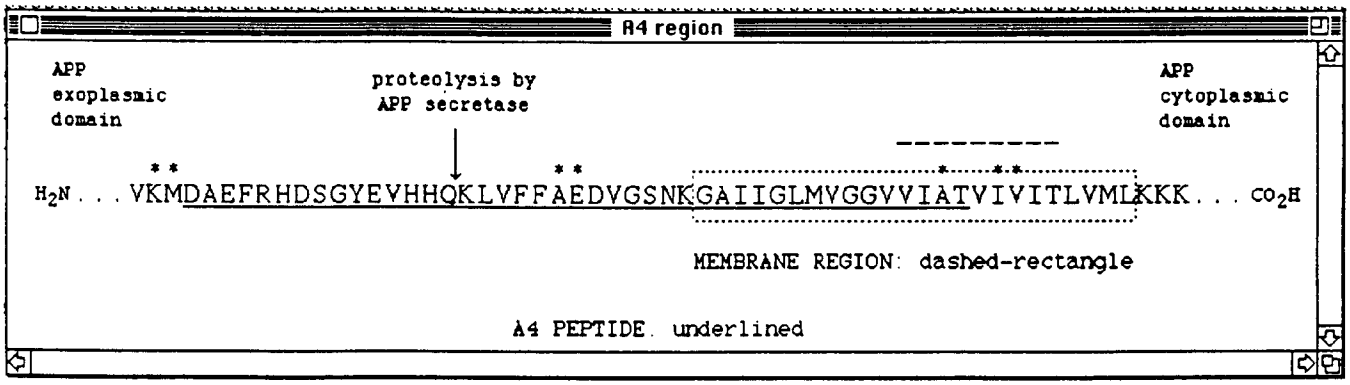


Figure 1a

Through alternative splicing of introns, APP is expressed in different cell populations in several isoforms. Two of these contain a Kunitz protease inhibitor domain. That part of the SENEX classification structure containing the 751 kD isoform APP is shown in figure 2.

---

1 Items presented in response to selection of a class during browsing default to subclasses and components, but may be specialized to include related items such as the gene encoding a polypeptide, the ligand(s) of a receptor, subunits of a multisubunit protein, pathway(s) in which a molecule participates, etc.

Visualizing microscopic findings is an important aspect of recognizing and understanding disease processes.

A logical progression of thought could easily flow from the amyloid precursor protein to the A4 amyloid peptide to senile plaque. Thus selections during browsing within SENEX are derived from estimates of logical thought progression in addition to progression in either forward or reverse directions through the classification structure. Sequential selection of A4 amyloid peptide and senile plaque from that point in the program shown in figures 1 or 1a produces a low power photomicrograph of senile plaques seen in the CA1 region of the hippocampus (figure 3). A region of the photomicrograph circumscribing a well-delineated plaque is mouse active and selection of this region displays a high power view of that plaque. The high power view itself contains a mouse sensitive region surrounding the core of the plaque, which upon selection displays textual information about the plaque core.

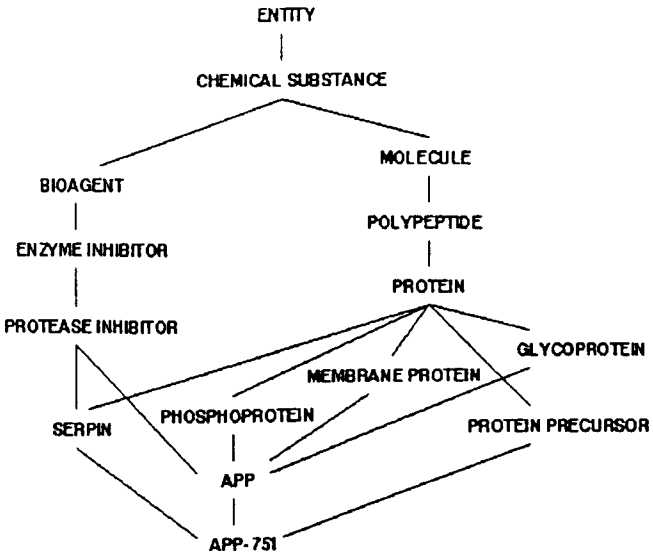


Figure 2

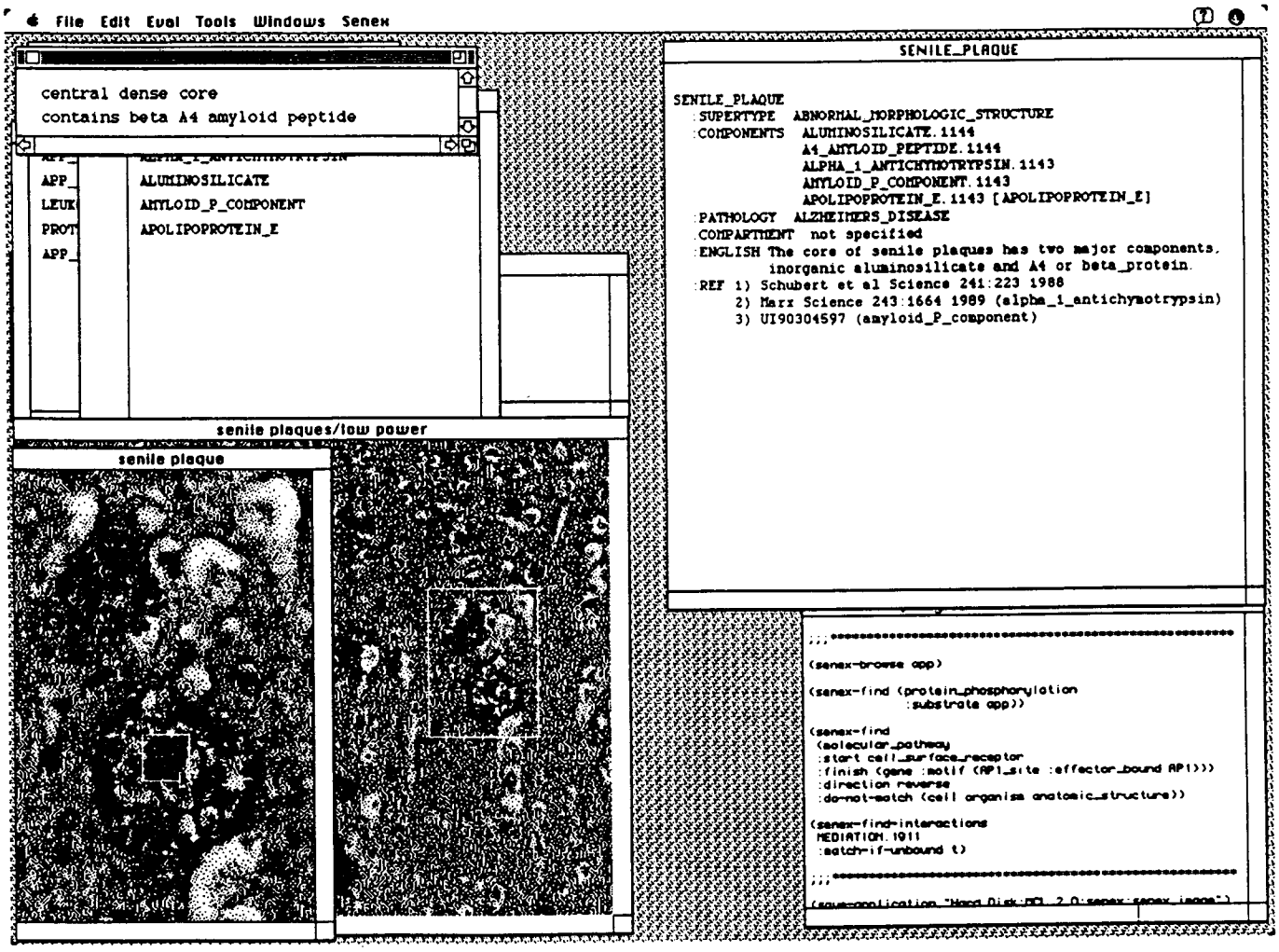


Figure 3

Methods defined on multiple classes provide a means of drawing objects in context of other objects.

Events and subtypes of events are represented in SENEX as objects. Finding events that satisfy particular specifications, for example, phosphorylations of the general class of protein APP (includes any subtype of APP) is facilitated with the command (senex-find (protein_phosphorylation :substrate APP)). SENEX collects all instances of the class PROTEIN_PHOSPHORYLATION testing for additional user defined target specifications (in this case :substrate APP). All instances in SENEX have associated unique instance id's. A concise symbolic representation of the phosphorylation of APP catalyzed by Ca+2/Calmodulin-dependent protein kinase-2 along with a schematic of the reaction is shown in figure 4. The ellipses in the schema represent proteins, with substrates on the left and products on the right. The enzyme for the reaction is shown above the arrow. APP in the context of a reaction now appears as a simple ellipse rather than as a stick figure illustrating all of its motifs as in figure 1. Thus, in figure 1 APP is drawn in context of itself, and in figure 4, it is drawn in context of a reaction it undergoes. We have also seen examples of a domain drawn in context of a protein and in context of itself (figures 1 and 1a). This feature serves to provide the user with an appropriate level of detail.

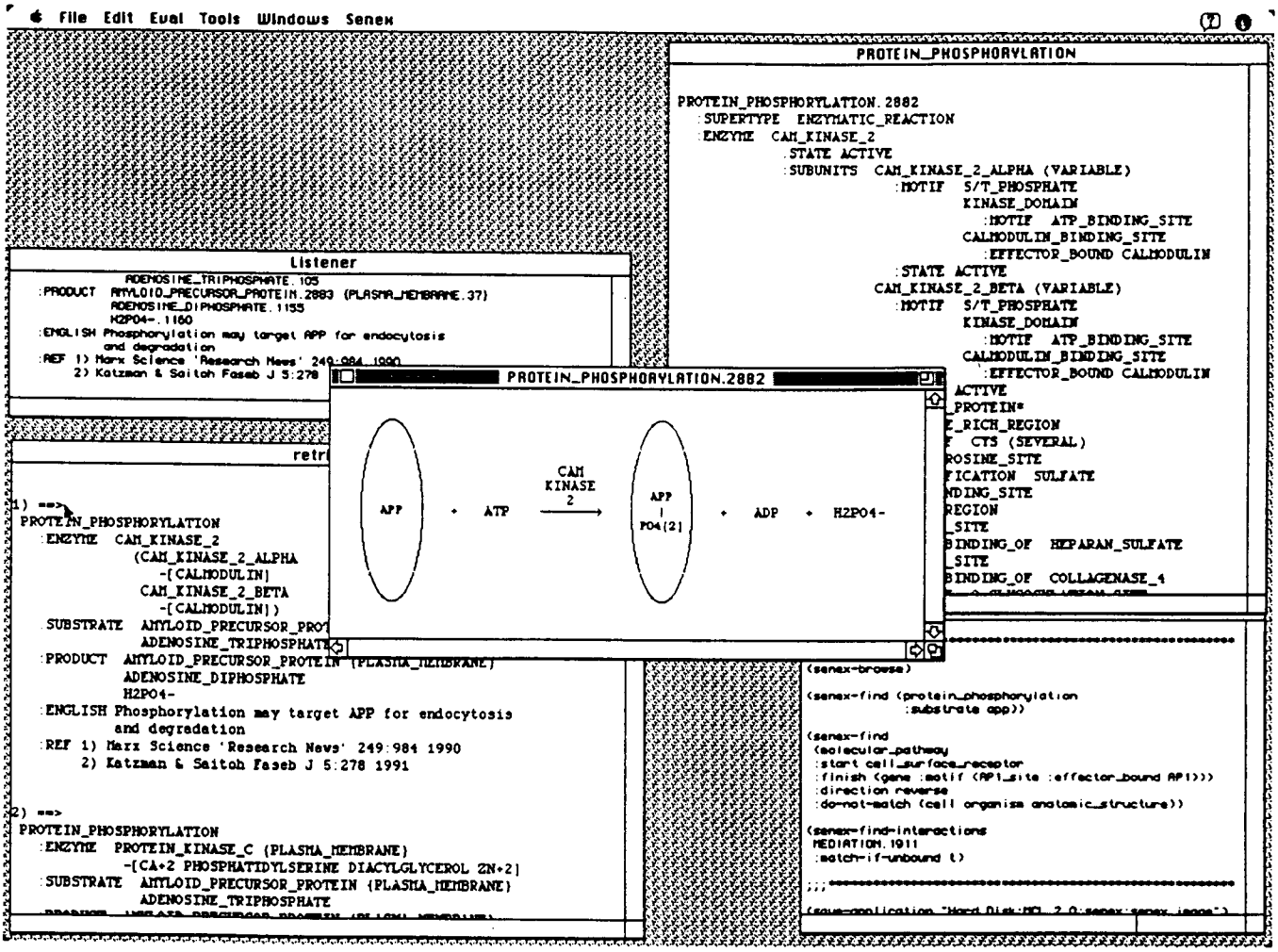


Figure 4

The objects in the window showing schematically the selected reaction are mouse-sensitive. Selecting any of the objects in the schema brings up further detail of the object. In the case of the ellipse representing APP, selection brings up a stick figure or line drawing along with a symbolic description of the molecule as shown in figure 1.

CLIM separates the internal representation of objects from the presentation data to users.

When information about a gene for a protein is available, the gene appears as a choice when that protein is selected during browsing (see figure 1). The symbolic representation of the APP gene (not shown) indicates that at least 6 different proteins are derived from this gene by alternative splicing of a single transcript. A chromosomal location for the gene (human if not otherwise specified) is shown as a value of the slot LOCUS. Internally, SENEX knows that genes are located in the cell nucleus but does not display this to the user since the fact is self-evident.

A graphical presentation is computed from the symbolic representation when APP_GENE is selected during browsing and displayed in a separate window (figure 5). The line drawing shows several gene regions including 5' enhancer and promoter regions, a coding region, and a 3' enhancer region. DNA regulatory elements or motifs which control expression of the gene are shown in the enhancer and promoter regions. Regulatory elements embedded within an exon or intron are revealed upon selecting the coding region with a mouse gesture.

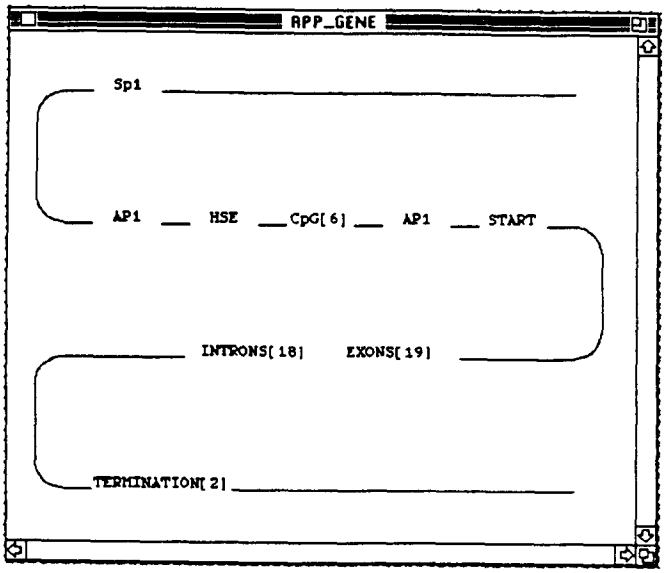


Figure 5. The gene encoding APP and its isoforms computed from symbolic representation. Note the 5'enhancer region containing a binding site for SP1, and the promoter region containing binding sites for AP1, a heatshock element, and 6 CpG islands. The gene contains 19 exons and 18 introns, details of which may be revealed through mouse-selection. There are two termination sites and no identified 3'enhancer elements.

Senex provides a means of exploring tangential issues and increasing levels of molecular detail.

It is noted that the gene for APP contains two AP1 sites, binding sites for the transcription factor AP1. We can use the browse facility to obtain information about AP1. A graphical presentation of AP1 is shown in figure 6 (pg 39). It is noted that AP1 is a heterodimer of p55fos and p39jun each of which contain motifs which hold the proteins together in a specific configuration. Thus, the cysteine-containing basic helix-loop-helix motif (CbHLH) on each protein is a half site which aligns with the analogous site on its companion protein to form a redox-sensitive DNA-binding motif. Similarly, the leucine zipper motifs on the two proteins are half sites which align to hold the proteins together by hydrophobic bonding of heptad repeats of leucine residues. The line drawing of AP1 thus shows the alignment of these motifs.

The PEST region of p55fos is selected with a mouse gesture to produce a new window which contains a symbolic representation of this motif. An English description and references may be the most useful aspect of this window.

Representation of events as objects facilitates identification of molecular pathways.

Cells communicate with their environment in part through interaction of extracellular molecules with receptors on the cell surface. Interaction of cell surface receptors with their ligands in turn induces intracellular events (referred to as signal transduction) which can lead to changes in gene expression. Transcription of genes is regulated through binding of specific nuclear proteins to regulatory elements within promoter or enhancer regions of the gene. Thus, we may be interested in signal transduction pathways that stimulate transcription of the APP gene.

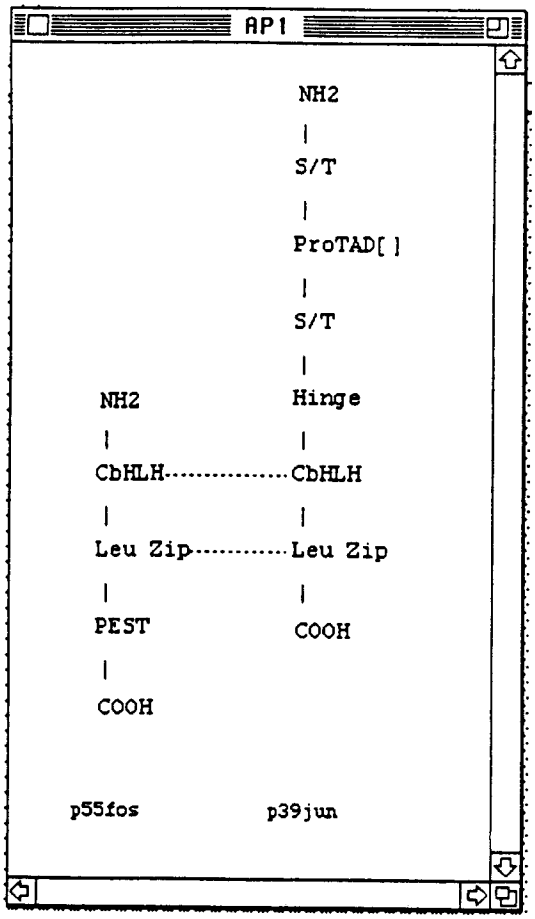


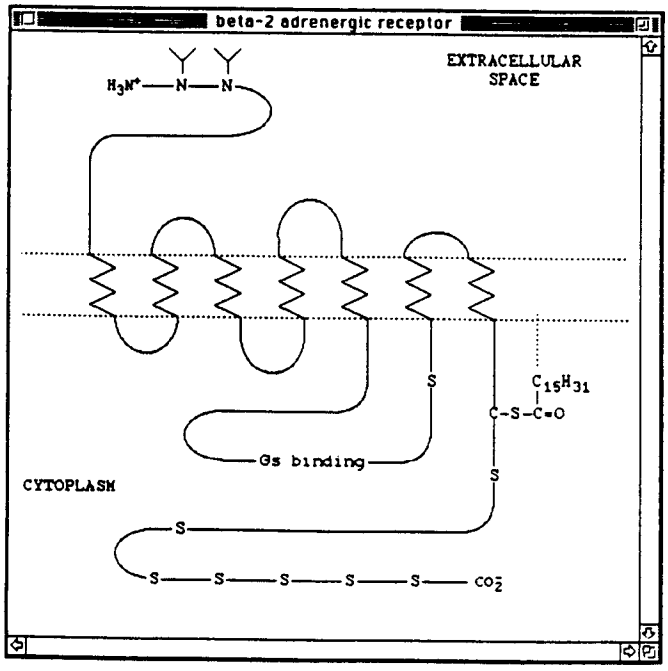
Figure 6. AP1 (Activating Protein 1) computed from symbolic representation. Note the alignment of the interacting motifs on the two subunits, p55fos and p39jun.

SENEX is a tool of discovery. Search options facilitate prediction of novel signal transduction pathways.

We can tell SENEX to ignore cell type, anatomic considerations, and organism type so that we might piece together reactions known to occur, but to occur in different cell types, or in different organisms, in the same reaction pathway. It is through queries of this type that SENEX may be used to predict novel signal transduction pathways, in essence generating hypotheses which may be tested in the laboratory.

We can also specify that SENEX start a pathway search from a molecule or molecular complex satisfying specified criteria, for example starting from a gene searching for a sequence of events which originate from a cell-surface receptor. This feature is useful if one is interested in examining the search queue for pathways which failed to find the target.

Such a query seeking regulation of APP gene expression by extracellular signals, yields twenty-two such possible pathways, one of these originating from the beta-2 adrenergic receptor (figure 7). The queue at each step of the search along with a summary of pathways found is printed to the listener window. Any one of the pathways as well as any of the events that comprise a particular pathway may be chosen for further examination. Figure 8 shows a screen shot showing translation of c-fos messenger RNA, obtained by selecting one of the events of a selected pathway.



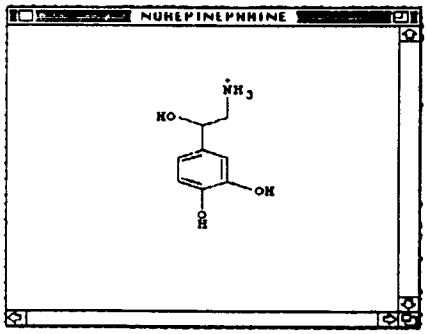


Figure 7. beta-2 Adrenergic receptor computed from symbolic representation with specializations of N-terminal domain, third cytoplasmic loop and cytoplasmic tail. Note the several serine phosphorylation sites and the palmitoylated cysteine in the cytoplasmic tail.

Figure 7a. Noradrenaline, the principle CNS ligand of beta-adrenergic receptors, drawn with CLIM tools.

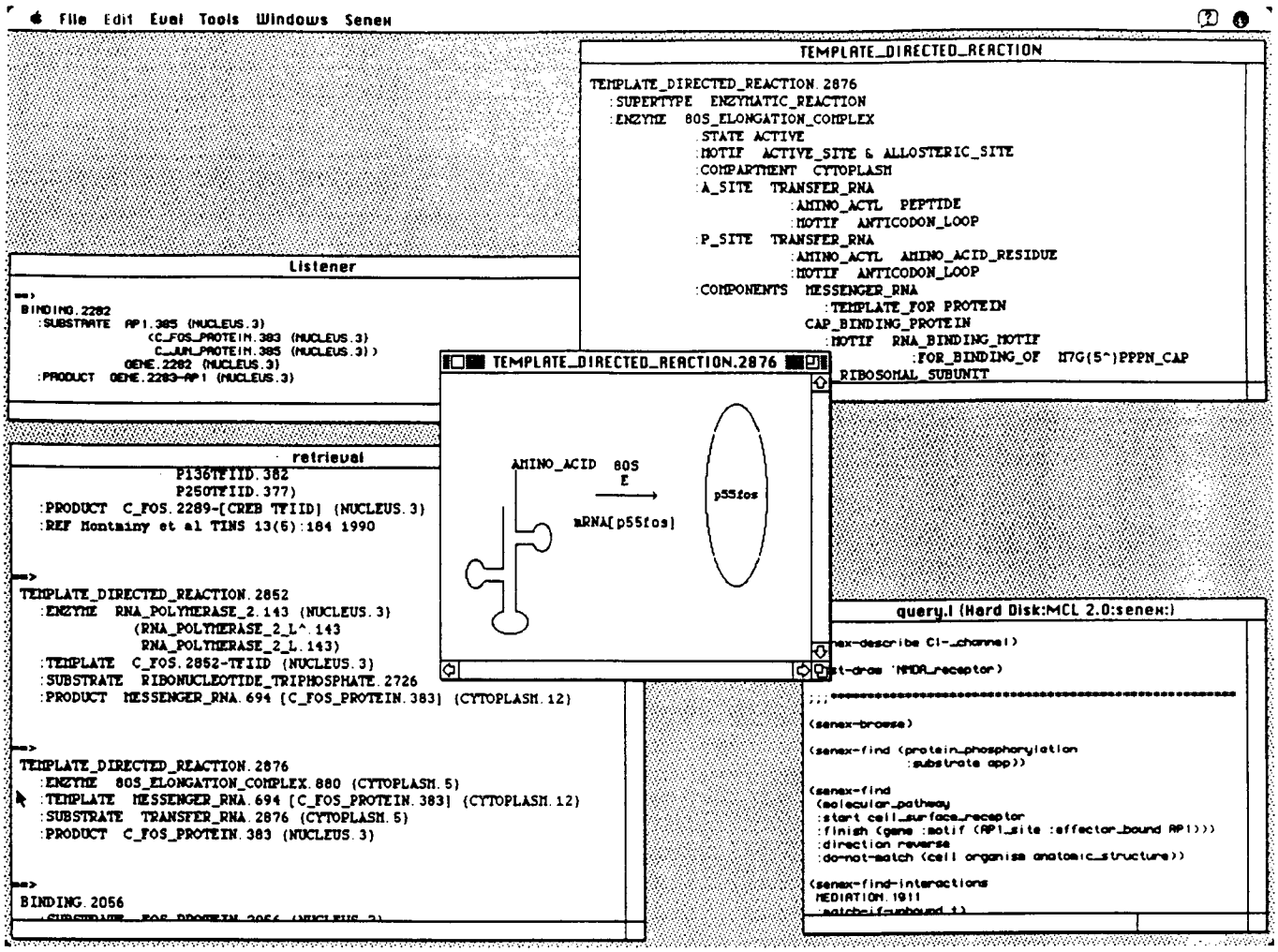


Figure 8

Regulation of signal transduction involves complex networks of interactions.

It is also possible to identify events which interact with other events or with molecular pathways. These interactions may be further specified as inhibitory or stimulatory. For example, in the pathway originating from the beta-2 adrenergic receptor and ending in binding of AP1 to the APP-gene, SENEX identifies several inhibitory interactions including: 1) inhibition of adenylate cyclase through binding of Gi-alpha; 2) proteolysis of protein kinase A (catalytic) by calpain; and 3) binding of AP1 to genes other than the APP-gene.

The algorithms employed in SENEX searches have been the most challenging aspect of building SENEX. These algorithms center about matching of objects subject to specified and contextual constraints. The matching of objects itself involves recursively decomposing objects and matching their component substructures. The difficulty in establishing criteria for matching objects is especially apparent in the contexts of molecular pathways and interactions of molecular events. However, it is in these algorithms that the real power of SENEX lies. Much of future work on SENEX will go into refining these algorithms.

Specialization of molecular presentations illustrates compartmental relationships of molecules.

Cell surface receptors which convey the signal of their extracellular ligand to the interior of the cell via activation of G proteins (serpentine receptors), have a common structure of 7 transmembrane

domains. The ligand binding regions and sites of interaction with G proteins lie within the membrane (see figure 7). Regulatory sites may be found within the cytoplasmic loops and cytoplasmic tail of the receptor. These features may be elucidated when the user clicks on one of the many mouse sensitive features in the line drawing.

Figure 7 shows a graphical presentation of the beta adrenergic receptor computed from the symbolic representation of the molecule. This is accomplished through specialized methods defined on the generic function *draw*. The common structure of serpentine receptors facilitates drawing regions of the generalized structure and specialized drawing of particular regions for particular receptors. In the case of the beta-2 adrenergic receptor, the exoplasmic N-terminus, cytoplasmic tail and 3rd cytoplasmic loop were drawn using specialized methods of the generic function *draw*.

Visualizing chemical structures of small molecules helps to understand their biological functions.

Ligands for receptors when known appear as choices when a receptor is selected during browsing. Selecting norepinephrine brings up a description of the principal CNS ligand for the adrenergic receptors (see figure 7a). Visually identifying the hydrophobic aromatic ring of norepinephrine with its polar substituents helps to understand how such a ligand might fit into the transmembrane pockets of the adrenergic receptors.

Presentation of molecular pathways and cascades is facilitated by generic functions which select methods appropriate for specialized objects and methods defined on multiple classes which provide a means of drawing objects in the context of other objects.

The observation that protease nexin 2 derived from APP secretase-mediated proteolysis of APP [5] inhibits coagulation factor XIa has turned attention to a possible involvement in the components of the components of the coagulation cascade in the pathogenesis of Alzheimer's disease [6]. Figure 9 shows a presentation of the coagulation cascade with its associated amplification loops. This figure represents a general level of detail. All of the objects in the presentation are mouse-active. Thus visualizing further detail about particular aspects (molecules, events, or disease manifestations) of coagulation are visualized by selecting appropriate objects in the sequentially revealed windows.

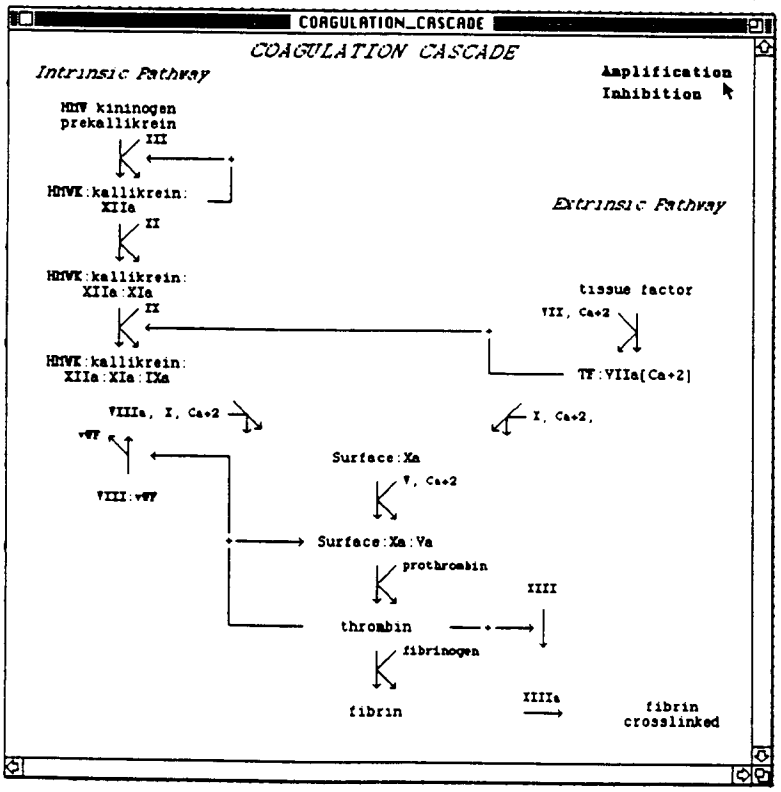


Figure 9. The coagulation cascade showing amplification loops, drawn with CLIM tools. Details regarding specifics of the cascade are revealed with mouse-selections.

Conclusion

SENEX uses many features of CLOS and CLIM to facilitate representation in a domain characterized by uncertainty and complex interactions of diverse structural elements and events. These features include CLIM presentations, classes defined on multiple supertypes, generic functions and methods defined on multiple supertypes, inheritance and specialization, and reflection on the CLOS metaobject protocol.

References

1. Ball, SS and Mah, VH (1992) Symbolic Representation in Molecular Pathology. Proceedings of the 2nd Annual Lisp Users and Vendors Conference, August 10-14, San Diego. CA, 1992.

2. Keene, SE (1989) Object Oriented Programming in COMMON LISP, Addison Wesley, Reading, MA.

3. Kiczales, G; des Riveres, J; Bobrow, DG (1991) The Art of the Metaobject Protocol, MIT Press, Cambridge, MA.

4. Bobrow, DG, Gabriel, RP & White JL (1993) CLOS in Context: The Shape of the Design Space. In: Object-oriented Programming: The CLOS Prespective. Paepcke, A (ed) MIT Press, Cambridge, MA, pg 29-61.

5. Esch, FS; Keim, PS; Beattie, EC; Blacher, RW; Culwell, AR; Oltersdorf, T; McClure, D; Ward, PJ (1990) Cleavage of amyloid beta peptide during consitutive processing of its precursor. Science 248:1122.

6. Smith, RP; Higuchi, DA; Broze, GJ (1990) Platelet coagulation factor Xla-inhibitor, a form of Alzheimer amyloid precursor protein. Science 248:1126.