SENEX: Symbolic Representation in Molecular Pathology

Sheldon S. Ball[1] and Vei H. Mah[2]

[1] Department of Pathology
University of Mississippi
Jackson, MS

[2] Department of Neurology
Thomas Jefferson University
Philadelphia, PA

ABSTRACT

SENEX is a computer program in the domain of molecular pathology. The application allows an individual to ask a sequence of questions in a single interactive session, thereby facilitating the development and testing of several hypotheses in a short period of time. Cartoon representations of molecules and molecular events enable individuals to grasp spatial and functional relationships among molecules, cellular compartments, and cell regions. Well-defined reasoning capabilities allow an individual to ask sophisticated questions and to predict novel molecular events or pathways. SENEX contains information about: 1) the molecules and the motifs that impart function to these molecules; 2) the molecular events; 3) cell-specific expression of genes, 4) disease processes. SENEX is being developed through object-oriented programming in a portable programming environment supported by COMMON LISP and the COMMON LISP INTERFACE MANAGER.

INTRODUCTION

SENEX is a computer system under development to explore issues related to representation of molecular information, presentation of data, and reasoning with molecular information. It is written entirely in 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. The authors are pathologists interested in aging and degenerative diseases of old age, especially malignancies and senile dementia.

Dementia is the global impairment of higher cortical functions including memory, the capacity to solve problems of day to day living, the performance of learned peceptuomotor skills, the correct use of social skills, and control of emotional reactions in the absence of gross clouding of consiousness. The condition is often irreversible and progressive. It has been estimated that as many as half of all individuals over 90 years of age may be considered clinically demented[1] with Alzheimer's disease contributing to the majority of these cases.

Alzheimer's disease is characterized by the presence of neurofibrillary tangles and senile plaques, insoluble accumulations of material largely composed of paired helical filaments and amyloid, respectively. The amyloid deposits contain large amounts of a 4200 dalton peptide A4, derived from one or more of several forms of a membrane protein referred to as the amyloid precursor protein (APP). This paper will illustrate features of CLOS and CLIM used in SENEX with APP and regulation of its expression serving as examples.

REPRESENTATION

Molecules and molecular processes are represented in SENEX by means of Common Lisp objects.[2-6] Objects in SENEX are organized in a classification structure with two main branches, classes of entities and classes of events. There are 4459 classes of entities and 25 classes of events currently in the SENEX classification structure. A rough breakdown of the different types of objects in the SENEX classification structure is shown below.

THE SENEX CLASSIFICATION STRUCTURE

	# of Classes
Entities	4459
Organisms	563
Anatomic structures	241
Cells	137
Compartments	107
Molecules	3014
Proteins	2096
Genes	203
Motifs	273
Diseases	318
Events	25

An object in SENEX is an instance of an object class. For example, Homo sapiens is a class of

organism and Sheldon S. Ball (author on this paper) 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. Objects have slots which provide a means of describing an object in detail. 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. There is a mapping of synonyms to the cannonical forms used as class names.

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 superytpes.

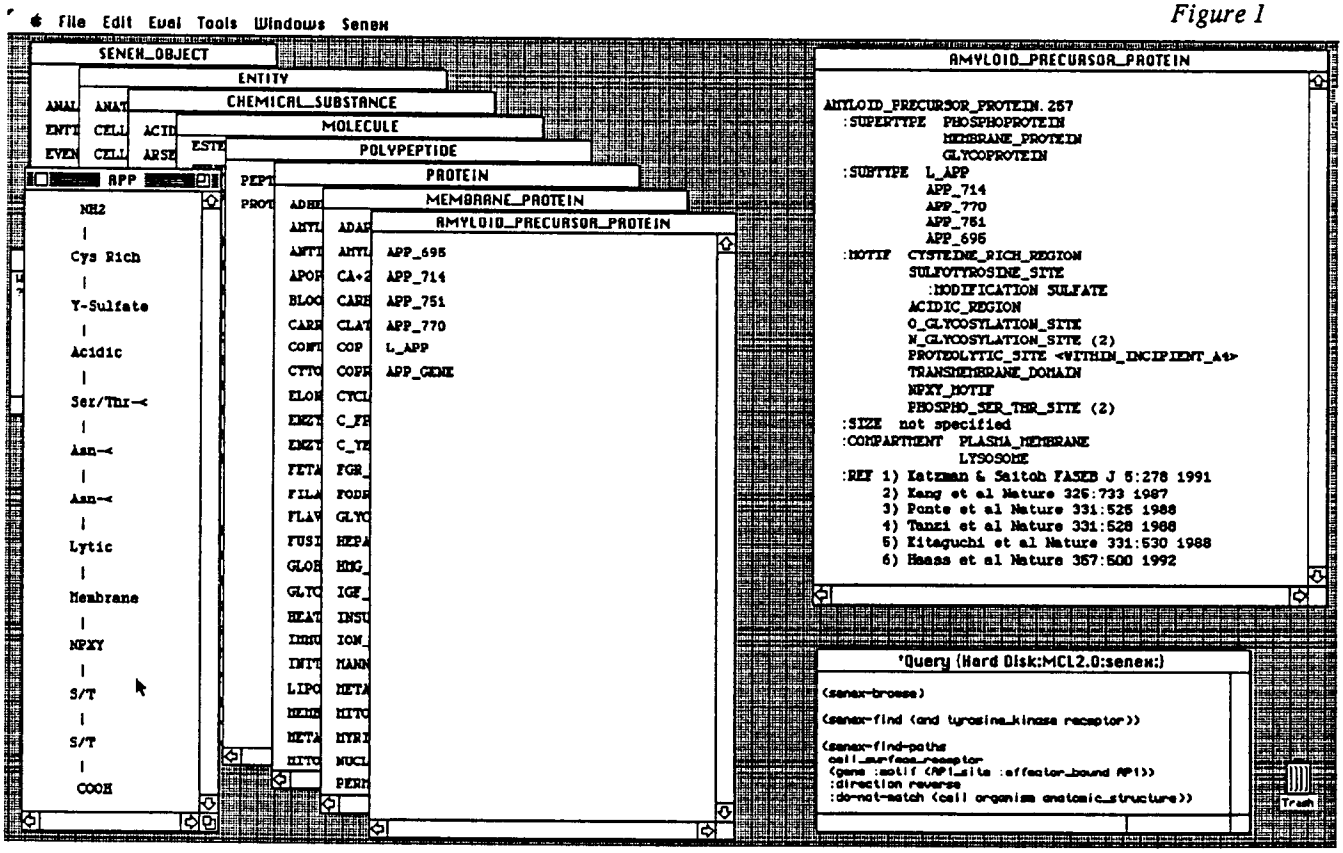
Inheritance of slot values is specialized depending upon the slot.

Proteins contain structural elements which give rise to the function(s) of the molecule. These structural element 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 1). 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.

Methods provide a means of programming biological knowledge into SENEX.

The inheritance of motifs by union neccessitates processesing of motif defaults to eliminate duplicates and identify specializations of generalized motifs. In addition, certain rules for ordering motifs can

Figure 1



be employed, and details regarding the structure of motifs may be dependent upon the state of the molecule. This type of information may be programmed into SENEX using methods defined on specific classes of molecules.

SENEX uses CLIM presentations for display of data.

Figure 1 shows a screen image obtained from browsing through the SENEX classification structure to find the amyloid precursor protein (APP). In the 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.

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 click 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 members of the SENEX classification structure, with each selection producing a new window showing subclasses of the selected class. 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.

From this symbolic representation, SENEX computes a cartoon representation of the object representing the class and displays this representation in the lower left hand corner of the screen. The cartoon 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 cartoon of APP (lower left) are mouse-sensitive, that is, a description of the motif will appear in a new window if the user selects the motif with a mouse click.

When the user has completed browsing, the screen may be cleared (leaving the menu window in lower right exposed) by selecting clear from the Macintosh menu bar at the top of the screen. Figure 2 shows that part of the SENEX classification structure containing APP.

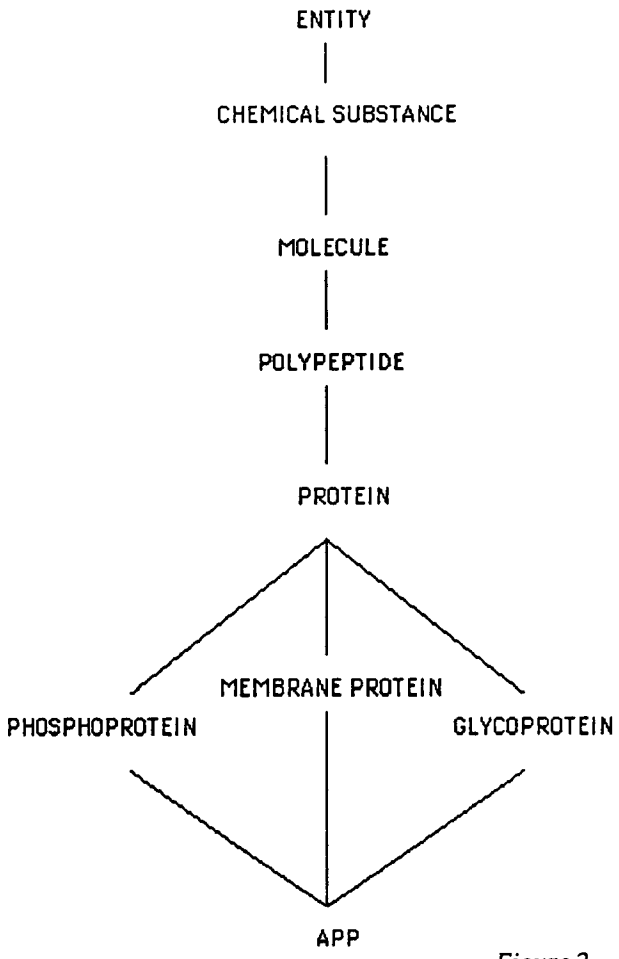


Figure 2

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, in this case, phosphorylations of the general class of protein APP (includes any subtype of APP) is facilitated with the command (senex-find). 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 reaction of Ca+2/Calmodulin-dependent protein kinase-2 along with a cartoon of the reaction is shown in figure 3. The ellipses in the cartoon 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 3, it is drawn in context of a reaction it

undergoes. This feature serves to provide the user with an appropriate level of detail.

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

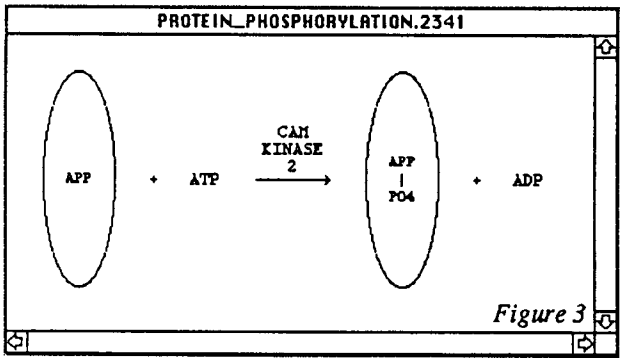


Figure 3

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 shown in figure 4 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 ar located in the cell nucleus but does not display this to the user since the fact is self-evident.

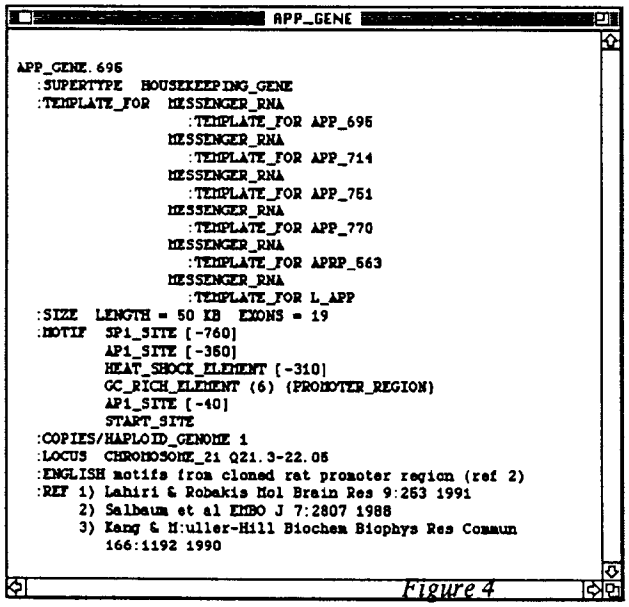


Figure 4

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 transcription factors (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. A query may be formulated to find a pathway linking a cell-surface receptor to any gene containing bound AP1, since we know that the APP gene contains 2 AP1 sites, and that in other genes, binding of AP1 to its recognition sequence in promoter regions stimulates gene transcription. We can specify that the search be run in reverse, that is starting from the gene and 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. We can also 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.

Twenty-two such possible pathways are found, one of these originating from the beta-2 adrenergic receptor (see figure 5). 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 may be chosen for further examination. It is also possible to examine events which interact with other events or with molecular pathways. These interactions may be further specified as inhibitory or stimulatory. The algorithms employed in SENEX searches are discussed in reference 1.

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, 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 5). Regulatory sites may be found within the cytoplasmic loops and cytoplasmic tail of the receptor. These features may be ellucidated when the user clicks on one of the many mouse sensitive features in the cartoon.

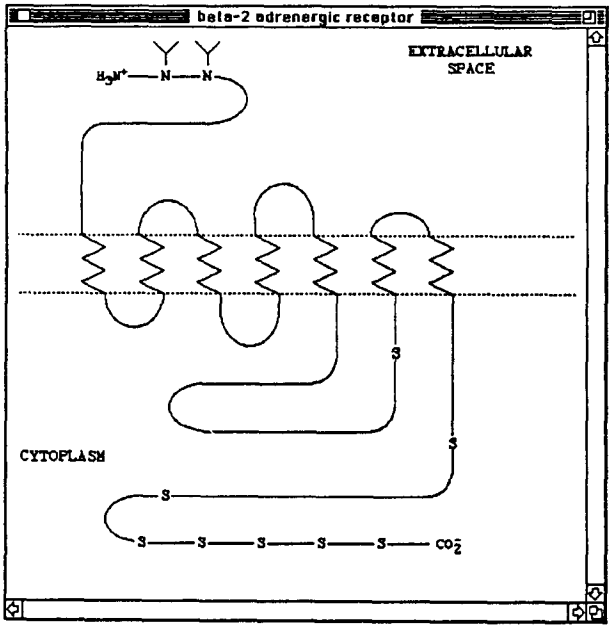


Figure 5

CONCLUSION

SENEX is an interactive LISP/CLOS/CLIM application in the domain of molecular pathology. SENEX allows a user to ask a sequence of questions in a single interactive session, thereby facilitating the development and testing of several hypotheses in a short period of time. Cartoon presentations of molecules and events help a user to grasp the spatial and functional relationships (and, in the future, temporal relationships) among molecules, cellular compartments, and cell regions. Simple, but well defined reasoning capabilities allow a user to ask relatively sophisticated questions and to predict novel molecular events or pathways. Much of SENEX development has been to explore issues related to representation of molecular information, presentation of data, and reasoning with molecular information.

SENEX contains information about: 1) the molecules and the motifs that impart function to these molecules; 2) genes and the regulatory elements that control their transcription; 3) molecular events, i.e. transcription, translation, receptor-ligand interactions, protein-phosphorylations etc.; 4) protein, mRNA, and gene sequences and gene loci contained in databases built and maintained elsewhere; 5) cell-specific expression of genes; 6) factors affecting the formation, degradation, and modification of molecules. Currently, SENEX is most useful for individuals wishing to explore issues tangential to a focused area of expertise, or for individuals wishing to explore a new domain. Its utility will progressively expand as the system grows in size and complexity.

REFERENCES

1. Evans, D.A.; Funkenstein, H,H.; Albert, M.S.; Scherr, P.A.; Cook, N.R.; Chown, M.J.; Hebert, L.E.; Hennekens, C.H.; Taylor, J.O. Prevalence of Alzheimer's disease in a community population of older persons. JAMA 262:2551-2556, 1989

2. Ball, S.S.; Mah, V.H.; Miller, P. The SENEX project: computer-based representation of cellular signal transduction pathways in the aging central nervous system. Computer Applications in the Biosciences. 7:175-187, 1991

3. Ball, S.S.; Mah, V.H. The Second Messenger System, Society for Neurosciences 1991 Annual Meeting, New Orleans, LA (Teaching of Neurosciences); pg 521

4 Ball, S.S.; Mah, V.H.; Miller, P. New SENEX directions, towards a laboratory workstation for molecular neurobiology. Proceeding of the American Medical Informatics Association, Mitchell J.A. (ed.) June 20-23 1990, Snowbird, Utah, 1990

5. Ball, S.S.; Mah, V.H.; Miller, P. New SENEX directions, towards a prototype gene expression map of the central nervous system. Symposium for Artificial Intelligence and Molecular Biology, March 27-29, 1990, Stanford University

6. Ball, S. S.; Wright, L.; Miller, P. SENEX, An object-oriented biomedical knowledge base. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care, L. C. Kingsland III, ed., IEEE Society Press, Washington, D.C., pg 85-89, 1989