LA-UR- 09-04478

Approved for public release; distribution is unlimited.

Title:	Hierarchical graphs for better annotation of rule-based models of Biochemical Systems
Author(s):	Hu, B., Z#: 223064, T-6/T-Division Hlavacek, B., Z#: 119605, T-6/T-Division
Intended for:	Meeting: Pacific Symposium on Biocomputing Fairmont Orchid, Big Island of Hawaii, Jan 4-8, 2010 Journal: Proceedings of the Pacific Symposium on Biocomputig

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

HIERARCHICAL GRAPHS FOR BETTER ANNOTATION OF RULE-BASED MODELS OF BIOCHEMICAL SYSTEMS

BIN HU
Department of Biology, University of New Mexico
Albuquerque, NM 87131, USA
Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory,
Los Alamos, NM 87544, USA

WILLIAM S. HLAVACEK
Theoretical Biology and Biophysics Group, Theoretical Division, and Center for Nonlinear Studies
Los Alamos National Laboratory,
Los Alamos, NM 87544, USA
Department of Biology, University of New Mexico
Albuquerque, NM 87131, USA

In the graph-based formalism of the BioNetGen language (BNGL), graphs are used to represent molecules, with a colored vertex representing a component of a molecule, a vertex label representing the internal state of a component, and an edge representing a bond between components. Components of a molecule share the same color. Furthermore, graph-rewriting rules are used to represent molecular interactions, with a rule that specifies addition (removal) of an edge representing a class of association (dissociation) reactions and with a rule that specifies a change of vertex label representing a class of reactions that affect the internal state of a molecular component. A set of rules comprises a mathematical/computational model that can be used to determine, through various means, the system-level dynamics of molecular interactions in a biochemical system. Here, for purposes of model annotation, we propose an extension of BNGL that involves the use of hierarchical graphs to represent 1) relationships among components and subcomponents of molecules and 2) relationships among classes of reactions defined by rules. We illustrate how hierarchical graphs can be used to naturally document the structural organization of the functional components and subcomponents of two proteins: the protein tyrosine kinase Lck and the T cell receptor (TCR)/CD3 complex. Likewise, we illustrate how hierarchical graphs can be used to document the similarity of two related rules for kinase-catalyzed phosphorylation of a protein substrate. We also demonstrate how a hierarchical graph representing a protein can be encoded in an XML-based format.

1. Introduction

Signal transductions in biological systems are typically built on protein-protein interactions, which often depend on the post-translational modifications of participating proteins. For example protein tyrosine phosphorylations have been found in various signaling pathways, including epidermal growth factor signaling pathway, T lymphocyte activation process, insulin signal transduction pathways, etc. Proteins containing the Src-homology 2 (SH2) domain can recognize phosphorylated tyrosine in itself and in other proteins. If the tyrosine in that protein is not phosphorylated, this molecular recognition cannot occur (Songyang, et al., 1993). Protein-protein interactions form both positive and negative feedback loops, which adds complexity to biological systems and makes it difficult even sometimes impossible to make intuitive predictions about behavior properties of such systems. Modeling and simulation have been used to help us understand those complex systems. Traditional modeling methods require all the species and reactions to be explicitly enumerated in a model. However, if all the posttranslational modifications, e.g. tyrosine phosphorylation, are included in a model, it is often impractical to list all the species and reactions that may arise even with a small number of initial species and reactions, a phenomenon known as combinatorial complexity.

In our previous work, we have introduced a rule-based modeling technique based on graph theory [1]. In the rule-based modeling approach, each protein is represented as a graph. Protein components are represented as color labeled vertices, and edges represent bonds between components. All the reactions are defined by graph-rewriting rules [2]. The benefit of using rule-based modeling over the traditional modeling method is that the combinatorial complexity can be efficiently solved in the rule-based modeling method so that modelers do not have to explicitly enumerate all the possible species and reactions that may be generated in a network [2]. Rule-based modeling has been implemented in BioNetGen [3]. The BioNetGen language (BNGL), a text-based language that encodes the graph structure of molecules, has been introduced and is used in BioNetGen [4].

Several different components types exit in proteins. For example, domains are components of proteins, and amino acid residues are components of proteins. When amino acid residues locate on the domains, they are components of those domains also. So components in proteins may contain other components and proteins observe hierarchical structural organizations. However, in non-hierarchical graph representations all components are considered equal in their relations to the protein and the hierarchical structure cannot be faithfully represented. This makes it difficult for a reader to understand a molecule, hence the model. Another limitation of using non-hierarchical graphs to annotate rule-based models of biochemical systems comes from lack of the ability to organize the reaction rules that share a same reaction center. In the rule-based modeling approach all reactants are screened to match reactant patterns defined in the rule set of a model. Only if all the patterns are matched to a given reaction rule, will that reaction rule be executed. It is common to have multiple reaction rules sharing the same reaction center. The difference between those reactions lies on the contexts of those reactions. The context information in reaction rules shows the scalability of rule-based modeling. In the coarsest form, a reaction rule has no context information in it and only a reaction center is included in that rule. This is the most generic form of reaction rules. When more context information is added into a coarse rule, that rule becomes more specific. With enough contexts added to the coarse rule, only one specific reaction may meet all the reactant pattern requirements, which is identical to the conventional modeling approach where every reaction must be enumerated. In text-based BNGL, all the reaction rules are listed together, which makes it difficult to sort which reactions may share the same reaction center. It would be nice if all the reaction rules sharing the same reaction center can be organized in a way that identifying the relationships among the reaction rules can be easily identified by modelers. This may also contribute to model reduction [5] and coarse-graining of rule-based models [6].

Some models with large numbers of interactions have been developed (Chen, et al., 2009). While these models have generated significant insights of those systems, representations of protein structures (e.g. domain information, amino acid modification, etc) still lack a format that allows easy understanding of a model (Hlavacek, 2009). In order to make a graph representation of a molecule informative of the underlying hierarchical organization of the molecule, we expand the graph-formalism to use hierarchical graphs for annotations of proteins and reactions that share the same reaction center. In this paper we provide examples showing how hierarchical graphs can annotate the internal structural information of proteins, which make models much easier to understand. We demonstrate that hierarchical graphs can help to illustrate the relationships among reactions that share the same reaction center. We also discuss how modelers can benefit from using hierarchical graphs and how to store hierarchical graph in an XML-based format. This XML representation of hierarchical graph allows researchers to annotate protein sequence, domain, post-translational modification and standard database links in a model directly.

2. Results

2.1. *Molecules*

In the following sections we demonstrate how simple proteins and multimeric protein complex can be represented using hierarchical graphs using two examples. The first example is the human lymphocyte cell-specific protein-tyrosine kinase (Lck, gene name: LCK, UniProt accession number P06239), which is an Src-family non-receptor tyrosine kinase that plays an important role in T lymphocyte activation process [7, 8]. It is composed of 509 amino acid residues, which forms one SH3 domain, one SH2 domain, and one protein kinase domain (Figure 1A) [9]. Several tyrosine residues, Y192, Y394 and Y505, in Lck have been shown to be phosphorylated [10]. Y192 in the SH2 domain of Lck is phosphorylated in T cells upon T cell antigen receptor triggering, which reduces the capacity of the SH2 domain to bind its ligands [11]. Autophosphorylation of Y394 in the kinase activation loop of Lck increases the kinase activity of Lck [12], while the phosphorylation of Y505 in the C-terminal allows Y505 bind to the SH2 domain in the same protein, which down regulates its kinase activity [13]. A non-hierarchical graph description of Lck only shows that there are 3 domains and 3 tyrosines in Lck without any information of the relationship among the tyrosines and the domains. It would be difficult to interpret the mechanism of how a

phosphorylation on Y505 inhibits its kinase activity and similarly how Y192 affects the affinity of the SH2 domain to its binding partners without reading related literature.

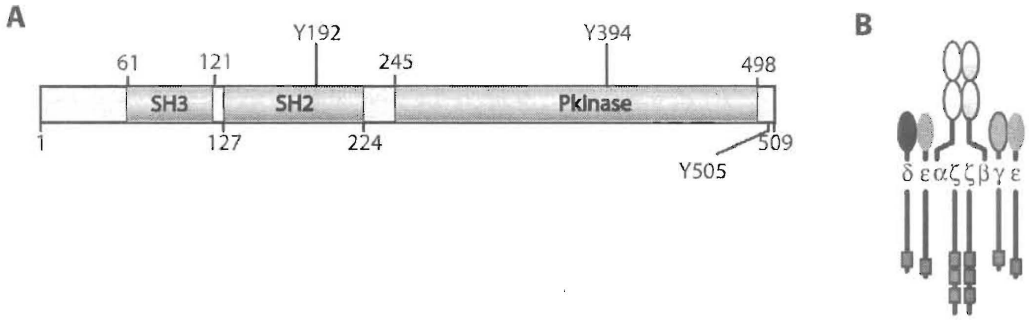

Figure 1. Representing a protein using hierarchical graph. a) The domain structure of Lck according to UniProt accession number P06239. Numbers indicate the starting and ending amino acid residue positions of domains in the protein. . Y192, Y394 and Y505 are tyrosine residuals at position 192, 394 and 505 respectively. SH3, SH2, and Pkinase are short names of Src homology 3, Src homolog 2, and protein kinase domain used in Pfam [24] respectively. b) The structure of the TCR/CD3 complex. The TCR is a heterodimer composed of two subunits α and β. The CD3 molecule has three different subunits: ε, δ, γ, and ζ. CD3ε can bind with either CD3δ or CD3γ, while CD3ζ binds itself. The heterodimers of CD3 subunits bind TCRαβ chains to form the TCR/CD3 complex. Black boxes represent ITAMs.

The second example is the TCR/CD3 complex (Figure 1B). The human T cell receptor complex (TCR/CD3 complex) is a molecule found on the surface of the T lymphocytes. It has two subunits TCRα and TCRβ. The extracellular part of these two subunits has a structure similar to immunoglobin (Ig), which allows TCR to recognize antigens bound to the major histocompatibity complex (MHC) molecules on an antigen presenting cell. The cytoplasmic parts TCRαβ are short and have no enzyme activities to further transfer the molecular recognition signal from the cell surface. To compensate this, the TCRαβ heterodimer bind to a protein named CD3. CD3 can bind some enzymes involved in the TCR activation process nonconvalently. The CD3 protein is consisting of CD3 εδ, εγ, and ζζ dimers. The basic residues within the transmembrane regions of TCRα and TCRβ may participate in assembly with CD3 and ζζ components [14-18]. Inside ε, δ, γ there are 2, 1, and 3 imunnoreceptor tyrosine-based activation motif (ITAM) respectively [19]. Those motifs are phosphorylated in seconds after TCR is activated. Within the ITAM in ε, there is a proline rich sequence (PRS) that may bind to SH3 domain containing proteins [20]. The Tyrosine at position 188 (Y188) is of special interest in that it belongs both to the PRS and ITAM. When phosphorylated, Y188 abolish the binding between the PRS to SH3 domain containing proteins. A non-hierarchical graph description of the TCRT/CD3 complex treats Y188 the same as PPS and ITAM. Modelers and readers will not be able to understand how Y188 may regulate which protein may bind to TCR/CD3 without enough prior knowledge. Such complex structural organizational information can be described in hierarchical graphs as we show in below.

A. Representation of a protein

Figure 2 demonstrates two equivalent hierarchical graph representations of Lck. Figure 2A uses a box representation for all the domains and sites. Figure 2B is a directed graph representation. In either graph, each domain is represented as a node. Boxes in Figure 2A and arrows in Figure 2B represent the containing or ownership relations to its contained boxes or linked objects respectively. From these graphs, it is easy to tell the relationships among various domains and residues in Lck. For example, from the nodes SH2 and Y192, one can tell Y192 locates in the SH2 domain of Lck, thus a phosphorylation on Y192 indicates the SH2 domain of the protein has been modified. Several conventions are used in Figure 2 to help different users to understand a molecule definition easily. One is that a gene name, instead of a protein name, is used because a protein may have many different protein names, yet in most cases only one common gene name. All domains are named after the corresponding domain names used in Pfam (http://pfam.sanger.ac.uk/) [24]. For example, Pkinase is the standard protein kinase family name used in

Pfam. Also domains and cites closer to the N-terminal are always on the left part of the graph while those close to the C-terminal are drawn on the right. Both of these two representations capture all the information of the protein sequence map showing in Figure 1A. Compared to a nonhierarchical BNGL representation of this molecule without tyrosine state specification is LCK(SH3, SH2, Pkinase, Y192, Y394, Y505), which tells us nothing about how these 3 different tyrosines relate to the three domains in Lck, a hierarchical graph defined by the BNGL string LCK(SH3,SH2(Y192),Pkinase(Y394),Y505) accurately reflects the internal relationships among those domains and tyrosines. Here, multiple braces are found inside Lck. Each closed braces is a site, a domain or a protein, from inner most to the outmost. In this representation, readers can tell that Y192 is inside a node named SH2, which is the SH2 domain and Y394 is part of the kinase domain. Readers can also tell that Y505 is a tyrosine located at the C-terminal side of the kinase domain, as long as we follow a convention of putting the N-terminal domains on left and C-terminal domains on right.

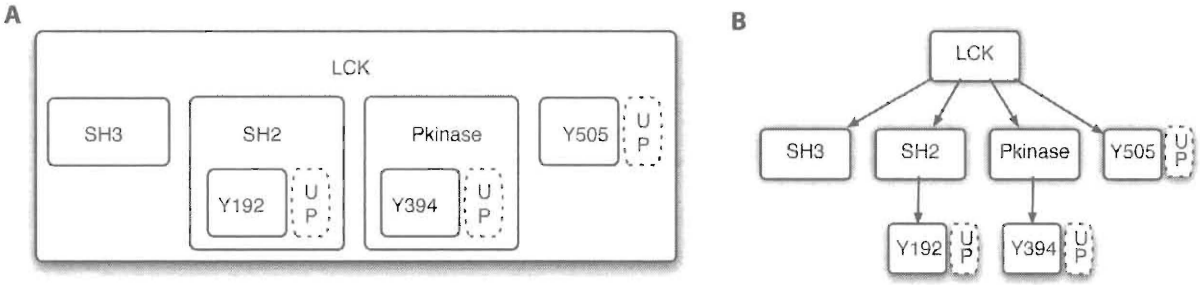

Figure 2. a) A box representation of Lck. Each domain or specific tyrosine sites is represented by a square with its name on the center. Boxes in dashed lines indicate possible states for the tyrosine sites on their left, where "U" means unphosphorylated and "P" means phosphorylated. b) A hierarchical representation of Lck . Arrows indicating the containing relationships. Each domain or site is represented as a node in the graph. Dashed squares surround possible states of specific tyrosine residues similar to a). These two graphs are equivalent. The BNGL string that encodes the hierarchical graph is: LCK(SH3, SH2(Y192~U~P),Pkinase(Y394(~U,~P)),Y505~U~P).

B. Representation of a multimeric protein

To represent molecule complex like TCR/CD3, we can decompose the complex into component molecules and each component molecule can be represented by a hierarchical subgraph, like Figure 2B. The hierarchical assembly of all these hierarchical graph into a bigger hierarchical graph become the final graph. The hierarchical graph of each component molecule becomes a subgraph of the final hierarchical graph. For example, Figure 3 shows a hierarchical graph representation of the TCR/CD3 complex. The top node indicates the name of this molecule complex is TCR/CD3. Nodes in the next layer show the names of each component molecules. In the third layer, each node represents a single polypeptide chain that forms the molecule on its owner node in the second layer. The fourth layer lists all the domains in those polypeptides and the fifth layer lists tyrosines that belong to their corresponding domains in the fourth layer. Thus complex molecules can be represented using hierarchical graphs. From this hierarchical graph it is obvious that Y188 appears in both the PRS and the ITAM domain of CD3ε chain. Thus it can be inferred that a modification on Y188 may directly regulates the biochemical properties of the PRS and the ITAM.

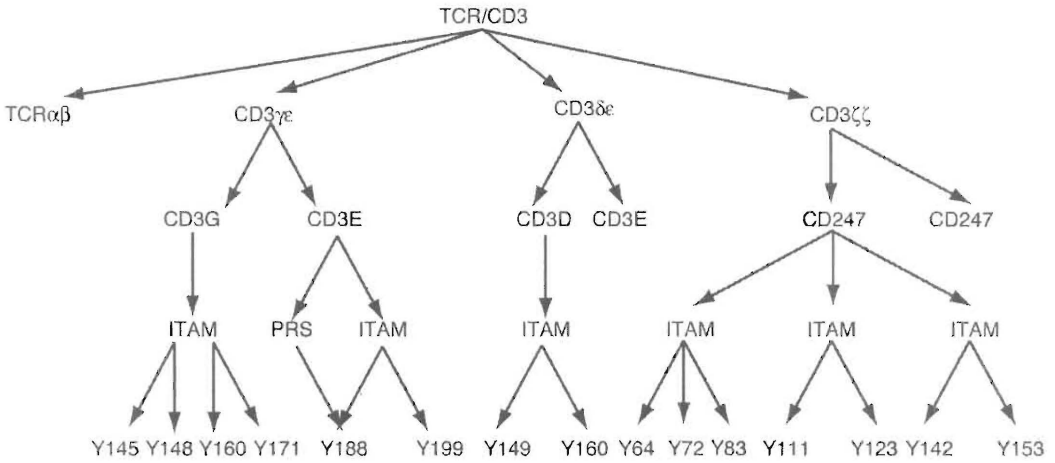

Figure 3. A hierarchical graph representation of the TCR/CD3 complex. Each of the first order components (two TCRs, CD3γε, CD3δε, CD3ζζ) is represented as a subgraph. The name of the first order component is shown in the top node of the corresponding subgraph. Arrows indicate components of lower order. Note that only one of each CD3 chain is drawn beyond the component molecules level. This is to clarify the graph.

2.2. *Representing relationships among reaction rules with a common reaction center*

In rule-based modeling modelers frequently define multiple reaction rules that share the same reaction center but have different contexts and different rate constants. The relationship among those reactions can also be captured in hierarchical graphs. For example, a protein may contain two tyrosine phosphorylation sites, S1 and S2. Coba et al [21] noticed that if the distance between S1 and S2 are within 10 amino acids of each other, the presence of phosphate on one site influences the phosphorylation on the other site, a phenomenon called priming effect. Two reaction rules are needed to represent the priming effect (R1 and R2 in Figure 4). The phosphorylation statuses of S2 are called the reaction context of these two reaction rules. If without the priming effect, that is the phosphorylation of S1 happens independently of the phosphorylation status of S2, this reaction can be called the context free reaction rule or the coarse reaction rule of the phosphorylation reaction center s1 (R0 in Figure 4). Three reactions (R0, R1 and R2) share the same reaction center S1. Priming effects are commonly found in biological signal transduction pathways. The reaction core (R0) may be treated as an abstract reaction pattern that helps modelers to sort reactions that share the same reaction centers in a model. For example, in Figure 4, the relationship between R1 and R2 is illustrated by a hierarchical graph. In this graph, both R1 and R2, are derivates from R0. Sorting reaction rules using hierarchical graphs makes the relationships among reaction rules clear and helps modelers to identify the effect of a certain context on the reaction center.

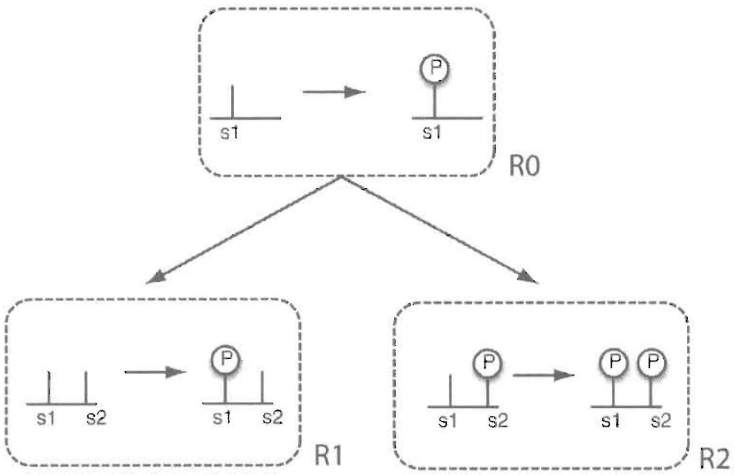

Figure 4. Relationships among reaction rules represented in hierarchical graph. S1 and S2 are phosphorylation sites. R0, R1 and R3 are three reactions that phosphorylates S1. Reaction R1 and R2 depend on the phosphorylation status of S2, while R0 does not. The BNGL code for these three reactions are: A(s1~u)-> A(s1~p), A(s1~u,s2~u) -> A(s1~p,s2~u), and A(s1~u,s2~p)->A(s1~p,s2~p) respectively. Tildes (~) indicate specific states of a site.

2.3. *XML-based Hierarchical Graph Representation*

A challenge in applying hierarchical graph for modeling biological systems is how to exchange models defined in hierarchical graphs to existing rule-based modeling software and non-rule-based modeling software. One convenient method of store and exchange hierarchical graph is via using XML. A prominent advantage of the XML-based model element definition over the traditional text-based BNGL format is that element definitions in XML are self-annotating and easy understand. In Appendix, we list an XML implementation of the hierarchical graph of LCK (Figure 1A). The BNGL XML with hierarchical graph extension contains a list of nesting objects, which can be either proteins or reaction rules. There is a <meta> tag to encode the minimum information requested in annotation of biochemical models (MIRIAM) [22]. Information about sequence length, and position information of domains and sites can also be extracted easily from this XML format so that they can be used by protein domain graph drawing software like DOG [23] and other bioinformatics tools. By referencing accession numbers of common biological database like Pfam [24], UniProt [25] and HPRD [26] in hierarchical graph, it is easier to integrate simulation environment to on-line database for visualizing computational models and easy access to external annotation pages. Information stored in XML files can also be mapped to a dedicated database so that model elements, including their documentation information, can be efficiently reused in different models. As we have discussed in our previous work [27], such an integration would be an ideal method to provide reusable model elements that are easy to understand and share.

3. Discussion

In this work we demonstrate an annotation improvement of rule-based modeling by adopting hierarchical graphs. Compared to non-hierarchical graphs, hierarchical graph representations keep the structural organization of protein domains and sites and are easy to understand. Relationships among reactions sharing a same reaction center can be identified easily if reaction rules are presented in hierarchical graph (Figure 4). We also propose some conventions in using hierarchical graphs to represent protein molecules. For example, UnirProt gene names are used to represent proteins. Domains closer to the N-terminal of a protein are presented by nodes on the left to those of the domains closer to the C-terminal. All protein domains, whenever a Pfam name is available, a Pfam name is used. A

specific amino acid site shall be represented by a 1-letter amino acid abbreviation followed by a number indicating the position of this site in the protein sequence. We show by examples how hierarchical graphs can be defined in BNGL, a text-based language for rule-based modeling software BioNetGen [4, 28].

Biological signal transduction pathways are often built upon common signal modules [29], and the same proteins may appear in different pathways. For example, many proteins that involved in the T-cell activation process are also involved in B-cell activation and platelet signaling. If a model element from a specific model can provide adequate information about its structural organization with annotation and links to various biological knowledgebase, it can be easily imported to different models to make future model development easier. Text-based language, like BNGL, is not suitable for annotation purpose. It shall be noted that the BNGL XML format is designed to be used for model element annotation and not a replacement of the current BNGL. Files in this XML format is not executable and shall be used as attachments to models defined in the text-based BNGL. There is a possibility to include hierarchical graphs and map it to a biological knowledgebase in future versions of GetBonNie, a web-based tool that allows user to build, analyze and share rule-based models (www.getbonnie.org) [30]. BNGL XML coded model elements may be exposed to users via web services in GetBonNie also. The Systems Biology Markup Language (SBML) with the multistate and multi-component species package, which is currently under development by the SBML community, may be used for model specification definition in future versions of BioNetGen and GetBonNie [31]. We are also considering provide a tool to exchange the BNGL XML to SBML so that modelers do not use the rule-based modeling technique may also benefit from the new annotation method that we are proposing here. The hierarchical graph-based BNGL XML model element annotation opens the potential for various modelers to share and reuse model elements and reuse.

Acknowledgments

This work was supported by NIH [grant GM076570] and DOE [contract DE-AC52-06NA25396].

Appendix

XML coding of protein LCK.

```
<?xml version="1.0" encoding="UTF-8"?>
<bngl extension="hierarchical graphs" version="0.1">
 <meta>...</meta>
 <listOfObjects>
  <!-- protein level -->
  <object type="protein, polypeptide chain"
  name="lymphocyte specific cytoplasmic kinase (Lck)" bngl="LCK">
   <sequence source="http://www.uniprot.org/uniprot/P06239" start="1" stop="509">
     MGCGCSSHPEDDWMENIDVCENCHYPIVPLDGKGTLLIRNGSEVRDPLVTYEGSNPPASPLQDNLVIA
     LHSYEPSHDGDLGFEKGEQLRILEQSGEWWKAQSLTTGQEGFIPFNFVAKANSLEPEPWFFKNLSRKD
     AERQLLAPGNTHGSFLIRESESTAGSFSLSVRDFDQNQGEVVKHYKIRNLDNGGFYISPRITFPGLHEL
     VRHYTNASDGLCTRLSRPCQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTK
     VAVKSLKQGSMSPDAFLAEANLMKQLQHQRLVRLYAVVTQEPIYIITEYMENGSLVDFLKTPSGIKLT
     INKLLDMAAQIAEGMAFIEERNYIHRDLRAANILVSDTLSCKIADFGLARLIEDNEYTAREGAKFPIKW
     TAPEAINYGTFTIKSDVWSFGILLTEIVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLMRL
     CWKERPEDRPTFDYLRSVLEDFFTATEGQYQPQP
   </sequence>
   <localization>Cytoplasm</localization>
   <externalLinks>
<url resource="uniprot">http://www.uniprot.org/uniprot/P06239</url>
   </externalLinks>

   <listOfObjects>
<!-- SH3 domain-->
```

```
8
 <object type="protein interaction domain" name="Src homology 3 (SH3) domain"
bngl="SH3">
 <sequenceRef ref="LCK" start="61" stop="121"/>
 <localization>Cytoplasm</localization>
 <externalLinks>
  <url resource="pfam">http://pfam.sanger.ac.uk/family?acc=PF00018</url>
  <url resource="prosite">http://www.expasy.org/cgi-bin/prosite-search-ac?PDOC50002</url>
 </externalLinks>
</object>

<!-- SH2 domain-->
<object type="protein interaction domain" name="Src homology 2 (SH2) domain"
bngl="SH2">
 <sequence ref="LCK" start="127" stop="224"/>
 <localization>Cytoplasm</localization>
 <externalLinks>
  <url resource="pfam">http://pfam.sanger.ac.uk/family?acc=PF00017</url>
  <url resource="prosite">http://www.expasy.org/cgi-bin/nicedoc.pl?PDOC50001</url>
 </externalLinks>

 <listOfObjects>
  <!-- Y192 in the SH2 domain-->
  <object type="amino acid residue"
  name="conserved tyrosine in the end of beta-strand E of the SH2 domain"
  bngl="Y192">
   <sequence ref="LCK" start="192" stop="192"/>
   <localization>Cytoplasm</localization>
   <listOfStates>
<state type="phosphorylation status" name="unphosphorylated"
    bngl="u"/>
<state type="phosphorylation status" name="phosphorylated"
    bngl="p"/>
   </listOfStates>
   <comment>
Phosphorylation of Y192 can be catalyzed by Syk and possibly Zap nonreceptor kinases, leads to a profound down-
regulation of the ligand binding capacity of the SH2 domain.
<listOfReferences>
 <reference pmid="8798764"/>
</listOfReferences>
   </comment>
  </object>
 </listOfObjects>
</object>

<!-- Pkinase domain-->
<object type="protein domain, catalytic" name="protein tyrosine kinase domain"
bngl="Pkinase">
 <sequence ref="LCK" start="245" stop="498"/>
 <localization>Cytoplasm</localization>
 <externalLinks>
  <url resource="pfam">http://pfam.sanger.ac.uk/family?acc=PF00069</url>
 </externalLinks>

 <listOfObjects>
  <!-- Y394 in the Pkinase domain -->
```

```
   <object type="amino acid residue" name="autophosphorylation loop tyrosine"
   bngl="Y394">
    <sequence ref="LCK" start="394" stop="394"/>
    <localization>Cytoplasm</localization>
    <listOfStates>
<state type="phosphorylation status" name="unphosphorylated" bngl="u"/>
<state type="phosphorylation status" name="phosphorylated" bngl="p"/>
    </listOfStates>
    <comment>
Fully activated Lck requires Y394 being phosphorylated
<listOfReferences>
 <reference pmid="17719247"/>
 <reference pmid="7486706"/>
</listOfReferences>
    </comment>
   </object>
 </listOfObjects>
</object>

<!-- Y505 -->
<object type="amino acid residue"
name="C-terminal tyrosine of Src-family kinase"
bngl="Y505">
 <sequence ref="LCK" start="505" stop="505"/>
 <localization>Cytoplasm</localization>
 <listOfStates>
  <state type="phosphorylation status" name="unphosphorylated" bngl="u"/>
  <state type="phosphorylation status" name="phosphorylated" bngl="p"/>
 </listOfStates>
 <comment>
  If phosphorylated, Y505 can bind with the SH2 domain of the same Lck molecule and inhibits its kinase activity.
Y505 is phosphorylated by Csk (Bergman et al 1992) and dephosphorylated by CD45 (Ostergaard et al 1989).
  <listOfReferences>
   <pubmed id="Bergman et al 1992" pmid="1639064"/>
   <pubmed id="Ostergaard et al 1989" pmid="2530588"/>
  </listOfReferences>
 </comment>
</object>

    </listOfObjects>
   </object>
 </listOfObjects>
</bngl>
```

References

1.Faeder, J.R., M.L. Blinov, and W.S. Hlavacek. *Graphical rule-based representation of signal-transduction networks.* in *The 20th ACM Symposium on Applied Computing*. 2005. Santa Fe, NM.

2.Hlavacek, W.S., et al., *Rules for modeling signal-transduction systems.* Sci STKE, 2006. **2006**(344): p. re6.

3.Blinov, M.L., et al., *Graph Theory for Rule-Based Modeling of Biochemical Networks*. Lecture Notes in Computer Science. Vol. 4230/2006. 2006, Springer Berlin / Heidelberg: Springer.

4.Faeder, J.R., M.L. Blinov, and W.S. Hlavacek, *Rule-based modeling of biochemical systems with BioNetGen.* Methods Mol Biol, 2009. **500**: p. 113-67.

5.Conzelmann, H., D. Fey, and E.D. Gilles, *Exact model reduction of combinatorial reaction networks.* BMC Syst Biol, 2008. **2**: p. 78.
6.Feret, J., et al., *Internal coarse-graining of molecular systems.* Proc Natl Acad Sci U S A, 2009. **106**(16): p. 6453-8.
7.Straus, D.B. and A. Weiss, *Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor.* Cell, 1992. **70**(4): p. 585-93.
8.Salmond, R.J., et al., *T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance.* Immunol Rev, 2009. **228**(1): p. 9-22.
9.Sefton, B.M., *The lck tyrosine protein kinase.* Oncogene, 1991. **6**(5): p. 683-6.
10.Oetken, C., et al., *Induction of hyperphosphorylation and activation of the p56lck protein tyrosine kinase by phenylarsine oxide, a phosphotyrosine phosphatase inhibitor.* Mol Immunol, 1994. **31**(17): p. 1295-302.
11.Couture, C., et al., *Regulation of the Lck SH2 domain by tyrosine phosphorylation.* J Biol Chem, 1996. **271**(40): p. 24880-4.
12.Veillette, A. and M. Fournel, *The CD4 associated tyrosine protein kinase p56lck is positively regulated through its site of autophosphorylation.* Oncogene, 1990. **5**(10): p. 1455-62.
13.Veillette, A., et al., *Regulation of the enzymatic function of the lymphocyte-specific tyrosine protein kinase p56lck by the non-catalytic SH2 and SH3 domains.* Oncogene, 1992. **7**(5): p. 971-80.
14.Morley, B.J., et al., *The lysine residue in the membrane-spanning domain of the beta chain is necessary for cell surface expression of the T cell antigen receptor.* J Exp Med, 1988. **168**(6): p. 1971-8.
15.Alcover, A., et al., *Lysine 271 in the transmembrane domain of the T-cell antigen receptor beta chain is necessary for its assembly with the CD3 complex but not for alpha/beta dimerization.* J Biol Chem, 1990. **265**(7): p. 4131-5.
16.Blumberg, R.S., et al., *Assembly and function of the T cell antigen receptor. Requirement of either the lysine or arginine residues in the transmembrane region of the alpha chain.* J Biol Chem, 1990. **265**(23): p. 14036-43.
17.John, S., et al., *Surface expression of the T cell receptor complex requires charged residues within the alpha chain transmembrane region.* Eur J Immunol, 1989. **19**(2): p. 335-9.
18.Call, M.E. and K.W. Wucherpfennig, *Common themes in the assembly and architecture of activating immune receptors.* Nat Rev Immunol, 2007. 7(11): p. 841-50.
19.Reth, M., *Antigen receptor tail clue.* Nature, 1989. **338**(6214): p. 383-4.
20.Kesti, T., et al., *Reciprocal regulation of SH3 and SH2 domain binding via tyrosine phosphorylation of a common site in CD3epsilon.* J Immunol, 2007. **179**(2): p. 878-85.
21.Coba, M.P., et al., *Kinase networks integrate profiles of N-methyl-D-aspartate receptor-mediated gene expression in hippocampus.* J Biol Chem, 2008. **283**(49): p. 34101-7.
22.Le Novere, N., et al., *Minimum information requested in the annotation of biochemical models (MIRIAM).* Nat Biotechnol, 2005. **23**(12): p. 1509-15.
23.Ren, J., et al., *DOG 1.0: illustrator of protein domain structures.* Cell Res, 2009. **19**(2): p. 271-3.
24.Finn, R.D., et al., *The Pfam protein families database.* Nucleic Acids Res, 2008. **36**(Database issue): p. D281-8.
25.Consortium, U., *The Universal Protein Resource (UniProt) 2009.* Nucleic Acids Res, 2009. **37**(Database issue): p. D169-74.
26.Keshava Prasad, T.S., et al., *Human Protein Reference Database--2009 update.* Nucleic Acids Res, 2009. **37**(Database issue): p. D767-72.
27.Hu, B., et al., *GetBonNie for building, analyzing and sharing rule-based models.* Bioinformatics, 2009.
28.Blinov, M.L., et al., *BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains.* Bioinformatics, 2004. **20**(17): p. 3289-91.
29.Hartwell, L.H., et al., *From molecular to modular cell biology.* Nature, 1999. **402**(6761 Suppl): p. C47-52.
30.Hu, B., et al., *GetBonNie for building, analyzing and sharing rule-based models.* Bioinformatics, 2009. **25**(11): p. 1457-60.
31.Hucka, M., et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.* Bioinformatics, 2003. **19**(4): p. 524-31.