Applications of Pathway Logic Modeling to Target Identification

Anupama Panikkar[1], Merrill Knapp[4], Huaiyu Mi[2], Dave Anderson[1], Krishna Kodukula[1], Amit K. Galande[1] and Carolyn Talcott[2]

[1]*Center for Advanced Drug Research, SRI International, Harrisonburg, Virginia 22802, the* [2]*Computer Science Lab, the* [4]*Biosciences Division, SRI International, Menlo Park, CA 94025*

{*firstname.lastname*}*@sri.com*

Abstract. To explore the role of proteases in pathogenesis and as potential drug targets we need to elucidate their function and effect on biological networks. In this paper, we describe the application of Pathway Logic (PL) (http://pl.csl.sri.com/) to the *symbolic* modeling of the interaction networks of proteases of Gram-positive bacteria and the use of Pathway Logic Assistant tool to browse and query these models. Pathway Logic is a systems biology approach to biological processes as integrated systems rather than isolated parts based on formal methods and rewriting logic. These models are developed using Maude, a formal language and tool set based on rewriting logic. We show how this approach can be used to represent and analyze systems at multiple levels of details. The Pathway Logic Assistant tool (PLA) enables us to identify key proteases and regulatory molecules – 'choke points' by comparing different pathways or networks within and across species and to predict how these molecules, if inhibited or avoided would affect the pathway or network.

1 Introduction

The emergence of Gram-positive drug resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA), *Streptococcus pneumoniae* and Enterococcus represent a serious public health problem. This might be resolved by using a new generation of antibiotics with a different mechanism of action. Proteases (also termed peptidases or proteinases) are enzymes that break down proteins by hydrolysis of peptide bonds in the proteins. Bacterial proteases are implicated in virtually every important biological process related to colonization and evasion of host immune defenses, acquisition of nutrients for growth and proliferation, or tissue damage during infection. Thus, bacterial proteases represent suitable drug targets and inhibition of these enzymes would retard the growth and proliferation of invading pathogens [1–3]. It is critical to understand the role of proteases by modeling them in the context of multiple components such as protein signaling networks and complex biochemical pathways that can influence their activity. In this study, we use the Pathway Logic (PL) [4–6] framework to *symbolically* model the protease networks and pathway interconnectivity of multiple Gram-positive bacteria including pathogenic and non-pathogenic species in an effort to develop a comprehensive computational model.

Pathway Logic is a *symbolic* systems approach to the modeling and analysis of molecular and cellular processes based on rewriting logic [7]. *Symbolic* systems biology is the *qualitative* and *quantitative* study of biological processes as integrated systems rather than as isolated parts. An important objective of Pathway Logic is to reflect the ways that biologists think about problems using informal models, and to provide bench biologists with tools for computing with and analyzing these models. Symbolic/logical models allow one to represent partial information and to model and analyze systems at multiple levels of details, depending on information available and questions to be studied.

Pathway Logic models are curated from the literature, and written and analyzed using Maude [8, 9] (http://maude.cs.uiuc.edu/), a rewriting-logic-based formalism. The Rewriting logic formalism is based on states of a system represented as elements of an algebraic data type and the behavior of a system given by local transitions between states described by *rewrite rules*. A rewrite rule has the form $t \Rightarrow t'$ if c where t and t' are patterns (terms possibly containing place holder variables) and c is a condition (a boolean term). Such a rule applies to a system in state s if t can be matched to a part of s by supplying the right values for the place holders, and if the condition c holds when supplied with those values. The process of application of rewrite rules generates computations (also thought of as deductions) and in case of biological processes these computations correspond to pathways. In Pathway Logic, algebraic data types

are used to represent concepts from cell biology needed to model signaling processes, including intracellular proteins, biochemicals such as second messengers, extracellular stimuli, biochemical modification of proteins, protein association, and cellular compartmentalization of proteins. Rewrite rules describe the behavior of proteins and other components depending on modification state and biological context. Each rule represents a step in a biological process such as metabolism or intra/inter-cellular signaling. A specific model is assembled by specifying an initial state (called a dish): the cells, their components, and entities such as ligands in the supernatant. Pathway Logic models are executable – hence they can be used for simulation. In addition, the Maude system provides search and model-checking capabilities. Using the search capability all possible future states of a system can be computed to show its evolution from a given initial state (specified by the states of individual components) in response to a stimulus or perturbation. Using model-checking a system in a given initial state can be shown to never exhibit pathways with certain properties, or the model-checker can be used to produce a pathway with a given property (by trying to show that no such pathway exists).

A Pathway Logic knowledge base includes data types representing cellular components such as proteins, small molecules, complexes, compartments/locations protein state, and post-translational modifications. Modifications can be as being activated, inhibited, phosphorylated, degraded or anchored. It also enable one to collect, store and retrieve curated information represented as metadata so that it can be understood and shared by a community of experimental biologists.

The Pathway Logic Assistant (PLA) [10] provides an interactive visual representation of PL models. Using the Pathway Logic Assistant (PLA) one can display pathways of interest, compare two pathways, search for cross talk between subsystems by exploring subnets, map gene expression data onto signaling networks and compute the effects of system perturbations by single or double knockouts (omission of individual or pairs of proteins that prevents reaching a specified state).

The remainder of the paper is organized as follows. The basic ideas of Pathway Logic are presented in §2, and illustrated with fragments from a model of heme transport in *Staphylococcus aureus*. Use of the Pathway Logic Assistant tool to browse and query models is discussed in §3. Applications of Pathway Logic in target discovery are shown by few examples in §4. The paper concludes with a discussion of future directions in §5.

2 Pathway Logic Basics

Pathway Logic models are structured in four layers: (1) sorts and operations, (2) components, (3) rules, and (4) dishes and queries. The *sorts and operations* layer declares the main sorts and subsort relations, the logical analog to ontology or class hierarchy. The sorts of entities include Chemical, Protein, Gene, Complex, Location (cellular compartments), and Cell. These are all subsorts of the sort, Soup that represents 'liquid' mixtures, as multisets (unordered collections) of entities. The sort Modification is used to represent post-translational protein modifications and gene regulations including up-regulation and down-regulation in the bacteria. They can be abstract, to specify that a protein is activated, inhibited, bound, anchored, degraded, phosphorylated, dephosphorylated, or more specific, for example, phosphorylation at a particular site. Gene up-regulation specifies increased expression of gene and their encoded protein and gene down-regulation indicates decreased gene and corresponding protein expression. Modifications are applied using the operator [–]. For example the term `[IsdA — anchored]` represents the iron-responsive surface determinant A (IsdA) protein in an anchored state and `[ClpP-gene — on]` represents *ClpP* gene in its 'on' state (upregulated).

A cell state is represented by a term of the form

`[cellType | locs]`

where `cellType` specifies the type of cell and `locs` represents the contents of a cell organized by cellular location. Each location is represented by a term of the form `{ locName | components }` where `locName` identifies the location. In gram-positive bacteria the locations defined are

`CLm` for cell membrane

`CLc` for cytosol
`CLw` for cell wall
`Xout` for outside of the bacterial cell and

`components` stands for the mixture of proteins, genes and other compounds in that location.

The *components* layer specifies particular entities (proteins, genes, chemicals) and introduces additional sorts for grouping proteins in families. The *rules* layer contains rewrite rules specifying individual reaction steps. In the case of signal transduction, rules represent processes such as activation, phosphorylation, complex formation, or translocation. The sorts and operations, components, and rules layers make up a Pathway Logic knowledge base. The *dishes and queries* layer specifies initial states, relative to which queries can be answered, and properties of states to be used in formulating queries. Initial states are in silico Petri dishes containing a cell, with its components, and ligands of interest in the supernatant.

We give a brief overview of the representation in Maude of bacterial intracellular processes, illustrated using a model of heme transport involving the membrane cysteine protease-transpeptidases Sortase A (SrtA) and Sortase B (SrtB) in the pathogenic bacterium *S. aureus* in the following §2.1.

2.1 Modeling Heme transport involving SrtA and SrtB protease-transpeptidases in Pathway Logic

Pathogenic bacteria require iron as a source of nutrient during the infection process. *S. aureus* utilizes heme (a non-protein chemical compound that contains an iron atom) as a source of iron for its growth during infection. It acquires heme from the host environment and transports it across the cell wall into the cytoplasm by the heme-binding Isd proteins. The passage of heme also requires two sortases namely SrtA and SrtB that anchor these heme-binding Isd proteins to the cell wall. SrtA anchors IsdA, IsdB and IsdH proteins and SrtB anchors IsdC protein to the cell wall. As shown in Fig. 2, the heme binds to IsdA, IsdB and IsdH proteins, which is then transported to the membrane transport system composed of IsdDEF into the cytoplasm. In the cytoplasm the heme is degraded by the IsdG and IsdI heme monooxygenases, releasing the free iron for use by the bacterium as a nutrient source. The Pathway Logic model of heme transport in *S. aureus* was curated based on papers from Olaf Schneewind's lab [11–13] and many other references (cited as metadata associated with individual rules). In the following we show an initial state for study of heme transport and examples of rules and briefly sketch some of the ways one can compute with the model. The initial state (called `Hemetransport`) is a dish `PD( ... )` with a single cell represented by the following

```
Dish: PD([Cell |
  {XOut | Heme}
  {CLm | IsdD IsdE IsdF SrtA SrtB}
  {CLc | IsdG IsdI}
  {CLw | IsdA IsdB IsdC IsdH}])
```

Here, the dish contains Heme in the outside environment (location tag `XOut`). The cell membrane (tag `CLm`) has proteins IsdD, IsdE, IsdF, SrtA and SrtB. The cell wall (tag `CLw`) contains IsdA, IsdB, IsdC, and IsdH. The cytosol (tag `CLc`) contains proteins IsdG and IsdI.

The following is an example of the rule representing heme uptake by IsdB.

```
rl[5]:
  {XOut | xout Heme  }
  {CLw | clw [IsdB - anchored]  }
  =>
  {XOut | xout                    }
```

```
    {CLw | clw ([IsdB - anchored] : Heme) } .
```

As shown in Fig. 2, applying rules 1-4 to the initial dish results in a dish with IsdA, IsdB, IsdC and IsdH anchored, and applying rule 5 to this dish results in

```
Dish: PD([Cell |
  {XOut | empty}
  {CLm | IsdD IsdE IsdF SrtA SrtB}
  {CLc | IsdG IsdI}
  {CLw | (Heme : [IsdB - anchored]) [IsdA - anchored] [IsdC - anchored]
         [IsdH - anchored]}])
```

Maude [8, 9] can be used to find some execution, or to search for a state; for example, a state with Fe3+ in the `CLc`. However, the textual representation of cell states and pathways quickly becomes difficult to use as the size of a model grows, and an intuitive graphical representation becomes increasingly important. In addition, it becomes important to take advantage of the simple structure of PL models when searching for paths and carrying out other analyses. In the next section we show how the Pathway Logic Assistant can be used to visualize a model as a network of reaction rules, to browse the network, and to specify and execute queries.

3 The Pathway Logic Assistant

The Pathway Logic Assistant (PLA) [10] provides an interactive graphical view of a PL knowledge base. A PL knowledge base uses the petri net transition representation of the Maude rules. A model is then generated by specifying a dish (initial state). Petri nets have a natural graphical representation, and additionally, there are very efficient tools for analyzing the Petri net models generated by PLA. Our Petri net models are a special case of Place-Transition Nets given by a set of occurrences (places in Petri net terminology) and a set of transitions [14]. Occurrences can be thought of as atomic propositions asserting that a protein (in a given state) or other component occurs in a given compartment. For example Heme outside a bacterial cell is represented by the occurrence `<Heme, Xout>` and IsdB anchored in the cell wall is represented by `<[IsdB - anchored], CLw>`. A system state is represented as a set of occurrences (called a marking in Petri net terminology), giving the propositions that are true. A transition is a pair of sets of occurrences. A transition can fire if the state contains the first set of occurrences. In which case the first set of occurrences is replaced by the second set. For example the rule labeled [5] shown in §2 becomes the transition

```
pnTrans[5]:
  <Heme,XOut> <[IsdB - anchored],CLw>
  =>
  <Heme :[IsdB - anchored],CLw > .
```

In PLA goal properties are Petri net properties expressed as occurrences that must be present (places to be marked) and avoids properties are occurrences that must not appear (places not to be marked) in a computation. Paths leading from an initial state to a state satisfying a set of goals can be represented compactly as a Petri net consisting of the transitions fired in the path, thus giving query results a natural graphical representation. Execution of the path net starting with the initial state, leads to a state satisfying the goals, and the net representation makes explicit the dependency relations between transitions: some can fire concurrently (order doesn't matter), and some require the output of other transitions to be enabled.

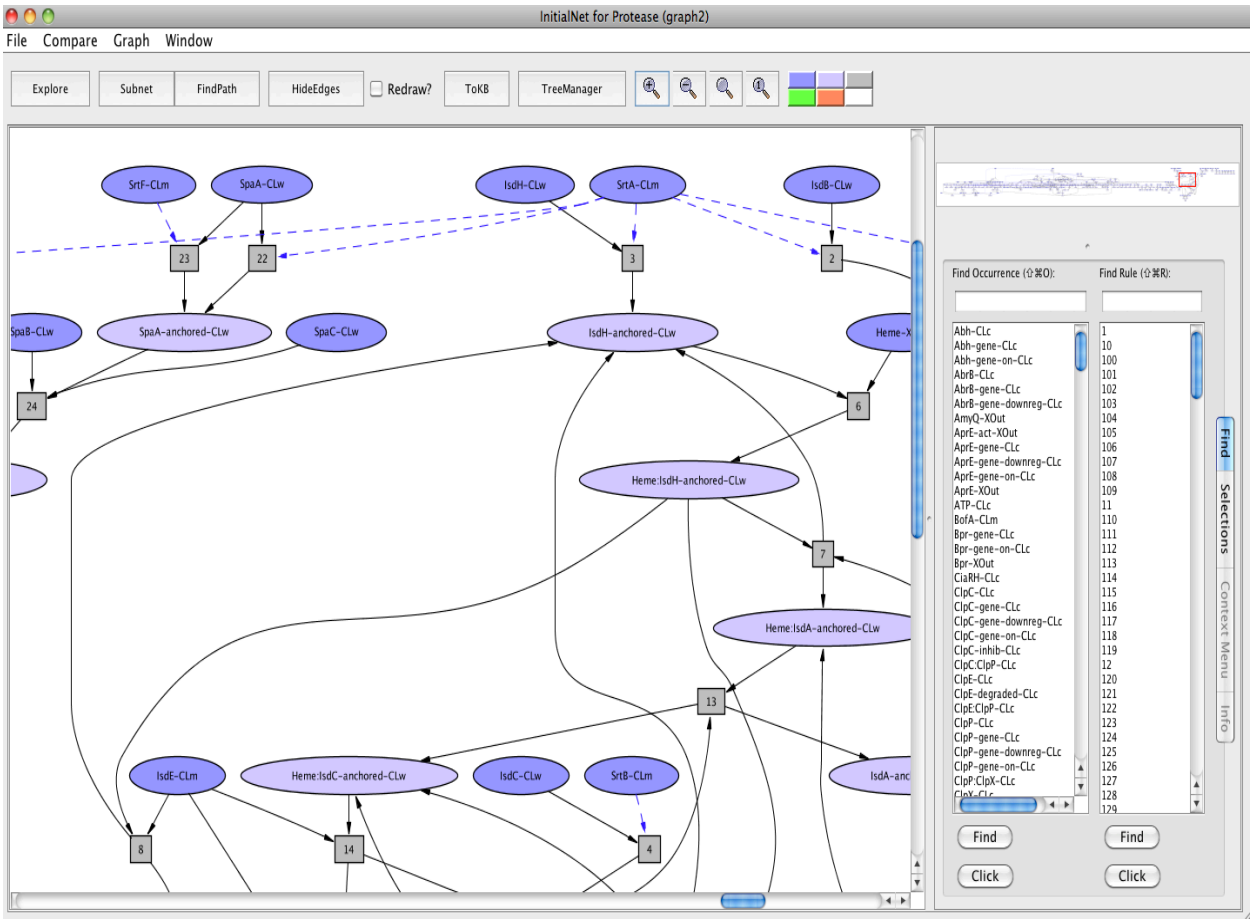

Fig. 1. Protease network model as Petri net viewed in PLA. Ovals are occurrences, with initial occurrences darker. Rectangles are transitions. Dashed arrows indicate an occurrence that is both input and output. The full protease net is shown in the upper right thumbnail. A magnified view of the portion in the red rectangle is shown in the main view.

Fig. 1 shows a screen shot of the Petri net model of the protease network generated by PLA. Ovals are occurrences, with initial occurrences darker. Rectangles are transitions. Dashed arrows indicate an occurrence that is both input and output. The thumbnail sketch in the upper right shows the full network. The main frame shows a magnified version of the portion of the network in the red rectangle. The Finder in the lower right allows one to locate occurrences and rules by name, and center the view on the selected node. To make a query, goals and avoids can be specified either by clicking on the occurrence and selecting goal or avoid in the selection window that appears, or by using the selection window directly. Once goals and avoids have been specified the user can ask to see the relevant subnet or to find a path. The relevant subnet contains all of rules needed for any (minimal) pathway satisfying the query, while the path is just the first path found by the analysis tool. Fig. 2 shows the path found in response to the query in which the goal is Fe3+ in the cytosol (`<Fe3+, CLc>`) and there are no avoids.

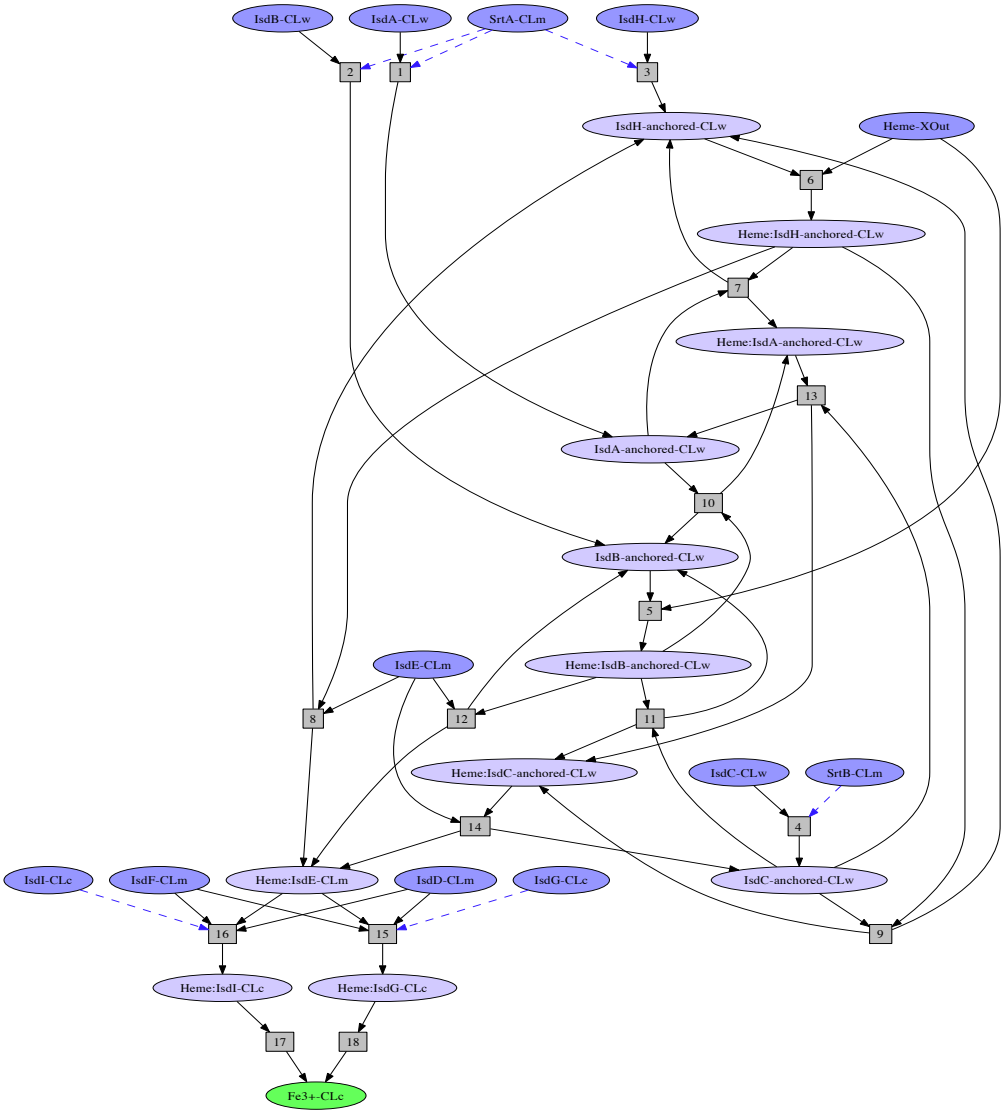

Fig. 2. Heme transport pathway in *S. aureus*. This pathway shows the transfer of heme by the Sortase anchored Isd proteins from outside of the bacterial cell to the cytosol and subsequent release of Fe3+ in the cytosol for use by the bacterium as a nutrient source.

4 Applications of Pathway Logic in target discovery

We describe here, few applications of Pathway Logic in target discovery. Identifying choke points or key molecules in the pathogens can accelerate the process of drug discovery. Choke points are critical points in a network and inactivation of choke points may lead to an organism's failure to produce or consume particular metabolites, which could cause serious problems for fitness or survival of the organism [15]. Potential drug targets are proposed based on the analysis of these choke points in the bacterial network.

4.1 Comparative network analysis and target inhibition

In Pathway Logic, one can identify choke points or key molecules by comparing different pathways or networks within and across species and target these molecules participating in pathogenesis. In addition

to generating subnets and pathways, two subnets and/or pathways can be compared. For this, the two networks are merged into one and color-coded. Fig. 3 shows the result of comparing two pathways from two bacterial species - the heme transport in *S. aureus* [12, 13] and the pilus assembly in *Corynebacterium diphtheriae* [16]: the heme transport (purple/darker color), and pilus assembly (blue-green/lighter color). The common part (SrtA) is peach colored.

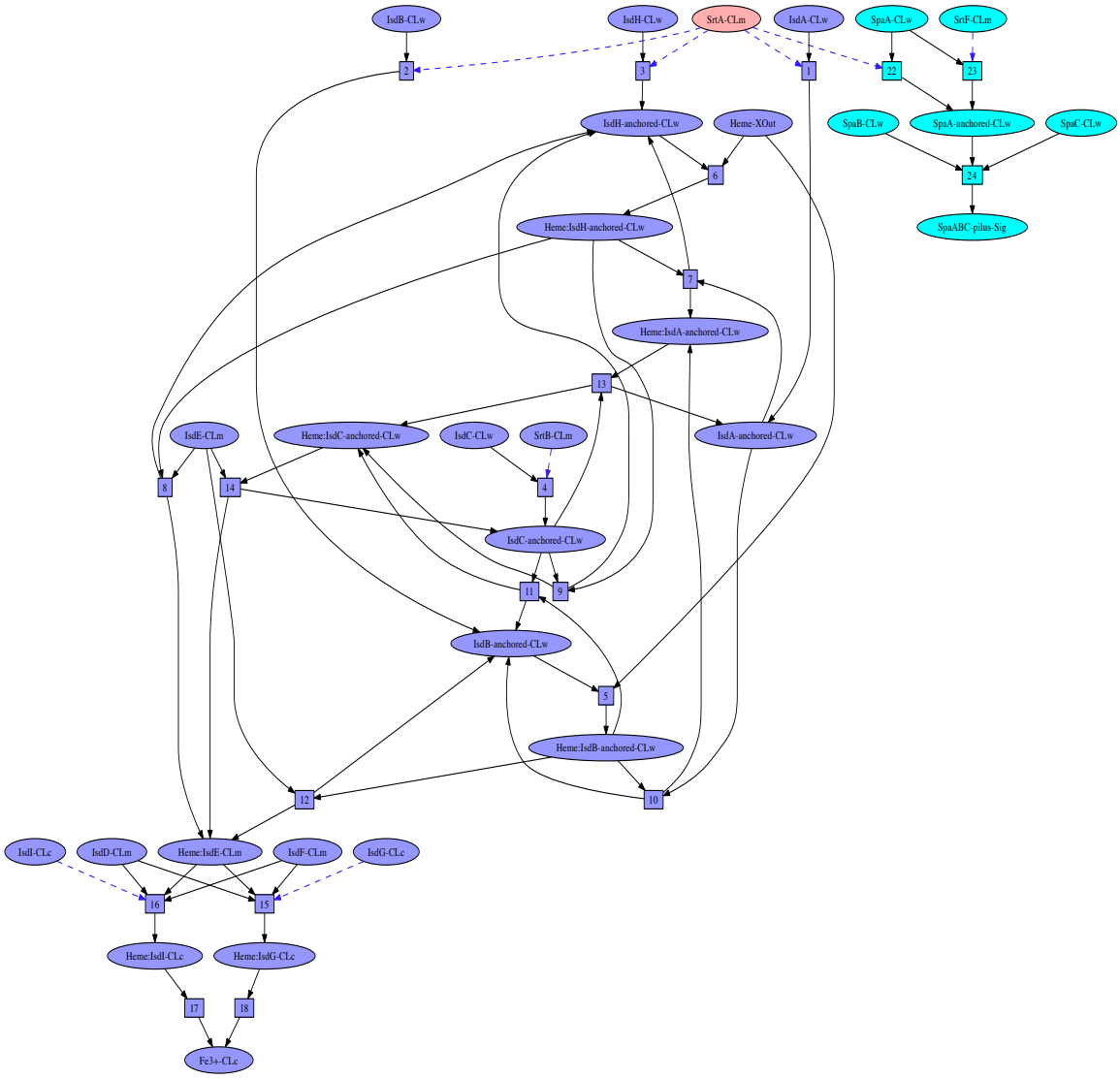

Fig. 3. Comparison of two pathways from *S. aureus* and *C. diphtheriae* - the heme transport (purple/darker color), and the pilus assembly (blue-green/lighter color). The common part (SrtA) is peach colored.

The observed sharing of SrtA suggests asking what happens if it is removed? Fig. 4 shows the subnet with SrtA avoided and viewing the result in the context of the heme-pilus subnet. The model shows inhibition of SrtA affects the heme transport pathway in *S. aureus.* This inhibition model corroborated with an earlier study showing that inactivation of SrtA lead to a decrease in the amount of iron associated with *S. aureus*

cells. Biological studies have shown that the knockout mutation of *SrtA* gene in *S. aureus* greatly reduces the capacity of the pathogen to establish an acute infection in mice [17, 18]. Thus, in the growing antibiotic resistant scenario, SrtA may prove exciting new target of anti-infective therapy [19]. SrtA did not inhibit the pili formation due to the occurrence of another redundant pathway in which SrtF, a house keeping sortase, is involved in the Spa pili assembly.

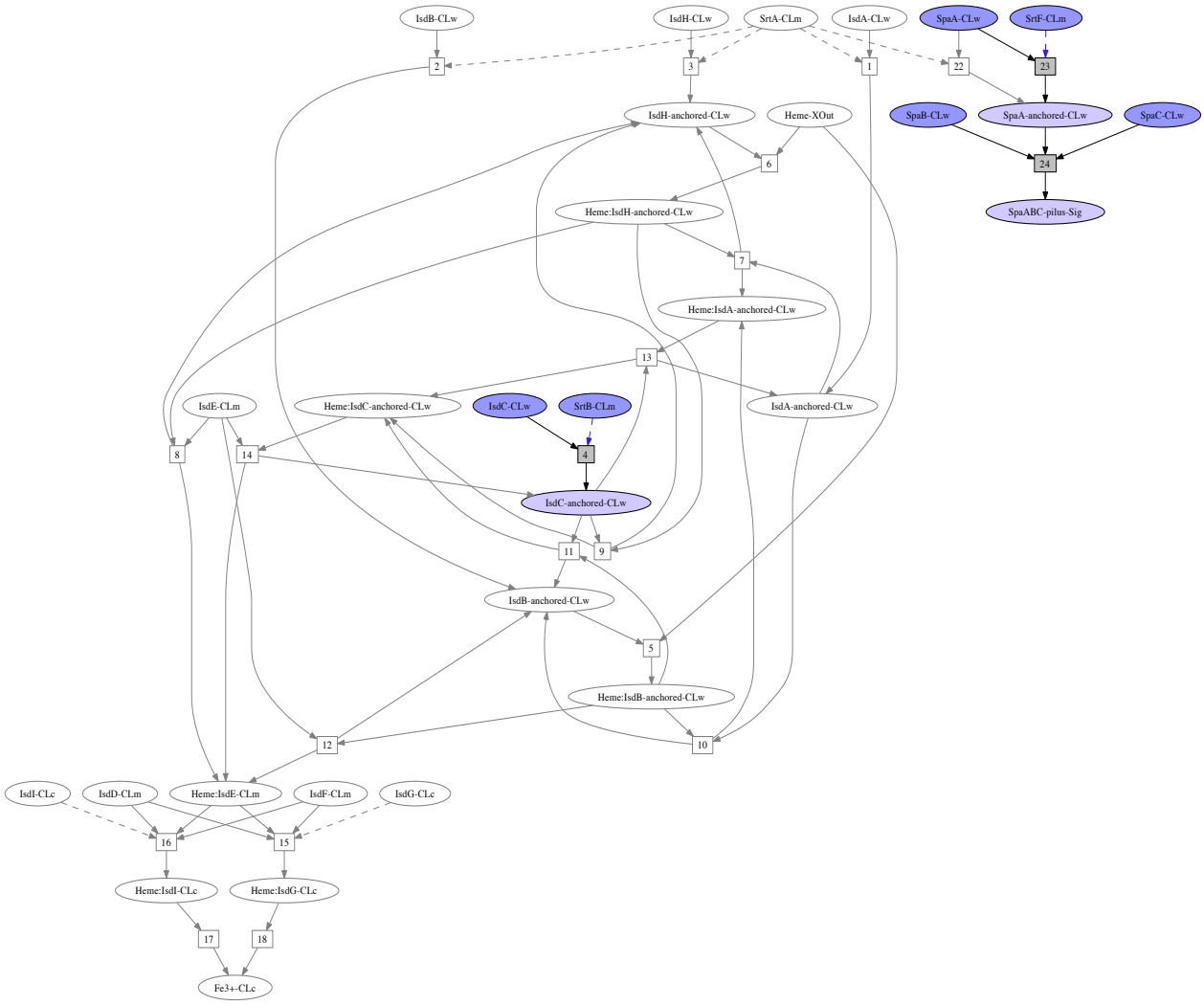

Fig. 4 Avoids query demonstrating single knockout of SrtA and its inhibitory effect on the heme transport in *S. aureus*.

4.2 Protease interconnectivity network

Pathway Logic models can be used to explore possible cross talks between proteases and identify key regulatory proteases. Insights into possible interactions among multiple processes involving proteases will give a better understanding to the underlying mechanisms and help develop specific inhibitors.

We show in Fig. 5, the interconnectivity between three proteases namely membrane serine protease HtrA, Group A streptococcus exotoxin B, SpeB (a cysteine protease) and membrane Type II SPase Lsp by setting these proteases as goals and generating the subnet (all the three protease are

green/lighter color). We are able to show the direct and indirect influence they may have on each other. We show the influence of HtrA and Lsp in the processing and activation of SpeB [20–24]. The network also shows CiaRH and RopB upregulating HtrA and proSpeB (inactive precursor form) expression respectively.

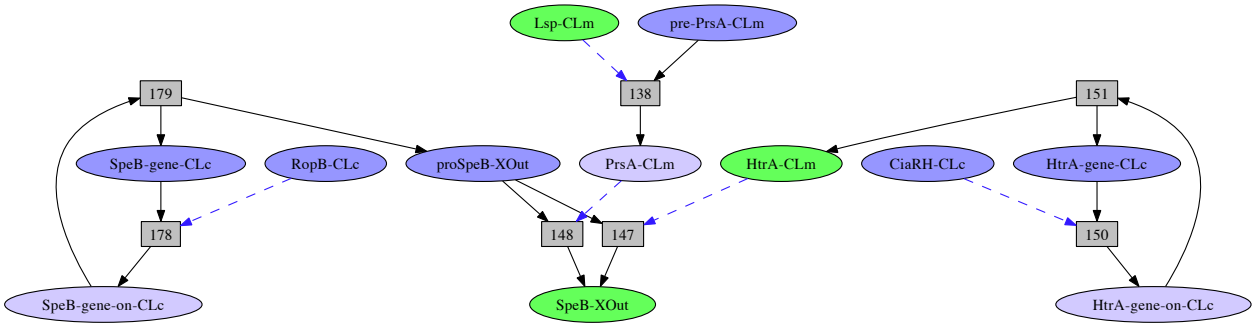

Fig. 5 The protease interconnectivity network. The three proteases (green/lighter color) within a larger protease network are shown to be interconnected in the process of SpeB activation.

5 Conclusions

We have described the Pathway Logic approach to modeling protease networks based on rewriting logic and the use of the Pathway Logic Assistant to browse and analyze these models. An important feature of PLA is the ability to generate pathways as query results. The current state of Pathway Logic is one step towards the grander vision of symbolic systems biology. Future challenges include developing such executable models of pathway interaction networks of proteases from both Gram-positive and Gram-negative bacteria to understand the underlying regulatory mechanisms for cell survival and pathogenicity leading to drug discovery. Another direction is to apply the basic approach to different types of systems, such as gene-regulation networks, or multi-cellular systems, and to integrate models of different types of systems to develop a systems level view.

Acknowledgements. The authors would like to thank Festschrift organizers for inviting this paper; and members of the Pathway Logic team for their contributions to the development of modeling techniques and the analysis and visualization tools. Encouragement and guidance from Dr. Walter Moos, Vice President of Biosciences at SRI, are gratefully acknowledged. This work is supported by funding to SRI International from the Commonwealth of Virginia.

References

[1] C. T. Supuran, A. Scozzafava and B. W. Clare. Bacterial protease inhibitors. *Med Res Rev*, 22(4): 329-372, 2002.

[2] J. Travis, and J. Potempa. Bacterial proteinases as targets for the development of second-generation antibiotics. *Biochim Biophys Acta*, 1477:35-50, 2000.

[3] S. I. Miyoshi and S. Shinoda. Bacterial metalloprotease as the toxic factor in infection. *J Toxicol Toxin Rev*, 16:177-194, 1997.

[4] C. Talcott, S. Eker, M. Knapp, P. Lincoln and K. Laderoute. Pathway logic modeling of protein

functional domains in signal transduction. In *Proceedings of the Pacific Symposium on Biocomputing,* 2004.

[5] S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer and K. Sonmez. Pathway Logic: Symbolic analysis of biological signaling. In *Proceedings of the Pacific Symposium on Biocomputing*, pages 400-412, 2002.

[6] S. Eker, M. Knapp, K. Laderoute, P. Lincoln and C. Talcott. Pathway Logic: Executable models of biological networks. In *Fourth International Workshop on Rewriting Logic and Its Applications*, Volume 71 of *Electronic Notes in Theoretical Computer Science*, 2002.

[7] J. Meseguer. Conditional Rewriting Logic as a unified model of concurrency. *Theoretical Computer Science*, 96(1):73-155, 1992.

[8] M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer and C. Talcott. All about Maude - A High-Performance Logical Framework. *Springer*, 2007.

[9] M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer and C. L. Talcott. The Maude 2.0 system. In *R. Nieuwenhuis, ed.: Rewriting Techniques and Applications (RTA 2003)*, Volume 2706 of *Lecture Notes in Computer Science*, 76-87, 2003.

[10] C. Talcott and D. L. Dill. The Pathway Logic Assistant. In *Proceedings of Computational Methods in Systems Biology,* 2005.

[11] L. A. Marraffini, A. C. DeDent and O. Schneewind. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev*, 70(1):192-221, 2006.

[12] E. P. Skaar, A. H. Gaspar, and O. Schneewind. IsdG and IsdI, heme-degrading enzymes in the cytoplasm of *Staphylococcus aureus*. *J Biol Chem*, 279(1):436-443, 2004.

[13] E. P. Skaar and O. Schneewind. Iron-regulated surface determinants (Isd) of *Staphylococcus aureus*: stealing iron from heme. *Microbes Infect*, 6(4):390-397, 2004.

[14] M. O. Stehr. A rewriting semantics for algebraic nets. In *C. Girault, R. Valk, eds. Petri Nets for System Engineering - A Guide to Modelling, Verification, and Applications,* 2000.

[15] I. Yeh, T. Hanekamp, S. Tsoka, P. D. Karp and R. B. Altman. Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res*, 14(5):917-924, 2004.

[16] H. Ton-That and O. Schneewind. Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol Microbiol*, 50(4):1429-1438, 2003.

[17] S. K. Mazmanian, G. Liu, E. R. Jensen, E. Lenoy and O. Schneewind. *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc Natl Acad Sci U S A*, 97(10):5510-5515, 2000.

[18] S. K. Mazmanian, H.Ton-That, K. Su and O. Schneewind. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc Natl Acad Sci U S A*, 99(4): 2293-2298, 2002.

[19] A. W. Maresso and O. Schneewind. Sortase as a target of anti-infective therapy. *Pharmacol Rev*, 60(1):128-141, 2008.

[20] J. N. Cole, J. A. Aquilina, P. G. Hains, A. Henningham, K. S. Sriprakash, M. G. Caparon, V. Nizet, M. Kotb, S. J. Cordwell, S. P. Djordjevic and M. J. Walker. Role of group A Streptococcus HtrA in the maturation of SpeB protease. *Proteomics*, 7(24):4488-4498, 2007.

[21] Y. Ma, A. E. Bryant, D. B. Salmi, S. M. Hayes-Schroer, E. McIndoo, M. J. Aldape and D. L. Stevens. Identification and characterization of bicistronic speB and prsA gene expression in the group A Streptococcus. *J Bacteriol*, 188(21):7626-7634, 2006.

[22] H. Tjalsma, V. P. Kontinen, Z. Pragai, H. Wu, R. Meima, G. Venema, S. Bron, M. Sarvas and J. M. van Dijl. The role of lipoprotein processing by signal peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. Signal peptidase II is required for the efficient secretion of alpha-amylase, a non-lipoprotein. *J Biol Chem*, 274(3):1698-1707, 1999.

[23] M. E. Sebert, L. M. Palmer, M. Rosenberg and J. N. Weiser. Microarray-based identification of htrA, a *Streptococcus pneumoniae* gene that is regulated by the CiaRH two-component system and contributes to nasopharyngeal colonization. *Infect Immun*, 70(8):4059-4067, 2002.

[24] A. Hollands, R. K. Aziz, R. Kansal, M. Kotb, V. Nizet and M. J. Walker. A naturally occurring mutation in ropB suppresses SpeB expression and reduces M1T1 group A streptococcal systemic virulence. *PLoS One*, 3(12):e4102, 2008.