Process Modeling and Automated Execution for Bio-Manufacturing

Zhiqi Shen[1], Robert Gay[1], Chunyan Miao[2], Tin Wee Tan[3], Chong Sean Kuay[1] and Hui Mien Lee[1]

[1]School of EEE, Nanyang Technological University, Singapore 639798
{zqshen, eklgay, kuay0001, leeh0015}@ntu.edu.sg
[2]School of Computer Engineering, Nanyang Technological University, Singapore 639798
ascymiao@ntu.edu.sg
[3]Department of Biochemistry, National University of Singapore, Singapore 117597
tinwee@bic.nus.edu.sg

Abstract

Recently there is a great demand for extending high throughput life science research to bio-manufacturing. However, there is a gap from life science research to bio-manufacturing. Most of existing Bio-workflow tools/systems only provide an isolated solution to help bio-scientist to orchestrate bio-R&D operations such as bio-database query, bio-computation and analysis etc for specific small bio-problems. They are static and lack of ability to adapt themselves in a dynamic changing environment. Bio-manufacturing involves many bioprocesses, each of which needs to choreographs a set of bio-workflows. Currently, none of the existing workflow systems support integrations of workflows from different workflow tools. Bio-scientists have to manually define and manage the workflows for bioprocesses through different workflow tools. This paper proposes a novel goal-oriented approach to modeling bioprocesses, to choreographing bio-workflows from different workflow tools, and to integrating agents, web services and workflows for automated execution. It also demonstrates how a multi-agent system can be further derived to adapt and automate complex bio-manufacturing workflow/processes in a dynamic changing environment.

I.   Introduction

Recently, there is a need to extend high throughput life science research to bio-manufacturing which aims to connect the current broad based of life science research to manufacturing areas and translate the technological know-how and research output into designs and manufactured products.

Collaborative life science R&D activities evolve from data integration to computation integration and to virtual bio-labs etc. Bio-scientists often need to perform experiments using shared data resources distributed worldwide. Such data resources are generated by deferent groups of researchers around the world. Many of the bio-labs/research groups have made their data sources available for sharing on the World Wide Web (WWW). In addition to raw data resources are shared on the web, many different bio-tools/applications that operate on it have been developed, most of them with restricted functionality and targeted at performing highly specific tasks. A bio-R&D activity is performed through a set of bioprocesses. These bioprocesses involve operations of different bio-applications/tools. Collaborative life science

R&D needs not only to share the bio-data resources but also to integrate bio-operations using various bio-tools. Therefore, it is both challenging and vital to integrate bio-tools/applications in various bioprocesses.

With the emergence of service-oriented technology, it is expected that any entity in a interconnecting bio-applications/tools will be viewed as a service, whether it is a bio-database, a bio-instrument operation, or a bio-application/tool. Web/grid services are fast emerging as the enabling technology for seamless bio-application integration. Workflows that orchestrate the bio-operations using different tools are promising to automate various bioprocesses. As a result, many attempts for using web service technology and workflow management systems to tackle the above issues have been made [1-4].

However, most of existing Bio-workflow tools/systems only provides an isolated solution to help bio-scientist to orchestrate bio-R&D operations such as bio-database query, bio-computation and analysis etc for specific small bio-problems. They are static and lack of ability to adapt themselves in a dynamic changing environment. Bio-manufacturing involves many bioprocesses, each of which needs to choreographs a set of bio-workflows. Currently, bio-scientists have to manually define and manage the workflows for bioprocesses through different workflow tools.

This paper proposes a novel goal-oriented approach to modeling bioprocesses and choreographs bio-workflows. It also demonstrates how an agent-oriented system can be further derived to automate complex bio-manufacturing processes in a service-oriented environment.

An agent is an encapsulated software entity that is situated in some environment and is capable of flexible, autonomous action in that environment in order to meet its objective. Agents are goal oriented. They act to achieve objective on behalf of human beings. Agents can invoke web services, workflows in a dynamic environment autonomously on behalf of bio-scientists. Integrating Web services, workflow and software agents will bring immediate benefit for bio-application integration and bioprocess automation. In this perspective, identifying a means of connecting agents, workflow and Web services is the motivation and central foundation of this work.

In this paper, a novel goal oriented methodology is proposed to model and schedule the bio-manufacturing processes based on the bio-workflows. The activities carried out according to a requirement are usually organized in groups of inter-related activities called processes that can be seen as a set of operations, rules and constraints specifying the steps that must be taken, and conditions that must be satisfied, in order to accomplish a given goal. The goal-oriented modeling method, Goal Net [5] will be explored for supporting user centric modeling, reasoning, bio-workflow abstracting, composition and automation. The bio-workflows will be defined for the data, computational and bioinstrumentation services. The bio-manufacturing processes will be defined for the integration of workflows, modeling of bioprocesses, and automatic execution based on Goal Net, semantic ontology and agent technology. The new methodology will lead to design of an integrated workflow infrastructure to straddle across manufacturing and services involving integration of bioinstrumentation, measurement systems, related databases, software and web-based services. The supply chain management will also be incorporated to promote more coherent and in-time operation.

Following the introduction section, Section 2 discusses some related work. Section 3 presents the bioprocess modeling method while Section 4 describes the automated execution of modeled bioprocesses. In Section 5, a real scenario is studied using the method proposed in this paper. Finally Section 6 concludes the paper.

II.   Motivation and Related Work

Most of current research on integration of the available bio-services is based on web service architecture. For example, Taverna [1] project (part of MyGrid project) is a leading research project of the UK government's e-Science programme. Taverna is a workflow workbench that provides a language and software tools to facilitate easy use of workflow and distributed services within the e-Science community.

As a growing number of bioinformatics resources such as computational tools and information repositories are being made available as Web services, the Taverna [1] project aims to provide a modeling tool with graphic user interface for the designing and constructing bioinformatics workflows on top of bio-services over the web. Taverna provides user-friendly interfaces for bio-scientists to select and compose web services in a sequential order that forms a workflow. Each operation in the defined workflow will be able to invoke a specific bio-web service. Integrated bio-services for a specific problem can be realized by running the corresponding workflows.

Some other research groups also focus on bioinformatics workflows on top of the available bio-services. They provide similar tools that allow users to compose and execute workflows, such as Pegasys [2], Wildfire [3] and BioWBI [4]. One of the common limitations is that most of them do not support dynamic adaptive workflows. More over, none of the existing workflows systems supports the integration of workflows defined by different workflow tools.

One research project "Integrated Bio-laboratory Manufacturing and Services System" is a national project of Singapore. It aims to investigate how an "Integrated Workflow Infrastructure" for offering manufacturing and services involving the integration of bioinstrumentation, measurement systems and related databases, as web services could be developed. KOOPrime has developed a suite of products and solutions, KOOPlatform that addresses the needs of life sciences processes, specifically in the area of genomics and proteomics research [6]. This suite is known as "Workflows for Life Sciences".

Although attempts have been made to offer descriptions by manually categorizing the services and sharing related workflows, these workflow management systems only provide a partial solution to the service integration. Existing Bio-workflow systems such as Pegasys, Taverna, Wildfire, BioWBI and KOOPlatform have limited user-centric modeling, abstracting, reasoning and automation capabilities. They are useful for composing partial, low-level processes, but lack of the ability to integrating and automating a complete pipeline from R&D to manufacturing.

To meet the above challenges, in this paper, we present a goal-oriented approach for modeling, integrating and automating bioprocesses as well as bio-workflows that are adaptive to dynamic changing environments.

III.   Bioprocess Modeling

A bioprocess is a specific ordering of activities with clearly identified inputs and outputs that achieve a certain goal. For example, a high level bio-manufacturing process takes a sample of an unknown infectious disease such as SARS or bird flu and produces a diagnostic material such as a DNA chip for rapid development of diagnostics and eventually for design of RNAi (miRNA/siRNA) therapeutics and DNA/peptide vaccines. The activities involved in a bioprocess can be bio-workflow executions, web service invocations, or other bio-application executions. The activities of a bioprocess and the order of the activities may be different time by time due to different situations. Currently most researchers of life science still manually manage the activities of a bioprocess according to the current situation and their expertise to adapt to the dynamic environment. In this paper, we propose a goal-oriented approach to modeling bioprocesses so that life science researchers are able to transform their expertise to the process models. Then we build an agent oriented system to automate the process execution based on the process models.

A bioprocess model facilitates the alignment of bioprocess specifications with the technical framework that IT development needs. The challenges for modeling a bioprocess include:

- The bioprocess model should capture relevant information consistently and thoroughly so that both life science researchers and the IT developers can understand the process requirements that are captured in the model.
- The bioprocess model should capture alternatives and exceptions to standard in addition to normal operations.
- The bioprocess model should be easily executed towards execution automation.

In this paper, we use Goal Net [5], a goal-oriented modeling method for bioprocess modeling and scheduling. The advantages to use Goal Net include:

- Goal Net is a novel goal oriented process modeling tool which can decompose a complex process (goal) into executable sub-processes (sub-goals) for achieving a common goal. The temporal relationships between processes (goals) are modeled. This is the key difference between Goal Net and other goal-oriented models.
- Goal Net has reasoning capability. An agent running a goal net can reason the next goal to pursue and the next task to take for achieving the selected goal based on the current situation. Therefore the agent can compose the low level workflows to a complete pipeline in a dynamic changing environment.
- Goal Net is also a multi-agent modeling tool by which a multi-agent system can be derived from the process modeling for automating the processes execution.

A goal of an agent is a desired state that an agent intends to reach. Goal Net is a composite goal hierarchy which is composed of *goals* and *transitions*. The *goals*, represented by round rectangle, are used to represent the goals that agent need to go through in order to achieve its final goal. The *transitions*, represented by arc and rectangle, connect one goal to anther specifying the relationship between goals it joins. Each transition must have at least one *input goal* and one *output goal*. Each transition is associated with a task list which defines the possible tasks that agent may need to perform in order to transit from the input goal to the output goal. Figure 1 shows a simple goal net.

There are two types of goals in Goal Nets, *atomic goal* and *composite goal*. An *atomic goal* accommodates a single goal which could not be split anymore; a *composite goal* may be split into sub-goals (either composite or atomic) connected via transitions.

Goal Nets can represent four types of basic temporal relationships between goals: *sequence*, *concurrency*, *choice* and *synchronization*. *Sequence* relationship represents a direct sequential relationship between one input goal and one output goal; *concurrency* relationship means one goal has more than one next goals, and all its next goals can achieve simultaneously; *choice* relationship specifies a selective connection from one goal to other goals; *synchronization* relationship specifies a synchronization point from different input goals to a single next goal. With different combination of the basic temporal relations, Goal Net supports a wide range of complicated temporal relations among goals. This is one of the major differences between Goal Net and other goal modeling methods.
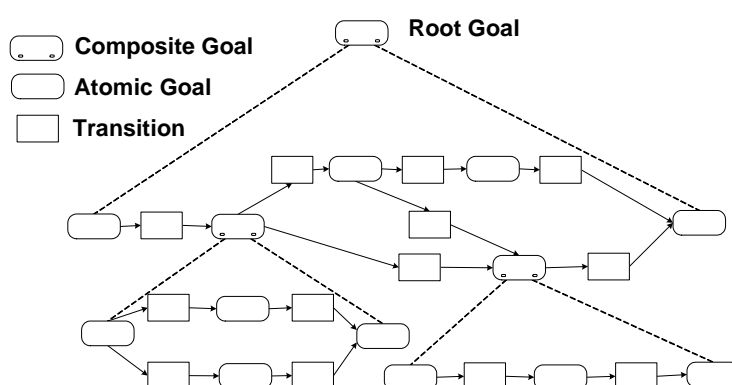

Figure 1. A simple goal net

With such a composite goal hierarchy and various temporal relations within the hierarchy, a complex system can be recursively decomposed into sub-goals and sub-goal-nets. In such a manner, a system can be easily modeled and simplified.

With Goal Net, the composition of each bioprocess is designed in order to achieve a specific goal. A bioprocess can be decomposed into a hierarchy of sub-processes and activities. Parts of the decomposition of this process are assigned to different recognized workflows. A combination of various processes takes place at different members of the virtual organizations in order to achieve the global goal of the high level process. The problem of the supervision or coordination of such a process at its various levels of decomposition is critical, in this context, where its definition and activity are not limited to a single organization, but to a set of autonomous, distributed, and heterogeneous nodes that need to cooperate. With Goal Net, the supervision and coordination are automatically derived during the process decomposition phase.

Goal Net provides a rich set of relationships and goal/action selection mechanisms in providing a dynamic and highly autonomous process integration model. This approach of process integration is viable, in which:

1. The interactions between bioprocesses, bio-workflows and web services are represented as a Goal Net.
2. A bio-workflow or web service operation is represented by a transition task and a goal shows that a particular goal is reached after the execution of the transition tasks.

3. Combinations of different relationships between goals, sub-goals, and transitions can be used to represent complex bioprocess logic.

4. The dynamic bioprocess flow is achieved by defining action selection and goal selection mechanisms.

The goal-oriented bioprocess modeling proposed in this paper can handle atomic Web services and other existing web service compositions such as the workflows defined in Taverna and Kooplatform. This is easily achieved through calling the APIs provided by the external composition models in transition tasks of Goal Nets.

For example, Figure 2 shows a bio-manufacturing process. Figure 3 shows the goal net which models the process.
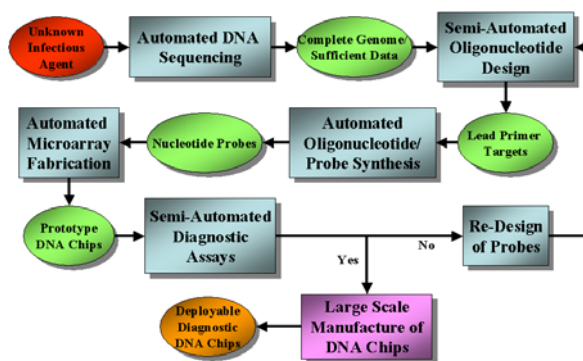

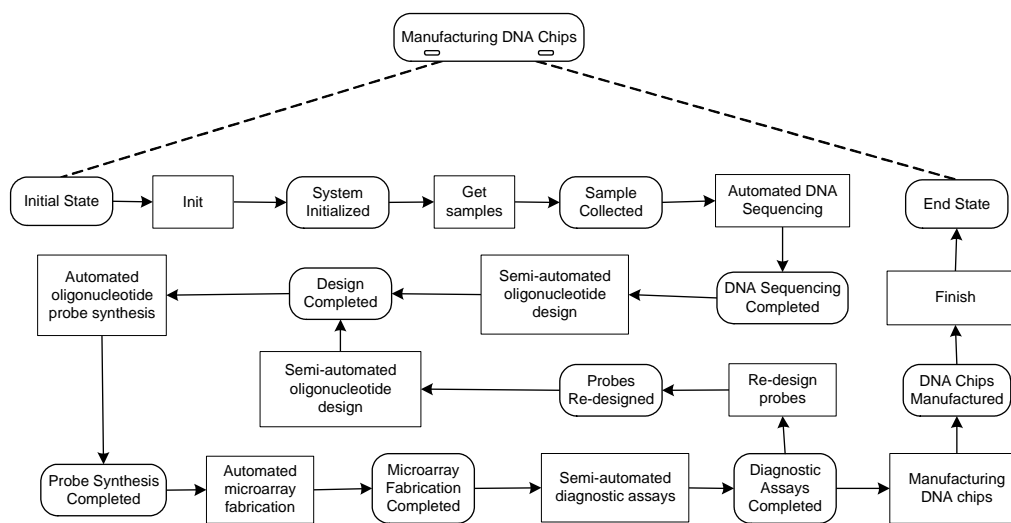**Figure 2. A bio-manufacturing process**

Figure 3. The goal net that models the bio-manufacturing process

This is a high level process. In the goal net, each goal is a composite goal which needs to be decomposed to sub-goals. Here we don't intend to make a complete bio-manufacturing process model. Rather, we want to illustrate how a process can be mapped or modeled using Goal Net.

IV. Automated Bioprocess Execution

From an unknown infectious agent such as a deadly virus, to the elucidation of its complete genome, from its genome to the complete analysis and design of specific diagnostic DNA reagents, from the designed candidate diagnostics to the manufactured biochemical products, to the testing of these products against clinical samples, to the fine tuning and optimization of the diagnostic material such as a DNA chip – each of these steps can be individually semi-automated for high throughput today. Yet no-one has attempted to connect these disparate and distributed steps into a complete workflow chain of manufacturing and design steps end-to-end. Each of these steps can be modeled on a software platform, to achieve and choreograph a specific high level goal using Goal Net. Each step can be residing on a grid of geographically distributed computers, databases and bioinstrumentation and machinery or a human operator or decision maker, represented digitally and callable by a workflow orchestration system (e.g. KOOPlatform, Taverna/MyGrid, etc). They can all be mapped ontologically and semantically for process integration and compatibility using Web service standards. Each of these steps can be mixed and matched and recombined into re-usable or re-purposable workflows that capture the business, operational and manufacturing processes at different levels of granularity.

In a bioprocess model represented using Goal Net, each step of operation is to pursue a goal. An invocation of workflow system or a web service indicates a transition from one achieved goal to the next goal. Workflows and web services are invoked in transitions of a goal net in the form of tasks of the transitions. In this way, the individual workflows and web services are integrated to bioprocesses using Goal Net. The execution of a goal net automates the execution of the bioprocess represented by the goal net.

Goal Net is an agent goal model. A goal net can be executed by an agent. We have proposed and developed an agent development framework based on Goal Net so that the created agent will take a goal net as its goal model to infer and guide its behaviors. When an agent is created, the agent has no goal and is running in an idle status. Therefore it is an agent body. A goal net that represents a bioprocess is then loaded to the agent as its brain. Now the agent will start goal pursuit based on the goal net from the initial state to achieve the final goal of the bioprocess.
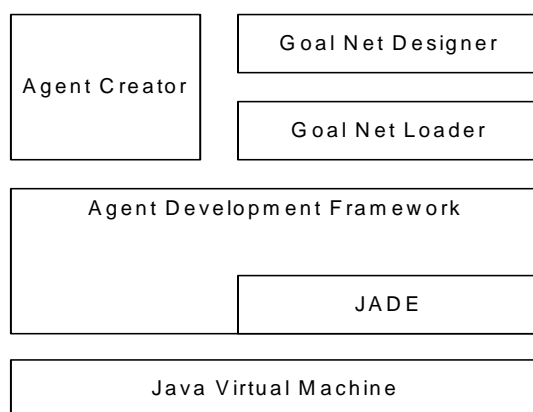

Figure 4. The agent development environment and execution platform

Figure 4 shows the multi-agent development environment (MADE) and the execution platform. In this figure, Goal Net Designer is a tool to design bioprocesses using Goal Net.

Agent Creator was built on top of the agent development framework [7] to provide an agent development environment. Goal Net Loader is an interface for users to load a goal net to a created agent. Users can use it to assign different goal nets to different agents according to the users' requirements. Although MADE has provided a complete agent development environment and execution platform, the agent development framework has been integrated with the popular agent development environment JADE [8]. JADE is compliance with industry standard FIPA [9], supports standard agent communication mechanism and provides multi-agent running platform. By the integration with JADE, MADE provides goal model development environment and at the same time, supports standard agent communication mechanism and agent running platform.

V.   A Prototype of Bio-manufaturing System

In this research, we have developed a prototype of bio-manufacturing system using the approach presented in this paper. In this system, we use Taverna and Kooplatform as the bio-workflow systems. Web services that wrap different bio-services are orchestrated through the two workflow systems respectively. Goal Net is used to choreograph the workflows for modeling different bioprocesses. Figure 5 shows the architecture of the prototype system.
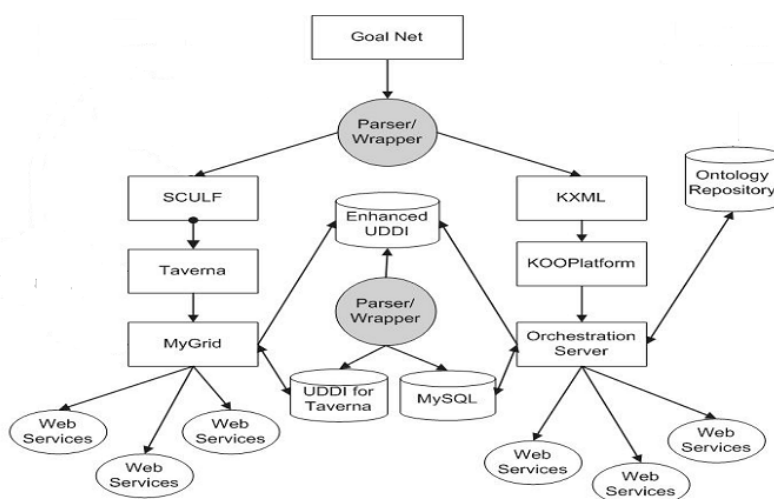

Figure 5. The architecture of the prototype system

As shown in this figure, in this system, Taverna and Kooplatform provide two sets of workflows. The extended UDDI provides common place for the services provided by the two systems. Goal Net provides a process integration platform for modeling bioprocesses and automating the bioprocess execution. The orchestration of existing bio-services needs not only a consistent definition of terminologies of bioprocesses but also the semantic linkage among various bioprocesses. The ontology repository is constructed in the system to define the consistent terminologies and concepts and the semantics between the concepts.

MADE allows easy insertion of additional task libraries into the framework. The task libraries are the entry point where different functionalities can be added to MADE. To enable agents created by MADE to invoke web services and workflows in the two workflow systems, we extended the MADE by adding web service invocation component, Taverna integration component and Kooplatform integration component. Figure 6 shows the structure of the extended MADE.
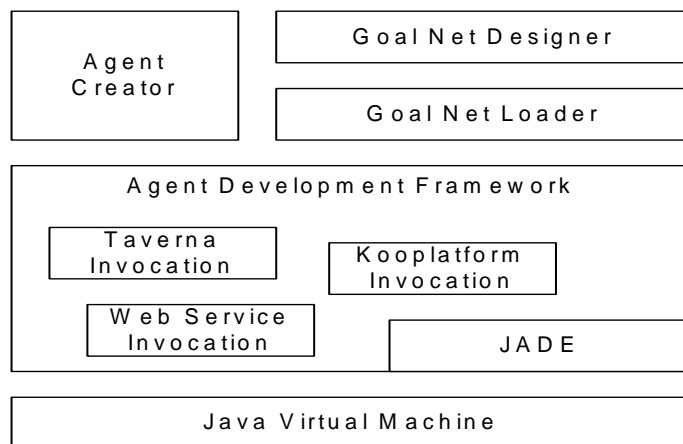
Figure 6. The structure of the extended MADE

The Web service invocation component actually provides the API calls to the existing AXIS Web service tool [10] provided by Apache. AXIS is chosen because it is the latest Web service tool that provides good features with reasonable performance. By providing the API calls to the Taverna and Kooplatform workflow systems, we have proven that the extended framework is able to act as a coordination framework for multiple atomic Web services and other existing Web services composition workflow models.
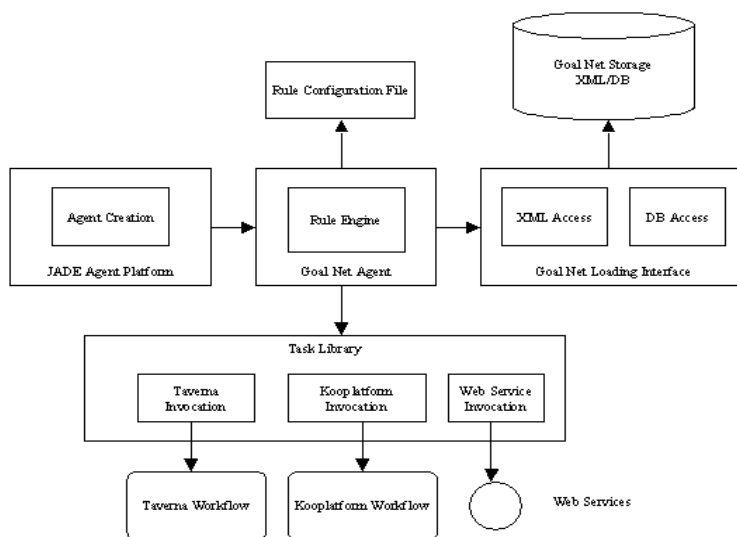

Figure 7. The architecture of the bioprocess execution system

Figure 7 shows the agent-oriented architecture of the bioprocess execution system. There are two types of storage methods for storing Goal Nets: XML-syntax description file and database storage. These two methods can be used to keep the goals, arcs, transitions, attributes and their interconnection information for dynamic Goal Net loading. As shown in the figure, there is a File/DB access library to access Goal Net storage system. User can specify the rules and configuration for running agents through the user rule configuration file. This user rule configuration file will keep the information for agents to make decisions in action selection and goal selection inferences. JADE platform acts as agent creator by which agents can be dynamically generated during run-time. Agents' deployment and un-deployment can be done through JADE services too.

The action selection and goal selection mechanisms in the Goal Net model have been combined to become a rule engine. As shown in the figure, Goal Net rule engine makes decisions in goal selection and action selection based on the conditions defined in the user rule configuration file. Users need to specify the desired conditions and variables in the rule configuration file. Besides, users need to register tasks and the mapping of rules for the goal selection and action selection mechanisms to the rule engine. Finally web services and workflows can be invoked through the task libraries, that is, the web service invocation component, Taverna workflow invocation component and Kooplatform workflow invocation component.

A typical scenario demonstrated by Taverna is that it provides a workflow to compare two genes X and Y. To illustrate our method we designed a goal net to represent a process in which a compareXandY workflow will be invoked according to the user designed goal net. We then created an agent using the extended MADE and loaded the goal net to the agent. Figure 8 shows the running result of an invocation to a Taverna workflow from the agent.
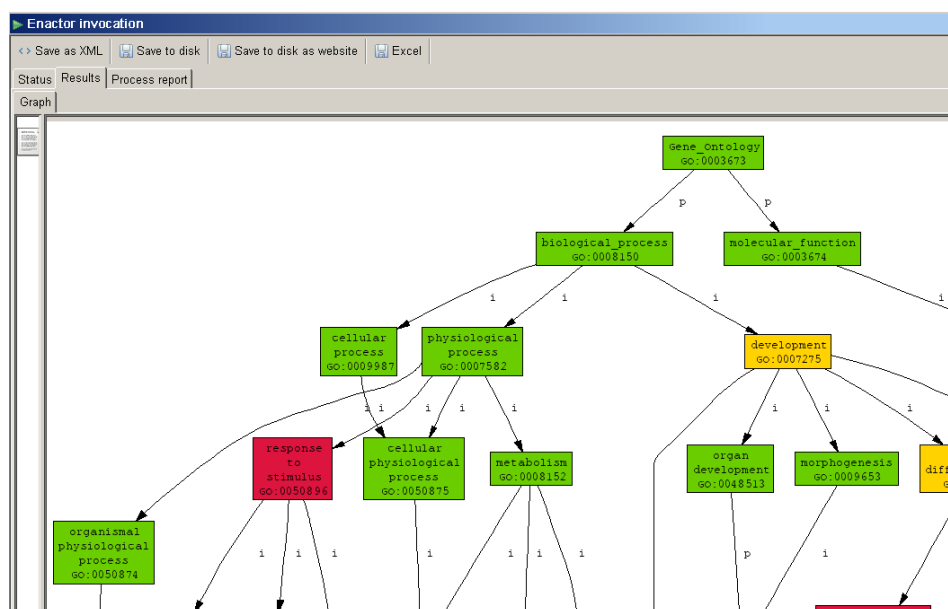

Figure 8. The running result of the agent

Originally a user needs to prepare the data and invoke the workflows manually through the Taverna workbench, a GUI tool for Taverna workflow operations. With our developed system, the designed and configured bioprocess can be stored in a database in the form of a goal net. A user needs only to create an agent and load the specific goal net to get the expected results. Further more the goal nets can be reused with different data for different requirements.

VI.   Conclusion

To integrate a complete pipeline from R&D to Diagnostic Kits manufacturing is not only an important leapfrog in terms of R&D value generated, but also vital in both economic and social context as it will allow us to build up a platform that can effectively response against new infectious outbreaks or bioterrorist attacks.

In this paper, we present a goal-oriented approach for bioprocess modeling and integration. A multi-agent platform for integrating bio-workflows and automating the bioprocess execution has been presented. The results generated by the developed prototype system shows our goal to integrate various existing workflows and to automate the process execution using proposed approach has been achieved.

In fact, the core of biopharmaceutical manufacturing of PCR diagnostics, RNAi (RNA interference) agents, peptide vaccines etc, are processes which assemble linear polymers of biochemical monomers from a linear sequence of genetic information, mimicking the way each living cell does it. Automated DNA/RNA sequencers, oligonucleotide synthesizer machines, peptide synthesizers etc are available bio-instruments today. Each can be called to produce specific sequences by sending the machines a text file generated from database searches and bioinformatics computation. Each system relies on standard sets of reagents and buffer solutions, which constitute the supply chain manufacturing and management system which when integrated with these systems will allow high throughput, automated or semi-automated manufacturing to take place.

References

1. T. Oinn, M.J. Addis, J. Ferris, D.J. Marvin, M. Greenwood, T. Carver, A. Wipat, and P. Li, "Taverna, lessons in creating a workflow environment for the life sciences", in Proceedings of GGF10, Berlin, Germany, 2004.
2. S. P. Shah, D. Y. M. He, J. N. Sawkins, J. C. Druce, G. Quon, D. Lett, G. X. Y. Zheng, T. Xu, and B. F. F. Quellette, "Pegasys: software for executing and integrating analyses of biological sequences", BMC Bioinformatics, 5:40, 2004.
3. Francis Tang, Ching Lian Chua, Liang Yoong Ho, Yun Ping Lim, Praveen Issac and Arun Krishnan, "Wildfire: distributed, Grid-enabled workflow construction and execution", BMC Bioinformatics 2005, 6(1):69, Mar 24 2005.
4. P. Leo, C. Marinelli, G. Pappadà, G. Scioscia and L. Zanchetta, "BioWBI: an Integrated Tool for building and executing Bioinformatic Analysis Workflows", BITS2004, Padova, Italy, Mar 26-27 2004.
5. Zhiqi Shen, Robert Gay, Chunyan Miao, Xuehong Tao, "Goal Oriented Modeling for Intelligent Software Agents", in Proceedings of the 2004 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'04), Beijing, China, September 20 - 24, 2004.
6. Kooplatform, *http://www.kooprime.com*.
7. Zhiqi Shen, Robert Gay, Chunyan Miao, Xuehong Tao, "Goal Autonomous Agent Architecture", in Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04), Hong Kong, China, September 28 - 30, 2004.
8. Fabio Bellifemine, Agostino Poggi and Giovanni Rimassa, "JADE: a FIPA2000 compliant agent development environment", in Proceedings of the fifth international conference on Autonomous agents, pp. 216 – 217, Montreal, Quebec, Canada, 2001.
9. Foundation for Intelligent Physical Agents, "FIPA Agent Management Specification", *http://www.fipa.org/specs/fipa00023/*, June 2002.
10. The Apache Foundation, "WebServices – Axis", *http://ws.apache.org/axis/*.

Dr. Zhiqi Shen obtained his BSc in Computer Science in Peking University, China, and PhD in Information Communication Institute, Nanyang Technological University (NTU), Singapore. His research interests include goal-oriented modeling, cognitive modeling, intelligent software agent, software engineering, semantic web/grid, sensor network and their applications in education innovation, integrated bio-manufacturing services, interactive games and various e-services. He has worked in both university and industry in many large funded R&D projects in China, Singapore and Canada.

Professor Robert Gay obtained his PhD in Electronics Engineering from the University of Sheffield in 1970. Since 1982 he has been a Professor at NTU. He was also the Research Director of Gintic Institute of Manufacturing Technology from, 1989 to 1999. From 2000 to 2003, he was Director of the ASP Centre and since 2003 he has been Director of the Managed Computing Competency Centre at NTU. He is currently also the Director for Research IT Resources at NTU. His current research interests and expertise are: Semantic Grid, Knowledge Based Systems, E-learning and Integrated Manufacturing Systems and Services.

Dr. Chunyan Miao received her PhD from School of Computer Engineering, Nanyang Technological University (NTU), Singapore. She is currently an Assistant Professor in the same school. Prior to joining NTU, she was with School of Computing, Simon Fraser University Canada. Her major research interest includes machine learning, intelligent software agent, agent mediated semantic web/grid, and agent oriented software engineering. She is particularly interested in infusing agent technology into various application domains.

Dr. Tin Wee TAN, Associate Professor, Dept of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, has been pioneering the use and development of bioinformatics over the past 15 years. He is also serving as board director of the INternational Society for Computational Biology, secretariat of the Asia Pacific BIoinformatics Network as well as the S* Life Science INformatics Alliance.

Mr. Kuay Chong Sean received his B.Eng. degree from Universiti Teknologi Malaysia in 2002, and the M.Sc degree from Nanyang Technological University, Singapore in 2005, all in electrical engineering. His master studies research work was mainly on communication network and agent system. Prior to joining NTU, he was a system validation engineer in the network processor division of Intel Microelectronics Sdn. Bhd., Penang, Malaysia. Since July 2005, he has been with Institute for Infocomm Research (I2R), Singapore. He is currently a research engineer with the Digital Wireless Department in I2R.

Ms Lee Hui Mien graduated from the School of Electrical and Electronic Engineering at Nanyang Technological University (NTU), Singapore, in 2003. She is currently a PhD candidate under the same school and funded by the Agency for Science, Technology and Research (A*STAR). Her current research interests are in product lifecycle management, knowledge and information management in sustainable product development, integrated manufacturing and services systems. She is also attached to the Singapore Institute of Manufacturing Technology where she is working on her thesis in a knowledge-based web-based adviser for sustainable product development.