The IGIPI Ontological Framework: Integrating Gene Interactions with Protein Interactions

Bill Andreopoulos, Aijun An, Xiangji Huang

York University, Department of Computer Science and Engineering, Toronto, Canada

Quotes from Quantum Physics

"*evidence obtained under different experimental conditions cannot be comprehended within a single picture, but must be regarded as complementary in the sense that only the totality of the phenomena exhausts the possible information about the objects*."

Contributions

We propose the IGIPI framework for:

- Integrating gene interactions with protein interactions at a low level
- Integrating with other biomedical knowledge at a higher level

Implement the IGIPI framework on the Semantic Web using the OWL Web Ontology Language

Motivation

Protein interactions produced from two-hybrid studies often are not mapped directly to gene interactions from gene expression or synthetic mutant lethality (SML) studies. The purpose of SML studies is to identify interactions between genes in the genome, by knocking out pairs of genes until a cell dies.

Sometimes a two-hybrid study may detect a protein interaction, although a gene expression or SML study fails to detect an interaction between the corresponding genes. Reasons may include:

Suppressor mutation: A mutation in one gene may restore (partially or fully) the function impaired by a mutation in a different gene, or at a different site in the same gene.

Nonallelic noncomplementation: Mutations in two genes may fail to complement, because the gene products are subunits of the same multi-protein complex.

Conditional-lethal mutation: Gene mutations may result in lethality under one environmental condition (e.g., high temperature) but not under another condition (e.g., lower temperature).

Motivation

Alternatively, if two genes exhibit synthetic lethality, this may not necessarily mean that their proteins also interact (and thus the genes may not have the same function).

A reason for this discrepancy could be that the gene mutations affect two different protein pathways, which perform different functions but lead to death when combined.

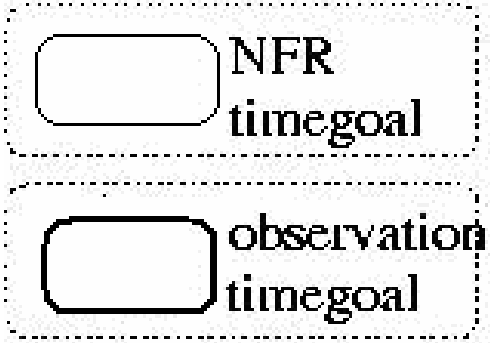
The IGIPI framework is based on the concept of *timegoals*.

Timegoal: a goal that needs to be satisfied at a specific time interval in an experiment, in order for a biological function to be observed (e.g., a network of protein interactions).

Each timegoal is represented as a node (cloud).

2 *types of timegoals*: *NFRs* (high level goals) and *observations* (low level goals). The term NFR is derived from the software engineering term “non-functional requirement”.

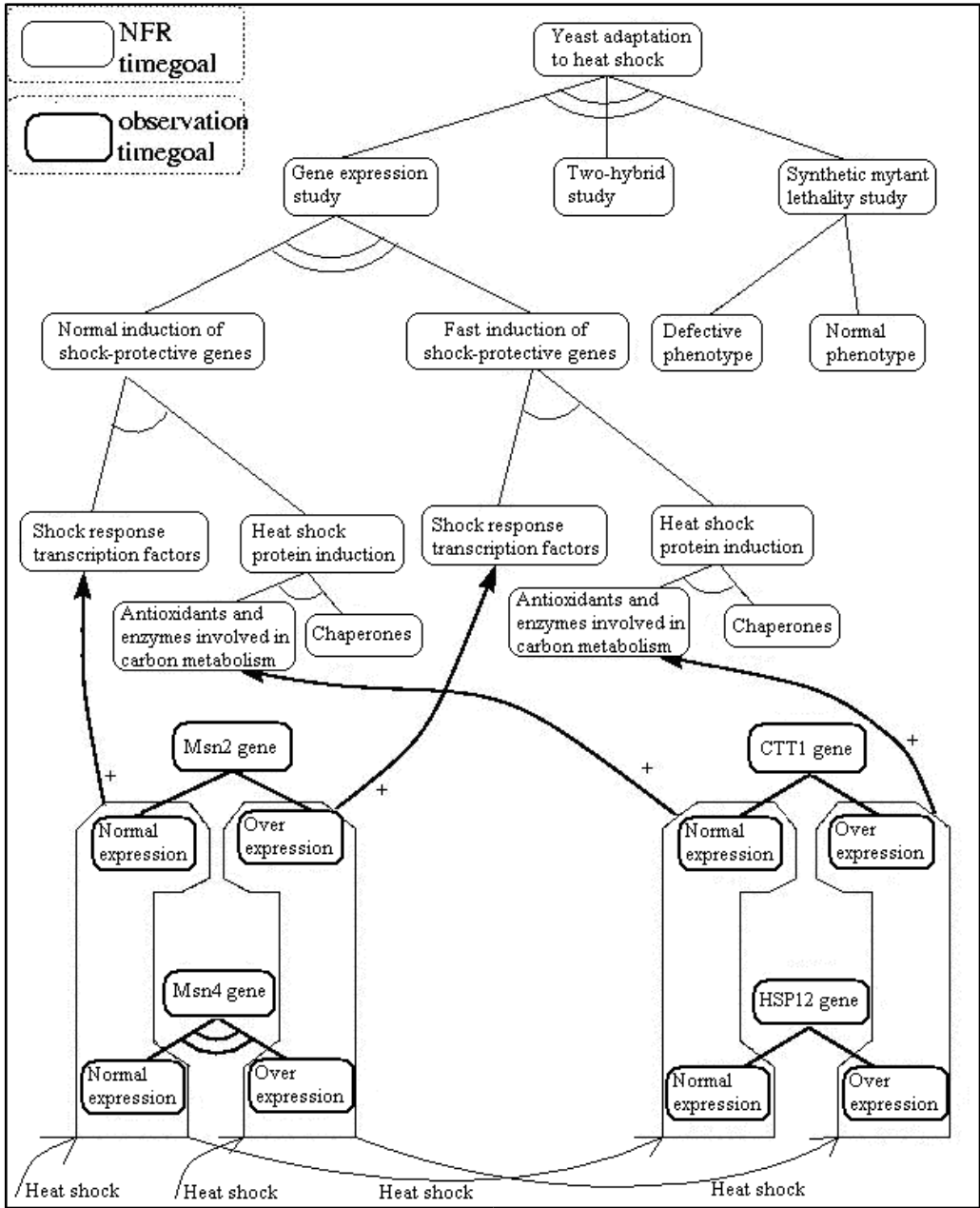
An NFR is a high level goal placed on a biological experiment, without stating anything about the precise means by which the goal will be satisficed in the experiment.



The NFR timegoals do not represent knowledge about the genomic-level events that need to occur for the biological function to be observed; this is the purpose of observation timegoals.

Timegoal Interdependency Graph (TIG)

The IGIPI framework represents information about timegoals using a graphical representation called the *timegoal interdependency graph* (*TIG)*. A TIG records all timegoals representing goals in experiments that, if satisficed, will lead to observing the root biological function. Each timegoal is represented as a node (cloud). The interdependencies between timegoals are represented as edges. This Figure shows observing the "yeast adaptation to a heat shock" in an experiment as a root NFR timegoal at the top of the TIG.



The IGIPI framework is based on the concept of *timegoals*.

Timegoal: goal with no clear-cut criterion for fulfilment. Instead, a timegoal may only contribute positively or negatively towards achieving another timegoal.

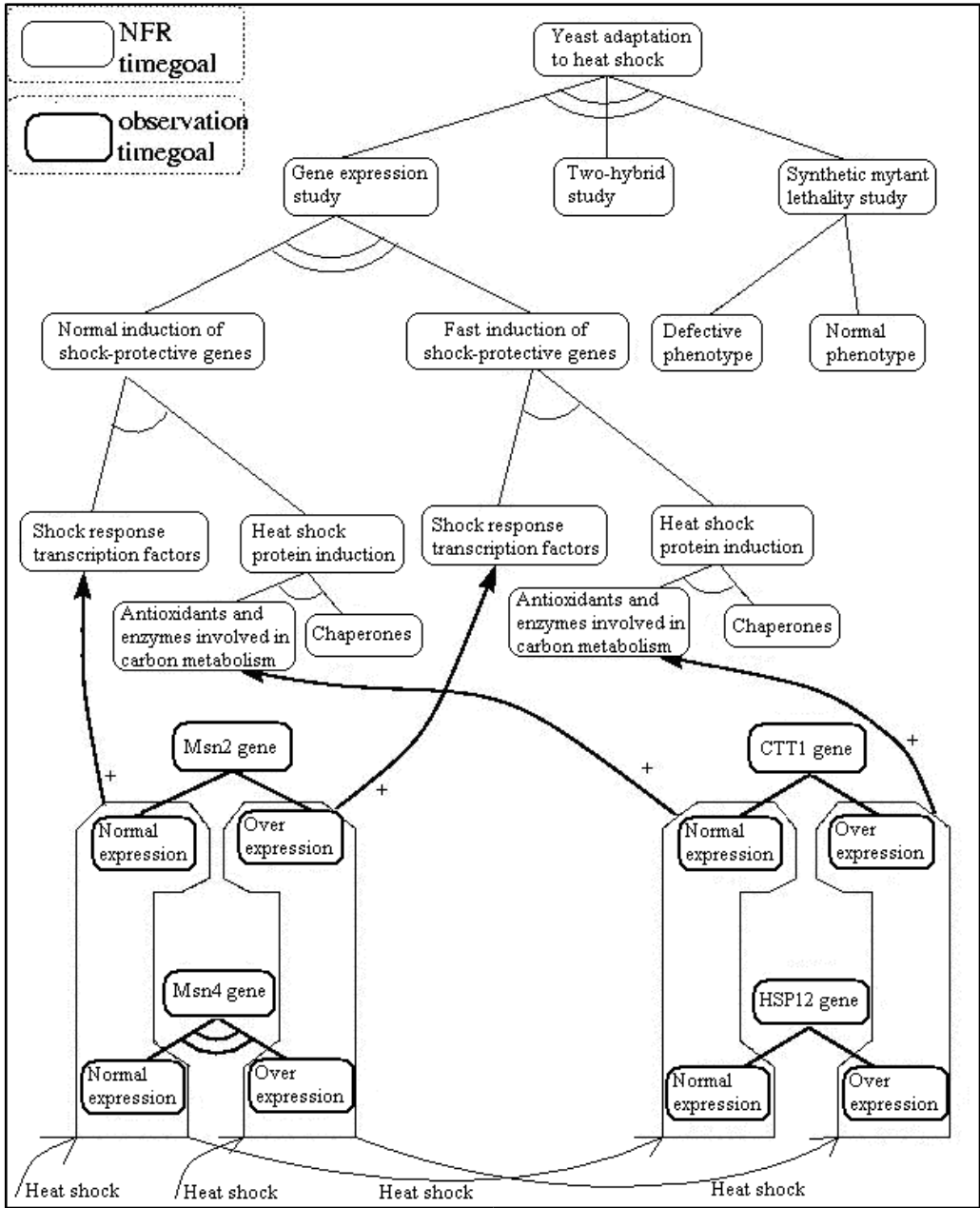
By using this logic, a timegoal can be *satisficed* or not.

In the IGIPI framework, *satisficing* refers to satisfying at some level a goal or a need, but without necessarily producing the optimal solution.

Timegoal Interdependency Graph (TIG)

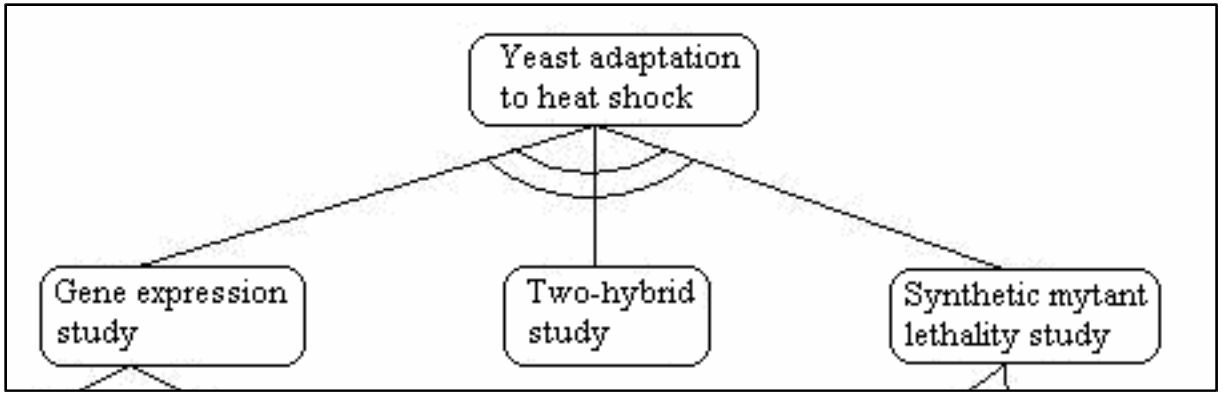
A developer can start constructing a Timegoal Interdependency Graph (TIG) by identifying the top-level biological function that is expected to be observed and sketching a root NFR timegoal for it. The root NFR timegoal of a TIG has a value taken from a domain of biological functions, such as the GO Gene Ontology.

The root NFR timegoal is decomposed into timegoals that represent knowledge about the biological function.



Decompositions of timegoals (AND/OR)

To represent the timegoals that need to be satisficed for the “yeast adaptation to a heat shock” to be observed experimentally, the root NFR timegoal is decomposed into the NFR timegoals “gene expression study”, “two-hybrid study” and “synthetic mutant lethality study”. This means that performing any of these studies leads to observing the yeast’s adaptation to a heat shock.



Timegoals are connected by interdependency links, which show *decompositions* of parent timegoals downwards into more specific offspring timegoals.

In some cases the interdependency links are grouped together with an arc; this is an *AND* contribution of the offspring timegoals towards their parent timegoal, and means that both offspring timegoals must be satisficed to satisfice the parent.

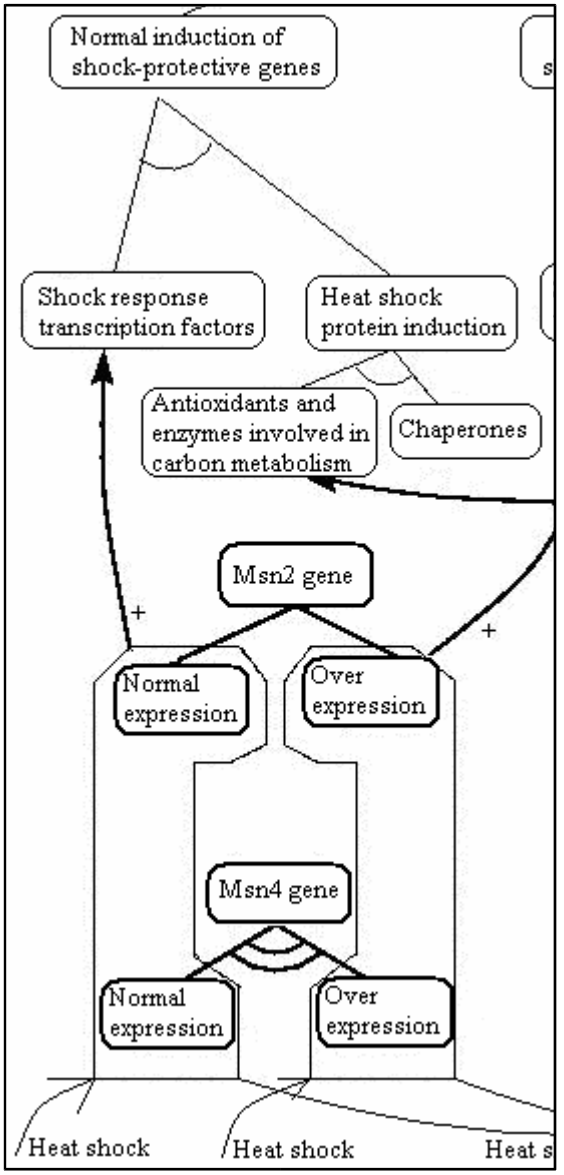
In other cases the interdependency links are grouped together with a double arc; this is an *OR* contribution of the offspring timegoals towards their parent timegoal and means that only one offspring timegoal needs to be satisficed to satisfice the parent.

Contributions of timegoals (positive/negative)

The bottom of a TIG consists of the *observation timegoals* that represent goals concerning the events that need to occur at a low genomic level, to satisfice one or more high-level NFR timegoals. An observation represents an observation or manipulation of a gene or protein at a low genomic level.

Observation timegoals make a *positive or negative* contribution towards satisficing one or more high level NFR timegoals. This Figure shows how interdependency links are used to represent an observation timegoal's contribution towards satisficing an NFR timegoal; such a contribution can be positive (“+” or “++”) or negative (“-“or “--“).

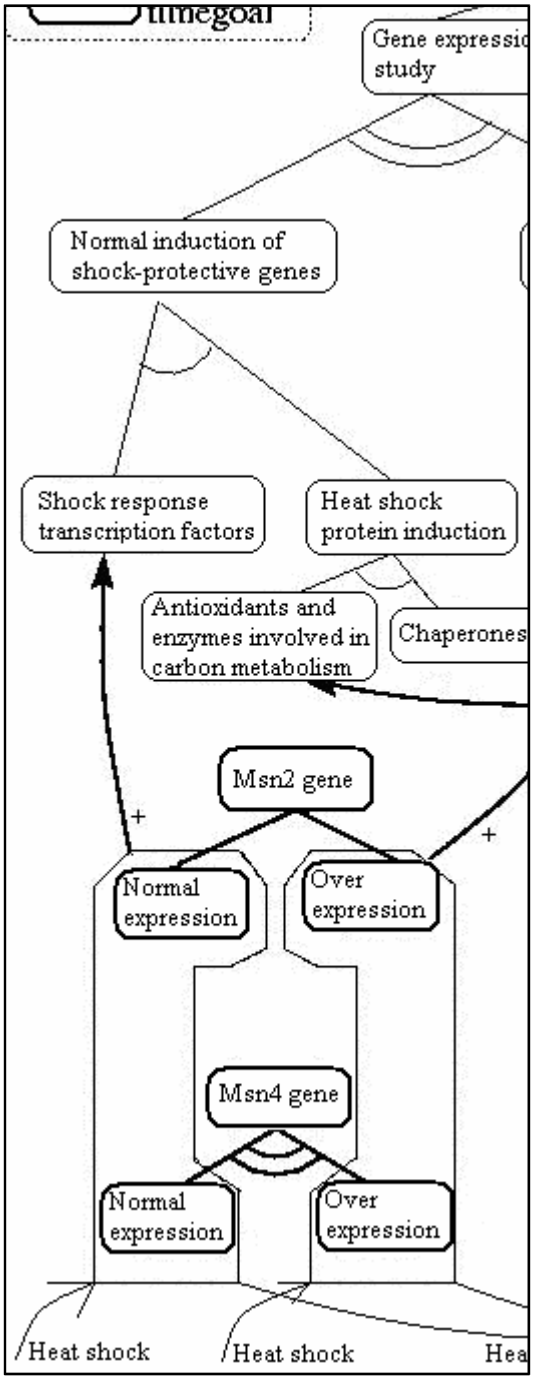
Since observations are considered timegoals they may be decomposed into more specific observations at a lower level. This Figure shows an observation timegoal representing the general goal of observing the Msn2 gene; this timegoal gets decomposed into the timegoals of *overexpressing* the Msn2 gene and observing the Msn2 gene at its *normal expression* level.



Complexes

An event at a time point may involve more than one participating genes or proteins in specific states of expression. The IGIPI framework offers a structural abstraction for grouping the participants at a time point. This abstraction is called a *complex*.

A complex joins several objects such as genes or proteins that participate in a transformation simultaneously. This Figure illustrates several examples of gene complexes. When a “normal expression” of Msn2 and a “normal expression” of Msn4 are joined in a complex, together they contribute towards satisficing the “shock response transcription factors” NFR timegoal, thus inducing the function of “yeast adaptation to a heat shock”.

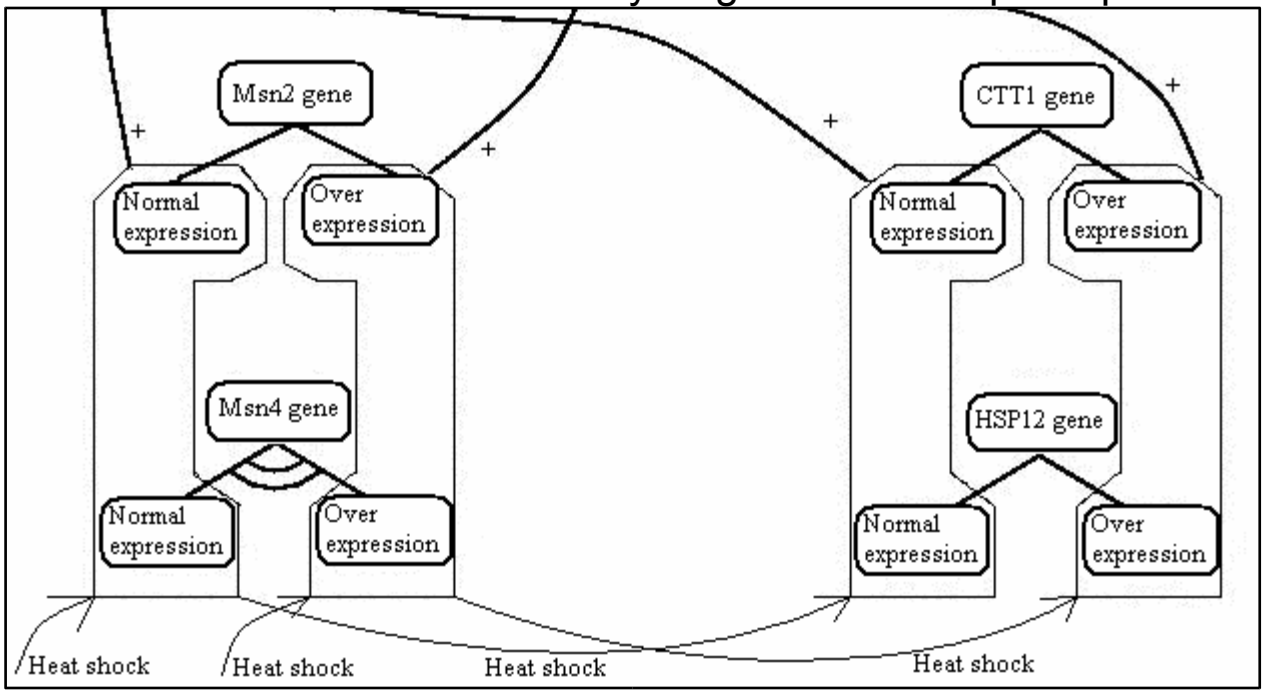


Time representation

The IGIPI framework deals with time and the changes that occur over time in a biological system. It allows representing processes that cause a change in the state of a biological system. The IGIPI framework refers to these processes as *transformations*. This figure shows the “Heat Shock” transformation as lines connecting observation timegoals.

A transformation consists of the participating timegoals, the environmental conditions involved (which may be preconditions for the transformation to occur) and the effects or changes induced by the transformation on the participating timegoals.

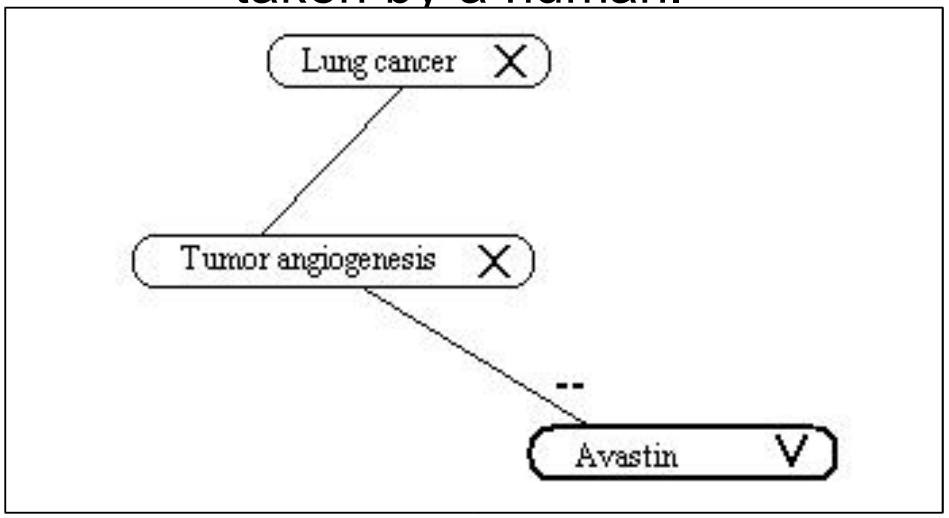
When a transformation precedes a timegoal's contribution to a high level timegoal, it means that the transformation and anything before it are prerequisites for the contribution to occur.



A gene expressed at a certain level at time *t* may be affected by a transformation, such that its expression at time *t+1* changes to a different level. This Figure shows a "heat shock" transformation being applied to the overexpressed Msn2 and Msn4 genes, which causes the CTT1 and HSP12 genes to be overexpressed at the next time point.

Contributions for Satisficing Timegoals

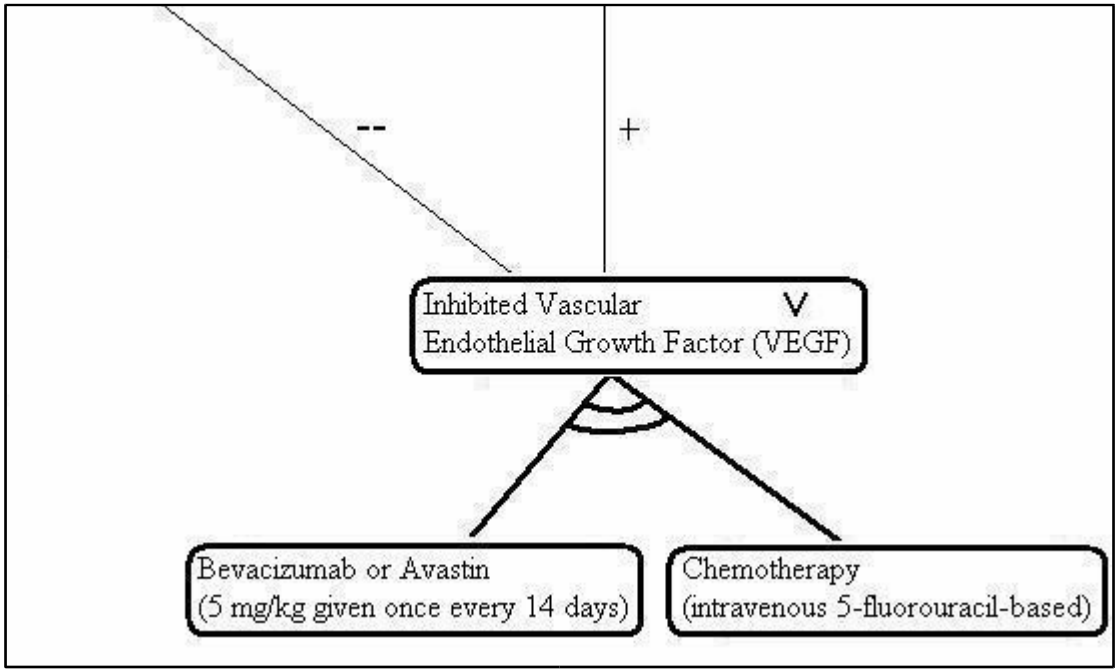
We use the notion of a timegoal being satisficed, as opposed to satisfied. The symbol “V” on a timegoal means that it is satisficed, while a symbol “X” means that it is not satisficed – for example, the timegoal “Avastin” is satisficed meaning that this drug has been taken by a human.



This Figure shows how contributions from lower timegoals are propagated upwards and contribute towards satisficing higher timegoals. The timegoal 'angiogenesis' contributes to timegoal 'lung cancer', but 'angiogenesis' receives a strong negative contribution from drug ‘Avastin’; thus timegoal 'lung cancer' is not satisficed.

Observation Timegoals Under the Influence of Drugs

An observation timegoal is decomposed to represent how it may be observed under the influence of drugs.



For example, this Figure shows the decomposition of the Vascular Endothelial Growth Factor (VEGF) into the timegoals “VEGF under Bevacizumab” and “VEGF under Chemotherapy”. This represents that the protein is in different states under the influence of Bevacizumab and

Integration of Biomedical Knowledge

There exist uncountable biomedical web sites containing bits and pieces of knowledge.

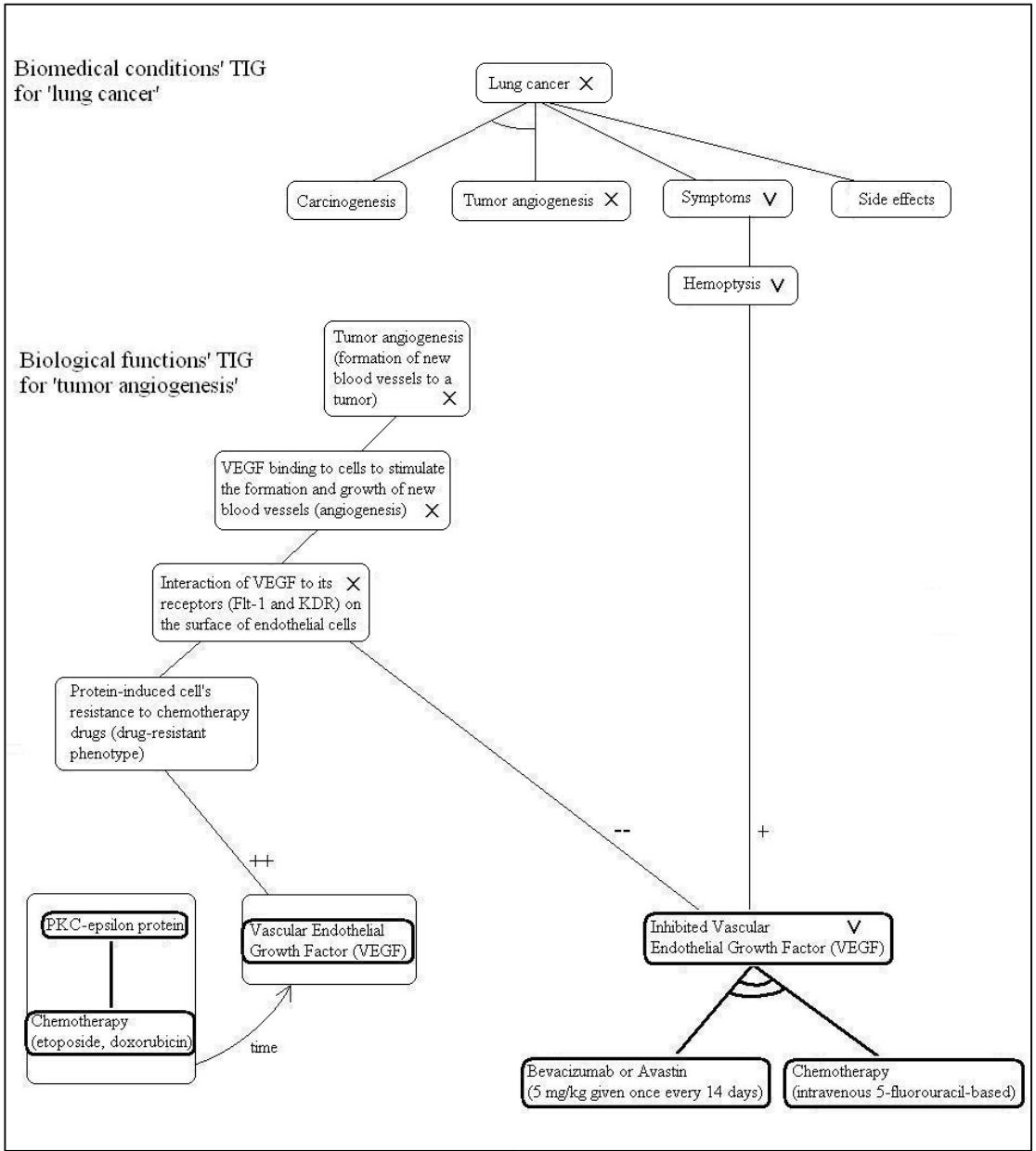
Besides building TIGs for biological functions as discussed previously, the IGIPI framework can also be used to build TIGs representing knowledge about how biomedical conditions are manifested.

These TIGs can help to integrate all of the web-based biomedical knowledge.

Integration of Biomedical Knowledge

To distinguish the timegoals of biomedical condition TIGs from the NFR and observation timegoals of a biological function TIG, we use the name *biomedical condition timegoals*.

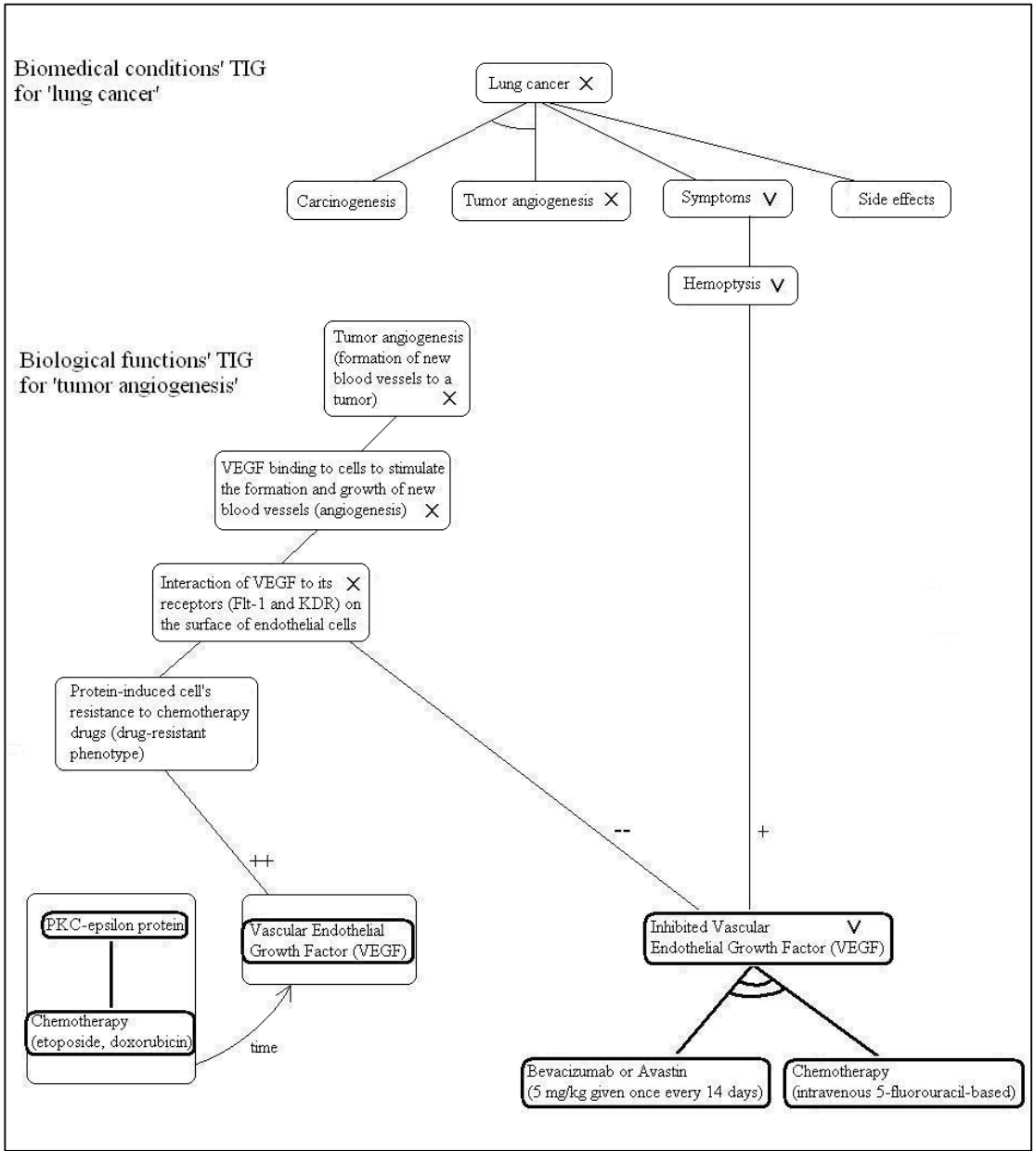
Like NFR timegoals, biomedical condition timegoals may be decomposed downwards into more specific offspring timegoals. The offspring biomedical condition timegoals make an AND or an OR contribution towards their parent timegoal.



Integration of Biomedical Knowledge

The root timegoal of a biomedical condition TIG has a value taken from a domain of biomedical conditions, such as “ischemic stroke”, “haemorrhagic stroke”, “lung cancer” etc. This domain could be the UMLS Unified Medical Language System (that integrates 100 biomedical vocabularies).

The root timegoal is decomposed into timegoals that represent knowledge about the biomedical condition.



Combining Different Ontologies

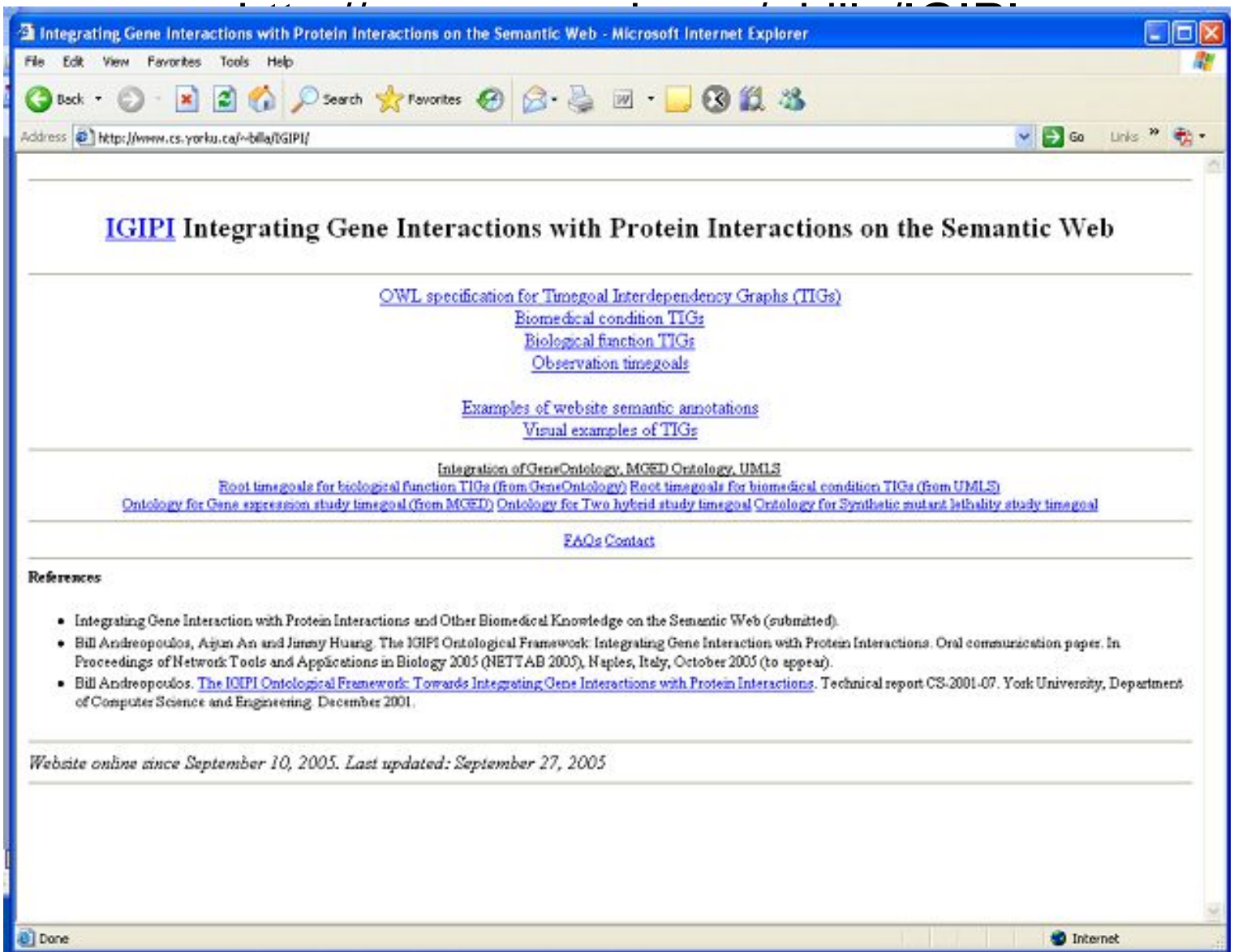
The IGIPI framework allows combining the GeneOntology, UMLS, MGED

GeneOntology is used to give values to the root timegoal of the biological functions TIG

UMLS is used to give values to the root timegoal of the biomedical conditions TIG

MGED gives values under the “Gene Expression Study” timegoal of a biological functions TIG

Website of IGIPI Ontologies in OWL



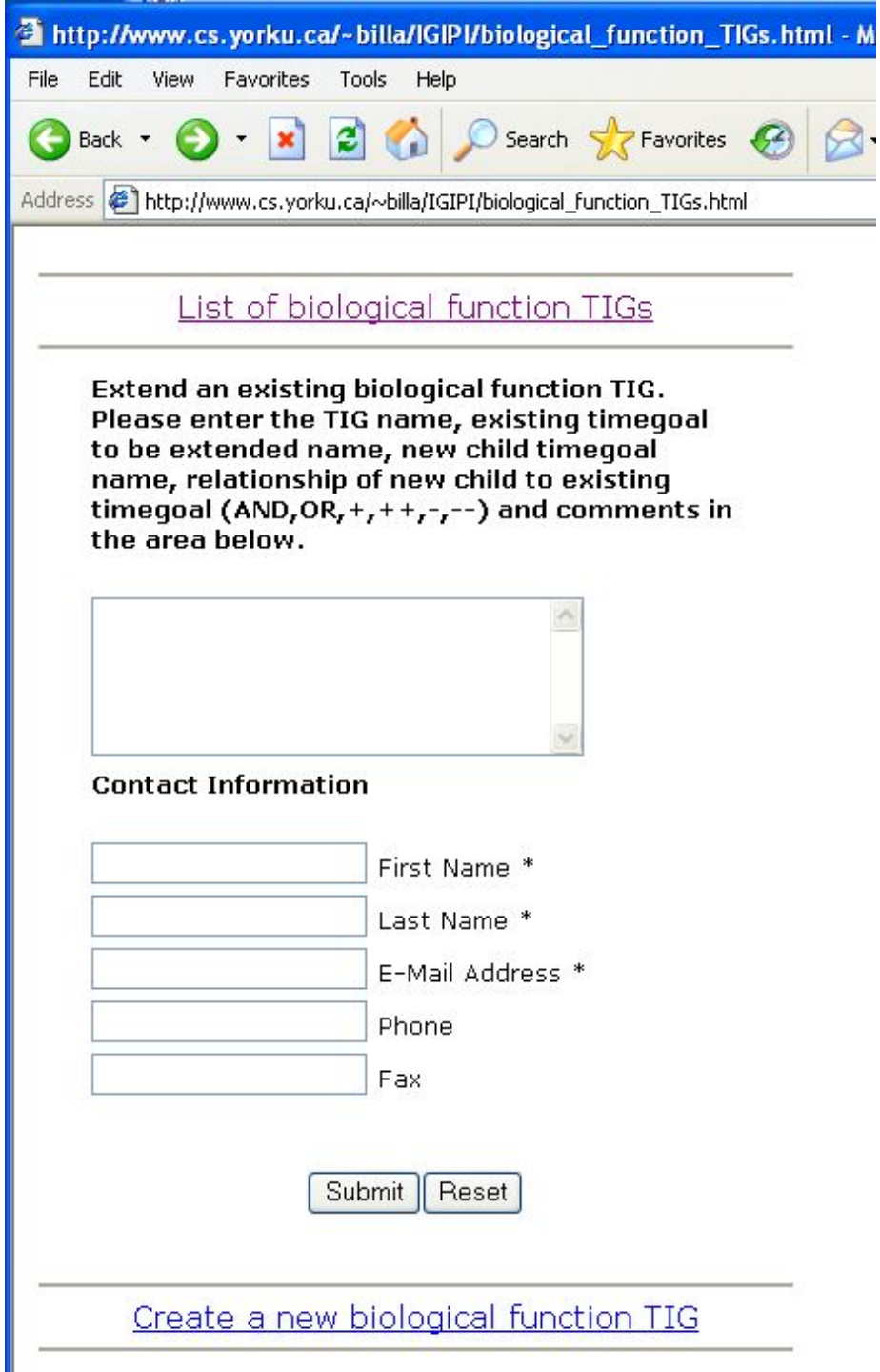
The OWL Specification – a sample

```
<owl:Class rdf:ID="#root_NFR_timegoal">
<rdfs:subClassOf rdf:resource=”#NFR_timegoal”/>
<rdfs:subClassOf>
        <owl:Restriction>
                <owl:onProperty rdf:resource=”#is_a”/>
                <owl:allValuesFrom rdf:resource=”#GO_molecular_function”/>
        </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
        <owl:Restriction>
                <owl:onProperty rdf:resource=”#gets_OR_contribution_by”/>
                <owl:hasValue>“#gene_expression_study”</owl:hasValue>
                <owl:hasValue>“#two_hybrid_study”</owl:hasValue>
                <owl:hasValue>“#synthetic_mutant_lethality_study”</owl:hasValue>
        </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
```

Example IGIPI ontology

<root_NFR_timegoal rdf:ID="#tumor angiogenesis">
<is_a rdf:resource="#tumor angiogenesis"/>
</root_NFR_timegoal>

<NFR_timegoal rdf:ID="#Interaction of VEGF to its receptors">
<contributes_OR_to rdf:resource="#tumor angiogenesis"/>
</NFR_timegoal>



Extending an IGIPI ontology

How does a user annotate his/her website with semantic information?

<root_biomedical_condition_timegoal rdf:ID="#lung cancer">
<is_a rdf:resource="#lung cancer"/>
</root_biomedical_condition_timegoal>

<root_NFR_timegoal rdf:ID="#tumor angiogenesis">
<is_a rdf:resource="#tumor angiogenesis"/>
</root_NFR_timegoal>

<observation_timegoal rdf:ID="#Avastin"/>

<NFR_timegoal rdf:ID="#Interaction of VEGF to its receptors">
<gets_negative_contribution_by rdf:resource="#Avastin"/>
</NFR_timegoal>

Future Work

Advertise the IGIPI website

Encourage users to annotate their biomedical websites using IGIPI OWL ontologies

Encourage users to contribute to extending the ontologies via the IGIPI website