Brain-Gene Ontology: Integrating Bioinformatics and Neuroinformatics Data, Information and Knowledge to Enable Discoveries

Nikola Kasabov, Vishal Jain, Paulo C. M. Gottgtroy, Lubica Benuskova, Frances Joseph
Auckland University of Technology, Auckland, New Zealand
nkasabov-, vjain-, pgottgtroy-, lbenusko-, fjoseph @aut.ac.nz

Abstract

The paper presents some preliminary results on the brain-gene ontology (BGO) project that is concerned with the collection, presentation and use of knowledge in the form of ontology. BGO includes various concepts, facts, data, software simulators, graphs, videos, animations, and other information forms, related to brain functions, brain diseases, their genetic basis and the relationship between all of them. The first version of the brain-gene ontology has been completed as a hierarchical structure and as an initial implementation in the Protégé ontology building environment.

1. Introduction

Ontology is a specification of a conceptualization of a knowledge domain [1]. That is, ontology is a description (like a formal specification of a program) of the concepts and relationships between them to support the sharing and reuse of formally represented knowledge among AI systems. For experimental purposes the medical ontologies [2], biomedical ontology (http://www.bioontology.org/) and the gene ontology (GO) have been created (http://www.geneontology.org/) [3]. The GO project provides a controlled vocabulary to describe gene and gene product attributes in any organism. Our project of Brain-Gene Ontology is focused on the brain and has a broader scope in that it integrates information from different disciplinary domains such as neuroscience, bioinformatics, genetics, computer and information sciences. The BGO is an evolving system that is changing and developing with the addition of new facts and knowledge in it. The system is designed to be used for research, simulation and teaching at different levels of tertiary education. Linking selected structured bodies of physiological, genetic and computational information provides a pathway for different types of users. Designing an interface that enables users with different levels of expertise, specialization and motivation to access the project – either through a familiar or specialist approach, or through a more general introduction – is a critical issue.

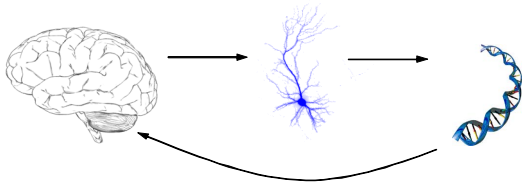

Figure 1. BGO is concerned with the relationships between the brain and genes

In this paper we describe the information structure of BGO, the environment in which the BGO is implemented, and how we can use the BGO for new discoveries and for teaching. We conclude with future directions for BGO development.

2. The BGO information structure

BGO is comprised of 3 main parts: 1. brain organization and function, 2. gene regulatory network, and 3. a simulation model. Brain organization and function contains information about neurons (e.g. their structure and spike generation), synapses and electroencephalogram (EEG) data for different brain states (e.g. normal and epileptic state). Gene regulatory network (GRN) part is divided into sections on neuro-genetic processing, gene expression regulation, protein synthesis and abstract GRN. The third large part, simulation model has sections on computational neurogenetic modeling (CNGM) [4], evolutionary computation, evolving connectionist systems (ECOS) [5], spiking neural network [6], simulation tool and CNGM results. The user can dive into these sections and subsections down to the genetic level and use the information for learning and research. Fig. 2 shows the overall structure of the BGO. Fig. 3 illustrates the detailed kind of information available in

Proceedings of the Sixth International Conference on Hybrid Intelligent Systems (HIS'06)
0-7695-2662-4/06 $20.00 © 2006 IEEE

the BGO about relations between genes, proteins, species and diseases. The user can again dive into each item to obtain the description, illustrations, and links to PubMed publications.

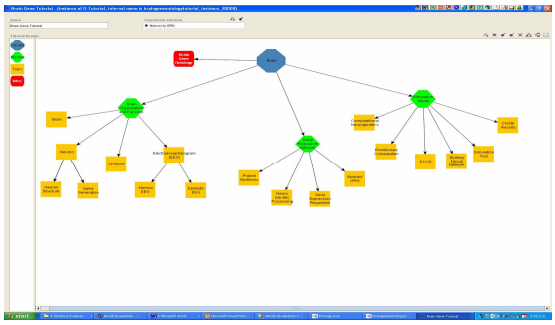

Figure 2. Snapshot of the BGO overall structure with three main sections and their subsections

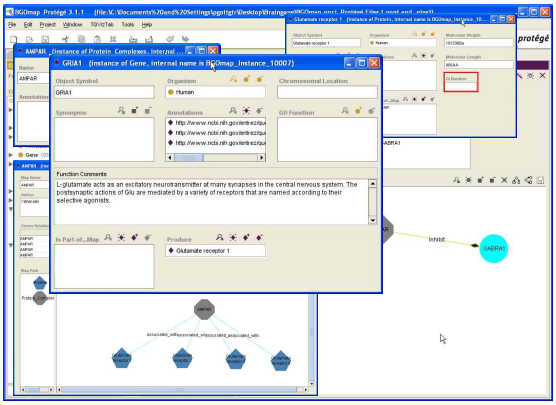

Figure 3. Snapshot of the BGO detail showing relations between genes, proteins, neuronal functions and diseases

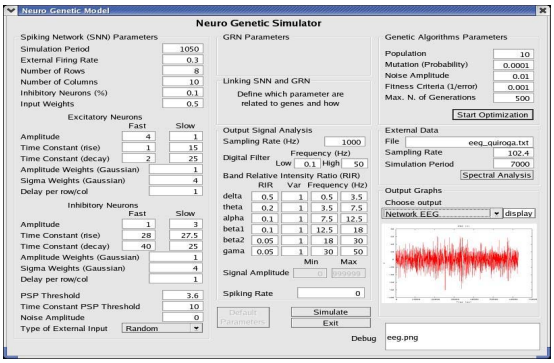

Figure 4. Snapshot of the BGO neurogenetic simulation tool

Data from the BGO can be used in simulation systems, such as computational neurogenetic simulation tool [7-9], NeuCom (www.theneucom.com), WEKA, and others. Results from the simulators can be added back to the BGO to update the BGO current "knowledge". Fig. 4 shows the main window of the simulation tool for neurogenetic simulations, which however is not available for users at the current stage of BGO. Only the theory of the computational neurogenetic modeling and some illustrative results are included.

3. Implementing BGO in a Protégé evolving ontology environment

Protégé is an open source ontology building environment developed by the Medical Informatics Department of the Stanford University (http://protege.stanford.edu/index.html). We have developed a set of plug-ins to enable to visualize, extract and import knowledge from/into different data sources and destinations. BGO is based on the two most used biological data sources, named Gene Ontology, and Unified Medical Language System – UMLS, along with knowledge integrated from gene expressions databases [10]. It also incorporates knowledge acquired from biology domain experts and from different literature databases such as PubMed. Another feature is the graphical presentation of relations by specific Protégé means (dynamic graphs, attached documents and pictures). BGO utilizes a novel evolving conceptual metadata structure which allows to incorporate new discoveries and adapt its structure. This evolving structure keeps track of change and provenance of source, date, among others [11]. Thus, the ontology framework that we have developed enables hierarchical representation of relationships between genes, proteins, neurons and brain functions in a complex evolving structure.

4. Using BGO for new discoveries

The developed BGO system provides conceptual links between data on brain functions and diseases, their genetic basis, experimental publications, graphical illustrations and the relationships between the concepts. Each instance in BGO represents experimental research and is traceable through a query language that allows us, for example, to answer questions such as "Which genes are related to the occurrence of epilepsy?" One can identify relationships between molecular weight, chromosome location, gene product, function in neurons, mutations and related diseases, and so on.

BGO also allows users to select and export the specific data of their interest like chromosomal

location or molecular sequence length, which can then be analysed in a software machine learning environment, such as WEKA and NeuCom to train prediction or classification models and to visualise relationship information. Such exported data can also be analysed in a different manner by standard bioinformatics softwares like BLAST and FASTA for revealing homology patterns for those genes/proteins of interest.

In future, the integral part of BGO will become the module of CNGM to aid discoveries of complex gene interactions underlying oscillations in neural systems (Fig. 5) [6, 7].

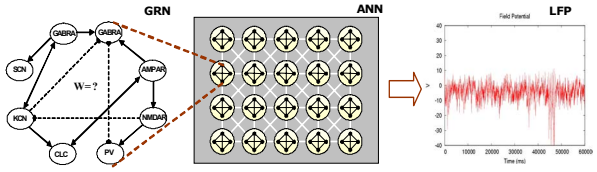

Figure 5. Computational NeuroGenetic Model (CNGM) as an abstract gene regulatory network (GRN) embedded in each neuron of the ANN with particular output LFP to match real data

In CNGM, interactions of genes in neurons affect the dynamics of the whole neural network model through neuronal parameters, the values of which depend on gene expression. Through optimization of the gene interaction network, initial gene/protein expression values and neuronal parameters, particular target states of the neural network operation can be achieved, and statistics about gene interaction matrix can be extracted. In such a way it is possible to model the role of genes and their interactions in different brain states and conditions that manifest themselves as changes in local field potentials (LFP) or EEG. At present, the method and simulation tool is introduced and described together with examples of simulation results for normal and epileptic LFP/EEG.

The BGO is an evolving ontology that evolves its structure and content so that new information can be added in the form of molecular properties, disease related information and so on. All of this information can be re-utilized to create further models of brain functions and diseases that include models of gene interactions. We hope that by linking and integrating simulation results from the CNGM simulations with genetic information in the BGO, we can facilitate better understanding of metabolic pathways and modelling of gene regulatory networks, and ultimately a more complete understanding of the pathogenesis of brain diseases.

5. Using BGO for teaching

The BGO can be used as an online teaching and learning tool for undergraduate and postgraduate students as well as researchers in bioinformatics, neuroinformatics, computer and information sciences and related areas. It exemplifies the importance of use of ontologies in nowadays knowledge management and interpreting relationships between molecules and brain functions. It enables teaching the basics of molecular biology and gene regulatory networks as well as introducing the area of computational neurogenetic modeling [4]. The BGO can be used to better understand and explain various topics related to brain, genes and their modeling, for example: the structure of the brain; main functions of the brain; the importance of gene mutation on brain functions and behavior; importance of gene regulatory networks in neurons; mental/neurological disorders and main receptor/ion channel genes/proteins involved; understanding neural signal propagation and the role of synapses; analysis of LFP/EEG data and its relevance to brain functions; neurogenetic modeling and the role of its parameters for the outcome.

The interface is built using textual, graphical, audio and visual media. The inclusion of 3D animation, gives both a dynamic narrative introduction and overview to the BGO (see Fig. 6).

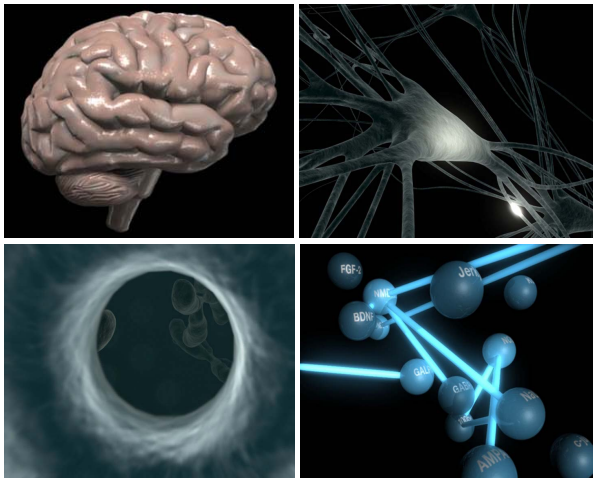

Figure 6. Snapshots from animations embedded in BGO (upper left) the brain (upper right) signal propagation in and within neurons (lower left) entering the neuron's nucleus (lower right) building an abstract GRN

The animation navigates and provides a sophisticated visual method of integration across the different domains of the BGO. This approach, drawn from the

aesthetics and immersive experience of computer games and special effects technologies is introduced as a way of engaging a younger or novice audience in this complex, emergent, cross disciplinary field of linking genes to brain functions.

6. Conclusion

The paper reports on a preliminary version of a world-first BGO that includes conceptual and factual information about the brain and gene functions and their relationships. BGO is a modern tool for teaching and research across areas of bioinformatics, neuro-informatics, computer and information sciences at different levels of education and expertise.

BGO allows users to navigate through the rich information space of brain functions and brain diseases, brain related genes and their activities in certain parts of the brain and their relation to brain diseases; to run simulations; to download data that can be used in a software machine learning environment, such as WEKA and NeuCom to train prediction or classification models; to visualise relationship information; and to add new information as the BGO has an evolving structure. The BGO contains also a computational model and later a simulation tool for modeling complex relationships between genes and neural oscillations. The BGO is designed to facilitate active learning and research in the areas of bioinformatics, neuroinformatics, information engineering, and knowledge management. Different parts of it can be used by different users, from a school level to postgraduate and PhD student level. In future, more data and information will be added, that will include both, higher level information on cognitive functions and consciousness, and lower level quantum information and models.

7. Acknowledgments

This work is supported by NERF AUT0201, TAD and KEDRI. We would like to thank several colleagues who participated at different stages of the BGO development, namely Ilkka Havukkala, Rene Kroon, Laurent Antonczak, Simei G. Wysoski, William Lu, John Eyles, and Mark Howden (sound). Scott Heappey is particularly acknowledged for creating all animations.

8. References

[1] T. R. Gruber, "A translation approach to portable ontologies", *Knowledge Acquisition,* vol. 5, 1993, pp. 199-220.

[2] D. M. Pisanelli, Ed., *Ontologies in Medicine*, IOS Press, Amsterdam, 2004.

[3] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler and J. M. Cherry, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium", *Nature Genetics,* vol. 25, 2000, pp. 25–29.

[4] L. Benuskova and N. Kasabov, *Towards Computational Neurogenetic Modeling*, Springer, New York, in press.

[5] N. K. Kasabov, *Evolving Connectionist Systems. Methods and Applications in Bioinformatics, Brain Study and Intelligent Machines*, Springer-Verlag, London, 2003.

[6] N. Kasabov and L. Benuskova, "Computational neurogenetics", *Journal of Computational and Theoretical Nanoscience,* vol. 1, 2004, pp. 47-61.

[7] N. Kasabov, L. Benuskova and S. G. Wysoski, "Biologically plausible computational neurogenetic models: modeling the interaction between genes, neurons and neural networks", *Journal of Computational and Theoretical Nanoscience,* vol. 2, 2005, pp. 569–573.

[8] L. Benuskova, V. Jain, S. G. Wysoski and N. Kasabov, "Computational neurogenetic modelling: a pathway to new discoveries in genetic neuroscience", *Intl. J. Neural Systems,* vol. 16, 2006, pp. 215-227.

[9] S. G. Wysoski, L. Benuskova and N. Kasabov, "Simulation of neurogenetic models", *Proc. Neuro-Computing and Evolving Intelligence,* N. Kasabov and Z. S. H. Chen, Ed., Auckland, New Zealand, 2004, vol. 1, pp. 30-31.

[10] P. Gottgtroy, N. Kasabov and S. MacDonell, "An ontology driven approach for knowledge discovery in Biomedicine," in *PRICAI 2004: Trends in Artificial Intelligence. Proc. VIII Pacific Rim Intl. Conf. AI, Lecture Notes in Artificial Intelligence, Vol. 3157*, C. Zhang, H. W. Guesgen and W. K. Yeap, Ed. Springer-Verlag, Berlin, 2004, pp. 53-67.

[11] P. Gottgtroy, N. Kasabov and S. MacDonell, "Evolving ontologies for intelligent decision support," in *Fuzzy Logic and the Semantic Web (Capturing Intelligence)*, E. Sanchez, Ed. Elsevier, Amsterdam, 2006.