Materials Research. 2014; 17(6): 1391-1396
DOI: http://dx.doi.org/10.1590/S1516-14392014005000098

DNA for Nano-bio Scale Computation of Chemical Formalisms Using Higher Order Logic (HOL) and Analysis Using an Interdisciplinary Approach

Nirmal Kumar*, Nilson Cristino da Cruz, Elidiane Cipriano Rangel

Laboratório de Plasmas Tecnológicos – LapTec, Universidade Estadual Paulista
"Júlio de |Mesquita Filho" – UNESP, Sorocaba, SP, Brasil

Received: March 18, 2014; Revised: May 5, 2014

Bio-molecular computing, 'computations performed by bio-molecules', is already challenging traditional approaches to computation both theoretically and technologically. Often placed within the wider context of ´bio-inspired` or 'natural' or even 'unconventional' computing, the study of natural and artificial molecular computations is adding to our understanding of biology, physical sciences and computer science well beyond the framework of existing design and implementation paradigms. In this introduction, We wish to outline the current scope of the field and assemble some basic arguments that, bio-molecular computation is of central importance to computer science, physical sciences and biology using HOL – Higher Order Logic. HOL is used as the computational tool in our R&D work. DNA was analyzed as a chemical computing engine, in our effort to develop novel formalisms to understand the molecular scale bio-chemical computing behavior using HOL. In our view, our focus is one of the pioneering efforts in this promising domain of nano-bio scale chemical information processing dynamics.

Keywords: *nano technology, HOL, DNA, unconventional computing, chemical computer*

1. Introduction

The idea that molecular systems can perform computations is not new and was indeed more natural in the pre-transistor age. Most computer scientists know of von Neumann's discussions of selfreproducing automata in the late 1940s, some of which were framed in bio-molecular terms based on bio-inspiration. Here the basic issue was that of bootstrapping: can a machine construct a machine more complex than itself? Important was the idea, appearing less natural in the current age of dichotomy between hardware and software, that the computations of a device can alter the device itself. "This vision is natural at the scale of molecular reactions, although it may appear "utopic", to those running huge chip production facilities. Alan Turing also looked beyond purely symbolic processing to natural bootstrapping mechanisms in his work on self-structuring in molecular and biological systems. Purely chemical computers have been proposed by Ross and Hjelmfelt extending Turing's approach"[1-7].

In biology, the idea of molecular information processing took hold starting from the unraveling of the genetic code and translation machinery and extended to genetic regulation, cellular signaling, protein trafficking, morphogenesis and evolution – all of this independently of the development in the lifesciences. For example, because of the fundamental role of bio-information processing in evolution, and the ability to address these issues on laboratory time scales at the molecular level, a number of alternative solutions exist indefinitely[5-7].

*e-mail: hmfg2014@gmail.com

2. Theoretical Background and Motivation

The unique properties of DNA make it a fundamental building block in the fields of supramolecular chemistry, nanotechnology, nano-circuits, molecular switches, molecular devices, and molecular computing. In addition to information processing, DNA acts as molecular scale heat engine, DNA stores energy, also available on hybridization of complementary strands or hydrolysis of its phosphodiester backbone[6-10].

"Bio-molecular computers are molecular-scale, programmable, autonomous computing machines in which the input, output, software, and hardware are made of biological molecules. Bio-molecular computers hold the promise of direct computational analysis of biological information in its native bio-molecular form, eschewing its conversion into an electronic representation to advance the nanoscale fabrication techniques for nanobio devices".

Nucleic acids are molecules of choice for both established and emerging nanoscale technologies. These technologies benefit from large functional densities of 'DNA processing elements' that can be readily manufactured[11-16].To achieve the desired functionality, polynucleotide sequences are currently designed by a process that involves tedious and laborious filtering of potential candidates against a series of requirements and parameters. Here, we present a complete novel methodology for the rapid rational design of large sets of DNA sequences using HOL - Higher Order Logic for nanoscale or nano-bio scale systems[15-16].

As we know, applied mathematics and computer science could provide the needed abstraction, for consolidating the knowledge of bio-molecular systems or bio-inspired

systems. Computer and bio-molecular systems both start from a smaller set of elementary components from which, layer by layer, as more complex entities are constructed with an ever increasing demanding applications based on sophisticated functions. "Nevertheless, the mathematical abstractions, tools and methods used to specify and study computer systems should illuminate our accumulated knowledge about bio-molecular systems. The exceptional ability of DNA to mediate charge transport (CT) is the basis of novel molecular devices and may be exploited by the cell for both redox sensing, signaling or other specified information processing"[8-11,17,18].

"Interpreting chemical reactions in terms of nano-bio scale interaction is yet another challenge. So far CMOS design and analog emulation of Reaction-Diffusion(R-D) systems have demonstrated the feasibility of mapping chemical dynamics onto silicon architectures. Semiconductor devices based on minority carrier transport may succeed in the upcoming designs of nano-scale R-D processors and single-electron R-D circuits"[15-17].

Inspite of numerous promising preliminary results obtained the in "R-D" computing domain, this particular field still remains an imaginary interdisciplinary art rather than science, most "Reaction-Diffusion" processors are produced on an ad hoc basis without structured top-down approaches, mathematical verification, rigorous methodology, relevant to other domains of advanced computing and computer hardware (this could be nanobio-wetware for implementation!) design. It is in this context, we have planned to consider HOL for rigourous analysis. There is a need to develop a coherent theoretical foundations for "Reaction-Diffusion" computing in chemical media or bio-chemical media, and adapt new computational substrates. As Einstein said," Imagination is more important than domain knowledge"[10-16,18].

The connection with nanomachines and nanosystems is very clear and will become more pervasive in the near future. In our view, DNA Computation is exciting for the following reasons[11-18]:

- opens the possibility of a simultaneous bootstrapping solution of future computer design, construction and efficient computation.
- provides programmable access to nanosystems and the world of molecular biology, extending the reach of computation.
- admits complex, efficient and universal algorithms running on dynamically constructed dedicated molecular hardware.
- can contribute to our understanding of information flow in evolution and biological construction.
- is opening up new formal models of computation, extending our understanding of the limits of computation.

2.1. *HOL as a simulation tool to develop bio-molecular computing systems*

HOL (Higher Order Logic) denotes a family of interactive theorem proving systems sharing similar (higher-order) logics and implementation strategies. Systems in this family follow the LCF approach as they are implemented as a library in some programming language. This library implements an abstract data type of proven theorems so that new objects of this type can only be created using the functions in the library which correspond to inference rules in higher-order logic. As long as these functions are correctly implemented, all theorems proven in the system must be valid. In this way, a large system can be built on top of a small trusted kernel. - source wiki and[5,6].

"Isabelle is a generic proof assistant. It allows mathematical formulas to be expressed in a formal language and provides tools for proving those formulas in a logical calculus. The main application is the formalization of mathematical proofs and in particular *formal verification*, which includes proving the correctness of computer hardware or software and proving properties of computer languages and protocols"[5,6].

2.2. *Sources*

1. http://isabelle.in.tum.de/overview.html {TU Munich, Munich, Germany}
2. http://www.cl.cam.ac.uk/research/hvg/Isabelle/Cambridge/ {Computer Science Laboratory ,Cambridge University, UK/}
3. http://www.wisdom.weizmann.ac.il/~tomr/ {Weizmann Institute of Science, Rehovot, Israel}

2.3. *An approximate HOL based simulation framework*

Theory Seq

theory Seq

imports Main

(* Title: HOL - Seq.thy 2014.

Author: Nirmal, LapTec, UNESP, Sorocaba, SP, Brazil.

DNA is considered as an abstraction, defined by a Mathematical String with four chemical bonds – {A,G,C,T}

A Simple Lemma is written to compute the DNA Sequence using "A', the rest could be derived easily for bio-molecular computation involving sensing, informatics or other computing tasks.

This Template based on HOL syntax is provided to encourage the reader in defining novel chemical formalisms to advance nano-bio computing platforms and devices for bio-molecular computing.

*)

header {* Finite sequences of the DNA Material System Using Higher Order Logic Syntax *}

theory Seq

imports Main

begin

datatype 'A seq = Empty | Seq 'A "'A seq"

fun compute :: "'A seq => 'A seq => 'A seq"

where

"compute Empty ys = ys"

| "compute (Seq A xs) ys = Seq A (compute xs ys)"

fun compute :: "'A seq => 'A seq"

where

"compute Empty = Empty"

| "compute (Seq A xs) = compute (compute xs) (Seq A Empty)"

lemma compute_A: "compute xs A = xs"

by (induct xs) DNA_A

lemma compute_G: "Describe the lemma here"

by (Write the Rule here/Left as an exercise to the reader) DNA_G

lemma compute_C: "Describe the lemma here"

by (Write the Rule here/Left as an exercise to the reader) DNA_C

lemma compute_T: "Describe the lemma here"

by (Write the Rule here/Left as an exercise to the reader) DNA_T

end

3. Results and Discussions

In this communication we have focused on the chemical formalisms of nano-bio system using DNA as the modeling element and shown some insights into the nano-bio scale formalisms. Further, we explain how to design and compute a simple bio-molecular sequence using HOL – higher order logic, as discussed in the abstract. A template showing the implementation of HOL syntax is also presented in one of the above sections, so as to acquaint the reader with HOL based design methodology. We are not going into the in-depth details of HOL based concepts as there are plenty of scientific papers already published and available with on line tutorials. Graphical views or flow charts of the design and methodology sequences are depicted in this paper to simplify the process of understanding the bio-chemical formalisms and computational concepts (Figures 1-5).

In the HOL-Template shown above, we discuss DNA as a nanoscale building block or as a bio-chemical tool to implement the nano-bio scale computation. DNA is considered as an abstraction and as a "Mathematical String", to showcase the theoretical model. As it is a known fact that DNA has 4 chemical bonds namely – A,G,C,T., we could further define ``DNA`` as {A,G,C,T} structure. The HOL-Template has a simple lemma and deduction methodology for Chemical bond ``A``,GCT bonds are not described on the basis of lemmas as we leave it as an exercise to the reader. For further understanding a "Reaction-Diffusion" computing processor could be easily deducted by proper sequencing and there by deriving the application. We draw inspiration from Adamatzky[15,16] to advance our research in reaction-diffusion computation.

3.1. *Considerations of mathematical and chemical computing formalisms for bio-molecular systems illustrated via Figures 1-5 as depicted below*

Since the computer science community has been developing, different approaches to support writing correct programs on a continuous basis, e.g. abstract interpretation , type systems, model checking and theorem proving. The art of theorem proving is devoted to provide tools to verify the correctness of a program by means of a formal mathematical proof. As large and complex programs necessarily require large and complex proofs, pen-and-paper based proofs become very difficult or even impossible to grasp. For this reason, the proofs are created with the assistance of an interactive or an automated theorem prover, in our case it is "Isabelle System"[1-11,18].

"Bio-chemical compounds which react are essentially parallel systems as per the existing and derived computing paradigms. Molecules of the same chemical compound will react in different ways at different moments. The high number of concurrent processes and parameters prevent

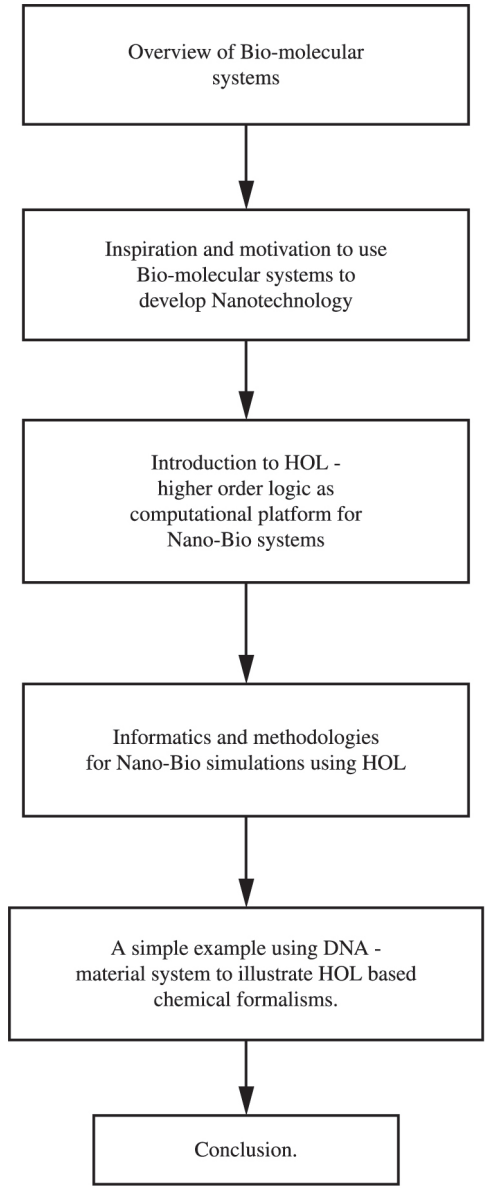

Figure 1. Total overview of the paper.

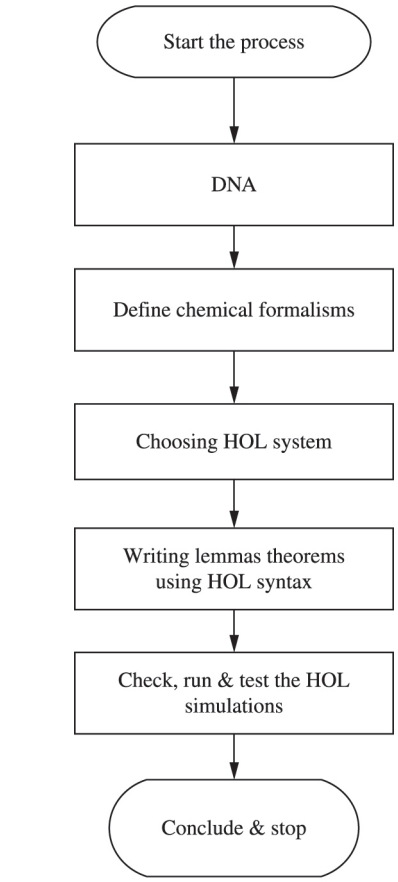

Figure 2. Chemical Formalism Flow Chart.

them from being simulated using older methods. Hence novel methods to simulate them are in fact essential". Here, we present a novel method for the rational design of optimized DNA sequences for a wide range of technological applications. The advantages of our new HOL based sequence design concept can be summarized as follows - the sets of contextually essential sequences exhibit extremely narrow ranges of melting temperatures, a requirement that is central to all applications. Furthermore, the mathematical tools of our HOL-based method allow us to impose very complex and detailed requirements on the sequences to be generated. These requirements are then automatically satisfied without an exception in every one of them[1-16,18].

"Reaction-Diffusion (R-D) chemical systems are well known now for their unique ability to efficiently solve combinatorial problems with natural parallelism. In R-D processors, both the data and the results of the computation are encoded as concentration profiles of the reagents. The computation per se is performed via the spreading and interaction of wave fronts. The R-D computers are parallel because the chemical medium's micro-volumes update their states simultaneously, and molecules diffuse and react in parallel"[15,16]. For more information on Reaction-Diffusion Computing Systems, we suggest Prof. Adam Adamatzky's website at UWE, Bristol, England, UK.

In our case, we are focusing on DNA – plasma interaction as the R-D Chemical system. We are not discussing the full-scale implementation here, as we intend to show the readers only a methodology to develop chemical formalisms based on HOL. Detailed discussion is beyond the scope of this paper and space constraints.

DNA – is General Sequence Genetic material made up of A,G,C,T. [dsDNA or ssDNA could be used]

Plasma – Non thermal.

Please see Figure 5 for simple explanation of concurrent mechanism implementation and algorithm design.

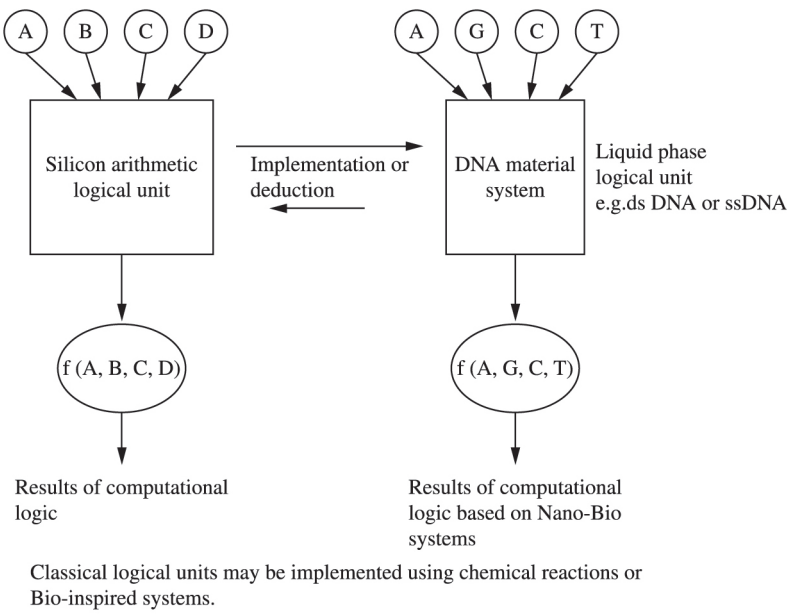

Figure 3. Deduction of bio-inspired systems from classical computing systems.

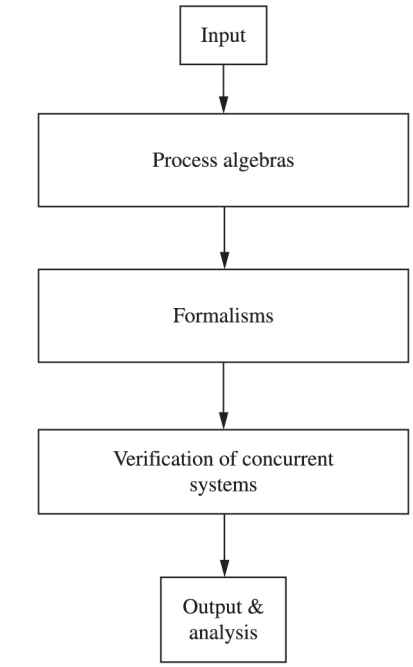

Figure 4. Mathematical Formalisms for applied informatics.

In our view and expectation, we are sure that the readers could easily adapt the methodology and HOL based framework to suit their research according to the situation. Figures 1-5 serve this purpose.

4. Conclusions with future perspectives

Promising concepts of DNA Computing operates in natural noisy environments, such as in a glass of water or even a simple test tube in a laboratory. It involves and includes an evolvable platform for computation in which the computer construction machinery itself is embedded. Bio-inspired "Embedded Computing", is possible without electrical power in microscopic, error prone and real time environments. Using these mechanisms and technology compatible with our own bio-inspired approach, DNA Computing is linked to molecular construction. These computations may eventually also be employed, to build three dimensional self-organizing partially electronic or more remotely even quantum computers. Moreover, DNA Computing opens computers to a wealth of applications in intelligent manufacturing systems, complex molecular diagnostics and molecular process control.

In future, we intend to focus on DNA based nano-bio computing platforms. For example using binding of DNA to Graphene and their interactions in plasma radiation environments. We are in the process of designing and developing computational tools based on mathematical methods using HOL as a pioneering effort. We hope to

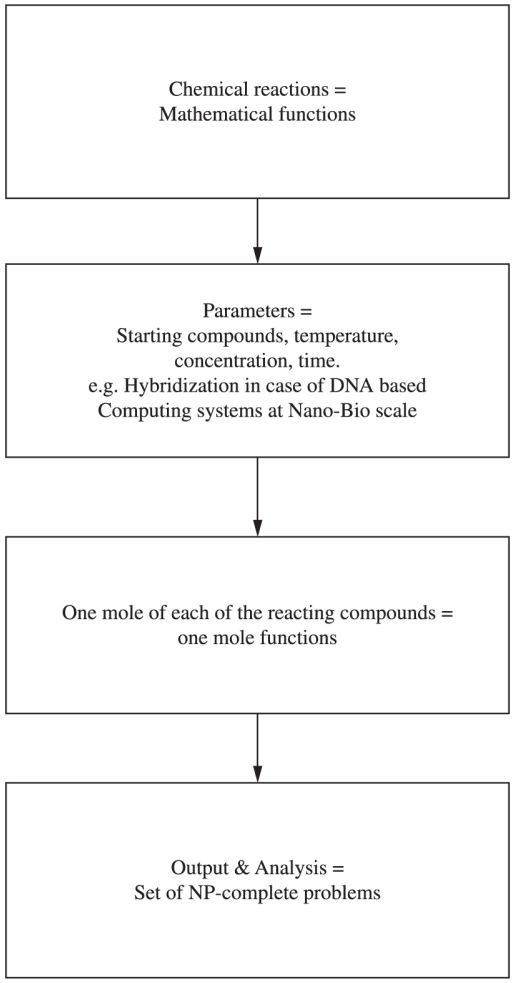

Figure 5. Implementation of concurrent systems and their algorithm design.

achieve remarkable progress in this new approach, to design better nano-bio computing systems using plasma processing technologies. Both thermal and non-thermal plasmas could be used in performing our experiments to check or verify our nanoscale mathematical models and chemical formalisms using HOL. However, we are focusing mainly on non-thermal plasmas and bio-materials interactions at the moment in our R&D efforts.

Acknowledgments

We, sincerely thank UNESP for providing a conducive environment and encouragement for research of novel concepts. Furthermore, we thank all those who have helped us directly and indirectly in producing the current paper for the POSMAT 2014 conference in Brazil. The authors strictly abide by open source software regulations, where applicable.

References

1. Cover TM and Thomas JA. *Elements of information theory*. New York: John Wiley & Sons; 1991. http://dx.doi.org/10.1002/0471200611
2. Kane BE. A silicon-based nuclear spin quantum computer. *Nature*. 1998; 393:133-137. http://dx.doi.org/10.1038/30156
3. Tribus M. *Thermostatics and thermodynamics*. New York: Van Nostrand; 1961.
4. Feynman RP, Leighton RB and Sands M. *The feynman lectures on physics*. Massachusetts: Addison-Wesley; 1964. vol 1.
5. Higher Order Logic. *Isabelle*. Cambridge: University of Cambridge. Available from: <https://www.cl.cam.ac.uk/research/hvg/Isabelle/>.
6. Kumar DNT and Wei Q. A general computational framework and simulations of branching programs of boolean circuits using Higher Order Logic (HOL) software: an insight into ECAD tool design paradigm. *Computer and Information Science*. 2012; 5(6):6-12. http://dx.doi.org/10.5539/cis.v5n6p6
7. Kumar DNT, Kost Y, Qiao H and Wei Q. Nucleic acids data sequencing using higher order logic: a suggestion of basic computational framework towards bio-sensors and gene-chips design, implementation and verification. *Journal of Applied Mathematics and Bioinformatics*. 2012; 2(2):65-79.
8. Kumar DNT, Jing L and Wei Q. An insight into calculus of communicating system and formalisms, for bioinspired logic device design using computational aspects of protocells. *International Journal of Applied Research on Information Technology and Computing*. 2012; 3(1):70-79. http://dx.doi.org/10.5958/j.0975-8070.3.1.007
9. Kumar DNT, Wei Q, Min Z, Jing L and Qiao H. computational analysis of logic gates and circuits derived from gene systems using quantified boolean formula methodology in CNF format. *International Journal of Applied Research on Information Technology and Computing*. 2012; 3(3):145-156. http://dx.doi.org/10.5958/j.0975-8070.3.3.014
10. Divaku NTK, Wei Q and Kost Y. RNA inspired genetic logic devices and functional verification based on Calculus for Communicating Systems (CCS) approach using formalisms. *Journal of Computational Intelligence in Bioinformatics*. 2011; 4(2):183-192.
11. Rozenberg G and Spaink H. DNA computing by blocking. *Theoretical Computer Science*. 2003; 292(3):653-665. http://dx.doi.org/10.1016/S0304-3975(01)00194-3
12. Penchovsky R and Ackermann J. DNA library design for molecular computation. *Journal of Computational Biology*. 2003; 10(2):215-229. PMid:12804092. http://dx.doi.org/10.1089/106652703321825973
13. Conrad M and Zauner KP. DNA as a vehicle for the self-assembly model of computing. *Bio Systems*. 1998; 45(1):59-66. http://dx.doi.org/10.1016/S0303-2647(97)00062-2
14. Phadke RS. Biomolecular electronics in the twenty-first century. *Applied Biochemistry and Biotechnology*. 2001; 96(1-3):269-276. http://dx.doi.org/10.1385/ABAB:96:1-3:279
15. Adamatzky A. *Computing in nonlinear media and automata collectives*. London: IoP Publishing; 2001.
16. Adamatzky A, De Lacy Costello B and Asai T. *Reaction-diffusion computers*. Amsterdam: Elsevier; 2005.
17. Pancoska P, Moravek Z, Moll UM. Rational design of DNA sequences for nanotechnology, microarrays and molecular computers using Eulerian graphs. *Nucleic Acids Research*. 2004; 32(15):4630-4645. PMid:15333695 PMCid:PMC516071. http://dx.doi.org/10.1093/nar/gkh802
18. Kumar DNT, Wei Q and Dan T. A framework of basics for sensor modelling, using Higher-Order Logic (HOL), as a sensing mechanics inference computational platform. *International Journal of Applied Research on Information Technology and Computing*. 2011; 2(3):54-60. http://dx.doi.org/10.5958/j.0975-8070.2.3.020