Role of the Computational Intelligence in Drugs Discovery and Design: Introduction, Techniques and Software

Geeta Yadav
Research Scholar, Dept. of Pharmaceutical Sciences, Birla Institute of Technology, Mesra, Ranchi, Jhrakhand, India

Yugal Kumar
Assistant Professor in CSE/IT Dept., Hindu College of Engineering, Industrial Area, Sonepat, Haryana, India 131001

G. Sahoo
Professor in Dept. of Information Technology, Birla Institute of Technology, Mesra, Ranchi, Jhrakhand, India

ABSTRACT
Drugs discovery & design is an intense, lengthy and consecutive process that starts with the lead & target discovery followed by lead optimization and pre-clinical in vitro & in vivo studies. This paper throws light on different computational techniques that play a vital role in the drugs discovery & design process. Earlier, computational techniques are use in the field of computer science, electrical engineering and electronics & communication engineering to solve the problems. But, now day's use of these techniques has changed the scenario in drugs discovery. & design from the last two decades. This paper present brief description of different computational techniques such as Particle Swarm Optimization, Ant Colony Optimization, Artificial Neural Network, Fuzzy logic, Genetic Algorithm, Genetic Programming, Evolutionary Programming, Evolutionary Strategy and also provide a tabular comparison of these techniques as well as a list of computational tools/ software.
Keywords: *Biological Inspiration, Computational Techniques, Fitness Function, Programming, Optimization*

1. INTRODUCTION
The traditional approach through which drugs were discovered mainly based on hit and trial method, like plant-based medicines, serendipity (penicillin) and chemical modification. But there has been major change in the field of drug discovery & design in the previous two decades. In the post-genomic era, rational drug discovery is a major approach for discovering and designing new drugs. Generally, experimental techniques are costly, time-consuming, and involve the use of animals for testing. Therefore, computer-based in silico models are alternate to experimental models. It is estimated that a drug from concept to market would take 12 years and cost more than US$800 million on an average [1]. For every 5,000-10,000 compounds that enter the research and development (R&D) pipeline, ultimately only one receives approval [2]. Several new technologies such as genomic data, chemical genomics, high throughput screening in the drug discovery process and computational tool/software have been developed and applied in drug discovery & design, to short the research cycle and to reduce the expenses. Computer-aided drug design (CADD) is one of such evolutionary technologies [3]. In the present time, drug design is based on disease models. Drugs discovery and design is a very complex process for the pharmaceutical companies because it is comprehensive, expensive, time-consuming and full of risk. The process consists of many steps that are shown in figure 1.

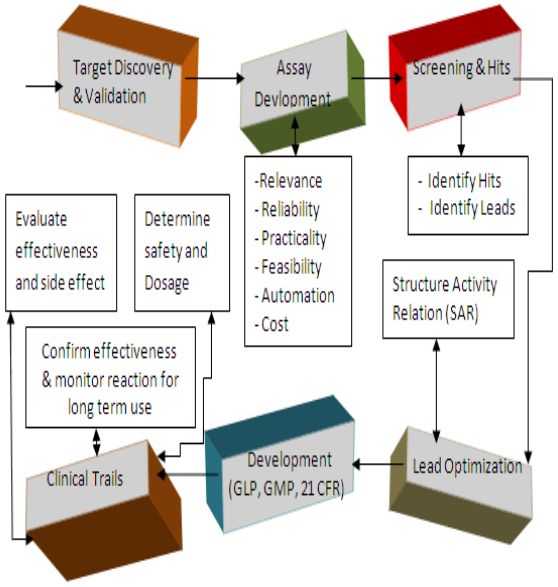

Figure 1: Key steps in Drug discovery and Design process

This paper provides an overview of computational technique use in drugs discovery and design. The next section describes the different computational technique such as genetic algorithms, genetic programming, evolution strategies, evolutionary programming, fuzzy logic, artificial neural network, PSO and ACO. Section 2 provides a comparative analysis of computational technique in the tabular form. Section 3 provides the brief summary of computational tools/software use in drugs discovery and design.

2. COMPUTATIONAL TECHNIQUES: AN OVERVIEW
Bacterial and viral infections like polio, small pox, tuberculosis, AIDS and related diseases that were once life threatening, now have become minor public health concerns only because of new computational techniques that help in discovery and designing of new potent drugs. Drug therapy has changed the fabric of society by improving both individual quality of life and life expectancy. Computational techniques shorten the discovery timeline and rationalize the design. Computational techniques play remarkable role in the drug design process. These techniques help in the designing of novel, innovative therapeutic agents that are both safe and effective. Computational techniques form the core of structure based drugs design. With the help of data management software, high performance computing and internet, it is possible to transform massive complex biological data into a

workable knowledge that can be used in modern drug discovery process. These techniques increase the possibility of success in the drug discovery process like from the identification of targets and elucidation of their functions, discovery & development of lead compound with desired pharmacological activity.

2.1. Genetic Algorithm

Genetic algorithm is a searching technique that is used to find approximate solutions for optimization and search problems. The genetic algorithm is proposed by John Holland in the 1960s [4, 5]. The genetic algorithms are the sub class of evolutionary computations. They are stochastic optimization methods and provide a powerful means to perform directed random searches in a large problem space as encountered in chemo metrics and drug design [4]. A genetic algorithm needs two things to be defined i.e. genetic representation of solutions and fitness function. To solve any problem, the first requirement is to draw the genetic representation. After the genetic representation the second requirement is to define a fitness function for the problem. The different problems have different type of fitness function. The Genetic Algorithm starts with the initialization of the population of solutions randomly and several individual solutions are randomly generated to form an initial population. The size of population relay on the nature of problem. It may consist hundreds or thousands of individual solutions. In the next, the fitness function is evaluated for each individual population. The last step is the reproduction of population. In this step the genetic operator such as selection, crossover and mutation are applied to generate the next generation of population. In drugs designing, a molecule is defined as input to GA and a binary string is used to code the molecule. A large number of the solution is generated by using genetic operator. The best population is selected and further used to generate the new population until a desired solution is reached.

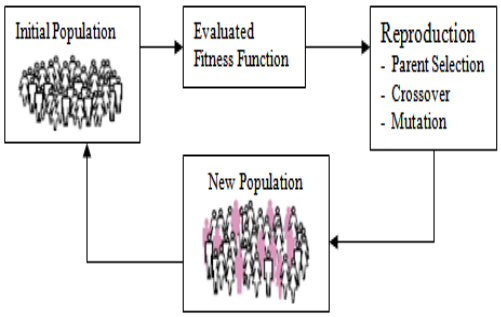

Figure 2: working of genetic algorithm

2.2. Genetic Programming

GP is the variant of the genetic algorithm. The GP is proposed by the John Koza [6, 7]. John Koza successfully applied genetic algorithms on LISP to solve a wide range of problems. Koza was defining the six steps to solving a problem using genetic programming.

1. Choosing the terminals.
2. Functions
3. Fitness function
4. Control parameters
5. Termination criterion
6. Determining the architecture i.e. program's automatically define functions (ADFs)

GP implements the Darwin theory of evolution as a computer program. These programs are written in the programming language. So the GP follows are the rule of the programming language and represent the solution in the form of parse tree that can be shown in the following figure. The GP is similar to the genetic algorithm but the possible solution can be coded as computer program rather than the binary strings. The mutation can be described as following

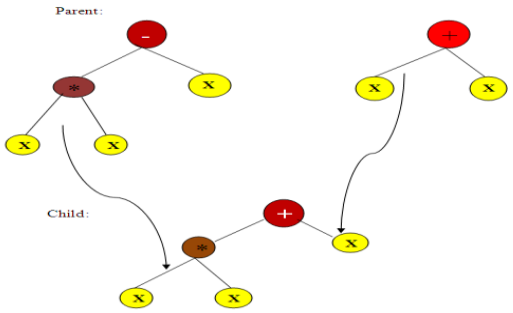

Figure3 shows the evaluation of new population from previous population

2.3. Evolutionary Strategy

ES was developed by Rechenberg and Schwefel [8, 9]. Basically ESs was designed for parameter optimization problems. During the developmental period of ES, a series of strategies were explored which can be listed as follows [10]:

- (┌, ┌) random walk
- (┌ + ┌) one parent → one child - select from all
- (μ+ ┌) multi-parent → one child - select from all
- (μ+ λ) multi-parent → multi-child - select from all
- (μ , λ) multi-parent → multi-child - select from children

The signification of the strategy can be defined as:

- ➢ ( ) → Represent the two successive generations.
- ➢ , → Represent the selection method from only one child.
- ➢ + → Represent the selection method from the pool i.e. parent and child.
- ➢ ┌ → Represent the only one individual.
- ➢ μ → Represent the populations of parent.
- ➢ λ → Represent the populations of children.

The above discussed strategies are used to define the successive generations of individuals. An individual is encoded through the real numbers. Each individual consist the strategy parameters. Each of these object parameters can be mutated individually by an amount based on the associated standard deviation recorded in σ. The strategy is further enhanced by the addition of rotation angles α which orient the direction of most extreme mutation based on the variances and covariance of the expected sets of mutations [5]. Mutation is the primary operator in the ES and works upon strategy parameters as well as object variables.

2.4. Evolutionary Programming

EP is a sub branch of EA. Fogel was proposed evolutionary programming as a means to develop artificial intelligence and argued that intelligent behavior requires both the ability to predict changes in an environment, and a translation of the predictions into actions appropriate for reaching a goal [11, 12]. In EP, a finite state machine is used to represent the individuals. A finite state machine processes the state with the input symbol and produced the new state and output symbol. Finite state machine consists of four tuple. These tuple can be defined as

- Q → Set of finite states
- ∑ → Set of input symbol
- Δ → Set of output symbol
- Δ→ Transition function

Transition function plays the important role in the finite state machine. It performed the mapping of the current state with an input symbol and produced the next state that can be act as next generation. The fitness of an individual is calculated by presenting sequentially to the finite state machine the symbols in the environment and observing the predicted output [13].

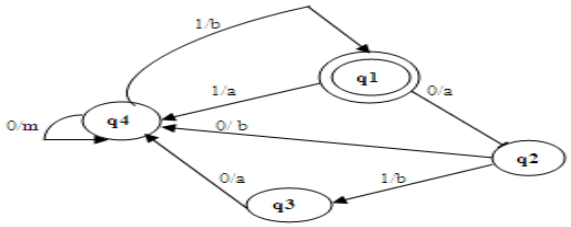

Figure:4 A Finite state machine with the state (q1, q2, q3, q4).

The input symbol belongs to the set (0, 1) and the output symbol belongs to the set (a, b, m). The labeled edge represents the transition from one state to another state. The double circle state represents the initial state.

2.5. Fuzzy Logic

Fuzzy logic is the science of reasoning, thinking and inference that recognizes and uses the real world phenomenon that everything is a matter of degree [14]. The original idea of fuzzy logic comes from a paper published by the Zadeh [15]. Fuzzy set is differing from traditional set theory i.e. fuzzy set has unsharp boundaries. So the traditional set theory has either value 0 or 1 but in fuzzy set the value is lie in between $0 \leq \mu \geq 1$ where μ is the membership function. Most important characteristic of fuzzy logic is fuzzy inference. Fuzzy inference systems based on fuzzy set theory are considered suitable for dealing with many real world problems, characterized by complexities, uncertainties, and a lack of knowledge of the governing physical laws [16]. The most important application of fuzzy set theory is the fuzzy rule-based models, where the relationships among system variables are modeled using linguistically interpretable rules [17]. Fuzzy logic can be especially useful in describing target properties for optimizations [18]. For example, the formulator might be seeking a tablet disintegration time of 200 s, i.e. any value less than 200 s has a desirability of 1 (i.e. 100%). But a tablet which disintegrates in 210 s is not entirely undesirable (as crisp logic would insist), and instead might be assigned a desirability value of 0.9. Fuzzy logic also used in process control with help of fuzzy if then rule. The figure 5 shows the membership function for temperature i.e. low, medium and high.

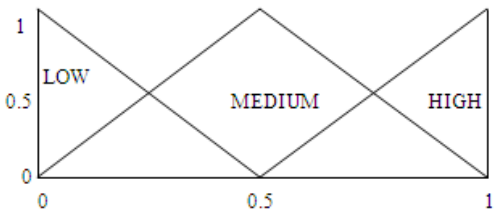

Figure 5: Fuzzy logic representation of temperature

The basic steps of the fuzzy set in the process modeling described as

- Arrange the input and output dataset.
- Clustering the output set
- Map the fuzzy inputs to the output
- Identify the significant variables
- Use the rule base in inference

The figure 6 shows the difference between crisp set and fuzzy set use in pharmacodynamic modeling [14].

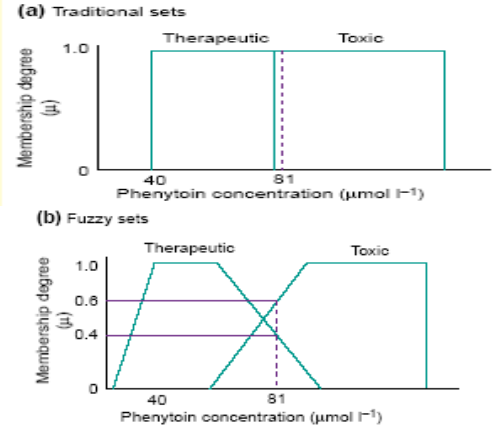

Figure 6: shows Difference between the crisp set and fuzzy set.

2.6. Artificial Neural Network

Neural network can be defined as networks of neuron. The biological inspiration of neural network comes from the nervous system of human being especially the human brain. Basically the neural network consist the two elements:-

- Processing Elements
- Connection Weight

Now, the term artificial neural network can be defined as a computer science paradigm that makes use of an abstract modeling of the neuronal structure of the brain as a tool for pattern recognition [19, 20, and 21]. Another way to describe the artificial neural network is a mathematical model to design for processing the information and knowledge acquisition such as human brain. The mathematical model of ANN consist large number of artificial neuron that are used to take the input signals and generate the appropriate output signal. The figure 7 showed a systematic architecture of two layers ANN.

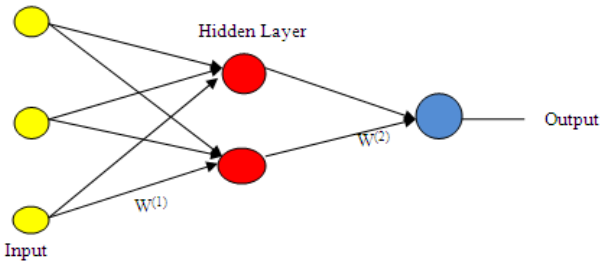

Figure 7: architecture of two layers ANN

Neural networks can be applied to four basic types of applications [4]:

• Association;
• Classification (clustering);
• Transformation (different representation);
• Modeling.

The ANN has large number of application that is used in pharmacy and life sciences. The table 1 provides a summarized way of ANN application [22].

Table 1: Summary of ANN applications:

Network Type/ Architecture	Main Applications
Supervised • Multilayer feed-forward (BP) • Recurrent networks • Encoder networks (ReNDeR)	➢ non-linear modeling of QSAR, prediction of molecule activity and structure, pattern recognition, classification, signal filtering, noise reduction, feature extraction ➢ sequence and time series analysis ➢ data compression, factor analysis, feature extraction Learning vector quantization auto-associative recall, data compression
Unsupervised • Kohonen self-organizing map • Hopfield networks • Bidirectional associative memory (BAM) • Adaptive resonance theory (ART) models	➢ clustering, data compression, visualization ➢ auto-associative recall, optimization ➢ pattern storage and recall (hetero-association) ➢ clustering, pattern recognition
Hybrid • Counter propagation networks • Radial basis function (RBF) networks • Adaptive resonance theory (ART) models	➢ function approximation, prediction, pattern recognition ➢ function approximation, prediction, clustering ➢ similar to ART and BP-networks

2.7. Particle Swarm Optimization

PSO is population based stochastic optimization method inspired by observation of swarms of insect, shoals of fish, bird flocking etc [23]. There are the million of insect and living creature on the earth as well as in the sea. Each creature posses the unique characteristic that characteristic makes the difference between these. The creature/insect has its own way to find food, save itself to the attack and survival for life. This collective and social behavior of living creatures motivated researchers to undertake the study of the insect/creature. Creatures such as fish schools and bird flocks clearly display structural order, with the behavior of the organisms so integrated that even though they may change shape and direction, they appear to move as a single coherent entity [24]. A swarm can be viewed as a group of agents cooperating to achieve some purposeful behavior and achieve some goal [25]. In principle, PSO is a multi agent parallel search technique. The particles are entities which fly in the multi-dimensional search space. Each particle has a position and a velocity in the multi dimensional search space that can be represented as P_i and Vi. The position a particle in the search space represents a trial solution of the problem. It considers that a particle/swarm moves over the solution space, and particles are evaluated according to some fitness criterion. The movement of each particle depends on two points:-

- Particle best position since the algorithm started (pBest).
- The best position of the particles around it (lBest) or of the whole group(gBest)

In each iteration , the particle changes its velocity towards pBest and lBest/ gBest. So the swarm explores the solution space looking for promising zones [26]. PSO is applied with success to difficult problems, such as feature selection for gene expression data [27, 28] identification of the global minimum geometry of chemical compounds [29], enzyme-inhibitor docking [30], QSAR [31], and protein motif discovery [32].

2.8. Ant Colony Optimization

Ant Colony Optimization (ACO) is a meta-heuristic algorithm inspired by the foraging behaviors of ants and developed by Marco Dorigo (Milan, Italy), and others in early 1990s. Basically the algorithm is based on a series of random decisions (by artificial ants) and probability of decisions changes on each iteration.

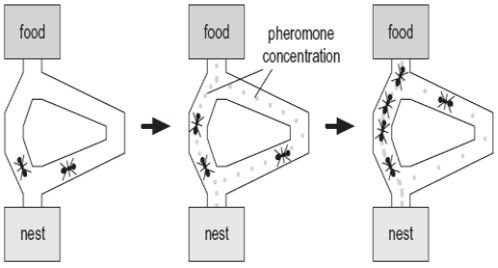

Figure8: shows the ant colony and behavior of ant

An ant's decision to perform a task depends on the physical state of the environment and the social interactions with other ants. The behavior of artificial ants is inspired from real ants: they lay pheromone trails (obviously in a mathematical form) on the graph edges and choose their path with respect to probabilities that depend on pheromone trails [26]. Basically the ant algorithm are use to solve the discrete search space problem such as TSP, Routing, Job shop scheduling. But

ACO is use in protein folding, protein ligand docking and conformational analysis of flexible molecules [33].
Tabular analysis of computational technique is performed in annexure-1 at the end of paper.

3. DRUGS DESIGN SOFTWARE & TOOL

The increasingly use of computational tools has brought the new revolution in drug discovery and design process. Computational approaches are designing the new drug faster and more economical as compare to traditional method. Due to the rapid advancement in hardware and software, large number of drugs discovery & design software and computational tools are developed. So, the use of computational tool/software in the drugs design and discovery, computers become an important part of drug design & discovery process and achieves their goals quickly & more efficiently. Software tool/packages are vary from modeling programs to virtual reality, explore more structural options and predictions of properties of new compounds. Computational tools are used in the target discovery, hit identification, hit-to-lead, and lead optimization phases of a drug discovery process. Here is a list of computational tools that are used in drug discovery and design process. A detailed list of various software with utilities, i/p format and required platform is provide in annexure-2 at the end of paper.

4. CONCLUSION

This paper deals with the scope of nature inspired optimization technique in drugs designing and discovery. The main objective of paper to describe the various computational techniques that are used in drugs discovery & designing. How these techniques applied in drugs discovery and what are the pros & cons of particular techniques in drug discovery & designing. Table 1 provides parametric comparison of computational techniques that are used in this paper. These techniques are compared on certain parameter such as methodology, fitness function, inspiration and application in pharmacy etc. This paper also deals with the list of various software tools that are used in Drugs Discovery & Designing. Table 2 provides a brief description of software tools that are used in pharmacy field and also provide information on utilities of these tools & input /output format of these tools.

REFERENCES

[1] DiMasi, J.A. et al., “The price of innovation: new estimates of drug development costs”, J. Health Economics 20023,22, 151–185.

[2] Drug Discovery and Development: Understand the R&D Process from innovative.org.

[3] Jorgensen W.L., “The many roles of computation in drug discovery”, Science 303, 2004, 1813–1818.

[4] Lothar Terfloth and Johann Gasteiger,” Neural networks and genetic algorithms in drug design”, Journal of Drugs discovery Today, 2001, Vol. 6, No. 12 (Suppl.) pp no. 102-108.

[5] Forschungsbericht erstellt von and Peter von Bülow,” A Survey and Comparison of Biological Genetics and Evolutionary Computation”, june 2007 (Thesis).

[6] Koza, JR., “Genetic programming: a paradigm for genetically breeding populations of computer programs to solve problems”, Technical Report STANCS-90-1314, Department of Computer Science, Stanford University, June 1990.

[7] Koza JR. “Genetic Programming”, Cambridge, MA: MIT Press, 1999.

[8] Rechenberg I., ‘Evolutions strategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution’, 1973(Phd Thesis).

[9] Schwefel H.P., ‘Numerical Optimization of Computer Models’, Wiley, Chichester, 1981.

[10] Back, Hoffmeister and Schwefel. "A Survey of Evolution Strategies", in Proceedings of the Fourth International Conference on Genetic Algorithms. Morgan Kaufmann Publ., San Mateo, California, 1992, pp. no. 2-9.

[11] Fogel L.J., Owens A.J., and Walsh M.J., ‘Artificial Intelligence through Simulated Evolution’, Wiley, New York, 1966 (Book).

[12] Fogel L. J., “Autonomous automata” Ind Res vol. 4, 1962, pp. no.14–19.

[13] Carlos Andre´s, Pena Reyes, Moshe Sipper,” Evolutionary computation in medicine: an overview”, Artificial Intelligence in Medicine 19 (2000), pp. no. 1–23.

[14] Beth A. Sproule, Claudio A. Naranjo and I. Burhan Türksen,” Fuzzy pharmacology: theory and applications”, TRENDS in Pharmacological Sciences Vol.23 No.9, 2002, pp. no. 412-417.

[15] Zadeh, L.A. "Fuzzy sets", Information and Control 8 (3), 1965, pp. no. 338–353.

[16] Zadeh L.A.,” Outline of a new approach to the analysis of complex systems and decision processes” IEEE Trans. Syst. Man Cybern. 3, 1973 , pp. no. 28 -44.

[17] Mohit Kumar, Kerstin Thurow, Norbert Stoll , Regina Stoll,” Robust fuzzy mappings for QSAR studies”, European Journal of Medicinal Chemistry published by Elsevier 42, 2007, 675-685.

[18] SVETLANA IBRIĆ, ZORICA DJURIĆ , JELENA PAROJČIĆ , JELENA PETROVIĆ ,” ARTIFICIAL INTELLIGENCE IN PHARMACEUTICAL PRODUCT FORMULATION: NEURAL COMPUTING”, Chemical Industry & Chemical Engineering Quarterly 15 (4), 2009, pp. no. 227−236.

[19] McCulloch W.S. and Pitts W., “A logical calculus of the ideas immanent in nervous activity”, Bull. Math Biophys (5), 1943, pp. no. 115-133.

[20] Rosenblatt, F. Principles of neurodynamics: perceptrons and thetheory of brain mechanisms, Washington, DC: Spartan 1962(Books).

[21] Haykin, S. Neural networks: A nomprehensive foundation, Upper Saddle River, NJ: Prentice Hall, 1998.

[22] Gisbert Schneider and Paul Wrede, “Artificial neural networks for computer-based molecular design”, Progress in Biophysics & Molecular Biology 70 (1998) 175- 222.

[23] Eberhart and Kennedy,” Particle swarm optimization”, Proc. IEEE international conf. on neural networks – IV, 1995, pp.no. 1942 - 1948.

[24] Couzin ID, Krause J, James R, Ruxton GD, Franks NR, "Collective Memory and Spatial Sorting in Animal Groups", Journal of Theoretical Biology, 2002, 218, pp. 1–11.

[25] Grosan C, Abraham A and Monica C, "Swarm Intelligence in Data Mining", SCI vol. 34, 2006, Springer, pp. 1–16.

[26] Swagatam Das, Ajith Abraham*, and Amit Konar," Swarm Intelligence Algorithms in Bioinformatics", Swarm Intelligence Algorithms in Bioinformatics, Studies in Computational Intelligence (SCI) 94, 2008, 113–147.

[27] Chuang LY, Chang HW, Tu CJ, Yang CH,"Improved bi nary PSO for feature selection using gene expression data", Comput Biol Chem32, 2008, 29–38.

[28] [28] Shen Q, ShiWM, Kong W, Ye BX,"A combination of modified particle swarm optimization algorithm and support vector machine for gene selection and tumor classification", Talanta 71, 2007, 1679–1683.

[29] Call ST, Zubarev DY, Boldyrev AI,"Global minimum structure searches via particle swarm optimization", J Comput Chem 28, 2007, 1177–1186.

[30] Namasivayam V, Günther R,"PSO@AUTODOCK: A fast flexible molecular docking program based on swarm intelligence", Chem. Biol. Drug Des 70, 2007, 475–484.

[31] Agrafiotis DK, Cedeño W (2002),"Feature selection for structure-activity correlation using binary particle swarms", J Med Chem.45, 2002, 1098–1107.

[32] Chang BCH, Ratnaweera A, Halgamuge SK,Watson HC, "Particle swarm optimisation for protein motif discovery", Genetic Programming Evolv. 5, 2004, 203–214.

[33] Marc De Jonge, Luc Koymans, Maarten Vinkers, "An Ant Algorithm for the Conformational Analysis of Flexible
Molecules", Journal of Computational Chemistry vol. 28 issue 5, 2007, pp. no. 890 - 898.

[34] Zhiwei Wang, Gregory L. Durst and Russell C. Eberhart," Particle Swarm Optimization and Neural Network Application for QSAR", 18th International Parallel and Distributed Processing Symposium (IPDPS'04) - Workshop 9, vol 10, 2004, pp 195.

[35] Hoffman, B.T. et al.,"2D QSAR modeling and preliminary database searching for dopamine transporter inhibitors using genetic algorithm variable selection of Molconn Z descriptors", J. Med. Chem. 43, 2000, pp. 4151–4159.

[36] Turner, D.B. and Willett, P, "Evaluation of the EVA descriptor for QSAR studies: The use of a genetic algorithm to search for models with enhanced predictive properties", J. Computational Aided Mol. Design 14, 2000, pp. 1–21.

[37] Kimura T. et al. (1998) ,"GA strategy for variable selection in QSAR studies: GA-based region selection for CoMFA modeling", J. Chem. Inf. Comput. Sci. 38, 1998, 276–282.

[38] Hasegawa, K. et al.,"GA strategy for variable selection in QSAR studies: application of GA-based region selection to a 3D- QSAR study of acetylcholinesterase inhibitors" J. Chem. Inf. Comput. Sci. 39, 1999, 112–120.

[39] James Cunha Werner and Terence C. Fogarty," Genetic programming applied to pharmaceutical drugs design", The Seventh ACM SIGKDD International Conference on Knowledge discovery and data mining, USA/ San Francisco, August 26-29,2001.

[40] Schneider G., Schuchhardt J., and Wrede P., Comput., "Artificial Neural network and simulated molecular evolution are potential tool for sequence oriented protein design", Appl. Biosci. vol. 6, 1994, pp. 635 645.

[41] Fogel D. B., 'Evolutionary Computation: Toward a New Philosophy of Machine Intelligence', IEEE Press, Piscataway, 2006 (Book).

[42] Luke B. T.,"Evolutionary programming applied to the devlopement of QSAR and QSPR", J. Chem. Inf. Comput. Sci., 34, 1994, 1279 1287.

[43] Gehlhaar D. K., Verkhivker G.M., Rejto P. A., Sherman C.J., Fogel D.B., Fogel L.G., and Freer S.T., "Molecular recoginition of the inhabibitor AG:1343 by HIV Protease: Conformationallhy flexible docking by evolutionary Programming", Chem. Biol. Vol 2, 1995, pp. 317 324.

[44] Huuskonen, J. et al.,"Aqueous solubility prediction of drugs based on molecular topology and neural network modeling", J. Chem. Inf. Comput. Sci. 38, 1998, 450–456.

[45] Schneider, G. et al.,"Combining in vitro and in vivo pharmacokinetic data for prediction of hepatic drug clearance in humans by artificial neural networks and multivariate statistical techniques", J.Med. Chem. 42, 1999, 5072–5076

[46] Jalali-Heravi M. and Parastar, F. (2000),"Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives", J. Chem. Inf. Comput. Sci. 40, 2000, 147–154.

[47] Burden, F.R. and Winkler, D.A. (1999), "Robust QSAR models using Bayesian regularized neural networks", J.Med. Chem. 42, 1999, 3183–3187

[48] Burden, F.R. et al., "Use of automatic relevance determination in QSAR studies using Bayesian neural networks", J. Chem. Inf. Comput. Sci. 40, 2000, 1423–1430.

[49] Angela Torres and Juan J. Nieto," Fuzzy Logic in Medicine and Bioinformatics", journal of Biomedicine and Biotechnology, 2006, Pages 1–7.

[50] Papageorgiou EI, Stylios CD, Groumpos PP. "An integrated two-level hierarchical system for decision making in radiation therapy based on fuzzy cognitive maps", IEEE Transactions on Biomedical Engineering 50(12), 2003, pp. 1326–1339.

[51] Oshita S, Nakakimura K, Sakabe T.,"Hypertension control during anesthesia - Fuzzy logic regulation of nicardipine infusion", IEEE Engineering in Medicine and Biology Magazine 13(5), 1994, 667–670.

[52] Johnson M, Firoozbakhsh K, Moniem M, Jamshidi M. ,"Determining flexor-tendon repair techniques via soft computing", IEEE Engineering in Medicine and Biology Magazine 20(6), 2001, pp. 176–183.

[53] Hassanien AE. ,"Intelligent data analysis of breast cancer based on rough set theory", International Journal on Artificial Intelligence Tools. 12(4), 2003, pp. 465–479.

[54] Seker H, OdetayoMO, Petrovic D, Naguib RN.,"A fuzzy logic based-method for prognostic decision making in breast and Prostate cancers", IEEE Transactions on Information Technology in Biomedicine 7(2) , 2003, 114–122.

[55] Schneider J, Peltri G, Bitterlich N, et al.,"Fuzzy logic-based tumor marker profiles including a new marker tumor M2-PK Improved sensitivity to the detection of progression in lung cancer patients", Anticancer Research 23(2A,) 2003, 899–906.

Annexure-1: Tabular Analysis of Computational Techniques

Sl. No.	Parameter	Genetic Algorithm	Genetic Programming	Evolution Strategies	Evolutionary Programming
1	Inspiration	Theory of Evolution	Darwin Theory of Evolution	Darwin Theory of Evolution	Darwin Theory of Evolution
2	Technique	stochastic optimization methods	stochastic optimization methods	Parameter optimization problems	variable optimization
3	Developer(s)/ Proposer(s)	Johan Holland	John Koza	Rechenberg and Schwefel	L. J. Fogel
4	Year	1975	1992	1973	1966
5	Methodology	• Set of solution Called population • Defined a fitness function • Select a population • Produce another population • Compare with Criterion Function • Stopping condition.	• Generate initial population for individuals. • Evaluate fitness function of population. • Compare with criteria function • Generate new fitness function using o Selection o Recombination o Mutation • Find Best individual.	• Each individual has a list of real numbers, are called the object variables of the problem. • Each individual contains a number of strategy parameters. • Strategy parameters are used to control the behavior of the mutation operator. • In each iteration, offspring are generated from a population. • Recombination operator produces one child and requires two parents for each object variable and strategy parameter in the child. • Parents are selected randomly from the current Population.	• A current population of m individuals is randomly initialized. • Fitness scores are assigned to each of the m individuals. • The mutation operator is applied to each of the m individuals in the current population to produce m offspring. • Fitness scores are assigned to the m offspring. • A new population of size m is created from the m parents and the m offspring using tournament selection. • If the termination conditions are satisfied exit, otherwise go to step 3.
6	Fitness function	• Fitness function varies problem to problem. For example, in multi-dimensional opti-mization the objective function is a weighted sum of the desirability of each of the properties. $F = \frac{\sum w_i f_i}{P \sum w_i}$,	• Fitness function :- = OK/OK+ contI*NOK/cont_A/contA+ NOK_I Where OK – right activity prediction NOK_A- is the wrong prediction of active compound. NOK_I- wrong prediction of inactive compound. ContI -is the total number of inactive compounds. ContA - is the total number of active compounds	• Each individual is coded in the form of a vector a , which has three component parts: $\vec{a} = (\vec{x}, \vec{\sigma}, \vec{\alpha})$ where; $\tan(2\alpha_{ij}) = \frac{2c_{ij}}{\sigma_i^2 - \sigma_j^2}$	• Payoff function, which measures the accuracy of the prediction.

Sl. No.	Parameter	Genetic Algorithm	Genetic Programming	Evolution Strategies	Evolutionary Programming
7	Application in Pharmacy	• Automated genera - - tion of small organic molecules, • Enable automated docking. • Assemble Drug-derived building blocks • Enables the flexible alignment of multiple molecules • Protein structure prediction	• In Cheminformatics and QSAR. • In Bioinformatics	……	• Design of molecule libraries. • In conformational analysis. • In molecule superposition and pharmacophore detection. • To finding quantitative structure-activity relation.
8	Advantage	• Improved the predictive value of a QSAR model by variable selection- • In comparative molecular field analysis. • Operate on fixed length strings which contain binary values.	• Combines the flexible problem representation with a powerful search mechanism. • Requires minimum assumptions. • Predict some other property of chemical by given known properties.	• In the field of parametric optimization. • Operate on fixed length strings which contain real values. • ES use a deterministic selection.	• Best suited to parametric optimization. • Operate on fixed length strings which contain real values.
9	Disadvantage	• If the population size is too small, the genetic algorithm may not explore enough of the solution space to consistently find good solutions. • Rate of genetic change is too high or the selection scheme is chosen poorly, beneficial schema may be disrupted. • Difficulty to dealing the problems with deceptive fitness functions. • Premature convergence	• No guarantee to find the suitable solution. • Large amount of Computer power required for large genetic program. • Reduce the genetic diversity contained in larger trees. • It is time consuming, and its application is less well understood in the formulation domain.	……..	• Rare in computational chemistry. • Evolution is wholly dependent on the mutation operator.
10	References	[4, 18, 34, 35, 36, 37, 38]	[6, 7, 13, 18, 39]	[8, 9, 11, 12, 40]	[11, 41, 42, 43]

Table 2: Parametric comparison of different computational techniques

Sl. No.	Parameter	ACO	Artificial Neural Network	Fuzzy Logic	Particle Swarm Optimization
1	Inspiration	Inspiration of real world Ant behavior.	• Characteristics of the biological neurons in the human brain and nervous system	Based on A is not -A	Behavior of Swarm in the nature such as bird, fish
2	Technique	population-based stochastic search method	……..	----------	Stochastic optimization Method
3	Developer(s)/ Proposer(s)	M Dorigo	Alexander Bain and William James 1873 and 1890	Lotfi Zadeh, 1965	Dr. Eberhert and Dr. Kennedy 1995
4	Methodology	……..	• Artificial neural network creates a model of neurons. • Provide connections	• Fuzzification of crisp input. • Rule evaluation on fuzzified input. • Aggregation of rule	• Initialize each particle • Calculate fitness function for each particle. • Compare the fitness value with another calculated

Sl. No.	Parameter	ACO	Artificial Neural Network	Fuzzy Logic	Particle Swarm Optimization
			between neurons. • Trains the model to associate output neurons with input neurons. • Generate relevant output for a set of input data.	output i.e. unification. • Defuzzification of output i.e. obtained crisp value from fuzzy output.	fitness value. The best value set as p Best. • Choose the particle with best fitness value to provide g best • Calculate particle velocity Update particle position
5	Fitness Function	• Probabilistic transition rule: $p_{ij}(t) = \frac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_k [\tau_{ik}(t)]^{\alpha}[\eta_{ik}]^{\beta}}$	• The o/p of neuron can be represented as: $z = \sum_{i=1}^{n} w_i x_i$ Thrre type of activation funcation use in NN. • Threshold Function • Linear function • sigmoid function	• Membership Function: $\mu_A(x) : X \rightarrow \{0, 1\}$ where $\mu_A(x) = 1$ if x is totally in A; $\mu_A(x) = 0$ if x is not in A; $0 < \mu_A(x) < 1$ if x is partly in A.	……
6	Application in Pharmacy	• Feature selection for QSAR models. • Identify optimum parameters for QSAR models. • In protein-legends docking. • To identify the best regression. • 2D-HP protein folding	• Classification and modeling of chemotherapeutic agents, anti-bacterials, anti-neoplastics and anti-fungal. • Prediction of hepatic drug clearance. • QSAR study of HIV-1 reverse transcriptase • lead discovery and analysis of multi-dimensional data • in the field of genomics is gene prediction	• Medicine (Drug Addiction) & Bio-Informatics. • Mechanical control of drug delivery devices. • Pharmacokinetic modeling. • Effectively putting clinical practice guidelines into Operation. • Classifying potential risk factors for strokes.	• To select a few (typically 3-7) features as inputs to QSAR. • Refine the cluster in Bio informatics.
7	Advantage	• Discover classification rules. • Provides good solutions if the simulations use a sufficient number of ants to evaluate all features in different combinations	• Accurately predicts results when the response variables are highly non-linear. • More accommodating to sparse and noisy data than statistical modeling packages. • Fast and can lead to saving in both time and cost of product development.	• To improve decision-making in radiation therapy. • To control hypertension during anesthesia. • To determine flexor-tendon repair techniques. • To detect breast cancer, lung cancer, or prostate cancer. • To assist the diagnosis of central nervous systems tumors(astrocytictumors).	• Near-optimal solutions could be found much faster than by using a random search. • Perform feature selection efficiently in data sets with large numbers of feature. • Solving a wide range of different applications without expensive human up front design.
`8	Disadvantage	• Probability distribution change by iteration. • Retains memory of entire colony instead of last one.	• Relationship cannot be expressed easily in mathematical form. • Primary risk in developing a model is overtraining.	• Fuzzy logic modeling is not appropriate in every situation. • Less able to deal with the imprecision associated with large amounts of missing	• Difficulty in choosing parameters for the back propagation. • Do not have a guarantee of success

Sl. No.	Parameter	ACO	Artificial Neural Network	Fuzzy Logic	Particle Swarm Optimization
				data.	
9	References	[3 ,33]	[19, 20, 21, 22, 44, 45, 46, 47, 48]	[1],[2];[3].[4],[5],[6],[7], [32],[39],[40],[41], [42]	[14, 15, 16, 17, 18, 49, 50, 51, 52, 53, 54, 55]

Annexure-2: List of Various Software that are used in Drugs Discovery & Designing.

S. no	Software	Company/ Institution	Utilities and URL	I/P format	Platform
1	InsightII, Discovery studio Cerius	Accelrys	• Molecular modeling and de novo drug design. http://www.accelrys.com/products/insight/ • Computational models for the prediction of ADME properties derived from chemical structures.. http://www.accelrys.com/products/cerius2/	.msi,. xtl, .car, .mdf, .dat, .fdat, .mol, .mdl , .pdb	IRIX, UNIX, LINUX, WINDOWS
2	Sybyl	Tripos	• Computational informatics software for drug discovery. http://www.tripos.com/	PDB, mol, mol2	Windows/Linux
3	Molecular Operating Environment (MOE)	Chemical Computing Group	• Bioinformatics, cheminformatics, protein modeling and structure-based design. • High throughput lead discovery and molecular modeling and simulations. http://www.chemcomp.com/	SD	Windows 2000/XP/Vista/ Windows 7, Linux, Mac OS X , Sun Solaris 10, Silicon Graphics Irix 6.5 , Unix
4	Glide, Prime, Maestro	Schrödinger Inc.	• Provides a complete suite of software's that addresses the challenges in pharmaceutical research http://www.schrodinger.com/	.PDB, .IN,INP	Linux, Windows or Mac
5	Bio-Suite	Tata Consultancy Services Ltd	• Genomics, protein modeling and structural analysis, simulation and drug Design. http://www.atc.tcs.co.in/biosuite/	…….	……
6	Sanjeevini	Indian Institute of Technology, New Delhi	• Active site directed drug design http://www.scfbio-itd.res.in/research/drugdesign.htm	PDB	……..
7	SQUAD	……..	• Constructs dynamic models of signaling networks, user-friendly graphical interface http://www.enfin.org/dokuwiki/doku.php?id=squad:start	XML, Text file SBML	Windows, Linux

8	CellNetAnalyzer	Developed by Steffen Klamt and Axel von Kamp at the Max Planck Institute for Dynamics of Complex Technical Systems.	• Analyze signaling, regulatory and metabolic networks. http://www.mpimagdeburg. mpg.de/projects/cna/cna.html	ASCII	Linux (32-bit), Windows XP (32-bit), Intel Mac
9	BioTapestry	Devloped at the Institute of Systems Biology in Seattle, collaboration with the Davidson Lab	• Analysis and modeling of large biological networks	CSV Tabular	Linux, Windows and Mac
10	SBMLSAT	Devloped by Zhike Zi, Yanan Zheng, Ann E. Rundell, Edda Klipp at Max Planck Institute for Molecular Genetics, Berlin, Germany.	• Computational modeling for systems Biology. • Sensitivity Analysis Tool. • Multi-Parametric Sensitivity Analysis. • Local sensitivity analysis. http://sysbio.molgen.mpg.de/ SBMLSAT/	SBML	Windows, Mac and Linux
11	Cytoscape	Cytoscape is a collaborative project between the Institute for Systems Biology, the University of California San Diego, Memorial Sloan-Kettering Cancer Center, the Institute Pasteur, Agilent Technologies and the University of California, San Francisco.	• Visualization and analysis of biological networks, data integration, allows export of network structures as images. http://www.cytoscape.org/	SIF, GML, XGMML, BioPAX, PSI-MI, SBML, OBO	Linux, Windows, and Mac OS
12	CellProfiler	Currently developed and maintained by the Carpenter Lab at the Imaging Platform of the Broad Institute.	• It has two components CellProfiler and CellProfiler Analyst • CellProfiler processed inage. • CellProfile Analyst is processed data produced by CellProfiler. http://www.cellprofiler.org/	DIB, .cp, .cpa	Macintosh, Windows (32-, 64-bit), Linux
13	FlexiDock	Developed by the Theoretical Biophysics Group at the University of Illinois	• Uses genetic algorithm for generation of configurations. • Simple, flexible docking of ligands into binding sites on proteins. http://www.tripos.com/software/ fdock.html	PDB	Red hat, Linux
14	DockVision	Developed at the University of Alberta by Trevor Hart, Steven Ness and Randy Read.	• Including Monte Carlo, Genetic Algorithm and database screening docking algorithms http://dockvision.com	PDB	Irix 5.3 or higher Linux
15	AutoDock	Molecular Graphics Laboratory, The Scripps Research Institute	• Uses a Monte Carlo (MC) simulated annealing (SA) technique for configurationally exploration with a rapid energy evaluation • Provide an automated procedure for predicting the interaction of ligands	PDBQ and PDBQT	Darwin, IRIX64, Linux, Mac OS X, Solaris and Windows/Cygwin

			with bimolecular targets.		
16	GACS	MolMo Services, Retie, Belgium	• Genetic algorithm based program for conformational search of flexible molecules. • Consistently generates all low energy conformers. http://www.molmo.be/software.html	.car, .mdf	All Platform
17	ANTCONF	MolMo Services, Retie, Belgium	• Searches the torsional degrees of freedom of a flexible molecule. • Determine the molecular mechanics energy of the lowest energy conformer. • Used to rapidly generate superior 3D geometries. http://www.molmo.be/software.html	.car, .prm,.ext	All Platform