ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Target Prediction Using Evolutionary Computation.

Sujitha George*, and E Nagarajan.

*P.G scholar, Sathyabama University, Chennai, Tamil Nadu, India.
Assisstant Professor, Sathyabama University, Chennai, Tamil Nadu, India.

ABSTRACT:

In modern drug research predicting the target site is the crucial step.In this model a new approach was introduced for predicting the target site. Clustering the attributes in the dataset is the incipient step in this work. Particle swarm Optimization (PSO), a population based optimization technique was used to develop the rules. Using these rules we can predict the binding site of a compound.

Keywords: Particle Swarm Optimization, Agglometrive Clustering, Rational Drug Design, Genetic Algorithm

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Drug design is also known as rational drug design which is a creative process of finding a new administration for medicine based on the awareness of a biological target. Traditional methods of drug design depends on testing of chemical substructure by trial and error methods on animals .In contrast to this ,rational drug design depends on a hypothesis that distortion of a biological target may have a salutary value .The different steps that involved in rational drug design are:

- Design candidate compound
- Study the 3-D structure
- Test the design

Discovering the candidate compound is the crucial step in drug discovery. Analysing the 3-D i is used to determine the degree of interaction with the target. The target may be biological pathways, receptors, cells etc.The rational drug design is used in pharmaceutical companies to discover new drug compounds.

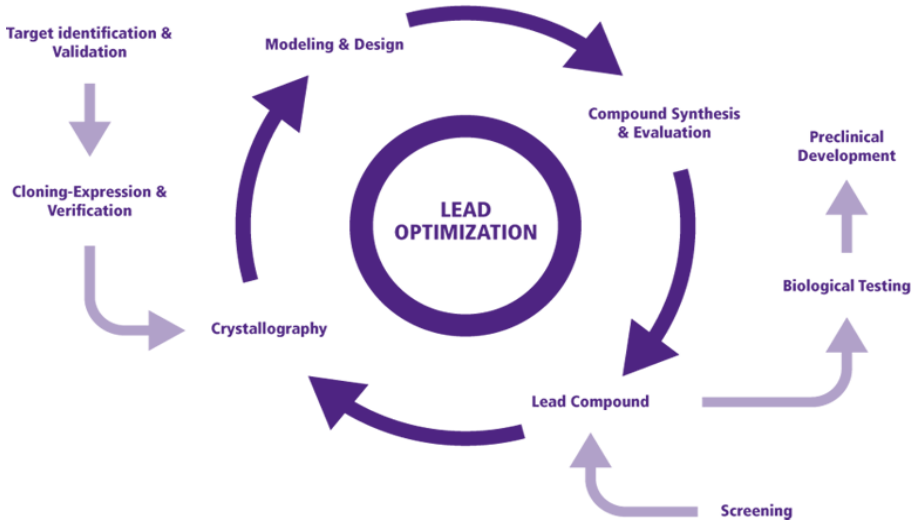

Figure 1.1 Steps in Rational Drug Design

Figure 1.1 shows the steps in rational drug design. Target identification and validation is the crucial step in drug design. In cloning and verification, analyse the cloning expression of the site. Crystallography is a technique used for the determination of the protein structure .From the analysed structure, develop the design and model of the compound. The next step is compound synthesis and evaluation. Here the compounds will synthesised and evaluate with the target. The outcome of this step will be a lead compound. In drug discovery a lead compound is a chemical compound which has a biological activity. This lead compound will undergo different screening and biological testing and then preclinical development.

LITERATURE SURVEY

Romanon T Kromer[1] proposed a method called in silico or virtual screening to search small molecule that bind the target site.The virtually screened compounds can stem from virtual compound libraries to place computer generated representations to target structure, docking program is used. Docking is helpful in reducing virtual compounds to manageable number of synthesized compounds. Charanpreet Kaur and Shweta Bhardwaj[2],proposed data mining technique in drug discovery. With the help of chemo informatics and data mining, speed of drug discovery can be accelerated. Clustering process divides database of unknown drugs based on similarity. Zhisond He and co-authors[3] suggested that proteins are encoded with biological featuers and drug compounds with functional groups. Here target proteins are divided into four different groups; nuclear receptors, ion channels, enzymes and G-protein coupled receptors. With the help of nearest neighbour algorithm, four predictors which are independent was established, to predict the intermediate employment between drug and one of the four protein group. D. Rognan[4],suggested that chemo genomics is

ISSN: 0975-8585

a new research field for studying systematically the biological effects of the ligand and macro molecular targets. With the help of this method ligand selection can be made effectively and thus novel targets can be addressed.

METHODOLOGY

The database of 503 compounds was use for analysis. Figure 3.1 shows a sample dataset Dataset consists of different attributes. Attributes includes compound name target, doses, activity data, IC50 values, EC50 values etc.

	CCLE Cell Line	Primary Cell Line Name	Compound	Target	Doses (uM	Activity Data (	Activity SD	Num Data	FitType	EC50 (uM)	IC50 (uM)	Amax	ActArea
2	1321N1_CENTF	1321N1	AEW541	IGF1R	.0025,.008	8.67,11.0,2.16	3.31,3.72,5.3	8	Sigmoid	8.717774	8	-42.558	0.7124
3	22RV1_PROST/	22Rv1	AEW541	IGF1R	.0025,.008	.94,12.5,-14,4	1.95,13.3,6.9	8	Sigmoid	8.165164	2.329924	-71.5893	1.6723
4	42MGBA_CENT	42-MG-BA	AEW541	IGF1R	.0025,.008	8.91,8.39,-3.5,	13.7,7.70,11.	8	Sigmoid	1.514508	2.68213	-63.4914	1.1852
5	5637_URINARY	5637	AEW541	IGF1R	.0025,.008	2.15,9.91,-3.5,	4.05,9.75,12.	8	Sigmoid	8.006595	5.002314	-62.3528	0.9948
6	639V_URINAR\	639-V	AEW541	IGF1R	.0025,.008	11.8,-7.3,-9.4,	.95,5.67,11.1	8	Sigmoid	0.931196	1.736181	-51.9598	1.5436
7	697_HAEMATC	697	AEW541	IGF1R	.0025,.008	18.4,4.65,8.64	25.4,1.92,15.	8	Sigmoid	8.700655	4.260822	-73.3338	1.7665
8	769P_KIDNEY	769-P	AEW541	IGF1R	.0025,.008	2.58,-4.1,6.18,	18.1,12.4,12.	8	Sigmoid	0.306243	8	-39.6881	1.4376
9	786O_KIDNEY	786-O	AEW541	IGF1R	.0025,.008	-1.2,-.53,.55,-	14.5,15.3,7.1	8	Sigmoid	5.267667	7.613148	-51.6863	0.5954
10	8305C_THYROI	8305C	AEW541	IGF1R	.0025,.008	1.00,-1.9,8.83,	6.49,3.67,4.6	8	Sigmoid	8.520771	4.950636	-66.2288	1.2929
11	8505C_THYROI	8505C	AEW541	IGF1R	.0025,.008	-6.0,-.21,-2.4,	.63,13.0,4.25	8	Sigmoid	1.84669	8	-22.2688	0.3189
12	8MGBA_CENTF	8-MG-BA	AEW541	IGF1R	.0025,.008	-6.0,-23,-17,9.	7.42,9.33,2.7	8	Sigmoid	8.212997	3.888341	-91.34	1.5025
13	A172_CENTRAI	A172	AEW541	IGF1R	.0025,.008	-.27,.084,-2.7,	5.84,1.59,2.7	8	Sigmoid	8.419621	8	-21.3668	0.3615
14	A204_SOFT_TI	A-204	AEW541	IGF1R	.0025,.008	3.12,2.35,2.69	5.56,3.35,2.2	8	Sigmoid	4.163712	5.150312	-67.1933	0.5525
15	A2058_SKIN	A2058	AEW541	IGF1R	.0025,.008	4.97,-2.9,1.30,	3.91,.85,4.79	8	Sigmoid	8.731297	5.513091	-92.3689	1.0006
16	A253_SALIVAR	A-253	AEW541	IGF1R	.0025,.008	2.93,3.40,.080	4.57,16.8,1.7	8	Constant	NA	8	-4.68877	0.6375
17	A2780_OVARY	A2780	AEW541	IGF1R	.0025,.008	6.95,1.20,-.26,	10.9,9.43,4.2	8	Sigmoid	8.735671	6.508904	-57.1806	0.6674
18	A375_SKIN	A-375	AEW541	IGF1R	.0025,.008	1.92,4.74,4.50	2.97,1.72,.44	8	Sigmoid	8.043994	2.755424	-79.6081	1.4484
19	A549_LUNG	A549	AEW541	IGF1R	.0025,.008	5.54,5.62,-.69,	2.46,.16,6.07	8	Sigmoid	3.299131	1.954787	-101.633	1.6406
20	A673_BONE	A-673	AEW541	IGF1R	.0025,.008	-3.6,5.42,-4.6,	9.41,4.84,6.6	8	Sigmoid	0.627646	1.127677	-65.0293	1.599

Figure 3.1 Sample Data Set

The attributes of the dataset are clustered agglometrive clustering. Clustering is the process of classifying a group of objects such that objects which belong to same group are more analogous to each other than to belong to other cluster (groups). Hierarchical clustering is the approach for analyzing the cluster which pursue to frame a hierarchy for cluster.

(1) Agglometrive :-Agglometrive clustering is a “bottom up” access each perception outset to its own clusters and then split recursively through down the hierarchy.
(2) Divisive :- Divisive clustering is a “ top up” access, all perception begin as single cluster and then recursively split as moves down the hierarchy.

In our proposed work we are using agglometrive clustering algorithm. The compounds in the dataset are clustered using the bottom up approach. The clustering of compounds are based on functional groups present in the compound, physiochemical property, hydrophobic property, hydrophilic property and spheric parameter and electron parameter.

Algorithm:

```
Workset ws = new Set(points);
KDTree kdtree = new KDTree(points);
while (true) {
  foreach (Element p in ws) {
    if (p.hasCluster()) continue;
    Point q = kdtree.findNearest(p);
    if (q == null) break; // stop if p is last element
    Point r = kdtree.findNearest(q);
    if (p == r) { // create new cluster e that contains a and b
      Element e = cluster(p, q);
      newWork.add(e);
    } else { // can't cluster yet, try again later
      newWork.add(p); // add back to worklist
    }
  }
  if (newWork.size() == 1) // we have a single cluster
    break;

  ws.addAll(newWork); //add new nodes to worklist
  kdtree.clear();
  kdtree.addAll(newWork);
  newWork.clear();
}
```

Using the hierarchical clustering algorithm the dataset has been clustered into different groups. The critical step in this work is to apply particle swarm optimization (PSO) to these clustered compounds. PSO is an optimization technique to find an optimal solution by repetitively demanding for a better candidate solution with a given range of quality. PSO was advanced by Dr, Kennedy AND Dr. Eberhart in 1995, motivated by the social behaviour of fish schooling or bird flocking. Traditional evolutionary computational approaches such as genetic algorithm have many similarities with PSO. In PSO, along a population of arbitrary solution, it searches for optimum solution by enhancing the generations. Contrary to GA, PSO has no operators like crossover and mutation. Particles which are the probable solution will fly over the problem space beyond the current best particles.

In the proposed work, Using the PSO algorithm, rule generation can be effectively implemented. In data mining, a common learning model is classification, which is the common way of discovering rules.

PSO Algorithm:

```
1: //initialize all particles
2: Initialize
3: repeat
4:    for each  particle i in S do
5:       //update the particle's best position
6:       if f(x_i) < f(pb_i) then
7:          pb_i = x_i
8:       end if
9:       //update the global best position
10:      if f(pb_i) < f(gb) then
11:         gb = pb_i
12:      end if
13:   end for
14:
15:   //update particle's velocity and position
16:   for each  particle i in S do
17:      for each  dimension d in D do
18:         v_{i,d} = v_{i,d} + C_1 * Rnd(0,1) * [pb_{i,d} - x_{i,d}] + C_2 * Rnd(0,1) * [gb_d - x_{i,d}]
19:         x_{i,d} = x_{i,d} + v_{i,d}
20:      end for
21:   end for
22:
23:   //advance iteration
24:   it = it + 1
25: until it > MAX_ITERATIONS
```

Rules and decision trees are the trivial form for the representationof a classifier. In this paper, classifier is described as a intent of IF-THEN rules. IF part of the rule specifies presence of antecedents and THEN part states the predicted class. A classical rule is of the form

IF(term1^term2^…termN) THEN class, using the attribute in the attribute in the database, generic rules so that target prediction will be easy

EXPERIMENT AND RESULT

The 503 records in the dataset are clustered using the hierarchial agglometrive clustering algorithm.PSO,the stochartic random search algorithm was applied to these clustered compounds and it will effectively predict the target site.Rule generation can be implemented using PSO to predict the target site.The target can be divided into different classes and the number of classes will depend on the binding site.Let C1,C2 are the two different classes representing the target site.C1 denote the target skin as a class and C2 denote the target lung as another class.The rule for the first class (C1->skin) can be generate as follows.

R1.1 If(Compound Name=Tetracycline)^(Amax=8.731)^(Actarea=5.5)
R1.2 If(Compound Name=Metronidizole)^(Amax=8.043)^(Actarea=2.75)
R1.3 (If Compound Name=Hydrocortisone)^(Amax=1.48)^(Actarea=8)

Thus compound with above specified attribute value was considered under evaluation,then we can easily predict the target site as skin.Similarly C2 denote the target lung as other class,(C2->lung) and rules can be generate as follows.

R2.1 If(Compound No=D0009)^(Compound Name=Adrenaline)^(Amax=8.55)
R2.2 If(Compound N0=DO9796)^(Compound Name=Hydrocortisone)^(Amax=1.042)
R2.3 If(Compound No=D05017)^(Compound Name=Metronidizole)^(Actarea=1.95)

Using thede attribute alues,PSO will classify compounds into different classes.Each class represent a target.Thus using PSO over clustered compounds,will help to predict the binding site of the compound and thus it will help to design the drug for the pharmacologists.

CONCLUSUION

In recent years,treatments are in the technical path so that users will get accurate results and can save time.But researchers found difficult to find an effective solution.Traditional drug design is a trial and reeor method on animals which will make huge loss in life and money.This proposed method will provide better preclinical trial and clinical trial of drugs for users and process it very accurately.Thus we can conclude that the proposed method, cluster based PSO in target prediction will find the attributes of the drug composition and cluster using agglometrive clustering.Rules will generate and can effectively predict the binding site using these rules

REFERENCES

[1] Romano T.Kromer, Structure based drug design: Docking and Scoring, Current protein and peptide science, 2007, 8,312-328
[2] Charanpreet Kaur and Shweta Bhardwaj, International Journal of Information and Computation Technology,ISSN 9074-2238,volume4,Number 4(2014),pp.335-342
[3] He Z,Zhang J,Shi X-H,Hu L-L,ong X,et al(2010),Predicting Drug Target Interaction Based On Functional Groups and Biological Features,PLOS ONE5(3):e9603. dui: 110.1371/journal.pone.0009603,March 11,2010
[4] D Rognan, Chemo genomic: Approaches to Rational Drug Design, British Journal of Pharmacology (2007).
[5] Thomas Engel, Basic Overview of Chemo informatics, J Chem Inf Model, 2006, 46, 2267-2277.
[6] Yu- Jun Zheng,Hai-Feng Ling,Jin-Yun Xue and Sheng –Yong Chen,Population Classification In Fire Evacuation:A MultiObjective Particle Swarm Optimization Approach,IEEE Transactions On Evolutionary Computations,Vol 18,No.1,February 2014.
[7] Amin Rostami and Maryam Lashakari,Extended PSO Algorithm For Improvement Problems For K Means Clustering Algorithm,Internatioanl Journal Of Managing Information Technology,Vol-6,No.3,August 2016.
[8] Manasi M Jahagirdar and Prof. S.M.Kamalapur, Swarm Intelligence Based Gene Classification, International Journal of Scientific and Research Publications, Vol-4, Issue-6, June 2014.