ELUCIDATING SIGNAL TRANSDUCTION MODULATORY DRUG TARGET NETWORK OF COLON CANCER- A NETWORK BIOLOGY APPROACH

[1] Deepak Ranjan Sethi, [1] Ibnul Hassan, [1]Sanjay Kundu, [2] Biplab Bhattacharjee, [2] Jayadeepa.R.M*, [3] Sushil Kumar Middha

[1] Amrita School of Biotechnology, Amrita Vishwa Vidya Peetham,Kerala,India
[2] Institute of Computational Biology,Bangalore,India
[3] Maharani Lakshmi Ammani College for Women, Bangalore, India

*Corresponding Author : jayadeepa.r.m@gmail.com

ABSTRACT-

Latest evaluation and validation of cancer drugs and their targets has demonstrated the lack and inadequate development of new and better drugs, based on available protocols. Even though the specificity of drug targets is a great challenge in the pharmaco-proteomics field of cancer biology, for eradicating such hurdles and paving the way for the drugs of future, a novel step has been envisaged here to study the relation between drug target network and the corresponding drug network using the advanced concepts of proteomics and network biology. The literature mining was done for the collection of receptors and the ligands. About 1000 natural compounds were collected and out of those 300 molecules showed anti-cancer activity against colon cancer. Ligand Vs multiple receptor docking was done using the software Quantum 3.3.0; the results were further used for the designing of a well connected Protein Ligand Interaction(PLI) network of colon cancer. The obtained network is then extrapolated to sort out the receptors expressed in the specific cancer type. The network is then statistically analyzed and represented by the graphical interpretation, in order to ascertain the hub nodes and their locally parsed neighbours. Based on the best docking scores, the graphs obtained from the docking analysis are statistically validated with the help of VisANT. In the network three hub nodes Neutrophil cytosol factor 2, UV excision repair protein RAD23 homolog A, & Receptor-type tyrosine-protein phosphatase eta were identified, which showed the highest interaction with the ligands. Butyrate and Farnesol showed highest interaction as ligands. Multiple Sequence Alignment was done of the binding site sequence of the drug targets to find out the evolutionary closeness of the binding sites. The phylogenetic tree was also constructed to further validate the observation. Further in-vitro and in-vivo studies needs to be done to analyse the receptor specificity and anti tumor activity of these compounds in Colon cancer.

Keywords: Signal transduction modulators, Colon Cancer, Hub Nodes, Natural Ligands, PLI (Protein Ligand Interaction)

Introduction to Study

Treatment of diseases and the application of scientific approach in the field of computer aided drug discovery and designing has become much more pronounced with the increased occurrences of genetic diseases in the recent past. Hence there is a lot of work that needs to be done in the field of in-silico drug designing in order to get rid of such new diseases using Network Pharmacology. The two methods by which a drug does its role is by either stimulating the target activity(agonist) or by down regulating its activity(antagonist) [1]. In this paper we have discussed an integrated approach of simultaneously focussing on Drug-Drug Target with Drug Target-Cancer protein interactions, that has the additional capability to act on cancerous proteins which are generally out of reach in normal circumstances in the classical analysis protocols, and also provides a model for the molecular mechanisms involved in the target prediction[2].

Cancer is a term used for describing uncontrolled and abnormal cells division along with the ability to invade other tissues. Cancer can be classified into various categories depending on its form and the cell type it is involved with [3]. Colon or Colorectal cancer is one of the most common cancers in the world and is one of the leading causes of cancer related deaths around the world. The development of Colon cancer begins with the abnormal cell growth & faulty repair mechanism in the colon cell lining and such cells are known as Polyps. Polyps are generally benign but do have the potential to turn cancerous. Like any other cancer type, Colon cancer is initiated by a signal to the receptor, loss of function of intermediary genes and hence resulting in erroneous cell behaviour and function [4].

Oncogenic proteins that participate in signal transduction pathways play a pivotal role in the transmission of a signal from the extracellular environment, through the cell membrane, into the cytoplasm and to the nucleus where transcription is initiated and even modulated to generate proteins that will eventually contribute to the cancerous phenotype. The function of these proteins is vital for homeostasis in the cells and tissues, moreover they control processes such as cell division, differentiation and apoptosis. All these molecules are potential targets for anti cancer drug designing due to the very fact that inhibition or activation

of their function will lead to the elimination of the tumor cells. Issues related to these compounds, such as mechanism-based toxicity and development of resistance, have to be addressed in order to ensure their successful development as chemo preventive agents. The continuation of the research in the development of these compounds as drugs for the future cancer will pave the way for adoption of novel strategies[5], [6].

Treatment of cancer involves a very high degree of complications and unseen factors like the possible side-effects, hence making successful treatment of cancer all the more difficult. Most cancers are highly invasive and there are problems of recurrence even after surgery, chemotherapy and radiation treatment. Butyrate is found to play a protective role in colon cancer by inhibiting the enzymes up regulated in colon cancer and responsible for the classic erroneous growth [7].The iso-prenoid alcohol Farnesol is an effective inducer of cell cycle arrest and apoptosis in a variety of carcinoma cell types. In addition, Farnesol has been found to function as a chemo preventative and anti-tumor agent in colon cancer [8].

Gene expression is the result of post transcriptional & translational modification of gene products. Therefore the up-regulated genes (highly expressed) in Colon cancer are sorted out for the further validation by experimentation and analysis by Hyper Geometric Test. The result obtained is fed into a well defined network, demonstrating the relationship between the various involved elements in order to ascertain their biological properties and functions. The centralities of the biological entities define their importance and biological hierarchy in the devised network.

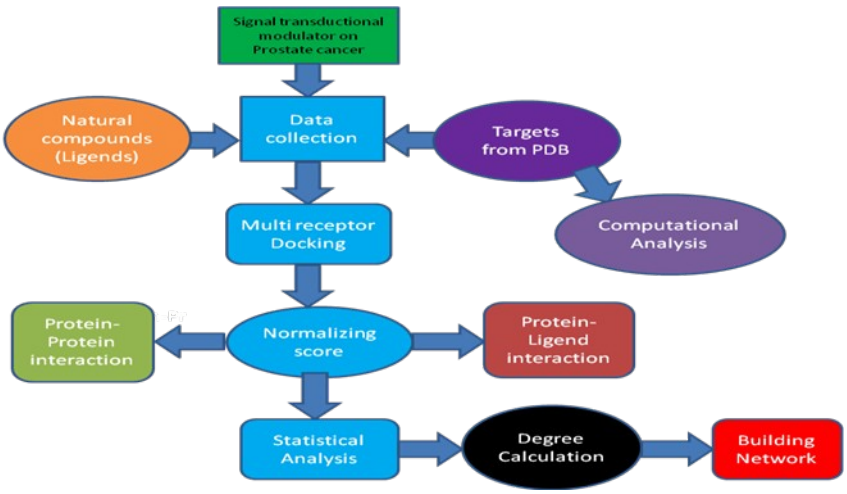

Figure 1: Overall work flow in the designing of a network.

Materials and Methods

A. Data mining, Target & Lead Identification:

The first step involved was the collection of Protein Drug Targets. From the PDB database, the disease proteins were collected in order to study the Colon cancer. The Colon cancer specific natural Signal Transduction Modulators were annotated from a wide range of publishers and databases like Wiley, Blackwell Synergy, Medline, Pubchem, Ingenta Connect, Chemfinder, Drug Bank, etc. The protein structures were collected from various online databases. The structures of the compounds i.e. the ligands were also collected from various sources.

Collection of Drug Targets

The colon cancer related proteins, extracted and validated leading to the 3 drug targets were considered based on their association with the specific cancer and the significance level of the protein in the disease as well as its involvement with the cellular processes, biological functions and molecular functions were assessed.

Table 1: Ligand Drug Target Interaction Scoring Matrix.

In Table 1 The Ligand- Drug Target Interaction between the 3 drug targets and the screened off 2 ligands (from about 300 ligands) is depicted. The possible interaction between the drug targets and the ligands is the building block of our study to obtain the best signal transduction modulators. The modulators have had been screened of based on the docking scores and only the best ones have made through into our study. Such method of screening is based on the fact that the signal modulators with higher scores are supposed to give greater interactions and the drugs based on this concept should give out better result.

In order to study the effects of colon cancer specific signal transduction modulators on the biological network of colon cancer, ligand Vs multi receptor docking was performed to elucidate the high grade specificity of the ligands for best fit. The analysis has been done on the basis of their docking score and RMS value. After the protein and ligand docking, the scores matrix was created from the scores obtained. Therefore a Structural Biology approach has been used for the classification and interpretation of the Systems Biology Network Model.

	DT1*	DT2*	DT3*
LD1*	65.97	58.94	NA
LD2*	NA	NA	82.13

	Cancer Protein 1	Cancer Protein 2	Cancer Protein 3	Cancer Protein 4	Cancer Protein 5	Cancer Protein 6	Cancer Protein 7	Cancer Protein 8	Cancer Protein 9	Cancer Protein 10
P.I *	1	1	14	2	7	7	15	8	3	21

Table 2: Protein Protein Interaction Scoring Matrix. The cancer proteins are plotted against the number of protein interactions that they are having and hence the Protein Protein Interaction Scoring Matrix was obtained. The scoring matrix so obtained, verifies the fact that indeed the signal transduction modulators having the highest docking scores have the highest interactions and hence provide an insight into the fact that the drugs conceptualised on this basis have better chances against the other drugs to carry out the signal modulation procedure, finally resulting in the state of homeostasis in the target drug targets.

B. Statistical Analysis:

The two sets (X, Y) of ligands and the target receptors respectively are taken, using their mean docking score & by help of the tool, VisANT [9] the highest degree of interaction was found to be with Neutrophil cytosol factor 2, UV excision repair protein RAD23 homolog A**,** Receptor-type tyrosine-protein phosphatase eta genes, where they shared degree distribution and it was found to be average score of 3.938 and the correlation coefficient was found to be 0.95, the graph was plotted for this statistical analysis data. Node degree is represented by its visible interactions. The distribution is calculated to see their fitness with power-law curve. The highest interacting proteins were found to be

Neutrophil cytosol factor 2, UV excision repair protein RAD23 homolog A**,** Receptor-type tyrosine-protein phosphatase eta [10].

*P.I= number of protein interactions.

*LD=ligands(signal transduction modulators)

*DT= drug targets

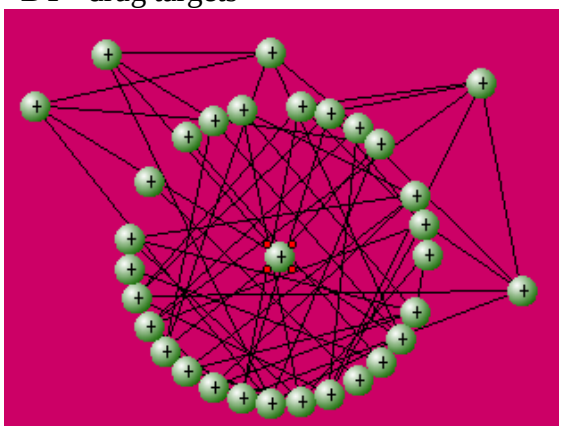

Figure 2: Highest Hub nodes show interaction with compounds

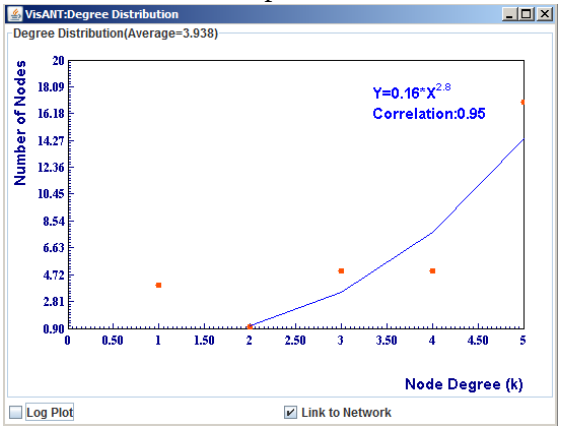

Figure 3: Degree distribution calculation

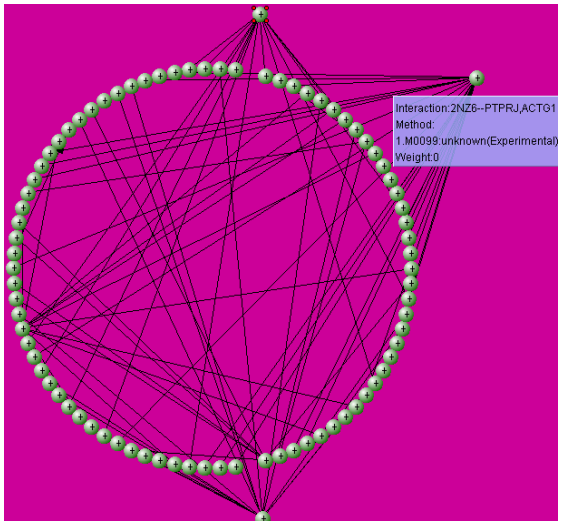

Figure 4: Protein –Protein interaction network

Fig 4: Shows the Drug Target Protein-Cancer Protein interaction. In this network Receptor-type tyrosine-protein phosphatase eta is the hub node. Many proteins related to the Receptor-type tyrosine-protein phosphatase eta and other targets show very weak interaction to the Drug Target Proteins which shares degree of 1.783, and the correlation coefficient was found to be 1.0.The graph distribution was found to be the average score & it was plotted for the statistical analysis data.

In this network Neutrophil cytosol factor 2, UV excision repair protein RAD23 homolog A**,** Receptor-type tyrosine-protein phosphatase eta are the proteins related to colon cancer. The scoring matrix of the Drug target Protein-Cancer Protein Interaction was done with the aid of PIPS: Human Protein Protein Interaction Prediction Server (http://www.compbio.dundee.ac.uk/www-pips/).

PIPs is a database of predicted human protein-protein interactions. The predictions have been made using a naïve Bayesian classifier to calculate a score of interaction. There are 37606 interactions with a Score ≥1 indicating that the interaction is more likely to occur than not to occur.

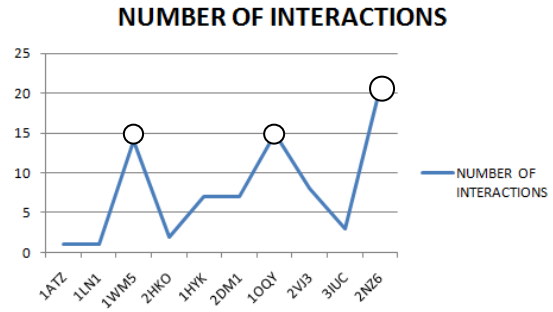

Figure 5: Graph between the drug targets and number of interactions.

In Fig 5 Blue line indicates the Protein-Protein Interactions scores graph between drug targets & cancer proteins found from the PIPs database. The peaks are represented by dots. 2NZ6 shows the highest interaction score.

C. Sequence and Evolutionary Analysis:

The identification of a good binding site and characterization of a drug target protein is of prime importance that leads it into its functional annotations. Therefore multiple sequence alignment is done to measure the evolutionary conservation of the binding sites of drug proteins that hold much evidence for the best fit of a ligand. The conserved binding site proved to be functionally enriched. The sequence analysis of the binding site was done for the

Signal transduction Drug Target Proteins-Neutrophil cytosol factor 2, UV excision repair protein RAD23 homolog A and Receptor-type tyrosine-protein phosphatase eta. The sequence analysis score shows a high degree of evolutionary conservation among the binding site sequences of the drug target proteins. The phylogenetic profile of the proteins depicts low substitution rate & less gap penalty which indicates that they belong to same protein family. The evolutionary tree was drawn using CLUSTALW2 [11]. A rooted Phylogenetic tree with a unique node corresponding to the most recent common ancestor was found using the evolutionary analysis study.

SeqA	Name	Len(aa)	SeqB	Name	Len(aa)	Score
1	1ATZ	445	2	1LN1	382	13
1	1ATZ	445	3	1WM5	360	21
1	1ATZ	445	4	2HK0	720	25
1	1ATZ	445	5	1HYK	622	16
1	1ATZ	445	6	2DM1	893	30
1	1ATZ	445	7	1OQY	1234	21
1	1ATZ	445	8	2VJ3	113	28
1	1ATZ	445	9	3IUC	704	22
1	1ATZ	445	10	2NZ6	383	23
2	1LN1	382	3	1WM5	360	19
2	1LN1	382	4	2HK0	720	21
2	1LN1	382	5	1HYK	622	21
2	1LN1	382	6	2DM1	893	32
2	1LN1	382	7	1OQY	1234	29
2	1LN1	382	8	2VJ3	113	21
2	1LN1	382	9	3IUC	704	24
2	1LN1	382	10	2NZ6	383	22
3	1WM5	360	4	2HK0	720	26
3	1WM5	360	5	1HYK	622	20
3	1WM5	360	6	2DM1	893	16
3	1WM5	360	7	1OQY	1234	30
3	1WM5	360	8	2VJ3	113	30
3	1WM5	360	9	3IUC	704	26
3	1WM5	360	10	2NZ6	383	14
4	2HK0	720	5	1HYK	622	18
4	2HK0	720	6	2DM1	893	23
4	2HK0	720	7	1OQY	1234	30
4	2HK0	720	8	2VJ3	113	17
4	2HK0	720	9	3IUC	704	21
4	2HK0	720	10	2NZ6	383	27
5	1HYK	622	6	2DM1	893	23
5	1HYK	622	7	1OQY	1234	28
5	1HYK	622	8	2VJ3	113	43
5	1HYK	622	9	3IUC	704	22
5	1HYK	622	10	2NZ6	383	23
6	2DM1	893	7	1OQY	1234	31
6	2DM1	893	8	2VJ3	113	13
6	2DM1	893	9	3IUC	704	27
6	2DM1	893	10	2NZ6	383	21
7	1OQY	1234	8	2VJ3	113	29
7	1OQY	1234	9	3IUC	704	24
7	1OQY	1234	10	2NZ6	383	27
8	2VJ3	113	9	3IUC	704	26
8	2VJ3	113	10	2NZ6	383	15
9	3IUC	704	10	2NZ6	383	26

```
1LN1        ----------------LLLLLLLYYYYYYYYYYYYEEYYYYYYYYRRRDDDYYYYYYYYY 51
2DM1        RRRRIIIIIIIIIIIIIIIIIIIYYYYYYYSSSSSSRRRRRRRRRRRRRRRWWWWWWWW 85
1HYK        CVVVVVVVVVVVVVVVVPPPPPPPCCCCCCCCCCCCDCCCCCCCYYYCCCCCCCCRRRLL 122
2VJ3        ------------------PPPPP----CCCCCCAATTTCCCC---------------- 30
1OQY        RRRRVVVVVVVVVVVVVVVVVVVVVVVVAAAAAALLLLLLLLLLLLLLLLLPPPPPL 180
1WM5        -------------------------GGCCCCCCMMTTTTTIIIIIFFFFQQQQQQQQRM 34
3IUC        ------------TTTTTTTTTTTTYYYYYYYYYYSSSSSCCCCIIIIIYKKKKKKKKKKR 58
2HK0        RRRRRRRVVVVLLGGGGAAAAMMMMMMVVVVVVVVVVVVVVTTTTTTNNNMMMFFFFFY 122
2NZ6        -----LLLLKKKLLLLLLLVVVVVVGGGGISSSSSQQQQQQQRRRRRGGGGPPPPPPLL 80
1ATZ        ------------QQQQQVVVVVPWWNNNNNRRRRPPPPPPPGGGGAAAKKKKKKNNRRR 67
```

Figure 6: Clustalw2 result showing the similarity in the Drug Target Gene Sequence

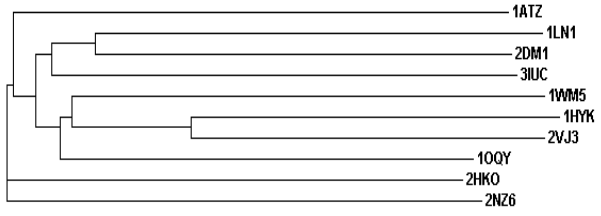

Figure 7: Target sequences show evolutionary relationship

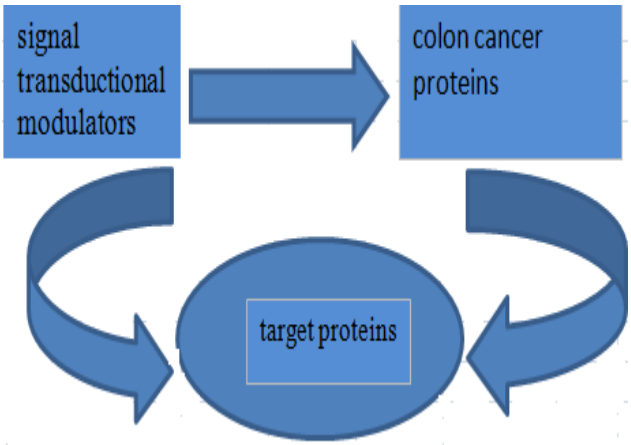

Figure 8: The relation between ligand and Drug Target Proteins. Small molecules i.e. the Signal Transductional modulators interact with their disease protein targets to generate a work flow, in which a drug and a protein are connected to each other through the medium of target proteins. As some of the small molecules cannot directly interact with the desired target due to their small size, they do require some other proteins that help them become fit to act upon the target proteins.

D. Matrix Analysis

The scoring matrix was plotted using the scores obtained from multiple receptor docking to identify the possible drug targets and were found to be Neutrophil cytosol factor 2, UV excision repair protein RAD23 homolog A, & Receptor-type tyrosine-protein phosphatase eta[Table no-1].

To obtain the best signal transduction modulators, the number of interactions obtained from the protein protein interactions were drawn in a scoring matrix to further validate the fact that the signal transduction modulators indeed have greater interactions and hence have better chances of serving the purpose as drugs for colon cancer[Table-2].

Results-

Thus, out of 300 natural molecules two molecules were selected for final study namely Butyrate and Farnesol that displayed the highest ligand and Drug Target Interactions with the obtained hub node proteins of the specific cancer. Both of the signal transduction modulators were filtered out from the available compounds, first by scores obtained from the multiple receptor docking and then by number of protein protein interactions. The

signal transduction modulators obtained from the screening procedure performed better than the other molecules. For the effective targeting and signal modulation of the involved target receptor sites multiple sequence alignment was performed to obtain the sequence conserved in evolution. The signal transduction modulators that efficiently and effectively target the conserved sequences can neutralize the abrupt signals and can bring the cell, back into the normal homeostatic condition.

Conclusion-

The interpretation of colon cancer network gives 3 hub nodes. The most prominent interactions were observed with Neutrophil cytosol factor 2, UV excision repair protein RAD23 homolog A & Receptor-type tyrosine-protein phosphatase. The compounds namely Butyrate and Farnesol exhibited the highest interaction with the up-regulated proteins in colon cancer. The phylo-genetic tree obtained also showed the evolutionary closeness of the binding sites in the Drug Target Proteins that has been conserved in evolution process. The significance of the work is to validate the target specificity of the naturally occurring anti oxidant drugs. Most of the drugs failed to work properly due to the lack of proper target identification. But here we have taken a predictive measure to map the natural oxidants to the colon cancer specific disease network so that an organized data can be available on the drug targets involved in the colon cancer and even their anti- oxidants agents. This study will further contribute to the biomedical field to develop new drugs with best targets.

Acknowledgment

We sincerely acknowledge Institute Of Computational Biology, Bangalore & DBT-BIF Centre, MLACW, Bangalore for the financial support for the research work

References

1.Laurent P. Rivory, Senior Scientist, Medical Oncology, Sydney Cancer Centre, and Clinical Senior Lecturer, Department of Pharmacology, University of Sydney, Sydney

2. B.Linghu. E.S Snitkin, Zhenjun Hu, Yu Xia and Charles delisi. Genome-wide priporitization of disease genes and identification of disease-disease associations from an integrated human functional linkage network,2009

3.http://www.cancer.gov/cancertopics/what-is-cancer

4.Marinus W. Lobbezooa, Giuseppe Giacconeb, Coenraad van Kalkena. Signal Transduction Modulators for Cancer Therapy NDDO Research Foundation, Amsterdam, The Netherlands; b VU Medical Center, Amsterdam, The Netherlands The Oncologist, April 2003,Vol. 8, No. 2, 210–213

5. Christos A. Tsatsanis and Demetrios A. Spandidos .Signal transduction pathways in cancer cells; novel targets for therapeutic intervention, 26 August, 2000; Accepted: 26 August, 2000; electronically published: February 2004

6. Barbacid M. ras Genes. Annu Rev Biochem 1987;56:779–82

7. Toni Baker Colon cancer shuts down receptor that could shut it down - 2009 April 13

8. Joung Hyuck Joo and Anton M. Jetten
Molecular Mechanisms involved in Farnesol-Induced Apoptosis
Cell Biology Section, LRB, Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709

9. http://vizant.sourceforge.net/

10.www.ncbi.nlm.nlh.gov

11. http://www.ebi.ac.uk/Tools/clustalw2/index.html