

₹ 200

ISSN - 2249-555X

**Volume : 1**

**Issue : 5**

**February 2012**



**Journal for All Subjects**

[www.ijar.in](http://www.ijar.in)

Listed in International ISSN Directory, Paris.



ISSN - 2249-555X

# Indian Journal of Applied Research

## Journal for All Subjects

### Editor-In-Chief

**Dr A Kumar**

Director, College Development Council (CDC)  
Director, Internal Quality Assurance Cell (IQAC)  
Professor in Management,  
Department of Business Administration, Faculty of Management,  
Bhavnagar University,

### Editorial Advisory Board

**Dr. S. N. Pathan**  
Maharashtra

**Dr. SM. Ramasamy**  
Gandhigram

**Dr. M. M. Goel**  
Kurukshetra

**Dr. S. Ramesh**  
Tamil Nadu

**Dr Ramesh Kumar Miryala**  
Nalgonda.

**Dr. B. Rajasekaran**  
Tirunelveli

**Dr. A. R. Saravankumar**  
Tamilnadu

**Dr. Roy M. Thomas**  
Cochin

**Dr. G. Selvakumar**  
Salem

**Dr. Apurba Ratan Ghosh**  
Burdwan

**Dr. Shrawan K Sharma**  
Uttarakhand

**Dr. Sudhanshu Joshi**  
Uttarakhand

**Prof. (Dr.) B Anandampilai**  
Pudhukottai

#### Advertisement Details

Position	B/W (Single Color)	Fore Color
Full Inside Cover	₹ 6000	₹ 12500
Full Page (Inside)	₹ 5000	-

#### Subscription Details

Period	Rate	Discount	Amount Payable
One Year (12 Issues)	₹ 2400	Nil	₹ 2400
Two Year (24 issues)	₹ 4800	₹ 200	₹ 4600
Three Year (36 issues)	₹ 7200	₹ 300	₹ 6900
Five Year (60 issues)	₹ 12000	₹ 600	₹ 11400

You can download the Advertisement / Subscription Form from website [www.ijar.in](http://www.ijar.in). You will require to print the form. Please fill the form completely and send it to the **Editor, INDIAN JOURNAL OF APPLIED RESEARCH** along with the payment in the form of Demand Draft/Cheque at Par drawn in favour of **INDIAN JOURNAL OF APPLIED RESEARCH** payable at Ahmedabad.

1. Thoughts, language vision and example in published research paper are entirely of author of research paper. It is not necessary that both editor and editorial board are satisfied by the research paper. The responsibility of the matter of research paper/article is entirely of author.
2. Editing of the Indian Journal of Applied Research is processed without any remittance. The selection and publication is done after recommendations of atleast two subject expert referees.
3. In any condition if any National/International University denies accepting the research paper published in IJAR, then it is not the responsibility of Editor, Publisher and Management.
4. Only the first author is entitle to receive the copies of all co-authors
5. Before re-use of published research paper in any manner, it is compulsory to take written permission from the Editor-IJAR, unless it will be assumed as disobedience of copyright rules.
5. All the legal undertaking related to Indian Journal of Applied Research is subject to Ahmedabad Jurisdiction.
7. The research journal will be send by normal post. If the journal is not received by the author of research papers then it will not be the responsibility of the Editor and publisher. The amount for registered post should be borne by author of the research paper in case of second copy of the journal.

**Editor,**

**Indian Journal Of Applied Research**

8-A, Banans, Opp. SLU Girls College, New Congres Bhavan, Paldi,  
Ahmedabad-380006, Gujarat, INDIA

Contact.: +91-9824097643 E-mail : [editor@ijar.in](mailto:editor@ijar.in)

# Index

Sr. No	Title	Author	Subject	Page. No.
1.	Assay Of Triphenylmethane Reductase Enzyme And PCR-Based Identification Of TMR Gene In Enterobacter Asbriae Strain XJUHX-4TM	Tina Mukherjee, Mounita Bhandari, Manas Das	Biotechnology	1-2
2.	An Analysis Of Growth Of Credit Card Industry	Dr. A. Vinayagamoorthy, K. Senthikumar	Commerce	3-5
3.	Impact Of Pre-Merger And Post Merger On Financial Performance (With Reference To Private Sector Banks)	Dr. Shital Vekariya	Commerce	6-8
4.	Relativity On Climate And Competencies In Human Resource Development With Reference To Neyveli Lignite Corporation Ltd,	S. Jayakumar. Dr. R. Ramachandran	Commerce	9-11
5.	Human Resource Outsourcing: A Strategy For Gaining Competitive Advantage	Dr. Santosh M. Singh	Commerce	12-13
6.	Relationship Between EVA And ROI And MVA (A Case Study Of Ten Manufacturing Industries In India)	Dr. Shivani Gupta	Commerce	14-15
7.	Modeling The Traits Of An Effective Teacher At Higher Education	Dr. Haridayal Sharma	Commerce	16-17
8.	Mahatma Gandhi National Rural Employment Guarantee Act (Mgnrega): Issues And Challenges	Dr. Mohd. Ashraf Ali, Mushtaq Ahmad	Commerce	18-20
9.	Standardisation And Grading	Viram. J. Vala, Dr. Vijay Kumar Soni	Commerce	21-22
10.	Profitability Of Selected Information Technology Companies In India	Dr. M. Jegadeeshwaran, C. Udaya	Commerce	23-25
11.	Emerging Trends In The Indian Media And Entertainment Industry	Dr Mahalaxmi Krishnan	Commerce	26-27
12.	Inventory Management Strategies And Control Techniques: An Empirical Investigation Of Small Scale Industries	Vipul Chalotra, Neetu Andotra	Commerce	28-30
13.	A Study On Performance Indicators Of Commercial Banks	Dr. G. Ganesan, P. Parthasarathy	Commerce	31-33
14.	Improved Approaches To Coreference Resolution In Machine Learning	Kuldeep Singh Raghuwanshi, Ashwini Kumar Verma	Computer Science	34-37
15.	Security Issues & Controls In Cloud Computing	V. Naga Lakshmi	Computer Science	38-40
16.	Human Development Index Of De-Notified Nomadic Castes In Maharashtra Division: A Study Of Jalna And Aurangabad Districts	Dr. Ashok Pawar	Economics	41-43
17.	Public Private Partnership In Rural & Urban Projects In India	Dr. Ashok S. Pawar, Dr. Shankar B. Ambhore	Economics	44-45
18.	Populace Insight On Development In Public Health Sector Of India Subsequent To Functioning Of National Rural Health Mission	Krishnakant Sharma	Economics	46-49
19.	Problems Of Rural Women Entrepreneurs In India: A Conceptual Overview	C. Jeyasri Usha N Devi, Dr. A. Sankaran	Economics	50-52
20.	Poverty Of Banjara And Vanjari Communities In India	Tidke Atish S., Dr. Pawar Ashok S.	Economics	53-54
21.	India And China: Economic Reforms And WTO	Dr. Surinder Kumar Singla, Dr. Kulwinder Singh	Economics	55- 56
22.	Implementing Life Skill Education Strategies In Teaching – Learning Process	R. Kalaiselvi, Dr. A. Palanisamy, Dr. A R. Saravanakumar	Education	57-59

23.	Utilisation Of Modern Technology By The Teachers In Pupil Processing Organisation	Dr. P.Paul Devanesan, Dr A. Selvan	Education	60-61
24.	Impact Of Vocational Training On Students	K.Sudha Rani, G.Umapathi, Dr. T. Ananda,	Education	62-63
25.	A Study On Emotional Intelligence Of Secondary School Teachers	Dr. Umme Kulsum, Prathima H.P.	Education	64-66
26.	The Efficiency Of Feedback Strategy Of Homework On The Development Of 10th Grade EFL Writing Skill In Al-Karak Educational Directorate	Majid Al- Khataybeh, Areej Al-Shourafa`	noitacudE	67-74
27.	Perspectives Of Stress Management In Education System	M. Meenakshisundaram, G. P. Raja, Dr. A R. Saravanakumar	Education	75-76
28.	Attention Regulation Of Meditators And Non-Meditators Of Class IX	G. Madhavi Kanakadurga, Dr. D. Vasanta Kumari,	Education	77-78
29.	Role Of Psychoeducation In Teaching – Learning Process	Dr. A R. Saravanakumar, Dr. A. Balu, Dr. S. Subbiah	Education	79-80
30.	Microcontroller Driven RGB Led System For Tristimulus Surface Colorimetry	T. N. Ghorude, A. D. Shaligram	Electronics	81-83
31.	Pmgsy And Rural Roads Development In India: Economic, Financial And Maintenance Issues	K.C. Manjunath	Engineering	84-86
32.	Routing Packets On A Chip.	Naren V Tikare	Engineering	87-89
33.	Finding The Nearest Neighbors In Biological Databases	Er. Pankaj Bhambri, Dr. O.P. Gupta, Er. Franky Goyal	Engineering	90-92
34.	Factors Affecting The Sustainability Of The Asphalt Roads: A Case Study Of Irbid Inner Ring Road, Jordan	Eng. Nasr Ahmad Dr. Mihai Iliescu	Engineering	93-94
35.	Physical And Chemical Testing Of Compounded PVC	Sapna Dabade, Dr. Dheeraj Mandloi, Deepak Khare	Engineering	95-96
36.	Impact Of Organic Farming On Yield Of Some Common Crops- A Case Study.	Namrata D. Awandekar	Environmental Science	97
37.	Hydrogeologic Settings Of The North And South Brahmaputra Plains In Upper Assam: A Comparative Study	Dr. Uttam Goswami	Geology	98-100
38.	To Study Staffing Pattern In Rajasthan Public Healthcare Delivery System.	Dr. Ashwin G. Modi, Sushman Sharma	Healthcare	101-105
39.	Work And Health: A Situational Analysis Of Factory Workers	Dr. S. S. Vijayanchali, Dr. E. Arumuga Gandhi	Home Science	106-108
40.	Performance Of Camel Kid Hair: Acrylic Blended Yarn And Knitted Fabric	Suman Pant, Anjali Sharma	Home Science	109-110
41.	Impact Of Holistic Nutrition Education Package On Diabetes Mellitus Control In Middle Aged Women	Dr. Anjali Rajwade	Home Science	111-112
42.	Assessment Of Relationship Between Ida And Personal Hygiene, Nutritional Knowledge And Dietary Practices In Adolescent Girls	Dr. Anjali Rajwade	Home Science	113-114
43.	Employee Attrition And Retention In Private Insurance Sector– A HRM Challenge	Dr. J. Senthil Vel Murugan, S.Bala Murugan	Human Resource Management	115-117
44.	A Study On Impact Of Unionism On Industrial Relations In Manufacturing Sector	Jaya Ahuja	Industrial Relations	118-120

45.	Augmentation Of India's Foreign Exchange Reserve: An Analysis	Dr.S P.Mathiraj, Ar.Annadurai	International Business	121-123
46.	Films – A Techno Literary Art Form	Dr. Dipti Mehta	Literature	124-125
47.	Indirect Models Of Reading To Develop Descriptive Writing	Dr. K. Madhavi	Literature	126-128
48.	Ramkrishna Mishra Ke Upanaso Me Rajnetaik Chetavni	Dr. Sanjay Rathod, Dilip Jhadav	Literature	129
49.	Hindi Kavita Me Nari Jivan Ka Badla Swarup	Dr. Sanjay Rathod	Literature	130
50.	Impact Of IPL Sponsorship On Consumer Buying Behavior With Reference To Nagpur City	Chandrima Das	Management	131-135
51.	Crowd Sourcing –A New Management Mantra	Devi Premnath, Dr. C. Nateson	Management	136-137
52.	Small Scale Industries In India: An Evaluation Of Productivity In The Post-Liberalized Scenario	Dr. Gaurav Lodha,	Management	138-139
53.	Comparative Analysis Of Milk Products With Respect To Its Competitors With Special Reference To Karnataka Milk Federation (KMF) – At Dharwada City, Karnataka, India	Dr. N. Ramanjaneyalu	Management	140-143
54.	A Study On Work Stress In Women Employees In Coimbatore District	R. Maheswari, N. Brindha	Management	144-145
55.	Accounting For Carbon Credits	Dr. Gaurav Lodha	Management	146-148
56.	A Literature Review On The Relationship Between Training (As A Core Responsibility Of HRM) And Firm Performance.	Priya Sharma, Dr. S. L. Gupta	Management	149-152
57.	A Study On Agricultural Marketing Practices And Constraints With Special Reference To Paddy / Rice.	CM Maran, Dr Raja Pranmalai	Management	153-156
58.	Performance Of Share Price Of Indian Public Sector Banks And Private Sector Banks - Comparative Study	V. Prabakaran, D. Lakshmi Prabha	Management	157-158
59.	Intuitionistic Fuzzy Primary And Semiprimary Ideal	Dr. M.Palanivelrajan, S.Nandakumar	Mathematics	159-160
60.	Significance Of Umbilical Artery Velocimetry In Perinatal Outcome Of Fetuses With Intrauterine Growth Retardation.	Dr G S Shekhawat	Medical Science	161-163
61.	Large Adult Sacrococcygeal Teratoma: A Case Report And Review Of Literature.	Dr.Yavalkar Pa, Dr. Naik Am.	Medical Science	164-165
62.	Epidural Steroid In Low Back Ache	Dr. B. L. Khajotia, Dr. Neelam Meena	Medical Science	166-167
63.	A Comparative Study Of Second Trimester MTP With Use Of Vaginal Misoprostol And Extra Amniotic Instillation Of Ethacridine Lactate.	Dr. Ketaki Junnare, Dr. Sameer Darawade, Dr. Priyamvada Shah, Dr. Swati Mali.	Medical Science	168-169
64.	A Novel Surgical Approach For Treatment Of Sui –TVT Obturator Tape	Dr. Ketaki Junnare, Dr. Durga Karne, Dr Neelesh Risbud.	Medical Science	170-171
65.	Advantage Of Fallopian Tube Sperm Perfusion Over Intra-Uterine Insemination When Used In Combination With Ovarian Stimulation For The Treatment Of Unexplained Infertility.	Dr G S Shekhawat, Dr Pushpalata Naphade	Medical Science	172-175

66.	"Bilateral Sertoli-Leydig Cell Tumor In Postmenopausal Female" A Case Report	Dr. Priyamvada Shah, Dr. Ketakijunnare, Dr. DurgaKarne	Medical Science	176-178
67.	Pretreatment With Ephedrine For Prevention Of Pain Associated With Propofol Injection.	Dr. Kavita U Adate, Dr. Jyoti A. Solanki	Medical Science	179-181
68.	Does The Structured Teaching Programme Influence The Knowledge About Physical Wellbeing Of School Children? A Quasi Experimental Study.	Dr. S. Valliammal, Dr. Ramachandra, Raja Sudhakar	Nursing	182-184
69.	An Approach For Information Retrieval For Bookstores Using Formal Ontology	Sumit Jain, C.S.Bhatia	Ontology	185-187
70.	Analgesic Activity Of Anacardium Occidentale	A. Devadoss, C. Aparna, K. Parimala, D. Sukumar	Organic Chemistry	188-190
71.	Behaviourism : Science Or Metaphysics	Dr. Jatinder Kumar Sharma	Philosophy	191-193
72.	Multi-Dimensional Perspectives Of Obesity And Its Management	S. Dhanaraj, Dr. A. Palanisamy	Physical Education	194-196
73.	Refractive Index, Density, Excess Molar Volume, Excess Molar Refraction For Liquid Mixtures (Ethyl Ethanoate + Benzene Derivatives) At Different Temperatures	Sheeraz Akbar, Mahendra Kumar	Physics	197-199
74.	Refractive Indices, Densities And Excess Properties For Liquid Mixtures (Cetane + Alkanols) At Different Temperatures	Sheeraz Akbar, Mahendra Kumar	Physics	200-202
75.	Capacity Building For Effective Local Governance: Indian Perspectives	Dr. Pralhad Chengte	Political Science	203-205
76.	Psychological Well-Being: A Study Of Non-Institutionalized Aged	Dr. Pankaj S. Suvera	Psychology	206-208
77.	Women Empowerment Through N R E G S (With Reference To State Of West Bengal)	Dilip Kumar Karak	Social Sciences	209-211
78.	Effect Of Selected Yogic, Aerobic And Laughter Exercises On Blood Pressure Of High School Boys	Dr.Manjappa.P, Dr.Shivarama Reddy. M	Sports	212-216
79.	Association Study Between Lead And Copper Accumulation At Different Physiological Systems Of Goat By Application Of Canonical Correlation And Canonical Correspondence Analyses	Partha Karmakar, Debasis Mazumdar, Seema Sarkar (Mondal), Sougata Karmakar	Statistics	217-219
80.	Development Of Silver -Silica Nanocomposite For Novel Humidity Sensing Application	Surender Duhan	Technology	220-221



## Finding The Nearest Neighbors In Biological Databases

\* Er. Pankaj Bhambri \*\* Dr. O.P. Gupta  
\*\*\* Er. Franky Goyal

\* Assistant Prof., Department of Info. Tech., Guru Nanak Dev Engg. College, Ldh.

\*\* Dy. Director, School of Info. Tech., Punjab Agriculture University, Ludhiana

\*\*\* Research Scholar, Department of Comp. Sci. & Engg., Guru Nanak Dev Engg. College, Ldh.

**Keywords : Biological, Finding**

### Introduction

#### Introduction to Bio-informatics

The term bioinformatics was given by Pauline Hogeweg and Ben Hesper in 1978 for the study of informatics processes in biotic systems. Bioinformatics is the application of computer science and information technology to the field of biology and medicine. Bioinformatics is the science of managing, mining, and interpreting information from biological sequences and structures. Bioinformatics deals with algorithms, databases and information systems, web technologies, artificial intelligence and soft computing, information and computation theory, software engineering, data mining, image processing, modeling and simulation, signal processing, discrete mathematics, for generating new knowledge of biology and medicine, and improving & discovering new models of computation. It is an interdisciplinary research area, which may be defined as the interface between biological and computational sciences. Java, XML, Perl, C, C++, Python, R, SQL and MATLAB are the programming languages popularly used in this field.

#### Scope

Bioinformatics consists of two subfields:

- The development of computational tools and databases
- The application of these tools and databases in generating biological knowledge to better understand living systems.

These two subfields are complementary to each other. The tool development includes writing software for sequence, structural, and functional analysis, as well as the construction and curating of biological databases. These tools are used in three areas of genomic and molecular biological research: molecular sequence analysis, molecular structural analysis, and molecular functional analysis. The analyses of biological data often generate new problems and challenges that use to development of new and better computational tools.

#### DNA

DNA (deoxyribonucleic acid) is the genetic material. This is profoundly powerful statement to molecular biologists. It is the information stored in DNA that allows the organization of inanimate molecules into functioning of living cells and organisms that are able to regulate their internal chemical composition, growth, and reproduction. All plants, animals, bacteria and some viruses contain genetic information in the form of DNA. A characteristic of living organisms is that DNA is reproduced and passed on to the next generation. DNA contains the instructions for making proteins. It is composed of linear array of nucleotides, each of which has a base plus a deoxyribose sugar and a phosphate. There were two main

types of bases, purines (R) and pyrimidines (Y). Each main type of base came in two flavours. Two types of purines, adenine (A) and guanine (G), and two types of pyrimidines, cytosine (C) and thymine (T). DNA is in the form of double helix.

Figure 1.1: Structure of DNA



DNA is the king of molecules.

The molecule forms a double helix in which two strands of DNA spiral about one other. Strands refer to two chains.

### Problem Formulation

The major problem being faced by biologists and researchers is huge amount of raw data but with a lack of means to effectively use this data.

#### Sequence Alignment

In bioinformatics, a sequence alignment is a way of arranging the primary sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationship between the sequences. Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix. Gaps are inserted between the residues so that residues with identical or similar characters are aligned in successive columns. In protein sequence alignment, the degree of similarity between amino acids occupying a particular position in the sequence can be interpreted as a rough measure. The absence of substitutions, or the presence of only very conservation substitutions in a particular region of the sequence, suggest that this region has structural or functional importance. Although DNA and RNA nucleotide bases are more similar to each other than amino acids, the conservation of base pairs can indicate a similar functional or structural role. Sequence alignment can be used for non-biological sequences, such as those present in natural language or in financial data.

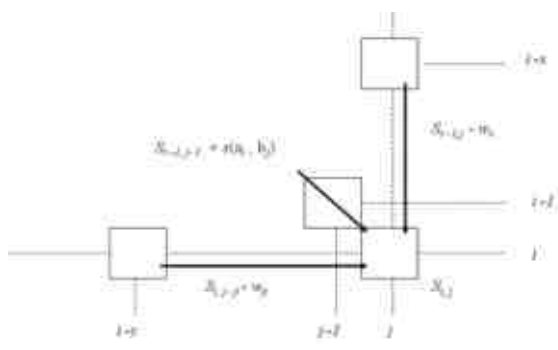


Global and Local Alignment

Global alignments, which attempt to align every residue in every sequence, are most useful when the sequences in the query set and of roughly equal size. A general alignment technique is called the Needleman-Wunsch algorithm and is based on dynamic programming. Local alignments are more useful for dissimilar sequences that are suspected to contain regions of similarity or similar sequence motifs within their larger sequence context. The Smith-Waterman algorithm is a general local alignment method also based on dynamic programming. With sufficiently similar sequences, there is no difference between local and global alignments. The work is focused on protein sequence alignment. The optimal alignment between two sequences will be computed. The optimal alignment is the one which gives the maximum alignment score for the input sequences. It is also possible to obtain multiple optimal alignments for the two sequences.

Implementation

Figure 2: Formal Description of dynamic programming

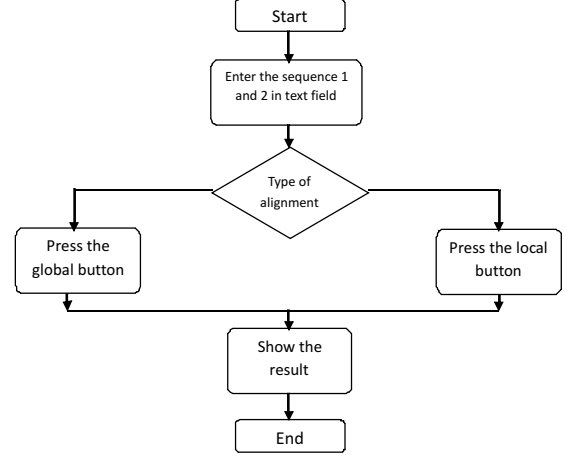


This diagram indicates the moves that are possible to reach a certain position (i,j) starting from the previous row and column at position (i-1, j-1) or from any position in the same row or column. Diagonal move with no gap penalties or move from any other position from column j or row i, with a gap penalty that depends on the size of the gap. For two sequences a = a<sub>1</sub>, a<sub>2</sub>, ..a, and b = b<sub>1</sub>, b<sub>2</sub>, ..b, where S<sub>i,j</sub> = S(a<sub>1</sub>...a<sub>i</sub>, b<sub>1</sub>...b<sub>j</sub>) then

$$S_{ij} = \max \left\{ \begin{array}{l} S_{i-1,j-1} + s(a_i, b_j), \\ \max_{x \geq 1} (S_{i-x,j} - w_x), \\ \max_{y \geq 1} (S_{i,j-y} - w_y) \end{array} \right\}$$

where S<sub>i,j</sub> is the score at position at i in sequence a and j in sequence b, s(a<sub>i</sub>,b<sub>j</sub>) is score for aligning the character at positions i and j, w<sub>x</sub> is the penalty for a gap of length x in sequence a, and w<sub>y</sub> is the penalty for a gap of length y in sequence b. Note that S<sub>i,j</sub> is a type of running best score as the algorithm moves through every position in the matrix.

Figure 3.2: Work flow diagram



Results And Discussion

Pairwise Sequence Alignment

The purpose of sequence alignment algorithms is to detect evolutionary, and thus structural and functional, relations among sequences. The successful sequence comparison would allow us to infer the biological properties of new sequences from data accumulated on related genes. In trying to characterize a newly discovered sequence (DNA, protein) the first step is always to check whether there are similar sequences in known databases and their annotation.

Scoring Matrices

The different scoring matrices used for homology prediction are PAM and BLOSUM.

Point Accepted Mutations (PAM)

PAM series is the first major amino acid substitution series of scoring matrices. The PAM matrices (also called Dayhoff PAM matrices) were first constructed by Margaret Dayhoff, who compiled alignments of seventy-one groups of very closely related protein sequences. Because of the use of very closely related homologs, the observed mutations were not expected to significantly change the common function of the proteins. Thus, the observed amino acid mutations are considered to be accepted by natural selection.

An ideal PAM matrix can be obtained as follows:

- Collect and align many pairs of homologous sequences that are known to be n PAM units diverged
- Calculate the number of times amino acid a<sub>i</sub> aligns opposite a<sub>j</sub> and divide by the total number of pairs in the aligned data, that gives us p(a<sub>i</sub>, a<sub>j</sub>)
- Calculate the number of times a<sub>i</sub> appears in both sequences and divide by their total length; this gives us p(a<sub>i</sub>) and then p(a<sub>j</sub>) for a<sub>i</sub>
- There are many different PAMs, which are representing different evolutionary scenarios. PAM 250 represents a level of 250% of changes expected in 2500 MY
- PAM is more suitable for studying quite distant proteins, BLOSUM is for more conserved proteins of domains

Table 4 PAM250 Substitution Matrix

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	L	I	V	F	Y	W	
C	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
S	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
T	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
P	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
A	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
G	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
N	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
D	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
E	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
Q	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
H	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	-4	0
R	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	-4	0
K	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	-4	0
M	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	-4	0
L	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	-4	0
I	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	-4	0
V	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	-4	0
F	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	-4	0
Y	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	-4	0
W	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	15	0

Amino acids are grouped according to the chemistry of the side group: (C) sulfhydryl, (STPAG)-small hydrophilic, (NDEQ) acid, acid amide and hydrophilic, (HRK) basic, (MILV) small hydrophobic, and (FYW) aromatic. Log odds values: +10 means that ancestor probability is greater, 0 means that the probabilities are equal, -4 means that the change is random. Thus the probability of alignment YY/YY is 10+10=20, whereas YY/TP is 3-5=-8, a rare and unexpected between homologous sequences.

BLOSUM Series of matrices

BLOSUM matrices were an alternative to PAM matrices. Their construction was based on a larger collection of over 2000 ungapped, aligned amino acid sequence blocks. Blocks are a short contiguous interval in a multiple alignment of amino acid sequences.

These matrices are derived from BLOCKS. It de-emphasis of pairs of rows that are too similar; for example if two rows that appear in a given block are identical, then one is removed;



Table 5 BLOSUM62 Substitution Matrix

	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Z
C	4	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
D	-1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	-1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	-1	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	-1	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	-1	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	-1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
K	-1	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
L	-1	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
M	-1	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
N	-1	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
P	-1	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
Q	-1	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
R	-1	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
S	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0
T	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
V	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
W	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
Y	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
Z	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4

Ideology of BLOSUM is similar but it is calculated from a very different and much larger set of proteins, which are much more similar and create blocks of proteins with a similar pattern.

Conclusion And Future Scope

Conclusion

The model is constructed for predicting different proteins sequence alignments along with the alignment scores. The protein alignment is useful in predicting the homologous sequences and thus useful for assigning the class for unknown protein. There are different sequence formats available from which plain text format is utilized. Global and local alignments are predicted along with the alignment scores. Needleman Wunsch algorithm is used for the global alignment problem and Smith Waterman algorithm is used for the local alignment problem. There are different applications of global and local alignments. The global alignment accesses the complete amino acid sequences without skipping any gap whereas the local alignment accesses the partial alignment skipping the initial and ending gap penalties. Scoring matrix assists in computing the alignment score. It consists of mutation scores for different amino acid alignments. The model is user friendly which provides different options to the user. The overall computation is also shown in the matrix along with the final sequence alignment.

Future Scope

Following improvements regarding the developed model of bioinformatics can be made:

- The model can be extended to implementing Gene expression.
- Protein structure prediction can be incorporated to further enhance the model.
- Database can be maintained for the known protein sequences and the search option can be provided to the user.
- Different scoring matrices can be included.

## REFERENCES

- [1] Bergeron, B. (2003), "Bioinformatics Computing", Pearson Education, pp. 110-160. | [2] Xiong, J. (2006), "Essential Bioinformatics", Texas A&M University, pp. 31-47. | [3] Krane, D. and Raymer, M. (2006), "Fundamental Concepts of Bioinformatics", Pearson Education Publishers. | [4] Li, J.J and Huang, De-S. (2005), "Characterizing Human Gene Splice Sites Using Evolved Regular Expressions", Proceedings of International joint Conference on Neural Networks, Montreal, Canada, July 31 August 4, 2005, pp. 493-498. | [5] Luscombe, N.M., Greenbaum, D., Gerstein, M. (2001), "What is Bioinformatics? A proposed definition and Overview of the field", Luscombe group publications, pp 346-358. | [6] Mount D. W. (2004), "Bioinformatics: Sequence and Genome Analysis" 2nd ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. ISBN 0-87969-6087. | [7] Altschul, F., S., Madden, L., T., Schaffer, A., A., Zhang, J., Zhang, Z., Miller, W. and Lipman, J., D. (2003), "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," Nucleic Acids Research, vol. 25, no. 17, pp. 3389-3402. | [8] Allcock, W., Bresnahan, J., Kettimuthu, R., Link, M., Dumitrescu, C., Raicu, I., Foster, I. (2005), "The globus striped gridFTP framework and server, in: Proceedings of Super Computing 2005", SC05. | [9] Arndt, P., Burge, C., Hwa, T. (2003), "DNA sequence evolution with neighbor-dependent mutation" Journal of Computational Biology 10: 313-322. | [10] Needleman, S.B. & Wunsch, C. D. (1970), "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins", J.Mol. Biol., vol. 48, pp. 443-453. | [11] Smith, T. F. & Waterman, M. S. (1981), "Identification of common molecular subsequences", J. Mol. Biol., vol. 147, pp 195-197. | [12] Pongor, S. and Landsman, D. (1999), "Bioinformatics and the developing world", Biotechnology-and-Development-Monitor, No. 40, pp. 10-13. | [13] Rastogi, S. C., Mendiratta, N. and Rastogi, P. (2005), "Bioinformatics Methods and Applications", Microsoft White pp. 10. | [14] Zhang Y, Skolnick J. (2005), "The protein structure prediction problem could be solved using the current PDB library", Proc. Natl Acad Sci USA 102: 1029-34. | [15] Mabrouk, M., S., Hamdy, M., Mamdouh, M., Aboelfotoh, M. and Kadah, Y., M., (2006), "BIOINFTool: Bioinformatics and sequence data analysis in molecular biology using Matlab", proc. cairo international biomedical engineering conference. | [16] Li, Z., Zhang, S., Wang, Y., Zhang, Sun X. and Chen, L. (2007), "Alignment of molecular networks by integer quadratic programming" Vol. 23 ,no. 13 2007, pp. 16311639. | [17] Joshi, Y. and Vadhiyar, S. (2009), "Analysis of DNA sequence transformations on grids" J. Parallel Distrib. Comput. 69, pp. 80-90.

## Screen Shots

Figure 6 : Specifying the type of alignment that operate on the specified sequences



Figure 6.2: PAM Scores for the sequences



Figure 6.3: BLOSUM Scores for the sequences





**Sara Publishing Academy**  
Indian Journal Of Applied Research  
Journal for All Subjects



**Editor,**  
**Indian Journal Of Applied Research**  
8-A, Banans, Opp. SLU Girls College,  
New Congres Bhavan, Paldi, Ahmedabad-380006.  
Contact.: +91-9824097643 E-mail : [editor@ijar.in](mailto:editor@ijar.in)

Printed at Unique Offset, Novatsing Rupam Estate, Opp. Abhay Estate, Tavdipura, Shahibaug, Ahmedabad