World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:7, No:1, 2013

Structure Based Computational Analysis and Molecular Phylogeny of C- Phycocyanin Gene from the Selected Cyanobacteria

N. Reehana, A. Parveez Ahamed, D. Mubarak Ali, A. Suresh, R. Arvind Kumar, N. Thajuddin

***Abstract*—**Cyanobacteria play a vital role in the production of phycobiliproteins that includes phycocyanin and phycoerythrin pigments. Phycocyanin and related phycobiliproteins have wide variety of application that is used in the food, biotechnology and cosmetic industry because of their color, fluorescent and antioxidant properties. The present study is focused to understand the pigment at molecular level in the Cyanobacteria *Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06. After extraction of genomic DNA, the amplification of C-Phycocyanin gene was done with the suitable primer PCβF and PCαR and the sequencing was performed. Structural and Phylogenetic analysis was attained using the sequence to develop a molecular model.

***Keywords*—**Cyanobacteria, C-Phycocyanin gene, Phylogenetic analysis, Structural analysis.

I. Introduction

CYANOBACTERIA are prokaryotic microorganisms with the unique ability to fix atmospheric nitrogen. They obtain their name from the bluish pigment phycocyanin, which are helpful to capture light for photosynthesis. In some Cyanobacteria, the color of light influences the composition of phycobilisomes. In green light, the cells accumulate more phycoerythrin, whereas in red light they produce more phycocyanin. Thus the Cyanobacteria appear green in red light and red in green light which is a complementary chromatic adaptation [1].

Molecular methods have become an indispensable tool for the characterization of Cyanobacteria. Utilizing the bioinformatic tools, the development of genomics and molecular technologies combined together to obtain more information related to molecular biology of these organisms [2]. Molecular assessment of Cyanobacterial biodiversity frequently uses markers like 16S rDNA, phycocyanin locus, *nif* gene, *rpo* gene, ITS region, introns, STRR, RAPD, M13 etc. [3]-[12]. A BLAST (Basic Local Alignment Search Tool) search enables a researcher to compare a query sequence with a library above certain threshold. Molecular phylogenetics is used to gain information on an organism's evolutionary relationships based on the database of taxonomy and biogeography using MEGA software. Protein has three main structures in which primary structure is the linear amino acid sequence usually represented by a one letter notation. α - helices, β - sheets and loops are formed in secondary structure, when the sequences of primary structures tend to arrange themselves in regular conformations. Protein folding is the process that results in a compact structure in which secondary structure elements are packed against each other in a stable configuration. This three dimensional structure of the protein is known as the protein tertiary structure [13]. Molecular modelling is helpful to know the assumed structure of molecules and also used to investigate the dynamics and thermodynamics of inorganic, biological, and polymeric systems [14].

N. Reehana is with the Department of Microbiology, Bharathidasan University, Tiruchirappalli – 620024, Tamilnadu, India (e-mail: reehanan@yahoo.in).

A. Parveez Ahamed, D. Mubarak Ali, A. Suresh, and R. Arvind Kumar are with the Department of Microbiology, Bharathidasan University, Tiruchirappalli – 620024, Tamilnadu, India.

N. Thajuddin is with the Department of Microbiology, Bharathidasan University, Tiruchirappalli – 620024, Tamilnadu, India (Phone: 9842379719; e-mail: nthaju2002@yahoo.com).

II. Materials and Methods

A. Morphological Identification of Cyanobacteria

The Fresh and Marine water Cyanobacterial strains were obtained as axenic cultures from germplasm of Department of Microbiology, Bharathidasan University, Tiruchirappalli and were examined carefully using light microscope. The strains were maintained with alternative illumination (i.e. 16 hr light and 8 hr dark conditions) in germplasm at 25°C and exposed to 2000 Lux light intensity. The Morphological characters were determined with their cell shape and cell size.

B. Molecular Characterization of Cyanobacteria

Two Cyanobacterial cultures namely *Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06 were harvested at their exponential growth phase used for the extraction of genomic DNA [15].

C. C-Phycocyanin Gene Amplification by Polymerase Chain Reaction (PCR)

PCR amplification was performed for the respective two samples of purified DNA in a thermal cycler (MYGene^{TM}Series Peltier model MG96G) using the universal primers [PCβF - 5' GGCTGCTTGTTTACGCGACA 3' and PCαR - 5'CCAGTACCACCAGCAACTAA 3'] [6]**.** The Initial denaturation was achieved at 94^{o}C for 2 minutes and further denaturation was carried out at 94^{o}C for 5sec; annealing at 47^{o}C for 10sec, elongation at 72^{o}C for 30sec and a final elongation at 70^{o}C for 7min for 40 cycles. The

Amplified products were isolated by electrophoresis on 1.2% agarose gel using 1X TBE buffer and the bands were observed under gel documentation system (Photonyx).

D. Sequencing of C-Phycocyanin Gene

The sequencing of C-Phycocyanin gene was done by Ocimum Biosolutions, Hyderabad. Primer sequences were checked for homology to other sequences deposited in the available databases using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

E. Phylogenetic Analysis

Based on the sequenced data, the phylogenetic tree was constructed using bioinformatics tool Mega 5.05 [16], which has been used for aligning the sequences by Neighbor joining method (http://www.megasoftware.net/mega.php).

F.Primary Structure Prediction

The nucleotide sequence was converted to amino acid sequence using transeq software and the primary structure was predicted by online software ProtParam, a tool which allows the computation of various physical and chemical parameters (http://www.expasy.org/tools/protparam.html).

G. Secondary Structure Prediction

The translated protein sequences were analyzed for secondary structure prediction. The query sequence was uploaded in alignment box and the query was submitted to GOR secondary structure prediction method version IV for structural analysis. The structure was predicted and compared with their models (npsa-pbil.ibcp.fr/cgi-bin/npsa automat.pl?page=npsa gor4.html) [17].

H. Molecular Modelling

The protein was performed for homology modelling by MODELLER9V8 (http://www.salilab.org/modeller/). The constructed model was minimized by CHIMERA. The overall stereochemical properties of the proteins were analysed in the RAMPAGE server. The three dimensional structure were further verified by VERIFY3D. RMS-Z score for bond angles of modeled protein structure was estimated by QMEAN server. The models are viewed in CHIMERA.

III. Results and Discussion

A. Morphological Identification of Cyanobacteria

The *Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06 were identified based on their morphological characters with standard Cyanophyta Monograph [18].

B. Molecular Characterization of Cyanobacteria

The Extraction of genomic DNA for *Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06 were amplified for C-Phycocyanin gene using the primers PCβF and PCαR. The range of C- phycocyanin gene was 700 bp and the sequence similarity was analysed using the online tool BLAST (Fig. 1).

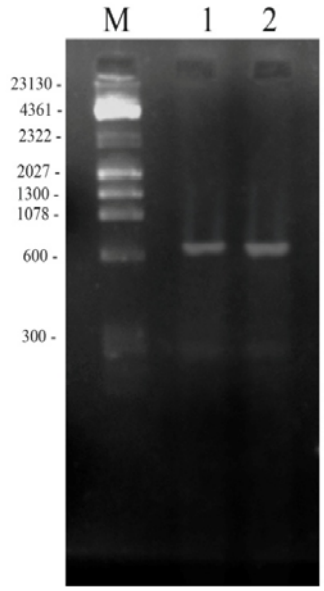

Fig. 1 Amplified Product of C - Phycocyanin gene from Cyanobacteria Lane 1: *Oscillatoria terebriformis* NTRI05 Lane 2: *Oscillatoria foreaui* NTRI06

The genetic material of an organism determines the character of a species. The application of DNA based amplification methods for the molecular typing of microorganisms in complex natural populations are making the study of Cyanobacterial systematics definitive rather than descriptive [2].

C. Phylogenetic Analysis

Phylogenetic analysis is the easiest way to depict any evolutionary relationship between the groups of organisms. The phylogenetic tree for the two isolates namely *Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06 were constructed using C-Phycocyanin gene sequence with that of other Cyanobacterial species by Neighbor Joining [19]. The sum of branch length of the tree is 1.90892857. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site (Fig. 2). The analysis involved 5 nucleotide sequences and codon positions included were $1^{st}+2^{nd}+3^{rd}$ +Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 70 positions in the final data set. Evolutionary analyses were conducted in MEGA5 [20].

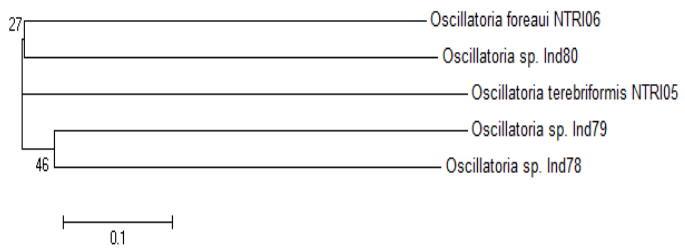

Fig. 2 Evolutionary Relationship of Taxa

D. Structural Analysis

The obtained nucleotide sequence for both isolates (*Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06) were analyzed for primary structure prediction. The physio-chemical parameters such as the molecular weight, isoelectric point (pI), extinction coefficient, half life, aliphatic index, amino acid property, instability index, Grand Average of Hydropathicity (GRAVY) were calculated using the Prot Param tool of the Expasy proteomics server.

Results showed that *Oscillatoria terebriformis* NTRI05 has 295 amino acid residues with the average molecular weight of 32909.6g mol $^{-1}$. Conversely, the *Oscillatoria foreaui* NTRI06 had an average molecular weight of 30007.1g mol $^{-1}$. The computed isoelectric point (PI) was above 7 indicating that the proteins are useful for developing buffer systems for purification by isoelectric focusing method. The total number of negatively charged residues (Asp+ Glu) was 25 and positively charged residues (Arg+ Lys) was 45 for *Oscillatoria terebriformis* NTRI05 where as for *Oscillatoria foreaui* NTRI06 (Asp+ Glu) was 18 and (Arg+ Lys) was 27. The Extinction coefficient of *Oscillatoria terebriformis* NTRI05 at 280 nm is 25035M^{-1} cm^{-1} and for *Oscillatoria foreaui* NTRI06 32860M^{-1} cm^{-1} with respect to the concentration of Cys residues. According to the Prot Param server, a protein whose instability index is larger than 40 may be unstable [21] therefore, the program predicted for both Cyanobacteria would be unstable in solution. The aliphatic index for the *Oscillatoria terebriformis* NTRI05 was 69.05 and for *Oscillatoria foreaui* NTRI06 90.63, which is regarded as a positive factor for the increase of thermal stability of globular proteins. The GRAVY value for both ranged from -0.464 to -0.068 indicating that the proteins will interact favorably with water (Table I).

TABLE I
PRIMARY STRUCTURE OF *OSCILLATORIA TEREBRIFORMIS* NTRI05 & *OSCILLATORIA FOREAUI* NTRI06

Physicochemical parameters	*Oscillatoria terebriformis* NTRI05	*Oscillatoria foreaui* NTRI06
Length (total no. of amino acids)	295	268
Isoelectric point (PI)	9.68	8.83
Molecular weight (MW)	32909.6	30007.1
Negatively charged residues (-R)	25	18
Positively charged residues (+R)	45	27
Extinction coefficient (EC)	25035	32860
Instability index (II)	41.96	48.16
Aliphatic index (AI)	69.05	90.63
Grand average of hydropathicity (GRAVY)	-0.464	-0.068

Secondary structure predicts the structure of proteins and nucleic acid sequences based only on the knowledge of their primary structure. The secondary structure helps in determining the exact structure of the gene which was predicted using GOR IV software. The alpha helix and beta sheet are made up of two networks: a sequence-to-structure network and a structure-to-structure network. The results of *Oscillatoria terebriformis* NTRI05 revealed that random coil (61.36%) dominated among secondary structure elements and alpha helices (20.00%) and extended strand (18.64%) are presented (Fig. 3). The results of *Oscillatoria foreaui* NTRI06 revealed that random coil (59.70%) dominated among secondary structure elements and alpha helices (14.55%) and extended strand (25.75%) are presented (Fig. 4). The secondary structure was employed to improve the sequence alignment. The secondary structures are more conserved than the nucleotide sequences; their analysis helps in understanding molecular evolution and increases the number of structural characters.

View GOR4:

```
        10        20        30        40        50        60        70
        |         |         |         |         |         |         |
YSPWPPFCVYRRRCECYRCGRKFGLSSLPNPSNMGSSVGISKSENQEELRQALKLAFRYDSRVLVEQGVN
cccccccceeeccccceecccccccccccccccccceeeeccchhhhhhhhhhhhhhhhcchhhhhcccc
AREIEVGLLGNYDVKSTLPGEVVKDVAFYDYDAKYIDNKITMDIPAKINDDVVAVMRQNAETAFRAIGGL
hhhhheeecccccccccccccccceeeccccceeecccccccccccchhhhhhhhhhhhhhhhhccc
GLSRCDFFLYRGERFFSTSSLPLAWFHPLCFPHHNKKNPGLKKFFCHPPKNGEAPQTLLKNGGGRTNNHD
cccccceeeccccceeeccccccccccccccccccccceeecccccccccceeeeecccceeecccc
GGVSTTPGGTHGKLHVRTRPKRNFFARGRPTHKILSEVASLFSNKAASACCVVGPDGTQQKICRRKDRSV
cceeeccccccceeeeccccccceecccccchhhhhhhhhhccccceeeeecccccchhhhhccccc
SLKPTGLIFCNFGVR
cccccceeecceeec
```

Sequence length: 295

GOR4:

Alpha helix	(Hh) :	59 is	20.00%
3_{10} helix	(Gg) :	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	55 is	18.64%
Beta turn	(Tt) :	0 is	0.00%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	181 is	61.36%
Ambigous states (?)	:	0 is	0.00%
Other states	:	0 is	0.00%

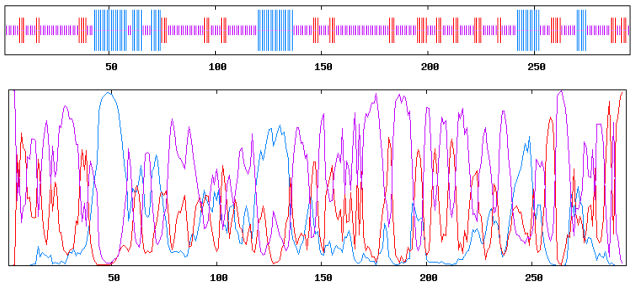

Fig. 3 GOR4 result for Secondary Structure of *Oscillatoria terebriformis* NTRI05

View GOR4:

```
        10        20        30        40        50        60        70
        |         |         |         |         |         |         |
EKRDNTPFMLMLSEVMCVLISEVEPFKFTSLLNPPHWVLCRFSKPEHEKSRRPNLCPHMTPRLGEQGVNA
ccccccccceeeeeeceeeeecccccccceeeeecccccceeeecccccccccccccccccccccccccc
PKIEVGLLGHLHVKNTPPKICQGRCLLSTTCLIYDINYYMIYSCPISQCGAVCAQKCKNTLPPLRPGAIS
cchhhhhhccccccccccccccccccccccceeecccceeeeeecceecccceecccccccccccccccc
CAISLHLRKEIFLIKLITLPGCPILSHLYLIRSPYAAESRGKVTHQNAGLYTLRGLLVADNQELYVTART
hhhhhhhhhhhhhhhhcccccccccccccccccccccccccceecccccchhhhhhhhhccccceeeeeeee
YYYAESCSQPCSCLYSSSYTRHADNYSTITRMIACSLASTAQLRLSSITDWSDQASGX
eeeecccccceeeeecccccccccceeeeeeeeecccccchhhhhhhccceeeeeeeeec
```

Sequence length: 268

GOR4:

Alpha helix	(Hh):	39 is	14.55%
3_{10} helix	(Gg):	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	69 is	25.75%
Beta turn	(Tt) :	0 is	0.00%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	160 is	59.70%
Ambigous states	(?) :	0 is	0.00%
Other states	:	0 is	0.00%

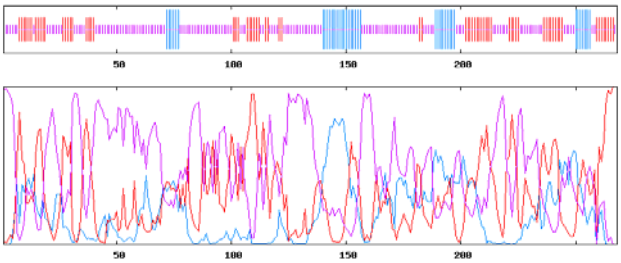

Fig. 4 GOR4 result for Secondary Structure of *Oscillatoria foreaui* NTRI06

The modelling of the three dimensional structure of the protein was performed by MODELLER9V8 for both the isolates *Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06 (Figs. 5 & 6). Molecular modelling methods are now routinely used to investigate the structure, dynamics, surface properties and thermodynamics of inorganic, biological and polymeric systems [4].

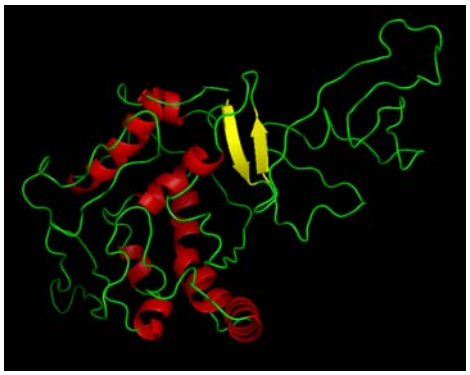

Fig. 5 Three dimensional structure of C-Phycocyanin gene - *Oscillatoria terebriformis* NTRIO5

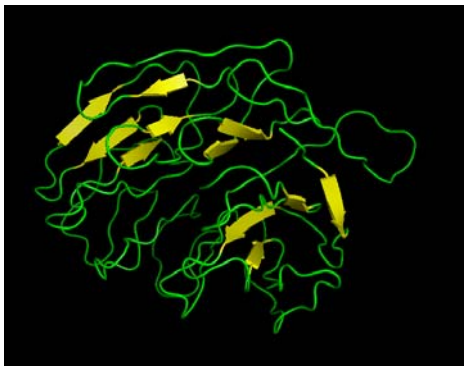

Fig. 6 Three dimensional structure of C-Phycocyanin gene - *Oscillatoria foreaui* NTRIO6

IV. Conclusion

In this study, C-Phycocyanin gene was amplified and sequenced from the Cyanobacterial strains, *Oscillatoria terebriformis* NTRIO5 and *Oscillatoria foreaui* NTRIO6. Phylogenetic analysis, primary & secondary structure analysis and modelling of three dimensional structure of the protein were also performed. Our future approach will step forward towards barcoding for taxonomic application.

Acknowledgment

The first author (N. Reehana) is thankful to UGC-MANF (Maulana Azad National Fellowship) for their financial assistance.

References

[1] N.T. Eriksen, "Production of phycocyanin – a pigment with applications in biology biotechnology foods and medicine," Applied Microbiology and Biotechnology, 2008, Vol. 80, pp. 1-14.

[2] W. Mount David, "Bioinformatics: Sequence and Genome Analysis," Spring Harbor Press, 2002, ISBN 0-87969-608-7.

[3] N. Thajuddin, and G. Subramanian, "Cyanobacterial Biodiversity and potential application in Biotechnology," Current Science., 2005, Vol. 89(1), pp. 47-57.

[4] N. Thajuddin, A. Sandra, Nierzwicki-Bauer, and G. Subramanian, "Cyanobacterial biodiversity and recent trends in their taxonomy," BDU J. Sci. & Tech., 2007, Vol. 1, pp. 1-14. (in Tamil), 15-28 (in English).

[5] N. Thajuddin, G. Muralitharan, G. Sundaramoorthy, R. Ramamoorthy, S. Ramachandran, M. A. Akbarsha and M. Gunasekaran, "Morphological and genetic diversity of symbiotic Cyanobacteria from cycads," J. Basic Microbiol., 2010, Vol. 50, pp. 254-265.

[6] N. Thajuddin, and G. Muralitharan, "Applications of PCR based fingerprinting in the phylogeny of marine Cyanobacteria," Indian Hydrobiol., 2008, Vol. 11(1), pp. 25-41.

[7] G. Muralitharan, and N. Thajuddin, "Evidence on the presence of $tRNA^{fmet}$ group I introns in the marine cyanobacterium *Synechococcus elongatus*," Journal of Microbiology and Biotechnology., 2008, Vol.18 (1), pp. 23-27.

[8] G. Muralitharan and N. Thajuddin, "M13 based genotyping of marine cyanobacterial strains from the Indian Subcontinent and maintained in the NFMC germplasm collection," Journal of Applied Phycology., 2010, Vol.22, pp. 709-716.

[9] G. Muralitharan, and N. Thajuddin, "Rapid differentiation of phenotypically and genotypically similar *Synechococcus elongatus* strains by PCR fingerprinting," Biologia, 2011, Vol. 66/2, pp. 238-243.

[10] D. Pandiaraj, D. Mubarak ali. R. Praveenkumar, S. Ravikumar and N. Thajuddin, "Molecular characterization and phylogeny of marine Cyanobacteria from Palk Bay Region of Tamil Nadu, India," Ecologia, 2012, Vol.2, pp. 23-30.

[11] D. Mubarak ali, J. Arunkumar, K. R. Suriya, K. A. Sheik syed ishack and N. Thajuddin, "Molecular modeling and Phylogenetic analysis of c-phycocyanin gene sequence from marine cyanobacterium *Phormidium tenue* NTDM05," Seaweed Res. Utiln., 2012, Vol. 34(1&2) pp.35-44.
[12] N. Kumari, A. K. Srivastava, P. Bhargava, and L. Rai, "Molecular approaches towards assessment of cyanobacterial biodiversity," African Journal of Biotechnology., 2009, Vol. 8(18), pp. 4284-4298.
[13] J.C. Kendrew, R.E. Dickerson, B.E. Strandberg, R.G. Hart, and D.R. Davies, "Structure of Myoglobin," Nature, 1960, Vol. 185, pp.422-427.
[14] M.J. Foster, "Molecular Modelling on Structural Biology," Micron, 2002, Vol. 33, pp.365 -384.
[15] J.A. Smoker, and S.R. Barnum, "Rapid small-scale DNA isolation from filamentous cyanobacteria," FEMS Microbiology Letters., 1988, Vol. 56(1), pp.119 – 122.
[16] B.A. Neilan, "The Molecular Evolution and DNA Profiling of Toxic Cyanobacteria," Curr. Issues Mol. Biol., 2002, Vol. 4, pp.1-11.
[17] J. Garnier, J. F. Gibrat, and B. Robson, "GOR Secondary structure prediction method version IV," 1996, Vol. 266, pp. 540-553.
[18] T. V. Desikachary, "Cyanophyta Indian Council of Agricultural Research New Delhi, India," 1959.
[19] N. Saitou, and M. Nei, "The Neighbor-joining method: A new method for reconstructing phylogenetic trees," Molecular Biology and Evolution, 1987, Vol.4, pp. 406-425.
[20] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods," Molecular Biology and Evolution., 2011, Vol. 28(10), pp.2731-2739.
[21] K. Guruprasad, B.V. Reddy, M.W. Pandit, "Correlation Between Stability of a Protein and its Dipeptide Composition: A Novel Approach For Predicting In Vivo Stability of a Protein From Its Primary Sequence," Protein Eng., 1990, Vol. 4(2), pp.155-61.