Journal of Proteomics & Bioinformatics - Open Access
www.omicsonline.com Research Article JPB/Vol.1/May 2008

Exploring the Interplay of Sequence and Structural Features in Determining the Flexibility of AGC Kinase Protein Family : A Bioinformatics Approach

Amit Kumar Banerjee, Neelima Arora, Varakantham Pranitha, U.S.N.Murty*,

Bioinformatics Group, Biology Division, Indian Institute of Chemical Technology, Hyderabad-500607, A.P., India

*Corresponding author: Dr. U.S.N Murty, Deputy Director/ Scientist "F" Head, Biology Division, Indian Institute of Chemical Technology, Hyderabad-500607, India
E-mail: murty_usn@yahoo.com; Tel: +91 40 27193134; Fax: +91 40 27193227

Received May 02, 2008; Accepted May 15, 2008; Published May 20, 2008

Citation: Amit KB, Neelima A, Varakantham P, Murty USN (2008) Exploring the Interplay of Sequence and Structural Features in Determining the Flexibility of AGC Kinase Protein Family : A Bioinformatics Approach. J Proteomics Bioinform 1: 077-089. doi:10.4172/jpb.1000013

Copyright: © 2008 Amit KB, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

In this study, data mining approach was used to generate association rules for predicting average flexibility from the various derived sequence and structural features. 21 parameters were calculated and their variable importance was calculated for 115 sequences of AGC kinase family belonging to mouse and human using Classification and Regression Tree (CART). Beta turns were found to have maximum influence on average flexibility while the total beta strands were found to exert minimum impact on average flexibility. Understanding the variable importance will prove useful as a simple pr edictor of flexibility from an amino acid sequence. This will aid in better understanding of phenomenon underlying the average flexibility and thus, will pave a way for rational design of therapeutics.

Keywords: AGC kinase; Protein flexibility; Data mining; Classification and regression tree (CART); Bioinformatics

Introduction

Every biological molecule is characterized and set apart from other biomolecules by a definite set of inherent intrinsic properties. Being the determinant of some vital functions like transport of metabolites (Anderson et al., 1990; Spurlino et al., 1991), catalysis (Bennett and Steitz, 1978; Remington et al., 1982) and regulation of protein activity (Perutz, 1970; Perutz, 1989) etc, average flexibility holds prime importance. Eukaryotic proteins demonstrate higher flexibility which influence conformational ability required in important biological processes like molecular recognition, interaction, assembly and modification. Moreover, protein flexibility is also known to influence stability and folding. There has been a sudden spur of interest in studies related to flexibility of proteins owing to discovery of role of some highly flexible proteins with implications in life threatening diseases like AIDS (HIV gp41) and scrapie (Chan et al., 1997). A comprehensive knowledge of fundamental nature of average flexibility will facilitate the unraveling of structure-function relationship and will also aid in development of novel therapeutics (Teague, 2003).

AGC protein kinase family, one among the eight ePK families defined in the Kinbase, includes many important enzymes such as cyclic nucleotide and calcium-phospholipid dependent kinases, ribosomal S6-phosphorylating kinases, G protein-coupled kinases, and few others. The AGC serine threonine kinases, known for phosphorylating sites surrounded by basic amino acids, are involved in many intra–cellular signaling pathways, critical cellular processes and control cell growth, differentiation and cell survival. Their crucial role in transmembrane signaling process hints on the importance of features of AGC kinases which may be responsible for membrane localization (Peterson and Schreiber, 1999). This group of protein kinases shares similarity within the catalytic domain and is characterized by similar mechanism of activation. Deregulation of AGC kinases is known to have implications in several diseases like Cancer, Diabetes, neurodegeneration, and thus, AGC kinases represent several attractive targets for small inhibitors of therapeutic significance (Breitenlechner et al., 2004).

Their stringent spatio-temporal regulation is attained through loop phosphorylation and repositioning of the key catalytic and substrate binding regions which indicates the importance of flexibility in these proteins (Kannan et al., 2007). There is preponderance of literature on flexibility of proteins but elucidating the effect of parameters influencing it is cumbersome. This study aims at exploring the importance of different parameters influencing the average flexibility of AGC kinase family using data mining approach.

Materials and Methods

Sequence Collection and Pre-Processing

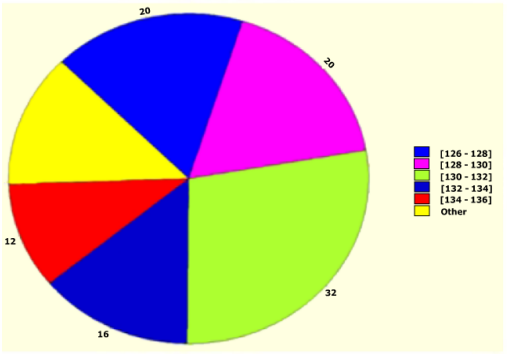
Protein sequences of the enzymes belonging to AGC family of protein kinase super family in FASTA format were collected from the non redundant (NR) protein database of NCBI (http://www.ncbi.nlm.nih.gov). Partial sequences were excluded from the study and sequences were again put to manual filtering so as to minimize the redundancy. This approach resulted in 600 sequences from the total 1259 sequences of AGC family available in the database were obtained. Out of these, sequences belong ing to *Homo sapiens* (59) and *Mus musculus* (56) were considered for this study.

Parameter	Mean	Standard Deviation	Skewness	Coefficient of variation	Variance	Kurtosis	Standard Error Mean
Accessible residues	5.8171	0.42102	5.0288	0.072376	0.17725	40.439	0.03926
Buried Residues	5.7892	0.72877	-4.2973	0.12588	0.5311	25.436	0.067958
Amino acid composition	5.786	0.19749	-0.034656	0.034133	0.039003	-0.15092	0.018416
Alpha helix	1.0192	0.031284	1.4608	0.030695	0.0009787	9.3437	0.0029173
Beta sheet	0.97093	0.025983	-0.20939	0.026761	0.00067513	1.077	0.0024229
Beta turn	1.02	0.027913	-0.11458	0.027365	0.00077915	-0.24003	0.0026029
Coils	1.0387	0.0309	0.39441	0.029749	0.00095484	-0.40818	0.0028815
Parallel Beta strands	1.0625	0.050085	0.045298	0.047139	0.0025085	-0.18584	0.0046704
Anti parallel beta strands	0.9799	0.033513	-0.38504	0.034201	0.0011231	-0.11993	0.0031251

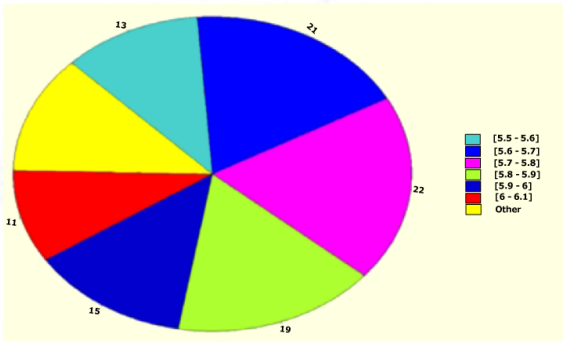
Trans-membrane Tendency	-0.5891	0.27052	5.5183	-0.45921	0.07318	45.421	0.02 5226
Total Beta strands	0.98868	0.030955	-0.56077	0.031309	0.00095818	0.31456	0.0028865
Relative mutability	76.674	2.9206	-0.085732	0.038091	8.53	-0.19263	0.27235
Refractivity	16.212	1.3109	0.12774	0.080856	1.7184	0.27699	0.12224
Recognition Factors	88.918	1.4693	0.43693	0.016525	2.159	-0.42356	0.13702
Polarity	19.936	1.9885	0.2598	0.099744	3.954	-0.022502	0.18543
Number of Codons	3.572	0.24312	-2.0097	0.068063	0.059107	11.473	0.022671
Molecular weight	130.19	3.7174	-0.33221	0.028553	13.819	1.203	0.34665
Hydrophobicity	-0.41118	0.35214	2.7344	-0.8564	0.124	15.724	0.032837
Bulkiness	14.261	1.1952	-6.3417	0.083806	1.4284	54.206	0.11145
Average Area buried	124.92	7.8686	-6.3319	0.062987	61.915	55.828	0.73375
Average Flexibility	0.44019	0.0060555	-0.11045	0.013757	3.6669e-005	0.57539	0.00056468

Table 1: Basic statistical features of parameters considered in the study.

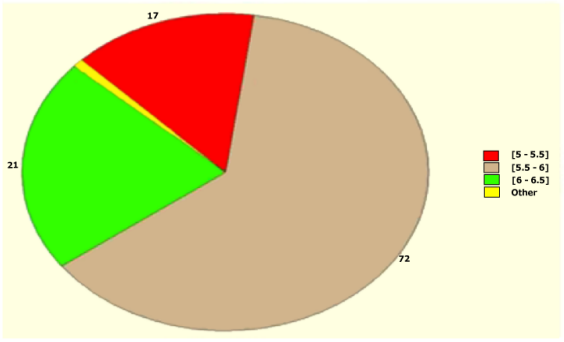
Molecular Weight



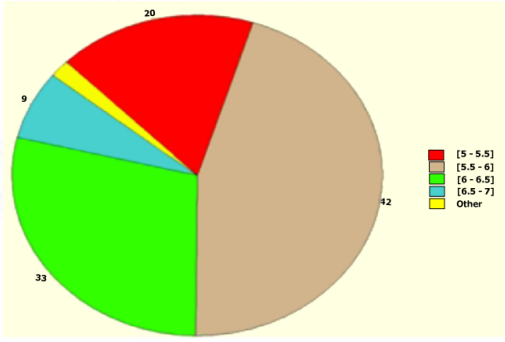
Amino Acid Composition



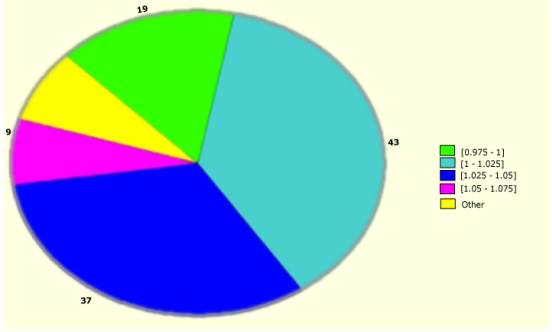
% Accessible Residues



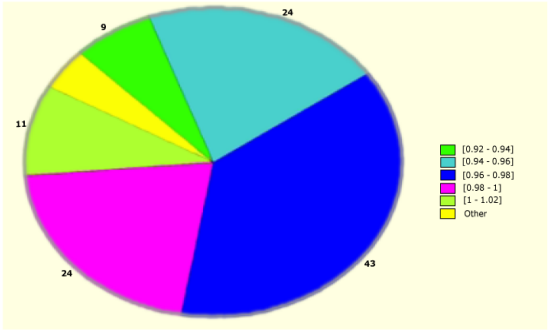
% Buried residues



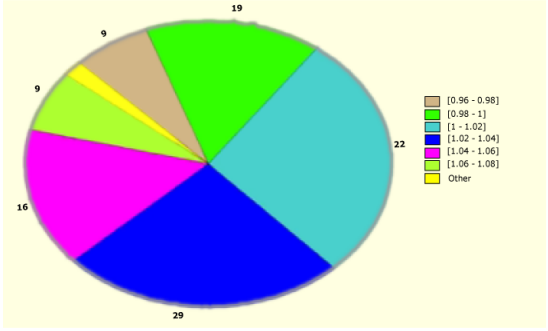
Alpha Helix



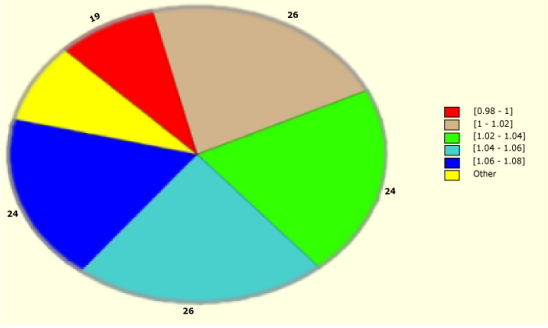
Beta sheet



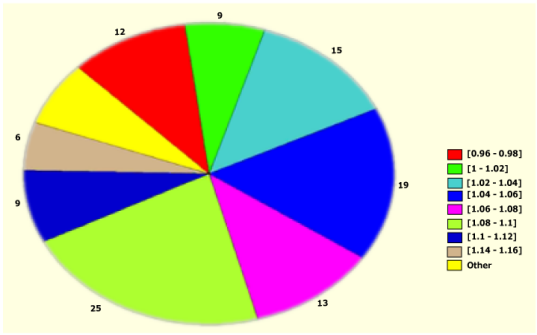
Beta turn



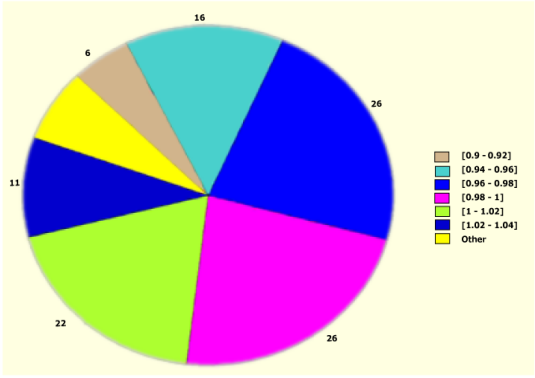
Coils



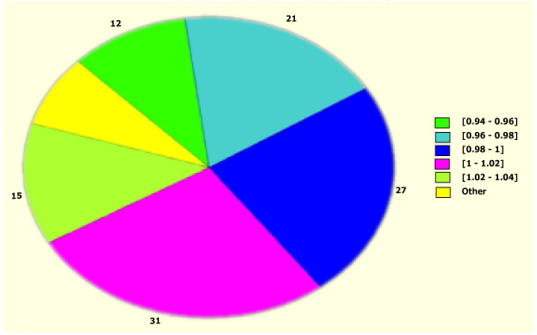
Parallel Beta strands



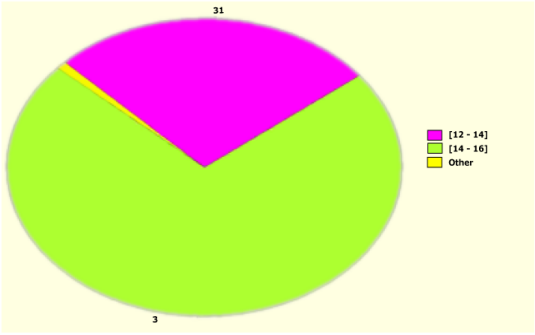
Anti-parallel Beta strands



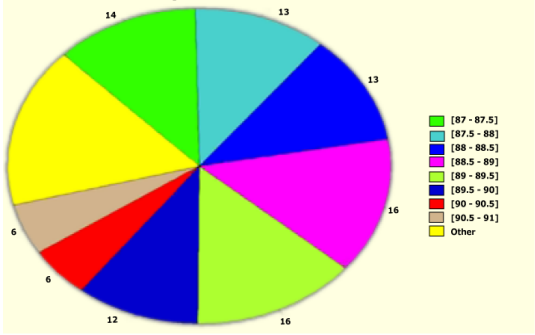
Total beta strands



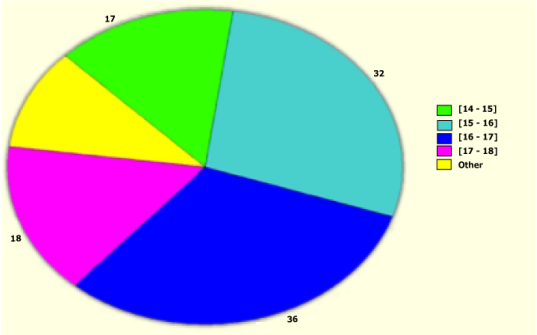
Bulkiness



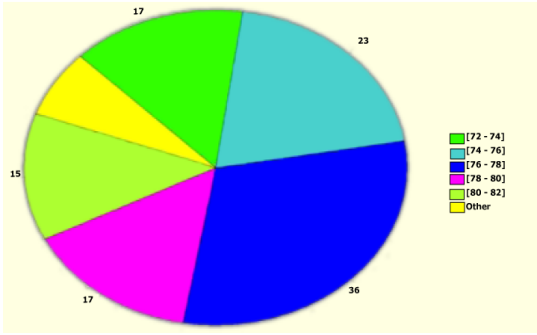
Recognition factors



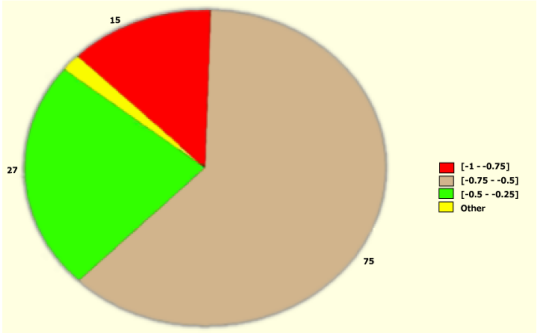
Refractivity



Relative Mutability



Trans-membrane tendency



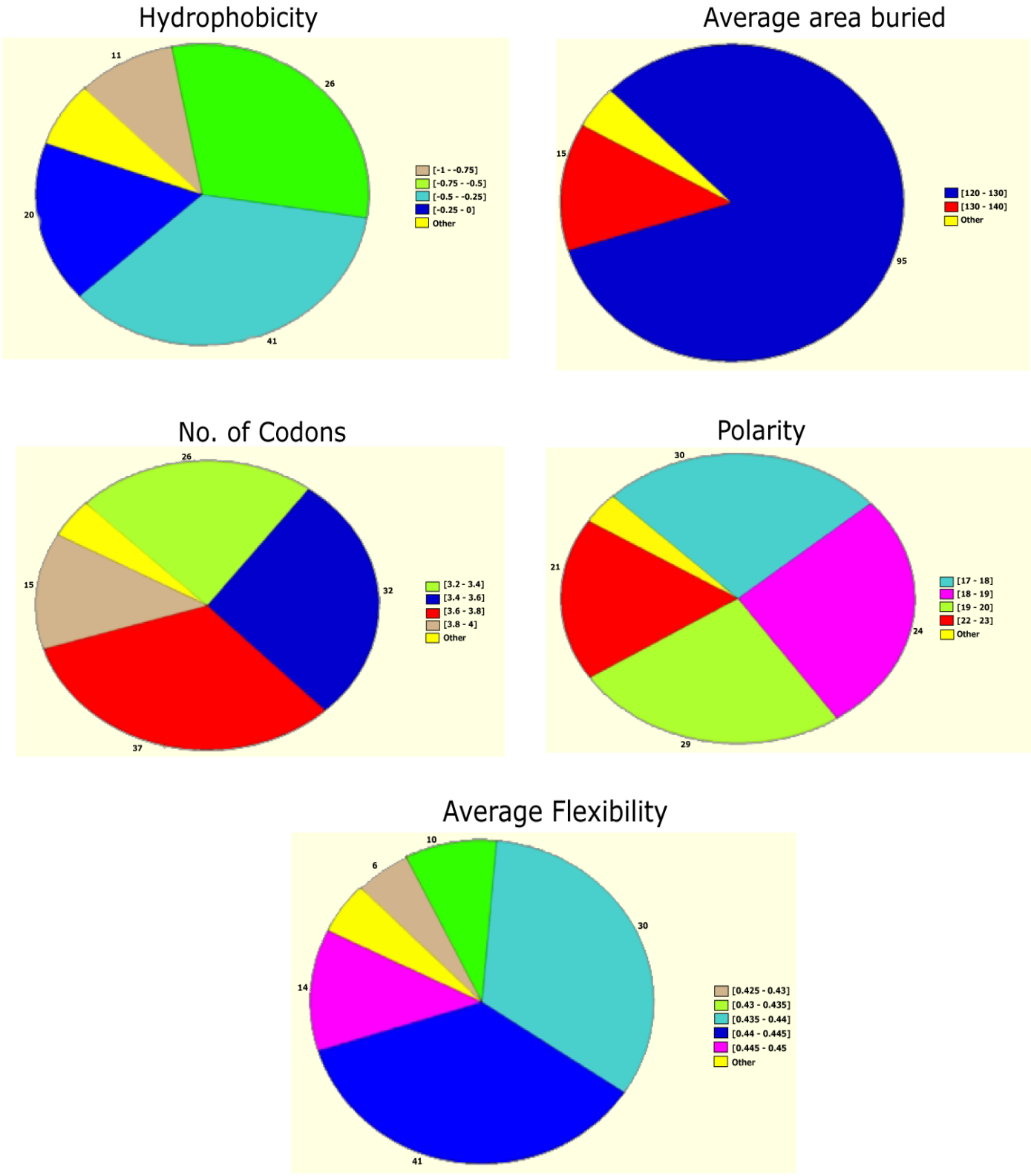
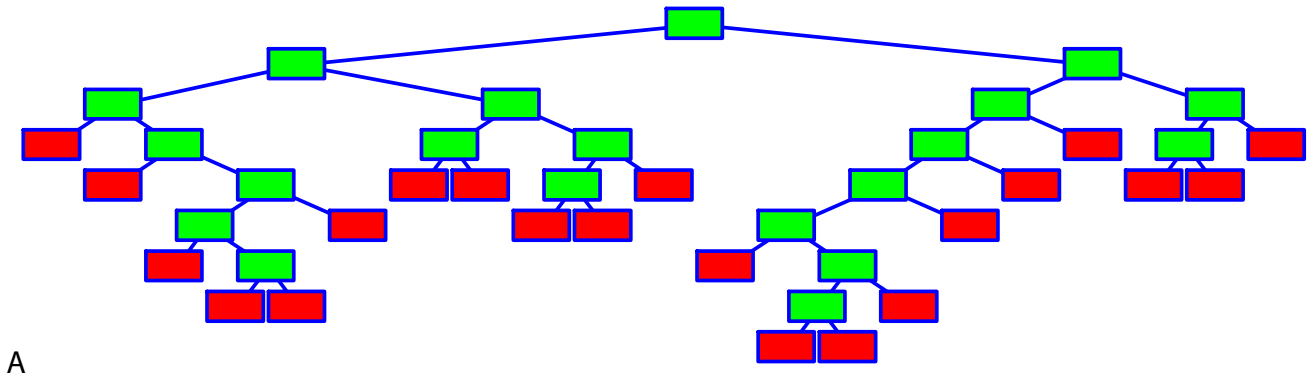


Figure 1: Frequency distribution chart for different parameters generated in CART 14 trees with different complexities and error values obtained using CART based on plitting criteria are reflected in table 2. Out of these trees, tree with 21 terminal nodes with minimum complexity and re-substitution relative error of 0.08501 and cross validated error of 0.72543 ± 0.12560 generated by Least Square splitting criteria was selected for generating decision rules. The topology of tree and error rate is represented in Figure 2. Splitters for the regression tree are shown in Figure 3. Decision rules obtained using CART are summarized in table 3(Supplement).

Tree No.	Terminal Nodes	Cross-Validated Error	Resubstitution Relative Error	Complexity
1	21	0.72543 ± 0.12560	0.08501	0.00000
2	20	0.71808 ± 0.12370	0.08653	1.00000E-005
3	19	0.71000 ± 0.11971	0.08899	0.00002
4	15	0.67935 ± 0.11594	0.11571	0.00003
5	13	0.66759 ± 0.11029	0.14635	0.00007
6	11	0.66746 ± 0.11162	0.18358	0.00008
7	9	0.65670 ± 0.11209	0.22481	0.00009
8	8	0.57881 ± 0.09948	0.25020	0.00012
9	6	0.60897 ± 0.08204	0.35804	0.00023
10	5	0.66411 ± 0.09268	0.41964	0.00027
11	4	0.89325 ± 0.08412	0.52601	0.00045
12	3	0.92470 ± 0.08126	0.65254	0.00054
13	2	0.91504 ± 0.07452	0.78894	0.00058
14	1	1.00139 ± 0.00159	1.00000	0.00089

Table 2: Details of trees generated in CART along with relative error and complexities



A

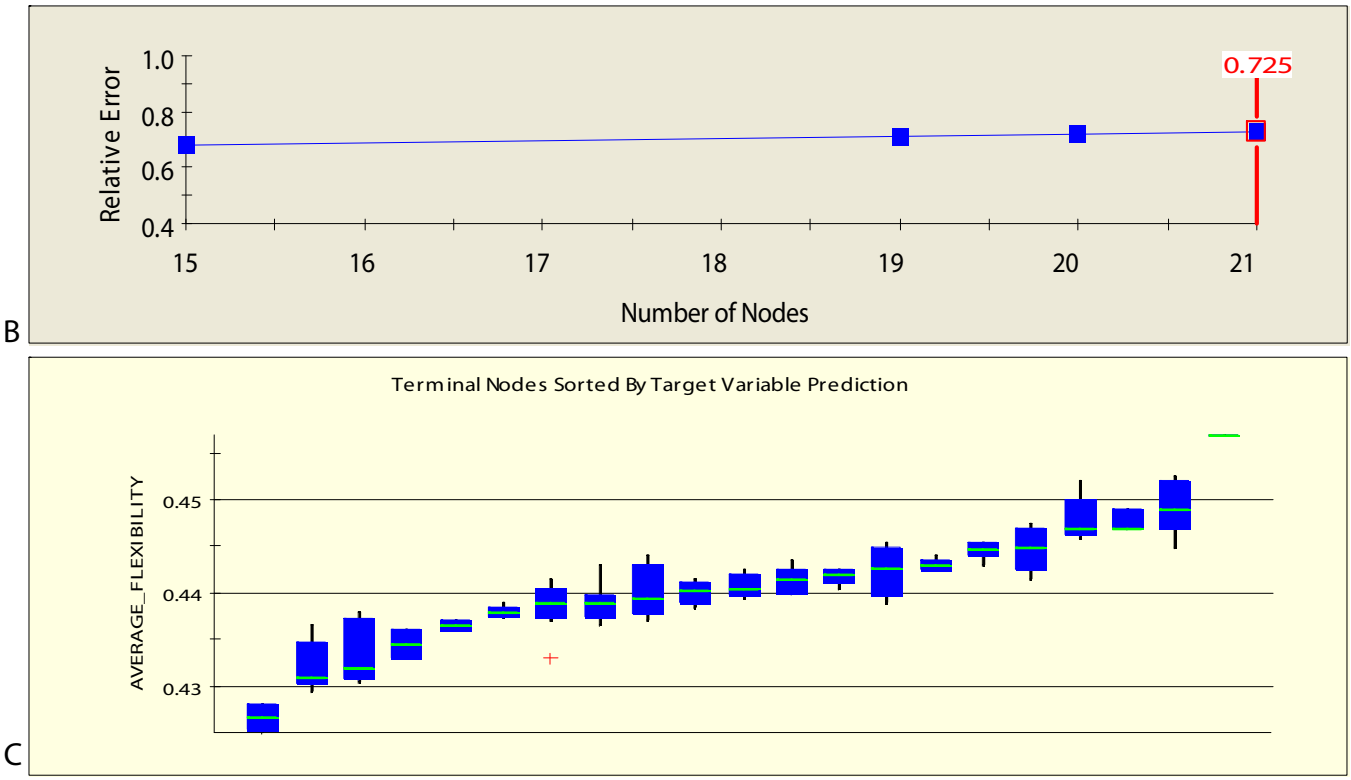


Figure 2: The tree sequence of lowest complexity which yielded 21 terminal nodes (A) with the cross validation error rate (B) and terminal node box plot(C).

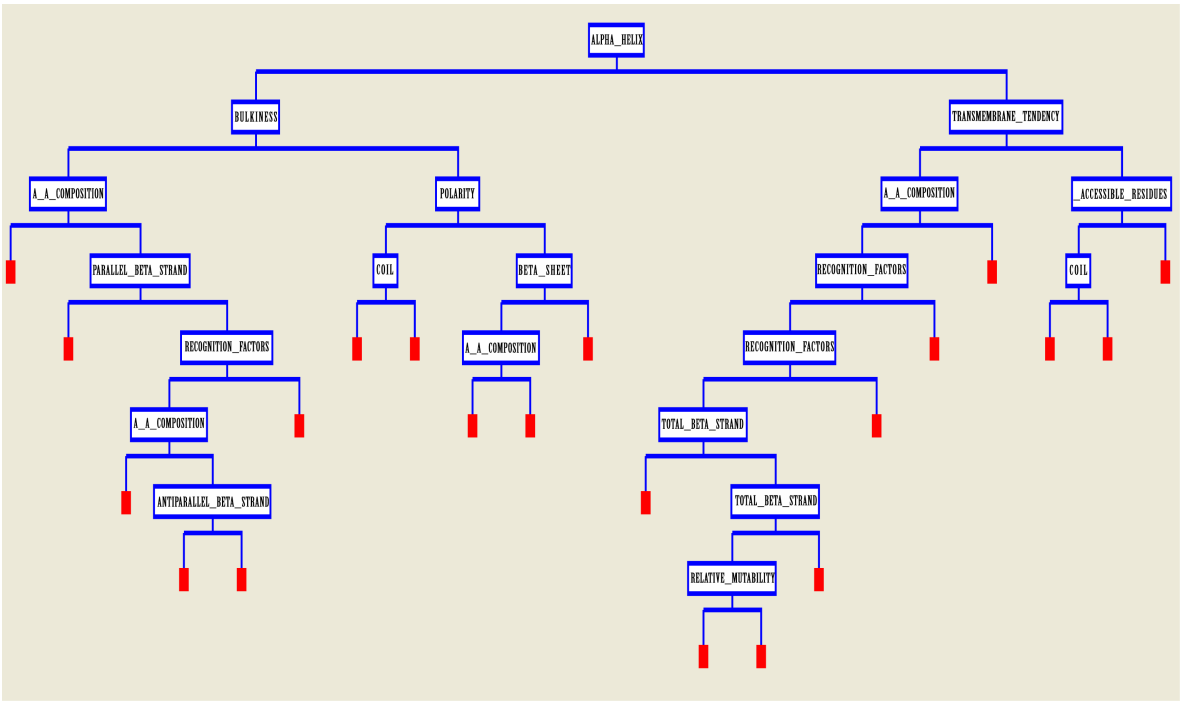


Figure 3: Details of splitter for the Decision tree

Rules derived from CART can be interpreted in simple context of "If "and "Then" based statement and thus are self-explanatory.

For example: Rule 1 can be interpreted as

Rule 1: IF "BULKINESS <= 14.2207" & "ALPHA -HELIX <= 1.01975" &" A.A COMPOSITION <= 5.55", THEN "AVERAGE FLEXIBILITY=0.457".

Rule 14: IF "RECOGNITION FACTORS<= 89.4723" & "TRANSMEMBRANE TENDENCY<= -54225" & "ALPHA -HELIX > 1.01975" & "TOTAL BETA-STRAND> 0.95975 &<= 1.018" & "A.A. Composition<= 6.0055" & "RELATIVE MUTABILITY<= 80.0835", THEN "AVERAGE FLEXIBILITY=0.436563".

Variable importance

Importance of different variables was calculated based on predefined scores in CART and summarized in Table 4.

S. No.	VARIABLE	IMPORTANCE
1.	BETA-TURN (CHOU & FASMAN)	100.00
2.	% ACCESSIBLE RESIDUES	93.57
3.	ALPHA HELIX (CHOU & FASMAN)	86.18
4.	TRANSMEMBRANE TENDENCY	78.43
5.	AMINOACID COMPOSITION	71.15
6.	BULKINESS	55.69
7.	COIL (DELEAGE & ROUX)	50.69
8.	PARALLEL BETA-STRAND	50.03
9.	RECOGNITION FACTORS	49.06
10.	MOLECULAR WEIGHT	34.84
11.	POLARITY (ZIMMERMAN)	33.05
12.	HYDROPHOBICITY (KYTE & DOOLITTLE)	32.08
13.	AVERAGE AREA BURIED	29.71
14.	REFRACTIVITY	29.16
15.	BETA SHEET (CHOU & FASMAN)	27.81
16.	NUMBER OF CODONS	21.31
17.	%BURIED RESIDUES	17.72
18.	RELATIVE MUTABILITY	2.37
19.	TOTAL BETA STRAND	1.14
20	ANTI-PARALLEL BETA STRAND	0

Table 4: Variable importance of parameters influencing average flexibility.

Discussion

Dynamic nature of proteins, conferred by their structural flexibility, is associated with function. Average flexibility, an innate property of proteins is being recognized with implications in many important physiological processes recently (Wright and Dyson, 1999; Bright et al., 2001; Dunker et al., 2001; Namba, 2001). Recognition of several highly flexibile proteins in some pathological conditions have led to the momentum in studies related to the flexibility of proteins. The huge gap in number of sequence and structures in PDB limits the utilization of 3-dimensional structure for deriving features affecting flexibility like B-factors. In unavailability of such data, sequence composition and secondary structure provides a rough estimation of structural properties. This warrants the need for an alternate and simplistic approach for determining the effect of various parameters on average flexibility in an easy to understand quantitative relationship. Data mining approaches based on decision tree based methods have been successfully exploited in elucidating importance of features affecting important biological processes (Banerjee et al., 2007). CART has been exploited in microarray studies (Boulesteix et al., 2003), ecological studies (De'ath and Fabricius, 2000), risk prediction (Gottschalk et al., 1998), diseases diagnosis (Hermanek and Holzmann, 1994) and social studies (Özge et al., 2006).

The dataset comprising of various derived features was used to elucidate decision rules by CART that can serve as rule of thumb for finding the effect of different parameters on average flexibility, which is virtually impossible to calculate in a lab simultaneously using conventional approaches. Among the secondary structure features, beta turn, alpha helix, coil, parallel beta strand, beta sheet and total beta strands were found to influence the average flexibility in descending order. Among sequence features, % accessible residues, trans-membrane tendency, amino acid composition, bulkiness, recognition factors, molecular weight, polarity, hydrophobicity, average area buried, refractivity, no. of codons, % buried residues, and relative mutability were observed to affect the average flexibility in decreasing order(Table 4). Beta turns were found to have maximum impact while total beta strand were found to have minimum effect on average flexibility of the proteins considered in the study. As more and more studies are advocating the inclusion of protein flexibility in docking algorithms, it will be interesting to gain an insight on features influencing the flexibility of proteins. It is speculated that an extensive knowledge of protein flexibility and the various parameters contributing towards is important for rational drug design. Such an approach will lead to better understanding of underlying biological phenomena and aid in enzyme engineering processes.

Acknowledgement

Authors thank Dr. J.S.Yadav, Director, IICT for his continuous support and encouragement. We thank anonymous reviewers for their critical suggestions for the improvement of the manuscript.

References

1. Anderson BF, Baker HM, Morris GE, Rumball SV, Baker EN (1990) Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins. Nature 344: 784–787. » CrossRef » Pubmed » Google Scholar
2. Banerjee AK, Arora N, Murty USN (2007) Stability of ITS2 Secondary Structure in *Anopheles*: What Lies Beneath? International Journal of Integrative Biology 3: 232-238. » CrossRef » Pubmed » Google Scholar
3. Bennett WS Jr, Steitz TA (1978) Glucose-induced conformational change in yeast hexokinase. Proc Natl Acad Sci USA 75: 4848–4852. » CrossRef » Pubmed » Google Scholar
4. Bhaskaran R, Ponnuswamy PK (1988) Positional flexibilities of amino. acid residues in globular proteins. Int J Pept Prot Res 32: 242-255. » CrossRef » Google Scholar
5. Boulesteix AL, Tutz G, Strimmer K (2003) A CART-based approach to discover emerging patterns in microarray data. Bioinformatics 19: 2465-2472. » CrossRef » Pubmed » Google Scholar
6. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Chapman & Hall New York NY.
7. Breitenlechner C, Gaßel M, Engh R, Bossemeyer D (2004) Structural Insights Into AGC Kinase Inhibition. Oncol Res 14: 267-278.
8. Bright JN, Woolf TB, Hoh JH (2001) Predicting properties of intrinsically unstructured proteins. Prog Biophys Mol Biol 76: 131–173. » CrossRef » Pubmed » Google Scholar
9. Chan DC, Fass D, Berger JM, Kim PS (1997) Core structure of gp41 from the HIV envelope glycoprotein. Cell 89: 263–273. » CrossRef » Pubmed
10. Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol 47: 45-148. » Pubmed » Google Scholar
11. Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in protein; in: M.O. Dayhoff (Ed.), Atlas of Protein Sequence and Structure, Washington DC 3: 345–352.
12. Death G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis, Ecology 81: 3178– 3192. » CrossRef » Google Scholar
13. Deléage G, Roux B (1987) An algorithm for protein secondary structure prediction based on class prediction. Protein Eng Des Sel 1: 289-294. » CrossRef » Pubmed » Google Scholar
14. Dunker AK, Lawson DJ, Brown CJ, Williams RM, Romero P, et al., (2001) Intrinsically disordered protein. J Mol Graph Model 19: 26-59. » CrossRef » Pubmed » Google Scholar
15. Fraga S (1982) Theoretical prediction of protein. antigenic determinants from amino acid sequences. Can J Chem 60: 2606-2610. » CrossRef » Google Scholar
16. Gottschalk KW, Colbert JJ, Feicht DL (1998) Tree mortality risk of oak due to gypsy moth. European Journal of Forest Pathology 28: 121-132. » CrossRef » Google Scholar
17. Hermanek P, Guggenmoos-Holzmann I (1994) Classification and regression trees (CART) for estimation of prognosis in patients with gastric carcinoma. J Cancer Res Clin Oncol 120: 309–313. » CrossRef » Pubmed » Google Scholar
18. Janin J (1979) Surface and inside volumes in globular proteins. Nature 277: 491-492. » CrossRef » Pubmed » Google Scholar
19. Jones DD (1975) Amino acid properties and side-chain orientation in proteins: a cross correlation appraoch. J Theor Biol 50: 167-83. » Pubmed » Google Scholar
20. Kannan N, Haste N, Taylor SS, Neuwald AF (2007) The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module. Proc Natl Acad Sci U S A 104: 1272-1277. » CrossRef » Pubmed » Google Scholar
21. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein, J Mol Biol 157: 105-132. » CrossRef » Pubmed » Google Scholar
22. Lifson S, Sander C (1979) Antiparallel and parallel-strands differ in amino acid residue preferences .Nature 282: 109-111. » CrossRef » Pubmed » Google Scholar

23. McCaldon P, Argo P (1988) Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. Proteins 4: 99-122. » CrossRef » Pubmed » Google Scholar

24. Namba K (2001) Roles of partially unfolded conformations in macromolecular self-assembly. Gene Cells 6: 1-12. » CrossRef » Pubmed » Google Scholar

25. Ozge C, Toros F, Bayramkaya E, Camdeviren H, Sasmaz T (2006) Which sociodemographic factors are important on smoking behaviour of high school students? The contribution of classification and regression tree methodology in a broad epidemiological survey. Postgrad Med J 82: 532-541. » CrossRef » Pubmed » Google Scholar

26. Parker PJ, Parkinson SJ (2001) AGC protein kinase phosphorylation and protein kinase C. Biochem Soc Trans 29: 860-863. » CrossRef » Pubmed » Google Scholar

27. Perutz MF (1989) Mechanisms of cooperativity and allosteric regulation in proteins. Q Rev Biophys 22: 139-237.» Pubmed » Google Scholar

28. Perutz MF (1970) Stereochemistry of cooperative effects in haemoglobin. Nature 228: 726-739.» CrossRef » Pubmed » Google Scholar

29. Peterson RT, Schreiber SL (1999) Kinase phosphorylation: Keeping it all in the family. Curr Biol 9: R521-4. » CrossRef » Pubmed » Google Scholar

30. Remington S, Wiegand G, Huber R (1982) Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. J Mol Biol 158: 111-152. » CrossRef » Pubmed » Google Scholar

31. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH (1985) Hydrophobicity of amino acid residues in globular proteins. Science 229: 834-838. » CrossRef » Pubmed » Google Scholar

32. Spurlino JC, Lu GY, Quiocho FA (1991) The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem 266: 5202-5219.» CrossRef » Pubmed » Google Scholar

33. Teague SJ (2003) Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2: 527-41.» CrossRef » Pubmed » Google Scholar

34. Wright PE, Dyson HJ, (1999) Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm. J Mol Biol 293: 321-331.» CrossRef » Pubmed » Google Scholar

35. Zhao G, London E (2006) An amino acid "transmembrane tendency" scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. Protein Sci 15: 1987-2001. » CrossRef » Pubmed » Google Scholar

36. Zimmerman JM, Naomi E, Simha R (1968) The characterization of amino acid sequences in proteins by statistical methods. Journal of Theoretical Biology 21: 170-201. » CrossRef » Pubmed » Google Scholar

Accession numbers of the considered AGC kinase protein sequences are as follows:

O70291.1, POC605.1, P16054.1, P18654.2, P23298.1, P31750.1, P54265.1, P68181.2, P70268.3, P70336.1, Q3UU96.2, O70293.1, P05132.3, P18653.1, P20444.3, P28867.3, P49025.3, P63318.1, P68404.3, P70335.1, Q3U214.2, Q3UYH7.1, Q7TPS0.2, Q7TSE6.1, Q7TSJ6.1, Q7TT50.1, Q8BSK8.1, Q8BWW9.2, Q8BYR2.2, Q8C0P0.1, Q8C050.2, Q8K045.1, Q8VEB1.2, Q9ERE3.1, Q9QZS5.1, Q9R1L5.3, Q9WUA6.1,Q9WUT3.1, Q9WVC6.1, Q9WVL4.1, Q9Z0Z0.1, Q9Z1M4.1, Q9Z2A0.2, Q9Z2B9.1, Q8OUW5.2, Q91VJ4.1, Q99MK8.2, Q811L6.2, Q922R0.1, Q02111.1, Q02956.1, Q60592.1, Q60823.1, Q61410.1, Q62074.2, P41743.1, P43250.2, P51812.1, P51817.1, Q02156.1, Q16513.1, Q16512.1, Q15835.1, Q15418.2, Q15349.2, Q15208.1, Q13976.3, Q13464.1, Q13237.1, CAE55958.1, NP_443073.1, O00141.2, O14578.2, O15021.2, O15530.1, O60307.2, O75116.3, O75582.1, O75676.1, O95835.1, P05129.3, P05771.4, P14619.1, P17252.3, P17612.2, P22612.3, P22694.2, P23443.2, P24256.1, P24723.2, P25098.2, P31749.2, P31751.2, P32298.3, P34947.1, P35626.2, Q09013.1, Q05655.1, Q05513.4, Q04759.3, Q96GX5.1, Q96BR1.1, Q9Y243.1, Q9Y5S2.2, Q9Y2H9.2, Q9Y2H1.3, Q9UK32.1, Q9UBS0.1, Q9NRM7.1, Q9HBY8.1, Q8WTQ7.1, Q6P5Z2.1, Q6P0Q8.2, Q6DT37.1, Q5VT25.1.

Node	Bulkiness	Polarity	Recognition factors	Trans membrane tendency	% Accessible residues	Alpha -helix	beta-sheet	Coil	Total beta-strand	Anti Parallel beta-strand	Parallel beta-strand	A.A. composition	Relative mutability	Average flexibility
1	<= 14.2207					<= 1.01975						<= 5.55		0.457
2	<= 14.2207					<= 1.01975					<= 0.977	> 5.55		0.4494
3	<= 14.2207		<= 90.611			<= 1.01975					> 0.977	> 5.55 & <= 5.63625		0.447667
4	<= 14.2207		<= 90.611			<= 1.01975				<= 0.98925	> 0.977	> 5.63625		0.443143
5	<= 14.2207		<= 90.611			<= 1.01975				> 0.98925	> 0.977	> 5.63625		0.441429
6	<= 14.2207		> 90.611			<= 1.01975					> 0.977	> 5.55		0.4479
7	> 14.2207	<= 19.9293				<= 1.01975		<= 1.0425						0.4336
8	> 14.2207	<= 19.9293				<= 1.01975		> 1.0425						0.438722
9	> 14.2207	> 19.9293				<= 1.01975	<= 0.97275					<= 5.68875		0.4419
10	> 14.2207	> 19.9293				<= 1.01975	<= 0.97275					> 5.68875		0.444667

11	> 14.2207	> 19.9293		<= -0.54225	<= 89.4723	> 1.01975	<= 1.01975	> 0.97275								0.4402
12				<= -0.54225	<= 89.4723	> 1.01975	<= 1.01975							<= 6.0055		0.44075
13				<= -0.54225	<= 89.4723	> 1.01975	<= 1.01975		> 0.95975 & <= 1.018					<= 6.0055	<= 80.0835	0.436563
14				<= -0.54225	<= 89.4723	> 1.01975	<= 1.01975		> 0.95975 &<= 1.018					<= 6.0055	> 80.0835	0.438
15				<= -0.54225	<= 89.4723	> 1.01975	<= 1.01975		> 1.018					<= 6.0055		0.440125
16			> 89.4723 &&<= 89.9445	<= -0.54225		> 1.01975	<= 1.01975							<= 6.0055		0.4425
17			> 89.9445	<= -0.54225		> 1.01975	<= 1.01975							<= 6.0055		0.4345
18				<= -0.54225		> 1.01975	<= 1.01975							> 6.0055		0.444714
19				> -0.54225		> 1.01975	<= 5.7975					<= 1.00675				0.432083
20				> -0.54225		> 1.01975	<= 5.7975					> 1.00675				0.426667
21				> -0.54225		> 1.01975	> 5.7975									0.439

Table 3: Association rules obtained in CART