Identification of a Novel Cysteine-stack Arrangement in Parallel β-helix Proteins using Computational and Knowledge-based Approach

Zeti AMH[1], Mohd Shahir S[2] and Gerloff DL[3]

[1]School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor. [2]Bioinformatics Research Lab, Biology Department, Faculty of Science, Universiti Teknologi Malaysia Skudai, Johor. [3]Institute of Structural & Molecular Biology, University of Edinburgh, Edinburgh, Scotland, UK.

Abstract

Parallel β-helices, a subclass of β-sheet proteins, represent the folding of a polypeptide chain into an elongated topologically simpler fold than globular β-sheets. However, the amino acid sequence rules that specify β-sheet structure in protein remain unelucidated. In this study, a combination of knowledge-based and computational analysis using the novel cysteine staple pattern found in pectin methylesterase A to predict the right-handed parallel β-helix structural motif in primary amino acid sequences is presented. The novel staple pattern was used to find further pectin lyase-like protein families displaying this pattern, and to produce high confidence 3D models for all detectable members of this superfamily. To achieve these goals, candidate sequences were retrieved from GenPept and SUPERFAMILY databases. Possible pectin lyase-like proteins were detected with two different fold recognition algorithms, and stringent criteria were designed to minimise false positive predictions. The filtered datasets were clustered, resulting into two main categories; Category A included all sequences with homologues in the PDB and Category B comprised all sequences with unknown homologues. Four families have been identified as new family, where prior to this study, the proteins in this new family had not been identified nor classified as pectin lyase-like proteins. Multiple sequence alignments were generated from each family in both categories, and were carefully inspected to finally produce highly accurate alignments to be used as input for large-scale automatic modeling. There were 298 high quality 3D models of pectin lyase-like proteins produced, which may be used to provide a valuable resource for biological research in this area. Models inspection has revealed numerous instances of possible cysteine stacks in the β-helix cores, where asparagine ladders are common as well as the existence of disulphide bridges.

Keywords: parallel β-helix; pectin lyase-like; disulphide bridge; fold recognition; multiple sequence alignment

Introduction

Yoder *et al.* reported the first right-handed parallel β-helix motif in 1993[1]. It is characterised by a series of progressive coils, each of which contributes to three elongated β-sheets that form a distorted triangular prism shape to comprise the fold. The cross section of the parallel β-helix is called a rung and consists of three β-strands connected by turn regions with variable length. The backbone folds up in helical with β-strands from adjacent cross-section stacking on top of each other in a parallel orientation. The cylindrical coil is predominantly composed of hydrophobic amino acids. The distinct feature in this fold is the presence of hydrophobic side chain stacks, as well as ladders of hydrogen bonding side chains, such as asparagines in some structures in this superfamily similar to other stacks such as serine stacks, aliphatic stacks, and ringed-residue stacks. This study is in agreement with the fact that a right-handed parallel β-helix motif is not common as demonstrated by the availability of only twelve such 3D structures in the Protein Databank (PDB) [http:www.rcsb.org/pdb] [2]. The parallel β-helix fold is observed in pectate lyase [1], pectin lyase [3], pectin methylestarase [4], insect antifreeze protein [5], *iota*-carrageenase [6], bacteriophage P22 tailspike adhesin [7], P.69 pertactin toxin [8], chondroitinase B [9], galacturonase [10] and recently after this study has been conducted, two more new families have been identified. Structural Classification of Proteins (SCOP) has classified these families into a superfamily known as pectin lyase-like (PLL) superfamily [11].

The unique protein architecture of β-helix motif mostly exhibit very low sequence identity, making them unsuitable to standard sequence searching such as PSI-BLAST [12] and HMMER[13]. However, in the pectin methylesterase A prediction, we managed to find a novel staple pattern of partially conserved cysteine, which later has helped us in finding more sequences that most likely adopt to this fold [14]. Consequently, this information becomes the basis of our knowledge-based approach. When the sequences were transformed into a 3D model, in some families, this novel pattern has shown that these cysteines were actually stacked replacing the used to be

Author for correspondence: Dr. Zeti Azura Mohamed Hussein, School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor. Tel: 603-89215993 Fax: 603-89252698 E-mail: zeti@pkrisc.cc.ukm.my

an asparagine ladder. Normally the existence of at least two residues of cysteines in a protein may form a disulphide bridge between various parts of the same protein or between two separate polypeptide chains [15,16], but interestingly not in this superfamily. Detailed inspection on this circumstance is actively underway with our collaborator to see whether this condition found *in silico* is in consensus with the findings in the wet lab experiment. If the result is in consensus with the prediction, we have made a very significant finding in the parallel β-helix proteins, where previously people believe that these proteins are very stable due to the existence of disulphide bridges, asparagines ladders and hydrophobic interactions [17-19].

Methods

This combinatorial approach was initiated by the success in finding the novel cysteine ladder pattern in the pectin methylesterase A prediction [14]. The computational approach was used in recognising the possible candidates of protein, which adopts the parallel β-helix fold, whilst the knowledge-based approach (based on the cysteine-anchoring position and conserved cysteine; Figure 1) was used in the multiple sequence alignments generated for each family to find the correct anchor for modeling purposes. Candidate sequences were retrieved from two databases, GenPept (934,648 sequences) and SUPERFAMILY (1305 sequences). To reduce the list of sequences to a manageable size, as well as to minimize the number of false positives, several reasonable limitations were incorporated into the target sequence selection. These are divided into three criteria, which are as follows: the occurrence of protein with parallel β-helix fold in certain organisms, the length of the protein region and the E-values of the 3D-Genomics [20] and SUPERFAMILY [21] fold assignment. An important point to consider once the above requirements had been fulfilled was the existence of cysteine in the sequences. These criteria were used to filter the unrelated sequences and the number of sequences remained were decreased significantly. These sequences were grouped based on the methods used, (3d-genomics and SUPERFAMILY algorithm), resulting into two groups abbreviated as 3D-dataset (5419 sequences; 3D is an abbreviation for 3d-genomics algorithm [20]) and SF-dataset (1035 sequences; SF is an abbreviation for SUPERFAMILY algorithm [21]). These two methods were used to recognise the sequences that likely to adopt the parallel β-helix fold, based their different algorithm. In order to have a very minimum rate of redundancy, BLAST [22] and IMPALA [23] were performed on each dataset of these potential predicted PLL proteins. BlastClust [24] was used to cluster the redundant-free datasets resulting to the classification of the sequences based on the PLL superfamily. Each cluster assigned was regarded as a family of its own, and each sequence in one particular cluster was carefully examined by crosschecking the sequence detail in its original database. The range of sequence similarity within each cluster were obtained using BLAST and DARWIN [25], where DARWIN also calculated the accepted point mutations distance between two sequences, and the sequence divergence for each cluster was shown in a phylogenetic tree (data not shown). For the clusters with no homologue, automated fold recognition approach was used, where one representative for each cluster was sent to Metaserver [26] and was fed into the 3dpssm algorithm [27]. Secondary structure prediction was obtained from PsiPred [28]. Each cluster represented one family, and a multiple sequence alignment was generated for each family by using T-Coffee [29]. Multiple sequence alignment for each family was analysed sequentially to identify both evolutionary and structural similarity among the proteins encoded by each member in the alignment. These alignments were cautiously inspected and where necessary were manually adjusted based on the knowledge of the PLL superfamily, the cysteine anchoring position and conserved cysteine in order to obtain the most accurate and reliable alignment for model construction. All satisfactory target-template alignments were used to construct high-confidence model for every sequence in each family. Models were automatically generated by MODELLER 6 version 2 [30] running on Silicon Graphics Origin200 processor. This automated modeling step was controlled by a specifically written PERL program, which calls for various programs to automatically perform the modeling procedure. For each sequence, ten models were generated, and the best model was selected based on the lowest objective function value.

Results and Discussion

The combined sequences from two datasets were analysed and assigned into two categories known as Category A (families with 3D homologue) and Category B (families with no 3D homologue). Twelve clusters represented seven PLL families have been identified in Category A, and nine clusters have been identified in Category B, including four new families, prior to this study the proteins in these new families had not been identified nor classified as PLL proteins. In addition, a small fraction of the predicted PLL sequences deemed to be probable false positives. The host organisms of these putative parallel β-helix proteins in these datasets were from three taxonomy kingdoms: bacteria, fungi and plants. The range of sequence similarity within each cluster is listed in Table 1 and consisted of HSPs and PAM values. Higher PAM width values indicate the presence of more divergent sequences in the cluster. The PAM width values for each cluster correlated with the percent sequence identity values obtained from BLAST HSPs. Where the PAM width value is low for a particular cluster, its percent sequence identity is high, meaning that it consisted of less divergent sequences. The results showed that cluster A_1GQ8, B_1IA5 and E_1IDK have higher percent sequence identity with relatively low PAM width value. This was consistent with the narrow species range present in these clusters, which are exclusively from plants and fungi, respectively.

Table 1: Range of sequence similarity within cluster. Finalised PLL families, number of clusters identified, sequence identity value and cysteine distributions in 3D models.

CATEGORY A	PLL Family	Number of Clusters
	Pectin methylesterase	2
	Galacturonase	3
	Pectate lyase	3
	Pectin lyase	1
	Pertactin	1
	Chondroitinase	1
	Iota-carrageenase	1

CATEGORY B	PLL Family	Number of Clusters
	Pectin methylesterase	2
	Galacturonase	1
	Pectate lyase	2
	New Family	**Number of Clusters**
	Copper-binding periplasmic protein	1
	Dextranase (as identified by SCOP)	1
	Glucanase	1
	Putative colanic acid	1

Cluster	PLL Family	PSI	PAM Width
A_1QJV	Pectin methylesterase	(25-52)	172
A_1GQ8	Pectin methylesterase	(43-99)	95
B_1HG8	Galacturonase	(29-83)	120
B_1IA5	Galacturonase	(48-85)	75
B_1BHE	Galacturonase	(24-37)	199
D_2PEC	Pectate lyase	(27-84)	127
D_1PCL	Pectate lyase	(25-94)	140
D_1EE6	Pectate lyase	(23-79)	153
E_1IDK	Pectate lyase	(51-71)	85

Cysteine Distribution	PLL Family	Total cysteine	No. of models with S-S	No. of models with cysteine stacks
	Pectin methylesterase	366	-	61
	Galacturonase	520	58	1
	Pectate lyase	864	23	-
	Pectin lyase	45	7	-

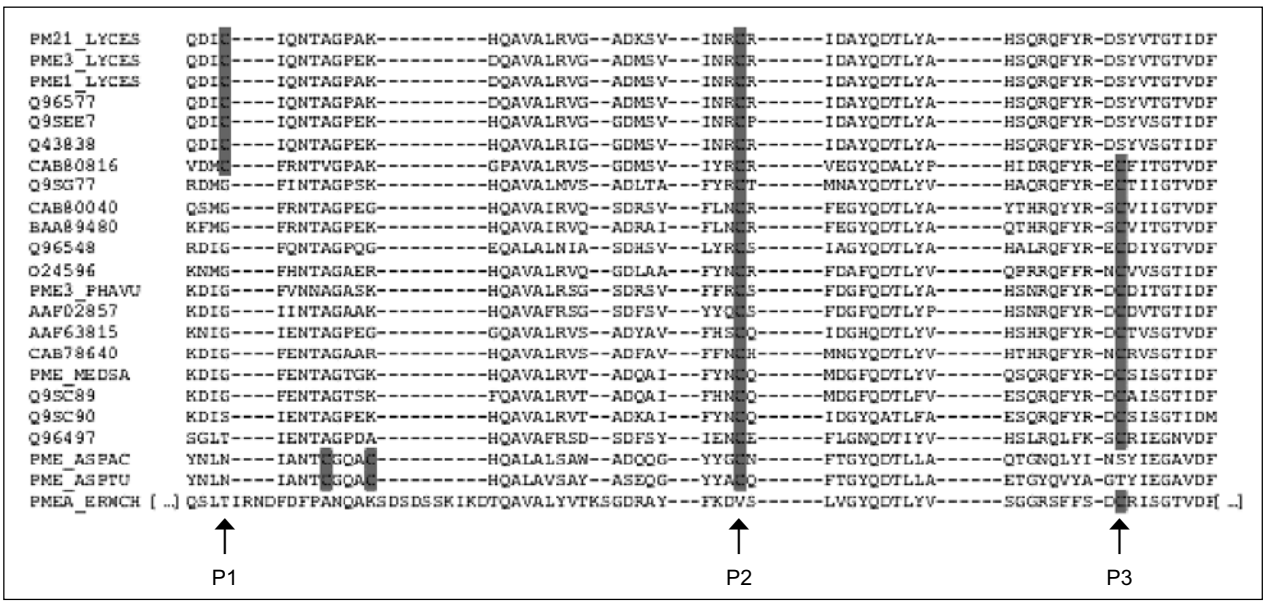

Figure 1: Cysteine stack pattern identified in a multiple sequence alignment of pectin methylesterase A prediction. P1, P2 and P3 are the cysteine-anchoring positions.

The multiple sequence alignment for each cluster in Category A has revealed interesting findings such as the identification of the putative active sites and the catalytic residues, the stacking organisations and the putative disulphide bridges. The cysteine stacks were identified in the pectin methylesterase clusters, galacturonase clusters and two clusters of pectate lyase. Full conservation of cysteine stacks have been identified in the pectin methylesterase (A_1QJV and A_1GQ8) and two of galacturonase (B_1HG8 and B_1IA5) clusters, whilst in D_1PCL and D_1EE6, the cysteine stacks were moderately conserved. Based on the analysis of the alignments of the families in the PLL superfamily, each family contains its functionally important residues within the motif related in one particular cluster. On the other hand, there were some important changes of the residues adjacent to the residues constituting the active site. The internal residues stacking in some of the clusters have also been identified. All these interesting features were clearly shown in the models generated from the modeling procedure based on these alignments.

In Category B, there were five clusters in the PLL superfamily and four clusters were identified as new family, prior to an initial identification that these sequences were not parallel β-helix proteins. Recently, SCOP has classified three new families (based on solved protein crystal structure) in the PLL superfamily, where one of these families was the same family predicted in this study (dextranase). Detailed analysis on the multiple sequence alignments have shown that there were no apparent cysteine-anchoring positions in these clusters, except in clusters Q and P, where these conditions have revealed the stacking organisation of cysteine, as well as predicted putative disulphide bridge formation in certain 3D models. Two types of disulphide bridges were observed in those models (Figure 2).

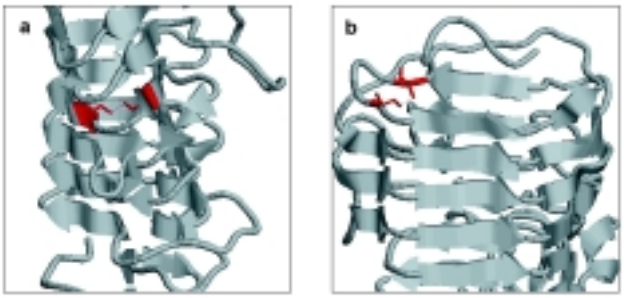

Figure 2: The disulphide bridge positions in the parallel β-helix structure. (a) Type 1 S-S bond: between consecutive turn. (b) Type 2 S-S bond: on the loop.

In most alignments (Categories A and B), glycine was found to be a well conserved residue, and this condition is a typical pattern for distantly related proteins with conserved functional properties. This finding was compatible as mentioned by Markovic & Jörnwall (1992) based on their *in vitro* experiment [31]. The high number of conserved glycines in the alignment indicates their crucial role in controlling conformational changes or avoiding steric clashes [32] and also their occurrence tight turns. Most notable was in clusters A_1GQ8, B_1HG8, B_1IA5 and E_1IDK. In pectin methylesterase cluster, aspartic acid, arginine, tryptophan and tyrosine are fully conserved, which characterised the active site [4,33]. Meanwhile, fully conserved lysine and arginine in galacturonase clusters indicated that these residues were involved in the substrate binding [34]. Cysteine was found to be fully conserved in this cluster, and structural study on the 3D models has shown that these cysteines were predicted to form disulphide bridges. In pectate lyase, the identification of conserved histidine at the vWiDH region (Figure 3) was consistent with the fact that histidine was part of the highly conserved region, which was believed to contribute to the folding or structural stability.

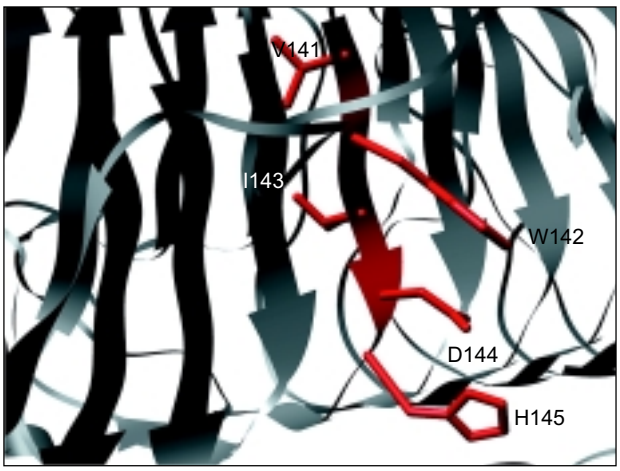

Figure 3: vWiDH region in pectate lyase protein. H145 is facing to the loop, and D144 is a catalytic residue.

All catalytic residues were well conserved in this family, which include the aspartic acid and arginine residues [35]. Conserved aspartic acid and glutamic acid in this family were important as it indicated the calcium-binding site, which was needed for this family for its function. No conserved cysteine was found in this family, indicating that the presence of disulphide bridges was not maintained in this family. Conserved polar residues such as tyrosine, asparagines, threonine and arginine could be of interest for providing reactive groups to the catalytic activity [36]. Conserved aspartic acid also indicated its involvement in the catalytic activity in all families. As far as the aromatic residues in these alignments were concerned, the interest in these residues was due to the fact that they may also participate in substrate binding, as well as pairwise interaction [18]. These important conserved residues were successfully identified from detailed multiple sequence alignment analysis, and were very useful in preserving the biological significant functions in PLL families.

Based on our detailed analysis on the cysteine-cysteine interaction, in models with no disulphide bridges, the

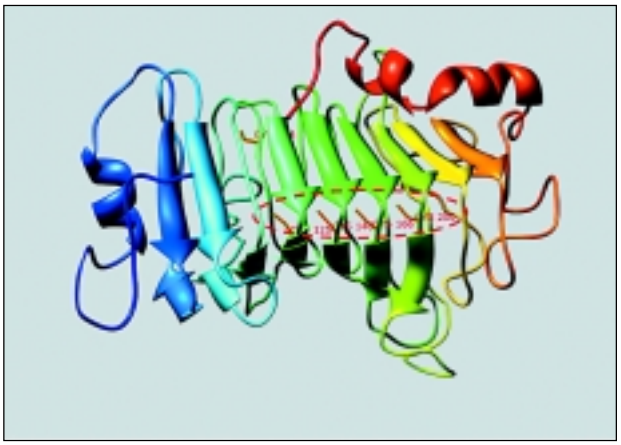

Figure 4: Example of cysteine stack in 3D model from pectin methylesetrase cluster. (Graphics produced with Chimera [38])

cysteine were likely to form the cysteine-stacks (Figure 3). The interaction study (data not shown) also suggested that this stack may provide the required stability by these proteins, which in some PLL families it was contributed by asparagine ladder or disulphide bridge at the concurrent positions in the structural core. For instance, in the pectate lyase family, the stability was mainly contributed by the presence of different types of stacks (aromatic, aliphatic etc.) [18,35,36], whilst in the galacturonase and pectin lyase families, a number of disulphide bridges were responsible for the structure stability [3,10,19,]. Meanwhile in the pectin methylesterase family, there was no disulphide bridge present in their native state. From our analysis, the conserved cysteines found in the alignments (when inspected in the structural analysis) were likely to form the cysteine stacks. Our *in silico* analysis and results have corroborated Pickersgill's findings, where in order to provide the stability in the absence of disulphide bridge, these conserved and unpaired cysteines were likely to form the stacks (Figure 4). These stacking will provide structural regularity and contribute to maintaining the relative disposition of the secondary structure elements. The regularity is significant for a modular structure that allows new turns of the helix to be incorporated into the parallel β-helix fold [37]. The observation that each turn interacts predominantly with the turns on either side strengthens this suggestion. Apart from the cysteines that were involved in the disulphide bridge or stacking, there were a number of cysteines which were found to be primarily solitary. Model observation showed that these cysteines were scattered in the loop region and occasionally in the middle core of the β-helix. These cysteines were most likely unpaired because there was no possibility for them to find a partner close enough to form a bridge.

Conclusion

With conscientious bioinformatic procedure, we managed to identify more sequences that are likely to adopt the parallel β-helix fold. Detailed inspection and manual editing on the multiple sequence alignments was a success in identifying conserved cysteines that formed an unusual structural feature termed as cysteine stack. Structural analysis on this feature has suggested that it could provide ample stability to the parallel β-helix structure and comparable with the ones (as mentioned in the introduction) that were usual in some of the families in the PLL superfamily. These findings proved to be very useful to provide valuable starting point for further analysis in the wet laboratory experiments. The identification of various conserved residue conditions could also be used in understanding the binding specificity of these proteins. Normally the binding specificity is determined by a small number of residues residing at or near the binding site, and variation of only a few residues allows these proteins to have very different binding specificities. A few cases have been seen in this study. With this observation, there is a need for further research either in the prediction aspect or in the experimental laboratory, so as to fully understand the above-mentioned phenomenon.

Acknowledgment

The authors would like to thank Lawrence Kelly, Kieran Fleming and Arne Mueller for providing the 3dpssm and 3d-genomics programs; and to Julian Gough for providing a raw PLL dataset from SUPERFAMILY database.

References

1. Yoder M, Keen N and Jurnak F. New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. Science 1993; 260(5113): 1503-1507.
2. Berman HM, Westbrook J, Feng Z. Gililand G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE. The protein data bank. Nucleic Acids Res 2002; 58: 899-907.
3. Mayans O, Scott M, Connerton I, Gravesen T, Benen J.A.E, Visser J, Pickersgill R, and Jenkins J. Two crystal structures of pectin lyase A from *Aspergillus* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyase. Structure 1997; 5: 677-689.
4. Jenkins J, Mayans O, Smith D, Worboys K and Pickersgill R. Three-dimensional structure of *Erwinia chrysanthemi* pectin methylesterase reveals a novel esterase active site. J Mol Biol 2001; 305: 951-960.
5. Liou YC, Toclij A, Davies PL and Jia Z. Mimicry of ice structure by surface hydroxyls and water of a β-helix antifreeze protein. Nature 2000; 406(6793): 322-324.
6. Michel G, Chantalat L, Fanchon E, Henrissat B, Kloareg B and Dideberg O. The i-carrageenase of *Alteromonas fortis.* J Biol Chem 2001; 276(43): 40202-40209.

7. Steinbacher S, Baxa U, Miller S, Weintraub A, and Seckler R. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. Proc Nat Acad Sci 1996; 93: 10584-10588.
8. Emsley P, Charles IG, Fairweather NF and Isaacs NW. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. Nature 1996; 381(6577): 90-92.
9. Huang W, Matte A, Li Y, Kim YS, Lindhart RJ, Su H and Cygler, M. Crystal structure of chondroitinase B from *Flavobacterium heparinum* and its complex with a disaccharide product at 1.7 Å resolution. J Mol Biol 1999; 294(5): 1257-1269.
10. Federici L, Caprari C, Mattei B, Savino A, Di Matteo A, De Lorenzo G, Cervone F and Tsernoglou D. Structural requirements of endopolygalacturonase for the interaction with PGIP (polygalacturonase-inhibiting protein). Proc Nat Acad Sci 2001; 98: 13425-13430.
11. Andreeva A., Howorth D., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. SCOP database in 2004: Refinements integrate structure and sequence family data. 2004; Nucleic Acids Res 32: D226-229.
12. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997; 25: 3389-3402.
13. Eddy S. Profile hidden Markov models. Bioinformatics 1998; 14(9): 755-763.
14. Zeti AMH, McCarthy-Troke MH, Mitchell BJ, Duffy, CRE, Leung SW, Cannarozzi GM and Gerloff DL. Incorporation of human-derived constraints from active/functional site models in protein tertiary structure assembly. Fourth Meeting on the Critical Assessment of Techniques for Protein Structure Prediction - CASP 4. 2000; A-33.
15. Miseta A and Csutora P. Relationship between the occurrence of cysteine in proteins and the complexity of organisms. Mol Biol Evol 2000; 17(8): 1232-1239.
16. Aslund F and Beckwith J. Bridge over troubled waters: sensing stress by disulfide bond formation. Cell 1999; 96: 751-753.
17. Akita M, Suzuki A, Kobayashi T, Ito S and Yamane T. The first structure of pectate lyase belongs to polysaccharide lyase family 3. Acta Crys Sect D 2001; 57: 1786-1792.
18. Yoder MD, Lietzke SE and Jurnak F. Unusual structural features in the parallel ß-helix in pectate lyases. Structure 1993b; 1: 241-251.
19. Jenkins J and Pickersgill R. The architecture of parallel β-helices and related folds. Prog Biophys Mol Biol 2001; 77(2): 111-175.
20. Fleming K, Muller A, MacCallum RM, Sternberg MJ. 3d-genomics algorithm. Personal communication.
21. Gough J, Karplus K, Hughey R and Chothia C. Assignment of homology to genome sequences using a library of Hidden Markov Models that represent all proteins of known structure. J Mol Biol 2001; 313: 903-919.
22. Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403-410.
23. Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L and Altschul SF. IMPALA: Matching a protein sequence against a collection of PSI-BLAST-constructed position-specific-scoring matrices. Bioinformatics 1999; 1: 1000-1011.
24. Dondoshansky I. BlastClust: BLAST score-based single-linkage clustering. www.ncbi.nlm.nih.gov. 2000.
25. Gonnet GH, Hallett MT, Korostensky C and Bernardin L. Darwin v. 2.0: an interpreted computer language for the biosciences. Bioinformatics 2000; 16(2): 101-103.
26. Ginalski K, Elofsson A, Fischer D and Rychlewski L. 3D-jury: a simple approach to improve protein structure predictions. Bioinformatics 2003; 19(8): 1015-1018.
27. Kelley LA, MacCallum RM and Sternberg MJE. Enhanced genome annotation using structural profiles in the program 3DPSSM. J Mol Biol 2000; 299: 499-520.
28. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999; 287: 797-815.
29. Notredame C, Higgins DG and Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000; 302(1): 205-217.
30. Sali A and Blundell TA. Comparative protein modelling by staisfaction of spatial restraints. J Mol Biol 1993; 234(3): 779 - 815.
31. Markovic O and Jornvall H. Disulfide bridges in tomato pectinesterase: Variations from pectinesterases of other species; conservation of possible active site segments. Prot. Sci. 1992; 1(10): 1288-1292.
32. Henrissat B, Heffron S, Yoder M, Lietzke S and Jurnak F. Functional implications of structure-based sequence alignment of proteins in the extracellular pectate lyase superfamily. Plant Physiol 1995; 107: 963-976.
33. Frenkel C, Peters JS, Tieman DM, Tiznado ME and Handa, AK. Pectin methylesterase regulates methanol and ethanol accumulation in ripening tomato *(Lycopersicon esculentum)* fruit. J Biol Chem 1998; 273: 4293-4295.
34. Pages S, Heijne WH, Kester HCM, Visser J and Benen JAE. Subsite mapping of *Aspergillus niger* endopolygalacturonase II by site-directed mutagenesis. J Biol Chem 2000; 275: 29348-29353.
35. Scavetta RD, Herron SR, Hotchkiss AT, Kita N, Keen NT, Benen JAE, CM, KH, Visser J and Jurnak F. Structure of a plant cell wall fragment complexed to pectate lyase C. Plant Cell 1999; 11: 1081-1092.
36. Yoder MD, Keen NT and Jurnak F. New domain motif - the structure of pectate lyase C, a secreted plant virulence factor. Science 1993a; 260: 1503-1507.
37. Jenkins J, Mayans O and Pickersgill R. Structure and evolution of parallel β-helix proteins. J Struc Biol 1998; 122(1-2): 236-246.
38. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE. UCSF Chimera - A visualization system for exploratory research and analysis. J Comp Chem 2004; 25(13): 1605-1612.