Motif Design for Nitrilases

Nikhil Sharma and Tek Chand Bhalla*

Sub-Distributed Information Centre, Himachal Pradesh University, Summer-Hill, Shimla-171005, India

Abstract

Nitrilase is one of the nitrile metabolizing enzymes that catalyses the conversion of nitriles to corresponding acids which has gained importance in green chemistry. Nitrilase being substrate specific yet it acts over a wide range of nitriles (aliphatic/aromatic) has drawn attention due to its utility in mild hydrolysis. Most of the nitrilases reported hitherto have been physically extracted characterized from the microbial/plant sources. In order to identify sequences for nitrilases two groups of motif were designed i.e. aliphatic nitrilase motif’s (MDMAl) and aromatic nitrilase motif’s (MDMAr) each with four motifs specifically belonging to nitrilase with conserved catalytic triad (Glu-48, Lys-131, Cys-165) which can be used as marker for nitrilase. Conserved regions were identified by performing Multiple Sequence Alignment (MSA) using Multiple EM for Motif Elicitation (MEME). The Manually Designed Motifs (MDM’s) were validated by ScanProsite and their presence is also confirmed by PRATT, Gblocks and MEME. The ScanProsite search against the MDMar exhibited some new sources of aromatic nitrilase from plant, animals and microbes whereas MDMAl only exhibited nitrilase from microbes. Besides identifying unique motifs in order to confirm their substrate specificity for nitriles, randomly selected sequences were validated by studying some important physiochemical parameters and position specific amino acids.

Keywords: Gblock; Manually designed motif (MDM); Multiple alignments; Nitrilase; PROSITE; Pratt; Multiple EM for motif elicitation (MEME); ScanProsite; Substrate specificity

Background

Nitrilases are responsible to catalyze the hydrolysis of nitriles into corresponding acids and ammonia which are used in chemistry for the production of industrially important acids. Nitrilases frequently exhibit enantioselectivity under mild reaction conditions, making them ideal tools for green chemistry from conversion of nitrile or in environment management for remediation of nitrile contaminated soil, water and air. Based on substrate specificity they are grouped as aromatic and aliphatic nitrilase.

A large number of microbial plant and animal genome have been sequenced in the past decade and genome sequence data have tremendously expanded over those years. Screening of genome/proteome databank will be worthwhile to find out novel sources of enzymes. Many tools and techniques such as BLAST, Hidden Markov Model (HMM), neural network classification are used for in silico screening of genome/proteome sequence database till date. Among these, motif design has been found to be one of the most reliable strategy for efficient screening of database [1-3] as motifs of a particular protein sequence signifies its specific structure and functionally which is useful to characterize and classify that protein.

Nitrile metabolizing microorganisms are mostly isolated from soil or water using enrichment culture method and these are cultured and tested for the nitrilase activity during screening of microbial isolates. This is a classical method for isolation of the nitrile metabolizing microorganisms and subsequent screening for nitrilases activity is a time consuming and cost intensive process [4-7]. Presently bioinformatics tools and techniques such as BLAST [1], HMM (Hidden Markov Model) [8-9] the neural network classification [10,11] and Motif Identification Neural Design (MOTIFIND) [12] are used for screening of genome and sequence database for searching gene coding for novel enzymes. The present communication aims at Manually Designed Motif (MDM) for aliphatic and aromatic nitrilase and their validation using Prosite, ScanProsite, BLAST, PRATT and G-block is being reported. On the basis of the earlier studies on in silico analysis of amino acid sequence of the aromatic and aliphatic nitrilases [13] and using bioinformatics approaches, motifs were manually designed to differentiate and identified aromatic and aliphatic nitrilases. Computational analysis of amino acid sequences and study pertaining to physiochemical properties [14] of nitrilases have revealed differences between aromatic and aliphatic nitrilases [13].

Methodology

Retrieval of sequences and designing of motifs

The protein sequences were obtained from the protein server UniprotKB/SwissProt (release 2011-12) database. To design motifs manually, six microbial aliphatic and aromatic nitrilases were selected on the basis of our earlier study (Table 1) [13].

Multiple Sequence alignment and phylogenetic analyses

Multiple sequence alignments were performed using CLUSTAL W [15] and CLUSTAL X (version 2.1) software with default settings and verified with MAFFT (Multiple Alignment with Fast Fourier Transform) and MEME (Multiple Em for Motif Elicitation). Nitrilase (aliphatic and aromatic) sequences were used for phylogenetic relationship inferences. Phylogenetic tree was generated by Neighbor Joining (NJ) using CLUSTAL X (Figure 2). After multiple alignments and also on the basis of presence of catalytic triad i.e glutamine, lysine, and cysteine (Glu-48, Lys-131, and Cys-165) motifs were manually designed (MDM) for aliphatic (MDMA) and aromatic nitrilases (MDM). In order to verify the manually designed motif and to eliminate poorly aligned or

*Corresponding author: Dr. Tek Chand Bhalla, Sub-Distributed Information Centre, Himachal Pradesh University, Summer-Hill, Shimla-171005, India, E-mail: bhallatc@rediffmail.com

Received August 06, 2012; Accepted September 24, 2012; Published October 01, 2012

Citation: Sharma N, Bhalla TC (2012) Motif Design for Nitrilases. J Data Mining Genomics Proteomics 3:119. doi:10.4172/2153-0602.1000119

Copyright: © 2012 Sharma N, 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.
In-silico physiochemical characterization and validation of motifs

After confirmation through MEME version 4.8.1 (Figure 1a and Figure 1b), Gblock tool, Manusively Designed Motifs (MDM’s) were subjected to database search through ScanProsite to validate and determine the specificity of these motifs in identifying nitrilase sequences. ScanProsite tool was used to search the databases both from UniprotKB/Swiss-Prot and UniprotKB/TrEMBL with match mode of “not greedy, not overlaps” and “no includes”. ScanProsite analysis resulted in protein sequences which were counted and were grouped into plants, bacteria and uncultured organisms (Table 2).

MDM were further verified by using Pratt tool to identify divergent regions of aligned protein Gblock software was used [16, 17]. Bootstrap value was calculated by using default value (1000).

![Figure 1a: Aliphatic Motifs for Nitrilase validated by MEME version 4.8.1(2012) (the height of the motif “block” is proportional to -log(p-value), truncated at the height for a motif with a p-value of 1e-10).]

![Figure 1b: Aromatic Motifs for Nitrilase validated by MEME version 4.8.1(2012) (the height of the motif “block” is proportional to -log(p-value), truncated at the height for a motif with a p-value of 1e-10).]

Table 1: Names of various microorganisms and accession number of the some microbial nitrilases (aliphatic and aromatic).
conserved pattern in protein sequences. In order to find the nature of the nitrilase sequences i.e. aliphatic/aromatic we also studied some of the physiochemical properties of randomly selected protein sequences. The physicochemical parameters studied included negatively charged residue (Asp+Glu), molecular weight, alanine content (%) and instability index. The values of these parameters for the predicted/reported nitrilase sequences were deduced from the ProtParam (http://web.expasy.org/protparam/) of Expert Protein Analysis System.

| Nitirilases | Manually Designed Motif | Protein Sequence Obtained | Hits obtained for MDM | Motif Presence | UniProtKB/TREMBL Entries |
|------------|-------------------------|---------------------------|-----------------------|----------------|-------------------------|
| Aliphatic  | ([FL]-[ILV]-[AV]-F-P-E-[VT]-[FW]-[IL]-P-[GY]-Y-P-[WY]) | 84 | F-17 B-38 UO-29 | 34-68 | 84 |
|            | R-R-K-[KRI]-[PA]-T-[HY]-[VAH]-E-R | 115 | F-31 B-53 UO-31 | 125-177 | 115 |
|            | C-W-E-H-[FLX]-[NQ]-[PT]-L | 248 | F-49 B-132 UO-67 | 157-215 | 248 |
|            | [VA]-A-X-[AV]-Q-[AI]-X-P-[VA]-X-[LF]-[SD] | 130 | F-06 B-87 UO-37 | 1-30 | 130 |
| Aromatic   | ([ALV]-[LV]-[FLM]-P-[E]-[AS]-[FLV]-[LV]-[AGP]-[AG]-Y-P) | 55 | F-01 B-42 P-10 UO-2 | 26-55 | 55 |
|            | [AGN]-[KR]-H-R-K-L-[MK]-P-T-[AGN]-X-E-R | 104 | F-09 B-68 P-18 UO-8 A-01 | 165-206 | 104 |
|            | C-W-E-N-[HY]-M-P-[LM]-[AL]-R-[KI]-A-[PT]-L-Y | 128 | F-3 B-67 A-01 UO-23 | 125-180 | 128 |

Table 2: Manually designed motif’s with number of hits obtained from UniProt KB and TrEMBL and their position. F-Fungus, B-Bacteria, UO-Uncultured Organism, P-Plants and A-Animal.

Pattern 1  ([FL]-[ILV]-[AV]-F-P-E-[VT]-[FW]-[IL]-P-[GY]-Y-P-[WY])

| S. No | Organism | Access. No. | Mol wt. | Amino Acid No. | NCR* | Alanine (%) | Instability index |
|-------|----------|-------------|---------|----------------|------|-------------|-------------------|
| 1     | Uncultured organism | Q6RWK6 | 378791 | 354 | 39 | 16.9 | 36.02 |
| 2     | Nocardia sp. | Q5D4V8 | 41971.9 | 381 | 51 | 11.5 | 30.61 |
| 3     | Sclerotinia sclerotiorum | A1F824 | 41160.5 | 368 | 49 | 6.8 | 35.31 |
| 4     | Nectria haematococa | C7YVF8 | 41018.6 | 366 | 53 | 7.7 | 33.84 |
| 5     | Silicibacter pomeroyi | Q5LLB2 | 37226.1 | 344 | 44 | 10.8 | 35.63 |

Pattern 2  (R-R-K-[KRI]-[PA]-T-[HY]-[VAH]-E-R)

| S. No | Organism | Access. No. | Mol wt. | Amino Acid No. | NCR* | Alanine (%) | Instability index |
|-------|----------|-------------|---------|----------------|------|-------------|-------------------|
| 1     | Uncultured organism | Q6RWS5 | 37242.6 | 345 | 40 | 15.4 | 37.02 |
| 2     | Uncultured organism | Q6RWS0 | 38048.3 | 349 | 36 | 12.3 | 35.09 |
| 3     | Rhodococcus rhodochrous | A4LA85 | 40370.4 | 386 | 46 | 10.9 | 39.80 |
| 4     | B. japonicum | Q89GE3 | 36203.3 | 334 | 39 | 13.5 | 33.69 |
| 5     | Aspergillus niger | A9QXEO | 40022.2 | 356 | 49 | 7.6 | 30.71 |

Pattern 3  (C-W-E-H-[FLX]-[NQ]-[PT]-L)

| S. No | Organism | Access. No. | Mol wt. | Amino Acid No. | NCR* | Alanine (%) | Instability index |
|-------|----------|-------------|---------|----------------|------|-------------|-------------------|
| 1     | Burkholderia multivorans | B98NU9 | 37461.8 | 345 | 42 | 11.9 | 36.48 |
| 2     | Uncultured organism | Q6RWF2 | 37809.2 | 344 | 44 | 11.0 | 39.69 |
| 3     | Uncultured organism | Q6RWF9 | 38029.4 | 353 | 40 | 17.0 | 37.41 |
| 4     | Methylbium petroleiphilum | A2SEG6 | 37969.1 | 357 | 39 | 16.0 | 42.17 |
| 5     | Paracoccus chlorophenolicus | C1HSH4 | 37871.1 | 356 | 39 | 10.1 | 37.47 |

Pattern 4  ([VA]-A-X-[AV]-Q-[AI]-X-P-[VA]-X-[LF]-[SD])

| S. No | Organism | Access. No. | Mol wt. | Amino Acid No. | NCR* | Alanine (%) | Instability index |
|-------|----------|-------------|---------|----------------|------|-------------|-------------------|
| 1     | Rhodococcus rhodochrous | Q03217 | 40189.2 | 366 | 46 | 11.2 | 39.36 |
| 2     | Uncultured organism | Q6RWK6 | 37879.1 | 354 | 39 | 16.9 | 36.02 |
| 3     | Sphingomonas wittichi | B4WRX6 | 35965.2 | 334 | 42 | 12.6 | 38.51 |
| 4     | Silicibacter pomeroyi | Q5LLB2 | 37226.1 | 344 | 44 | 10.8 | 35.63 |
| 5     | Providencia alcalfaciens | B6XD88 | 37552.7 | 343 | 43 | 10.8 | 30.32 |

Table 3a: In silico analysis of some physiochemical properties of some aliphatic nitrilase (sequences obtained from ScanProsite) from Manually Designed Motif analysis.
Pattern 1  ([ALV]-[LV]-[FLM]-[P-E-[AS]]-[FLV]-[LV]-[AGP]-[AG]-Y-P)

| S. No | Organism                     | Access. No. | Mol wt. | Amino Acid No. | NCR\(^{2}\) | Alanine (%) | Instability index |
|-------|------------------------------|-------------|---------|----------------|-------------|-------------|------------------|
| 1     | Burkholderia capacia Hi2424  | A0K4N0      | 3273.0  | 307            | 36          | 11.4        | 41.50            |
| 2     | Microscilla marina           | A1ZD79      | 3358.0  | 302            | 31          | 7.9         | 42.07            |
| 3     | Magnoparthe grisea           | A4R7D5      | 3487.0  | 324            | 42          | 9.0         | 44.68            |
| 4     | Populus trichocarpa          | B9H6W3      | 28464.0 | 266            | 28          | 10.5        | 41.42            |
| 5     | Pseudomonas aeruginosa       | A6V5Q2      | 33147.2 | 310            | 35          | 10.6        | 40.34            |

Pattern 2  ([AGN]-[KR]-H-R-K-L-[MK]-P-T-[AGN]-X-E-R)

| S. No | Organism                     | Access. No. | Mol wt. | Amino Acid No. | NCR\(^{2}\) | Alanine (%) | Instability index |
|-------|------------------------------|-------------|---------|----------------|-------------|-------------|------------------|
| 1     | Shewanella sediminis         | C8Q5J0      | 3273.0  | 307            | 31          | 7.5         | 43.43            |
| 2     | Actinobacillus minor NM305   | C5RYV4      | 3399.0  | 307            | 39          | 7.5         | 43.43            |
| 3     | Pirellula staleyi DSM 6068   | D2R9H8      | 3252.3  | 302            | 35          | 10.6        | 49.18            |
| 4     | Pantoea sp. At-9b.           | C8Q5J0      | 3273.0  | 306            | 35          | 10.5        | 40.97            |
| 5     | Burkholderia cenocepacia     | Q1BZ21      | 3273.7  | 307            | 36          | 11.4        | 41.50            |

Pattern 3  ([C-W-E-N-[HY]-M-P-[MLM]-A[L]-R-X-X-[ML]-Y])

| S. No | Organism                     | Access. No. | Mol wt. | Amino Acid No. | NCR\(^{2}\) | Alanine (%) | Instability index |
|-------|------------------------------|-------------|---------|----------------|-------------|-------------|------------------|
| 1     | Aspergillus niger            | A2QYH5      | 34964.7 | 320            | 36          | 8.8         | 41.75            |
| 2     | Shewanella peletana          | A8H6J4      | 34786.8 | 314            | 39          | 9.2         | 39.82            |
| 3     | Syntrophobacter fumaroxidans | A0LKP2      | 3601.14 | 328            | 44          | 11.0        | 41.02            |
| 4     | Burkholderia ambitaria       | B1FFB0      | 32904.0 | 307            | 38          | 11.1        | 40.05            |
| 5     | Algoriphagus sp. PR          | A3HT3       | 34734.8 | 305            | 43          | 4.9         | 41.30            |

Pattern 4  ([A-X-E-G-R-C-[FWV]-[LV]])

| S. No | Organism                     | Access. No. | Mol wt. | Amino Acid No. | NCR\(^{2}\) | Alanine (%) | Instability index |
|-------|------------------------------|-------------|---------|----------------|-------------|-------------|------------------|
| 1     | Syntrophobacter fumaroxidans | A0LKP2      | 36011.4 | 328            | 44          | 11.0        | 41.02            |
| 2     | Talaromyces stipitatus       | B8MN39      | 35560.5 | 325            | 41          | 10.2        | 46.51            |
| 3     | Uncultured organism          | Q6RWR5      | 33711.5 | 310            | 36          | 11.0        | 41.86            |
| 4     | Microscilla marina           | A1ZH2Q2     | 34937.1 | 311            | 38          | 9.0         | 41.02            |
| 5     | Rhodopseudomonas palustris   | D2MB94      | 33698.4 | 317            | 40          | 13.9        | 39.93            |

Table 3b: In silico analysis of some physiochemical properties of some aromatic nitrilase (sequences obtained from ScanProsite) from Manually Designed Motif analysis.

Results

Manually Designed Motifs were validated for the two groups of nitrilases i.e. aliphatic and aromatic and the results are shown in Table 1. Hits obtained from each manually designed motif (MDM) of aliphatic and aromatic nitrilases in SwissProt and TrEMBL databases are obtained through a motif search. These motifs are conserved for aliphatic and aromatic nitrilases (Figure 1a and Figure 1b). MEME version 4.8.1 (2012) software also confirmed the presence of all the MDM which were analyzed through ScanProsite and resulted in protein sequences only for nitrilase families. These sequences have conserved catalytic triad i.e. glutamine, lysine, and cysteine (Glu-48, Lys-131, Cys-165) which are essential for the nitrilases activity [18-21].

Designing and validation of motifs for aliphatic nitrilase

In the present study manually designed motifs for aliphatic nitrilases (MDMAr) were (i) [FL]-[ILV]-[AV]-F-P-E-[VT]-[FW]-[IL]-[PY]-P-[YW]; (ii) R-R-K-[L]-[KKR]-[PA]-T-[HY]-[VAH]-E-R; (iii) C-W-E-N-[HY]-[MLM]-A[L]-R-X-X-[ML]-Y and (iv) A-X-E-G-R-C-[FWV]-[LV]; ScanProsite analysis of these MDMAr resulted in 55, 104, and 128 and 105 protein sequences respectively. Out of these 392 protein sequences, 251 were of microbes, 28 of higher plants and 7 from animals. Forty one sequences belonged to uncultured organisms. Plant protein sequences were only obtained from first and second MDMAr and not from (iii) and (iv) manually designed motifs (Table 3). The designed motif pattern was found to be in between 40-60; 130 to 165; 160-180 and 1-30 amino acid numbers respectively (see additional file 1). Presence of MDMAr through Pratt analysis showed that the first, second and third MDMAr (design motif pattern) were in between 34 to 68, 125 to 177 and 157 to 215 numbered amino acids, respectively (see additional file 1).

Total amino acid numbers of all resulted sequences were found to be in between 330-382. Molecular masses were from 37226 to 41971 daltons, alanine content was in between 10-16%, negative charged residue were in between 37-52 and instability index ranged from 30.22 to 44.13 for the aliphatic nitrilase sequences (Table 3a).

Designing and validation of motifs for aromatic nitrilase

Manually designed motifs for aromatic nitrilase (MDMAl) were (i) [ALV]-[LV]-[FLM]-[P-E-[AS]]-[FLV]-[LV]-[AGP]-[AG]-Y-P; (ii) [AGN]-[KR]-H-R-K-L-[MK]-P-T-[AGN]-X-E-R; (iii) C-W-E-N-[HY]-[MLM]-A[L]-R-X-X-[ML]-Y and (iv) A-X-E-G-R-C-[FWV]-[LV]. ScanProsite analysis of all the four MDMAr resulted in 55, 104, 128 and 105 protein sequences respectively. Out of these 392 protein sequences, 251 were of microbes, 28 of higher plants and 7 from animals. Forty one sequences belonged to uncultured organisms. Plant protein sequences were only obtained from first and second MDMAr and not from (iii) and (iv) manually designed motifs (Table 3). The designed motif pattern was found to be in between 40-60; 130 to 165; 160-185; 200-220 amino acid respectively (see additional file 2). In order to confirm the ScanProsite results, Pratt analysis regarding the presence of MDMAr was done and it was found that first, second, third and fourth designed motif pattern (DMP) were in between 26 to 52, 125 to 180 and 206 and 191 to 216 amino acids, respectively.

In silico physiochemical analysis of the protein sequences retrieved from ScanProsite using first, second, third and fourth MDMAr exhibited the total number of amino acid ranged between 306-331.
molecular sizes were found to be between 32273–36940 dalton's, alanine content was between 4.9-11.0%, negative charged residues ranged from 34–44 and instability index was found to be 39.82-48.68 for the aromatic nitrilase sequences (Table 3b).

Phylogenetic analysis

Twelve protein sequences for nitrilases (aliphatic/aromatic) were subjected for the phylogenetic tree construction. Tree topology generated by Neighbor Joining Method (NJ) revealed that aliphatic and aromatic nitrilases were clustered with well distinct groups and well supported bootstrap value of 1000 (Figure 2).

Discussion

The expansion of molecular sequences and genomic databases has made the area of genome sequence data analysis more challenging and interesting to develop tools for rapid and reliable searching and analysis of the data. In this endeavor, database searching against gene/protein families with motifs has emerged as important strategy for efficient similarity searching [22]. While several domain/motif databases are being compiled, it is important to develop database search tools that fully utilize the conserved structural and functional information embedded in those sequence data to enhance the reliability of the search. Sequence motifs typically occur in a specific and known order in a sequence family. The ordering and spacing of motifs therefore, provide powerful additional criteria for classifying sequences into families. In this paper, we have manually designed groups of motifs (MDM) i.e. MDMal and MDMar each with four motifs for rapid and reliable nitrilase identification and compared it to the currently available methods, including the BLAST search, the PROSITE pattern search, Gblock [23] MEME version 4.8.1 (Multiple Em for Motif Elicitation) and the HMM method. All the designed motifs have one amino acid residue of active site, therefore when MDM were analyzed and validated through ScanProsite, resulting protein sequences were specifically belonging to nitrilase with conserved catalytic triad Glu–48, Lys–131, Cys–165 [18,24,20,15] and total number of hits were found to be significantly higher (578 aliphatic & 392 aromatic) when compared to databases such as BLAST and PROSITE.

Gblock eliminated poorly aligned and divergent region of the protein sequences for better phylogenetic analysis. Pratt results confirmed that the Manually Designed Motif for aliphatic nitrilase (MDMAI) ranged between 40-55 amino acid for pattern 2; 125-141 amino acid for pattern 3; 160-172 amino acid for pattern 4 (see additional file 1 & 2). This revealed that Pratt analysis for MDM are reliable and accurate tool for conducting search of similar sequence from databases as its results were similar to the Scan Prosite analysis as other similar sequence from databases as its results were similar to the ScanProsite analysis. One motif i.e. [VA]-A-X-[AV]-Q-[AI]-X-P-[VA]-X-[LF]-[SD] was not confirmed by the Pratt analysis as Pratt started the analysis after 30th amino acid and we have designed this MDM between 1-30 amino-acid (Table 2).

Multiple sequence alignment (MSA) in present study revealed that there are specific amino acids present at specific position which are responsible for the activity of nitrilases. Phylogenetic analysis by the Neighbor Joining (NJ) method clearly makes a distinction between the two groups of nitrilases i.e. aliphatic and aromatic nitrilases (Figure 2).

In our previous communication [13] we have reported that some of the physiochemical properties play a significant role in the substrate specificity of aliphatic and aromatic nitrilases. The physicochemical parameters analysis of these sequences further confirmed that all the sequences having a total number of amino acid between 330-382, molecular weight between 37226–41971 daltons, alanine content 10-16%, negatively charged residue (Asp+Glu) 37-52 and instability index 30.32-44.13 belonged to aliphatic nitrilases. Similarly sequences having a total number of amino acid between 300-330, molecular weight between 32273–36940 daltons, alanine content (%) 8-11%, negatively charged residue (Asp+Glu) 37-52 and instability index 39.82-48.68 were of aromatic nitrilases (Table 3a and 3b).

It is well known that uncultured organisms are the new sources of enzymes, antibiotics and drug discovery [25,26] In the present study, we have found nitrilase sequences belonging to uncultured organisms from the database (aliphatic-164, aromatic-41). The protein sequences of these uncultured organisms have all the specific properties responsible for the nitrilase activity. Further studies on these sequences may lead to find out specific nitrilases needed for transformation of nitriles in organic chemistry and industry.

Conclusion

Nitrilases are distributed among microorganisms, plants and some plant species. We here present computational analysis of two classes of nitrilase enzyme which includes the study of motifs, physiochemical and phylogenetic analysis which could be used as tool to predict and differentiate nitrilases on basis of their substrate specificity. The present analysis predicts new sources of nitrilases with study of common and important physiochemical/biochemical properties as they share a common ancestry and use of these observations will be useful to predict
molecular function of the randomly selected nitrilases. Computational analysis of various properties of nitrilases revealed differences and also MEME, Pratt and Gblock analysis have confirmed the presence of four motifs. Additionally this approach has also led us to find new sources of nitrilase across SwissProt/TrEMBL databases. MDM for aliphatic/aromatic nitrilases, have been validated with the results and these motifs will be of potential application for screening genome/protein sequence databases to find novel sources of nitrilases.

Acknowledgement
The authors are grateful to Department of Biotechnology, Ministry of Science and Technology, Govt. of India for funding.

References
1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.
2. Henikoff S, Henikoff JG (1991) Automated assembly of protein blocks for database searching. Nucleic Acids Res 19: 6565-6572.
3. Attwood TK, Beck ME, Bleasby AJ, Parry-Smith DJ (1994) PRINTS—a database of protein motif fingerprints. Nucleic Acids Res 22: 3590-3596.
4. Steele HL, Jaeger KE, Daniel R, Streit WR (2009) Advances in recovery of novel biocatalysts from metagenomes. J Mol Microbiol Biotechnol 16: 25-37.
5. Lefevre F, Jarrin C, Ginothac A, Auriol D, Nalin R (2007) Environmental metagenomics: An innovative resource for industrial biocatalysis. Informa Healthcare 25: 242-250.
6. Cowan DA, Arslanoglu A, Burton SG, Baker GC, Cameron RA, et al. (2003) Metagenomics, gene discovery and the ideal biocatalyst. Biochem Soc Trans. 32: 298-302.
7. Ferrer M, Martinez-Abarca F, Golyshin PN (2005) Mining genomes and 'metagenomes' for novel catalysts. Curr Opin Biotechnol 16: 588-593.
8. Krogh A, Brown M, Mian IS, Sjölander K, Haussler D (1994) Hidden Markov models in computational biology. Applications to protein modeling. J Mol Biol 235: 1501-1531.
9. Eddy SR, Mitchison G, Durbin R (1995) Maximum discrimination hidden Markov models of sequence consensus. J Comput Biol 2: 9-23.
10. Wu C, Whitson G, McLarty J, Ermongkonchai A, Chang TC (1992) Protein classification artificial neural system. Protein Sci 1: 667-677.
11. Yu C, Berry M, Shrivakumar S, McLarty J (1995) Neural networks for full-scale protein sequence classification: Sequence encoding with singular value decomposition. Machine Learning 21: 177-193.