Molecular Evolution of Miraculin-Like Proteins in Soybean Kunitz Super-Family

Purushotham Selvakumar · Deepankar Gahloth ·
Prabhat Pratap Singh Tomar · Nidhi Sharma ·
Ashwani Kumar Sharma

Abstract Miraculin-like proteins (MLPs) belong to soybean Kunitz super-family and have been characterized from many plant families like Rutaceae, Solanaceae, Rubiaceae, etc. Many of them possess trypsin inhibitory activity and are involved in plant defense. MLPs exhibit significant sequence identity ($\sim$30–95%) to native miraculin protein, also belonging to Kunitz super-family compared with a typical Kunitz family member ($\sim$30%). The sequence and structure–function comparison of MLPs with that of a classical Kunitz inhibitor have demonstrated that MLPs have evolved to form a distinct group within Kunitz super-family. Sequence analysis of new genes along with available MLP sequences in the literature revealed three major groups for these proteins. A significant feature of Rutaceae MLP type 2 sequences is the presence of phosphorylation motif. Subtle changes are seen in putative reactive loop residues among different MLPs suggesting altered specificities to specific proteases. In phylogenetic analysis, Rutaceae MLP type 1 and type 2 proteins clustered together on separate branches, whereas native miraculin along with other MLPs formed distinct clusters. Site-specific positive Darwinian selection was observed at many sites in both the groups of Rutaceae MLP sequences with most of the residues undergoing positive selection located in loop regions. The results demonstrate the sequence and thereby the structure–function divergence of MLPs as a distinct group within soybean Kunitz super-family due to biotic and abiotic stresses of local environment.

Keywords Rutaceae · Kunitz super-family · Miraculin-like proteins · Sequence analysis · Evolutionary relationships · Positive selection

Introduction

The groups of genes showing similarity with each other are referred to as a gene family, reflecting the assumption that all arise from a common ancestor. Gene families arise essentially by gene duplication, either by wholesale duplication of entire genes or by duplication and shuffling of exons from different genes (Conant and Wolfe 2008). Functional differences between duplicate genes can originate in several different ways, including mutations that directly impart new functions, subdivision of ancestral functions, selection for changes in gene dosage and gene regulation (Louis 2007). It has been established that the positive selection plays an important role in retention and diversification of duplicate copies in plant defense genes (Moore and Purugganan 2005).

The Soybean Kunitz super-family inhibitors are one of the important arsenals in plant defense mechanism. These proteins inhibit the enzymatic activity of digestive proteases of pests and pathogens important for their survival (Ryan 1990). Kunitz family inhibitors have been established as a rapidly evolving multigene family, and they are characterized by the presence of conserved N-terminal signature sequence (Talyzina and Ingvarsson 2006). Over the long period of time, these inhibitors have evolved from a common ancestral gene by gene duplication and gene conversion events and adapted themselves for defense role

P. Selvakumar · D. Gahloth · P. P. S. Tomar · N. Sharma ·
A. K. Sharma (✉)
Department of Biotechnology, Indian Institute of Technology
Roorkee, Roorkee 247 667, India
e-mail: aksbsfbs@yahoo.co.in

in different plant species as per biotic and abiotic stresses. In this process, sequence and thereby the structural and functional evolution in their specificities have occurred as per the local environmental conditions. In general, plant gene families are largely conserved. Much of the diversity is thought to have been caused by gene duplication and adaptive specialization of pre-existing genes. Neofunctionalization and subfunctionalization are considered to be important processes responsible for retention of duplicate genes (Flagel and Wendel 2009).

The miraculin-like proteins (MLPs) are a group of proteins which exhibit significant sequence identity (~39–55%) to miraculin protein. The native miraculin, a 24.6-kDa plant protein purified from red berries of *Richadella dulcifica*, possesses unique taste-modifying property (Theerasilp and Kurihara 1988; Hirai et al. 2010). Till date, there are no reports of any MLP having the taste-modifying property. Many of the characterized MLPs have been shown to have an important role in plant defense and possess trypsin inhibitory activity (Tsukuda et al. 2006). Recently, remarkable up-regulation of MLPs was observed at different stages of citrus huanglongbing (HLB) disease development (Fan et al. 2011). HLB is the most destructive citrus pathosystem threatening citrus production worldwide (Gottwald 2010; Callaway 2008).

Both miraculin and MLPs belong to Kunitz super-family and have shown amino acid sequence similarity (~30%) to soybean Kunitz family trypsin inhibitors. The plant Kunitz-type soybean trypsin inhibitors (STIs) are the proteins with a molecular mass of approximately 20 kDa with four cysteine residues arranged into two disulfide bridges (Richardson 1991). They play an important role in plant defense against pathogens and predators and are known to be involved in many biological functions, such as blood coagulation, platelet aggregation, and anti-carcinogenesis (Kennedy 1998; Oliva et al. 2000; Ryan 1978). The notable variations, however, have been reported among Kunitz inhibitors. Crystal structures of Kunitz-type inhibitors having single or no disulfide bridge have been reported (Cavalcanti et al. 2002; Hansen et al. 2007). The variations in sequence and structure with functional diversifications indicate that members of Kunitz family have undergone rapid evolutionarily changes because of varying selective pressures. Plant proteinase inhibitors are encoded by multigene family and have rapidly evolved in response to different biotic and abiotic stresses (Talyzina and Ingvarsson 2006).

The comparison of known MLPs with classical Kunitz family members has shown significant variations in sequence, structure, and function. The present study elucidates the molecular evolution of MLPs in soybean Kunitz super-family. The analysis was performed on 34 gene sequences which included newly cloned Rutaceae MLP type 1 and 2 genes, known MLPs, native miraculin, typical Kunitz family inhibitors which showed considerable similarity with MLPs, and classical STI. Our results demonstrate the sequence and thereby the structural and functional divergences of MLPs which have evolved into a distinct group within Kunitz super-family.

Materials and Methods

Genomic DNA was isolated from leaves of different members of Rutaceae family using CTAB plant genomic DNA isolation kit (Bangalore Genei, Bangalore, India). Two sets of primers were used to amplify orthologous and paralogous genes. The first set of primers designated MLP type-I were designed on basis of *Murraya koenigii* miraculin-like protein (MKMLP) gene sequence [MLP-1 Forward: 5′-AAT ACC ATG GGA TCC TTT GCT TGA TAT CAA TG-3′; MLP-1 Reverse: 5′-AAT ACT CGA GTC AAG ACA CGC ATG AG-3′]. PCR conditions were 94°C (4 min) 94°C (1 min)/ 62°C (1 min)/72°C (1 min) for 30 cycles, followed by 72°C (10 min). The second set designated as MLP type-II was designed targeting the conserved sites of MLPs that are available in the database [MLP-2 Forward 5′-CCT TCT TTC CTT ATC CTT (AG)CC TTG (AG)CC (AT)CA-3′; MLP-2 Reverse 5′-AAC CA(GC) AA(GC) CAG ACG T(AC)G AAC GCC ATC-3′]. PCR conditions were 94°C (4 min) 94°C (1 min)/66°C (1 min)/72°C (1 min) for 30 cycles, followed by 72°C (10 min). PCR products obtained were purified from 1% agarose gel using a gel elution kit from Zymo research and sequenced three times from Ocimum Biosciences, Hyderabad, India. The obtained sequences were submitted to NCBI Gene databank.

Multiple sequence alignments were made using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) taking default parameters. Coding sequences lacking signal sequences and few residues of C-terminus were taken for analysis. Sequences showing significant similarities to MLPs were taken from NCBI database (http://www.ncbi.nlm.nih.gov/) (Table 1). Phylogenetic trees were constructed by means of the MEGA version 5 program from amino acid alignments using the Maximum Likelihood method based on the JTT matrix-based model (Tamura et al. 2011). The reliability of the branching was tested by bootstrap statistical analysis (1,000 replications). Synonymous and non-synonymous substitution rates were calculated from a set of codon-aligned nucleotide sequences through SNAP server (Korber 2000). Selecton version 2.4 server was used for the identification of site-specific positive selection and purifying selection in a protein (Stern et al. 2007).

Structural modeling of a representative member of type 2 MLP protein of *Citrus jambhiri* (Cj2) was done using Modeller9v8 program (Sali and Blundell 1993), taking

Table 1 Sequences retrieved from NCBI database for phylogenetic analysis with accession numbers

Name of plant	GenBank accession number
M. koenigii MLP-I	FJ468002.1
Citrus × paradisi MLP-2	AF283533.1
Citrus × paradisi MLP-3	AF283534.1
Citrus × paradisi mRNA	AF283532.1
Citrus unshiu putative MLP-2	FJ532031.1
C. jambhiri Rlem MLP-1	AB213395.1
C. jambhiri Rlem MLP-2	AB213396.1
Populous trichocarpa Kunitz trypsin inhibitor	GQ184804.1
Solanum palustre MLP	AY395691.1
Nicotiana tabacum tumor related protein	U66263.1
N. tabacum Kunitz trypsin inhibitor	FJ494920.1
Theobroma bicolor trypsin inhibitor	AY753565.1
CoMir	ABK01288.1
Talisin	FJ436002.1
Lemir	U70076.1
RdMir	D38598.1
Arabidopsis lyrata trypsin inhibitor	XM_002890206.1
Citrus cv *shiranuhi* MLP-2	EF122397.1
C. unshiu putative MLP-2	FJ532030.1
Glycine max STI	X64447.1

Table 2 MLP sequences from different Rutaceae members submitted to NCBI database with accession numbers

Name of plant	GenBank accession number	Fragment size obtained in base pairs
Citrus limonia MLP-I	HM627638	516
C. limonia MLP-II	HM627644	519
Citrus aurantifolia MLP-I	HM627639	516
C. aurantifolia MLP-II	HM627645	519
Citrus maxima MLP-I	HM627640	516
Citrus maxima MLP-II	HM627646	519
Aegle marmelos MLP-II	HM627647	522
Citrus reticulata MLP-I	HM627642	516
Citrus reticulata MLP-II	HM627648	522
Citrus sinensis MLP-I	HM627641	516
Murraya paniculata MLP-I	HM627643	516
M. paniculata MLP-II	HM627649	522
M. koenigii MLP-II	HM627650	519
Citrus aurantium MLP-II	JF713054	519

MKMLP structure (PDB code-3IIR chain A) (Gahloth et al. 2010) as template. The best model was selected using Procheck after evaluating the stereo-chemical quality. Protein–protein docking of MKMLP and porcine trypsin (PDB code-1AVW chain A) was done through ZDOCK server (Chen et al. 2003), In Trypsin, active site residues (His57, Asp102, and Ser195) and in MKMLP putative reactive loop and N-terminal region are chosen to be present in interaction region. Various types of protein interactions between MKMLP and trypsin complex were predicted by PIC server (Tina et al. 2007). Tools from Expasy server were used to identify signal peptides, conserved domains, and various types of sequence motifs (Gasteiger et al. 2003).

Results

The genomic DNA was used for amplification of new MLP genes as they do not contain introns (Gahloth et al. 2010). Fourteen new MLP genes were amplified from members of Rutaceae family plants. Among them, six genes were obtained from the first set, and eight from the second set of primers. Around 516 and 519–522 base pair fragments were obtained after sequencing MLP type 1 and 2 genes,

respectively. All the 14 gene sequences have been submitted to the NCBI Gene databank (Table 2).

The dataset used for sequence and phylogenetic analysis included new Rutaceae MLP type 1 and 2 genes, known MLPs, native miraculin, typical Kunitz family inhibitors which showed considerable similarity with MLPs and classical STI. In total, 34 sequences were used in the analysis. Maximum Likelihood tree constructed from the complete dataset divided the above sequences into four major groups where group I and II cluster together suggesting that they belong to a common gene family (Fig. 1). Group I included type 1 MLPs, group II included type 2 MLPs, and group III consisted of native miraculin, and stress-induced MLPs. Talisin shared the common ancestor of Group I and II but formed separate group. Classical soybean Kunitz inhibitor did not feature in these three groups and formed a separate group. MLPs in group I and II represent orthologous gene copies separately in the respective groups.

Amino acid sequence analysis of all the new and old sequences presented interesting results. Multiple sequence alignments of new Rutaceae type 1 and 2 MLP sequences separately showed 98 and 75–95% identities, respectively among themselves, while 49–58% identity was observed between the two groups (Fig. 2). The sequence comparison of I and II groups with known MLPs and native miraculin protein showed ~40–95 and ~35% identities, respectively. With classical Kunitz inhibitor like STI, ~30% identity was observed with both group I and II MLP sequences. Group III MLPs among themselves showed 42–95% identity. Comparison with group I and II showed ~39–45 and ~30–43% identities, respectively (Fig. 3).

Fig. 1 Phylogenetic tree of MLP sequences was constructed by the maximum likelihood method dividing them into three groups. The *numbers above and below the branch points* indicate the confidence levels for the relationship of the paired sequences as determined by bootstrap statistical analysis. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site
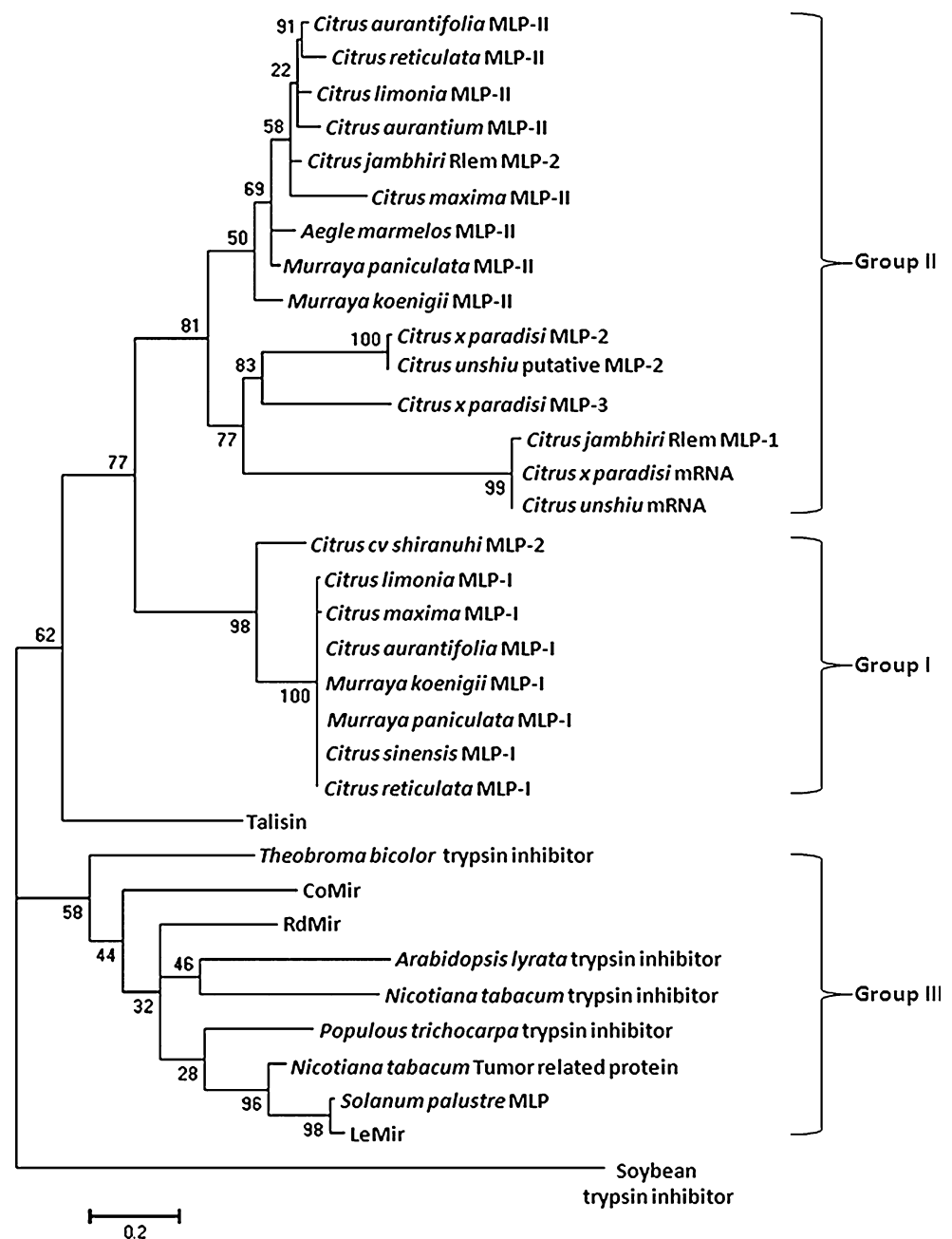
Except for native miraculin, no other group III MLPs showed the presence of two histidines (His30 and His60 miraculin numbering), which is considered to be responsible for taste modifying property.

The primary amino acid sequences of the most of the MLPs possess a conserved Kunitz signature motif [L/I/V/M]-x-D-x-[E/D/N/T/Y]-[D/G]-[R/K/H/D/E/N/Q]-x-[L/I/V/M]-(x)5-Y-x-[L/I/V/M] at the N-terminal of the sequence (Laskowski and Kato 1980) (Fig. 3). The presence of three disulfide bonds is a typical feature of MLPs, and their positions in the most of the sequences of type 1 and 2 are conserved. In classical Kunitz inhibitors, only two conserved disulfides are present. Multiple sequence alignment comparative analysis of sequences at reactive site loop revealed major changes at both the primary and secondary specificity sites in groups I and II MLPs as compared with classical Kunitz inhibitors. Compared to native miraculin and classical Kunitz inhibitors like STI, the conventional Arginine/Lysine at P1 position has been replaced by the conserved Asn65 residue in both groups I and II MLPs (Fig. 2). The differences in secondary specificity sites at all the positions were observed in MLPs when compared with native miraculin and soybean Kunitz trypsin inhibitor. In MLP type 1, Tyr63 (P3), Asn64 (P2),

Fig. 2 Multiple sequence alignments of new Rutaceae type 1 and 2 MLP sequences. Multiple sequence alignments of sequences (submitted to NCBI GenBank) were done by ClustalW, deleting the signal sequences and few C-terminal residues. The predicted Glycosylation site (*bold*), STI (Kunitz) protease inhibitors family signature (*gray shade*), and phosphorylation site (*underlined*) are indicated. The putative P1 residue is shown by the *arrow*, and conserved cysteines are shown in *boxes*. Abbreviations used: ClI, *C. limonia* MLP-I; ClII, *C. limonia* MLP-II; CaI, *C. aurantifolia* MLP-I; CaII, *C. aurantifolia* MLP-II; CmI, *C. maxima* MLP-I; CmII, *C. maxima* MLP-II; AmII, *Aegle marmelos* MLP-II; CrI, *C. reticulata* MLP-I; CrII, *C. reticulata* MLP-II; CsI, *C. sinensis* MLP-I; MpI, *M. paniculata* MLP-I; MpII, *M. paniculata* MLP-II; MkII, *M. koenigii* MLP-II; and CamII, *C. aurantium* MLP-II

```
ClI    DPLLDINGNVVEASRDYYLVSVIGGAGGGGLTLYRGRNELCPLDVIQLSSDLHKGTRLRF 60
CaI    DPLLDINGNVVEASRDYYLVSVIGGAGGGGLTLYRGRNELCPLDVIQLSSDLHKGTRLRF 60
CmI    DPLLDINGNVVEASRDYYLVSVIGGAGGGGLTLYRGRNELCPLDVIQLSSDLHKGTRLRF 60
CsI    DPLLDINGNVVEASRDYYLVSVIGGAGGGGLTLYRGRNELCPLDVIQLSSDLHKGTRLRF 60
CrI    DPLLDINGNVVEASRDYYLVSVIGGAGGGGLTLYRGRNELCPLDVIQLSSDLHKGTRLRF 60
MpI    DPLLDINGNVVEASRDYYLVSVIGGAGGGGLTLYRGRNELCPLDVIQLSSDLHKGTRLRF 60
ClII   EPLLDVNGNKVESTLQYYIVSAIWGAGGGGVSLHGGRNGYCPLDVIQLPSDTQNGIQLTL 60
CaII   EPLLDVHGNKVQSTLKYYIVSAIWGAGGGGVSLHGGRNGYCPLDVIQLPSDMQNGIELTL 60
CmII   EPLLDVNGNKVESTLQYHIVSAIWGAGGGGVSLHGGRNGYCPLDVIQLPSAIQRGKKLTL 60
AmII   EPLLDVNGNKVESTLQYYIVSAIWGAGGGGVSLHGGRNGYCPLDVIQLPSDMQKGIKLTL 60
CrII   EPLLDVNGNKVESTLKYYIVSAIWGAGGGGVSLHGGRNGYCPLDVIQLPSDMQNGIELTL 60
MpII   EPLLDVNGNKVESTLEYYIVSAIWGAGGGGVSLHGGRNGYCPLDVIQLPSDMQNGIKLTL 60
MkII   EPLVDVYGNKVESTLEYYIVSAIWGAGGGGVSLHGGRNGYCPLDVIQLSSDMENGIRLTL 60
CamII  EPLPDVNGNKVESPLQYYIVSAIWGGGGGGVSLHGGRNGYCPLDVIQLPFDKQNGIKLTL 60
       :** *: ** *::. .::**.* *.****::*: ***  ********.   ..* .* :

            P1
ClI    AAYNNTSIIHEAVDLNVKFSTETS----CNEPTVWRVDNYDPSRGKWFITTGGVEGNPGA 116
CaI    AAYNNTSIIHEAVDLNVKFSTETS----CNEPTVWRVDNYDPSRGKWFITTGGVEGNPGA 116
CmI    AAYNNTSIIHEAVDLNVKFSTETS----CNEPTVWRVDNYDPSRGKWFITTGGVEGNPGA 116
CsI    AAYNNTSIIHEAVDLNVKFSTETS----CNEPTVWRVDNYDPSRGKWFITTGGVEGNPGA 116
CrI    AAYNNTSIIHEAVDLNVKFSTETS----CNEPTVWRVDNYDPSRGKWFITTGGVEGNPGA 116
MpI    AAYNNTSIIHEAVDLNVKFSTETS----CNEPTVWRVDNYDPSRGKWFITTGGVEGNPGA 116
ClII   SPYNNSTIVRESTDLNLRFSVLLSGRDYCNEQPLWKVDNYDAASGKWFITTGGLDGHPGA 120
CaII   SPYNNSTIVRESTDLNLRFSVLLSGRDYCNEQPLWKVDNYDAASGKWFITTGGLDGHPGA 120
CmII   SPYDNSTIVRESTDLNLRFSGLLSESNRCNELPLWKVDNYDASSGKWFITTGGLDGHPGA 120
AmII   SPYDNSTIVRESTDLNLRFSVLLSGRDYCNEQPLWKVDNYDASAGKWFITTGGLDGHPGA 120
CrII   SPYNNSTIVRESTDLNLRFSVLLSGRDYCNEQPLWKVDNYDAASGKWFITTGGLDGHPNY 120
MpII   SPYDNSTIVRESTDLNLRFSVLLSGREYCNEQPLWKVDNYDASLGRWFITTGGLDGHPGA 120
MkII   TPYDNSTIVRESTDLNLRFSVLLSGREYCNEQPLWKVDNYDASSGKWFITTGGLDGHPGA 120
CamII  SPYNNSTIVRESTDLNLRFSVLLSGRDYCNEQPLWKVDNYDAASGKWFITTGGLDGHPGA 120
       :.*:*::*::*:.***::**   *    *** .:*:*****.: *:*******::*:*.

ClI    QTLKN-WFKLERVGTDQGTYEIVHCPSVCKSCVFLCNDVGV-SYDYR-RRLAF 166
CaI    QTLKN-WFKLERVGTDQGTYEIVHCPSVCKSCVFLCNDVGV-SYDYR-RRLAL 166
CmI    QTLKN-WFKLERVGTDQGTYEIVHCPSVCKSCVFLGNDVGV-SYNYR-RRLAL 166
CsI    QTLKN-WFKLERVGTDQGTYEIVHCPSVCKSCVFLCNDVGV-SYDYR-RRLAL 166
CrI    QTLKN-WFKLERVGTDQGTYEIVHCPSVCKSCVFLCNDVGV-SYDYR-RRLAL 166
MpI    QTLKN-WFKLERVGTDQGTYEIVHCPSVCKSCVFLCNDVGV-SYDYR-RRLAL 166
ClII   ETLLN-WFKLEKIGNFPGTYKIVHCPSVCESCVNICNNVGR-SFEDGVRRLLV 171
CaII   ETLLN-WFKLEKIGNFPGTYKIVHCPSVCESCVKLCNNVGR-SFEDGVRRLLV 171
CmII   ETLLN-WFKLEKTKNSRASY-IVHCPVCYSSFVLSSNIVGLSFEDGVRRLLV 171
AmII   ETLQN-WLKFEKIGNFPGIYKIVHCPSVCESCVKLCNSVGRFSYKEGVRRLVV 172
CrII   DAASGKWFKLEKIGNFPGTYKIVHCPSVCESCVKLCNNVGR-SFEDGVRRLLV 171
MpII   ETLLN-WFKLEKIGNLPRTYKIVHCPSVCESCVKLCNSVGRFSYEDGVRRLGL 172
MkII   ETLLN-WFKFEKLGTLLSAYKIVHCPSVCESCVRLCNRVGI-SSEDGVRRLLV 171
CamII  ETLLN-WFKLEKIGNFPGTYKIVHCPSVCKSCVKLCNNVGR-SFEDGVRHLLC 171
       ::  .  *:*:*:  .   *  ****.** **.. .: . :  *  .   *:*
```

Thr66 (P1′), and Ser67 (P2′); and in MLP type 2 Tyr63 (P3), Asp64 (P2), Ser66 (P1′), and Thr67 (P2′) are present in place of Pro61 (P3), Tyr62 (P2), Ile64 (P1′), and Arg65 (P2′) in STI. Likewise, native miraculin showed differences at secondary specificity sites where Asn66 (P3), Pro67 (P2), Glu69 (P1′), and Asp70 (P2′) are present in native miraculin. Interestingly, Asn13 which plays an important role in stabilizing the reactive loop conformation in STI is replaced by Ala13 and Ser13 in group I and II MLPs, respectively, whereas in group III MLPs, this position shows variability consisting of any one of Thr/Val/Pro/Ala (Fig. 3).

The sequences were compared for the presence of functional motifs. A significant feature is the presence of a phosphorylation motif, created by insertion of four amino acids at position 85 in type 2 MLP sequences, and its absence in type 1 MLPs. Also, differences in glycosylation sites were observed when group I and II MLPs were compared with native miraculin and STI. Glycosylation motif is located in the putative reactive loop of both type 1 and 2 MLPs (Fig. 2).

The three groups were analyzed separately as well as among themselves to assess the rates of synonymous and non-synonymous substitutions. The average dN and dS values compared within groups as well as for all pairwise comparisons across groups were compared, and Z-test results from MEGA 5 and SNAP server suggested that dS level was higher than dN implying that sequences are

```
MkI      DPLLDINGNVVEASRDYYLVSVIGGAG-GGGLTLYRGRNEL---CPL-DVIQLSPDLHKGTRLRFAAYNNT----SIIHEAVDLNVKFS----TETSCN---EPTVWRVDNYDPSRGKWF 104
MkII     EPLVDVYGNKVESTLEYYIVSAIWGAG-GGGVSLHGGRNGY---CPL-DVIQLSSDMENGIRLTLTPYDNS----TIVRESTDLNLRFSVLLSGREYCN---EQPLWKVDNYDASSGKWF 108
CrI      DPLLDINGNVVEASRDYYLVSVIGGAG-GGGLTLYRGRNEL---CPL-DVIQLSPDLHKGTRLRFAAYNNT----SIIHEAVDLNVKFS----TETSCN---EPTVWRVDNYDPSRGKWF 104
CrII     EPLLDVNGNKVESTLKYYIVSAIWGAG-GGGVSLHGGRNGY---CPL-DVIQLPSDMQNGIELTLSPYNNS----TIVRESTDLNLRFSVLLSGRDYCN---EQPLWKVDNYDAASGKWF 108
Ccvs     DPLVDVNGNKVESRDYYLVSAIRGAG-GGGLTLFRGRNEL---CPL-DVVQLSSDSERGTRLRFSMSDKT----SIINEDVDLNVRFS----TETRCN---EPTVWRVDSYDPSRGKWF 104
Cxp2     EPVLDIIGNKVKSTLEYKLIPVLDG---AGGFSTHLGTNGQ---CPL-DVIQLSSPSERGDNLRLLPYDNY----KIVKVSTDVNLRFPVKVSVK-PCN---EPLWKVDSYDDPL---- 101
Cxp3     EPVLDVNGNKVESNLEYKLLEVKNGT---SGGFSIHGGTSGE---CPL-DVVQLSSPTERDHYVRLLPFDNS----TVVRESTDLNL---------------IEPVWTVGNYNDSLGKWF 94
Cxpm     EPILDVYGNQVDSSHRYYLVSALWGVKTGGGISADKGKNGQ---CPT-DVIQLSPKDKRGKNLGLLPYDNS----TIVRESTNIKLKFS-RVSSLQQCN---KDSLWKVDNDNASLGKQF 108
Cu2      EPVLDIIGNKVKSTLEYKLIPVLDG---AGGFSTHLGTNGQ---CPL-DVIQLSSPSERGDNLRLLPYDNY----KIVKVSTDVNLRFPVKVSVK-PCN---EPLWKVDSYDDPL---- 101
Cum      EPILDVYGNQVDSSHRYYLVSALWGVKTGGGISADKGKNGQ---CPT-DVIQLSPKDKRGKNLGLLPYDNS----TIVRESTNIKLKFS-RVSSLQQCN---KDSLWKVDNDNASLGKQF 108
Cj1      EPILDVYGNQVDSSHRYYLVSALWGVKTGGGISADKGKNGQ---CPT-DVIQLSPKDKRGKNLVLLPNDNS----IIVRESTNIKLKFS-RVSSLQQCN---KDSLWKVDNDNASLGKQF 108
Cj2      EPLLDVNGNKVESTLQYYIVSAIWGAG-GGGVSLHGGRNGY---CPL-DVIQLPSDTQNGIKLTLSPYNNS----TIVRESADLNLRFSVLLSGRDYCN---EQPLWKVDNYDAASGKWF 108
RdMir    NPVLDIDGEKLRTGTNYYIVPVLRDH--GGGLTVSATTPNGTFVCPP-RVVQTRKEVDHDRPLAFFFENPKE---DVVRVSTDLNFSAF----MPCRWT-SSTVWRLDKYDESTGQYF 109
Pt       EAVLDTDGEKLRAGTEYYILPVFRGR--GGGLTMASTRDE--TCPL-DVVQDPLEISKGLPLTFTPVNPKK---GVIRVSTDLNIKFSAS----SIC--V-QSTVWKIQKSVNSEIQWF 104
All      EPVKDINGKPLLTKYNYYILPVFRGR--GGGLTMSNFK-NE--TCPR-SVIQDQFEVSQGLPVKFSPYD-KS---RIVRVSTDQNIKFSP---------TSIWELANFDERTSQWF 97
Sp       -EVVDIDGKILRTGVDYYILPVFRGR--GGGLTMDSIG-DK--ICPLDAVVQEHQEIDQGLPLTFTPVNPKK---GVIRVSTDLNIIFSAN----SIC--V-QTTQWKLDDFDETTGQYF 104
Nt-T     -AVVDIAGKKLRTGIDYYILPVFRGR--GGGLTLDSTG-NE--SCPLDAVVQEQQEIKNGLPLTFTPVNPKK---GVIRVSTDLNIKFSAA----SIC--V-QTTLWKLDDFDETTGKYF 104
Nt       SPVLDINGDKVKVGLNYFVLPVIRGR--GGGLLPSNVKQNN--TCPR-DIIQNSDEVQEGLPVVFAPFNTKK---GVVRLSIDLNVRFFTP----TIC--A-RETIWKLGTYDDDKLKQYF 105
Tb       SPVLDTDGDELRTGVRYYVVSSIWGAG-GGGLALGRATGQS---CPE-IVVQRRSDLDNGTPVIFSNADSK--DGVVRLSTDINIEFVP---IRDRLCS---TSTVWKLDNYDNSAGKWW 107
Talisin  EPLYDINGDEVVTGTKYYIVSAIRGAG-GGGLNLLPGRNEL---CPM-DVFQQSSDLQRGRPVMFFPVEYNGEKDNVVRTSTDMNIQFT---DPPNSCN--QYSAVWKVDNYDKKEGAWF 110
Lemir    PEVVDIDGKILRTGVDYYILPVFRGR--GGGLTMDSIG-DK--MCPLDAVVQEHNEIDQGLPLTFTPVDPKK---GVIRVSTDLNIIFSAN----SIC--V-QTTLWKLDDFDETTGQYF 105
CoMir    YPVLDINGEEIRPGVEYYIGTTFRP---GGGVTYGKGPGNE--ICPL-AVAQAW--LQRGLPVTFTPVNPEE---GVVRVSTDLNIKFAEP--PVGRICS---GSNVWKVH-FNDLFEKHF 104
Sti      DFVLDNEGNPLENGGTYYILSDITAF---GGIRAAPTGNER---CPL-TVVQSRNELDKGIGTIISSPYRIR----FIAEGHPLSLKFDSF----------AVIMLCVGIPTEWSVVED 98
             :  *   *.:      *  :            **.          **    : *         ..     :               .:      .:            :
```

```
MkI      ITTGGVEGNPGAQTLKN-WFKLERVGTD----------QGTYEIVHCPS-VCK-SCVFLCNDVGVSY---DYRRRLAL 166
MkII     ITTGGLDGHPGAETLLN-WFKFEKLGT--L---------LSAYKIVHCPS-VCE-SCVRLCNRVGISSE--DGVRRLLV 171
CrI      ITTGGVEGNPGAQTLKN-WFKLERVGTD----------QGTYEIVHCPS-VCK-SCVFLCNDVGVSY---DYRRRLAL 166
CrII     ITTGGLDGHPNYDAASGKWFKLEKIGN--F---------PGTYKIVHCPS-VCE-SCVKLCNNVGRSFE--DGVRRLLV 172
Ccvs     ITTGGVEGNPGAQTLKN-WFKFERIGRD----------RATYKIVHCPS-VCE-SCVKLCNDVGVSN---DHARRLAL 166
Cxp2     IVTGGSEGHPGAETLIN-WFKIEKAGNFSY--------GGVYNIVHCPS-VCL-SCPRRCNKIGVSTK--DGVRRLLL 166
Cxp3     LTTGGIIGHPGAKTLLN-WFKLEKVS-----------VSIYNLVHCPS-VCD-SCVKLCNKVGIFYV--DGVRRLVL 155
Cxpm     ITIGE------GKTCQN-FFKLEKVSASIFDMKIALDIPCLYKIVHCSTLVNG-SCDTLCKDVGVSNV--DGVQRLVV 176
Cu2      IVTGGSEGHPGAETLLN-WFKIEKAGNFSY--------GGVNIVHCPS-VCL-SCPRRCNKIGVSTK--DGVRRLLL 166
Cum      ITIGE------GKTCQN-FFKLEKVSASIFDMKIALDIPCLYKIVHCSTLVNG-SCDTLCKDVGVSNV--DGVQRLVV 176
Cj1      ITIGE------GKTCQN-LFKKLEKVSASIFDMKIALDIPCLYKIVHCSTPVNG-SCDTLCKDVGVSNV--DGVQRLVV 176
Cj2      ITTGGLDGHPGAETLLN-WFKLEKIGN--F---------PGTYKIVHCPS-VCE-SCVKLCNNVGRSFE--DGVRRLLV 171
RdMir    VTIGGVKGNPGPETISS-WFKIEEFCGSGF---------YKLVFCPT-VCG-SCKKVKCGDVGIYIDQ-KGRRRLAL 172
Pt       VTTGGVEGNPGIETITN-WFKIEKAGDD----------YKLVFCPT-VC--DCGALCRDVGIYIHD-NGVRTLSL 164
All      ISTCGVEGNPGQKTVGN-WFKIDEFEKD----------YKIRFCPT-VCN-FCKVICRDVGVFVQD--GKRRLAL 157
Sp       ITLGGNQGNPGRETISN-WFKIEKFDRD----------YKLLYCPT-VCD-FCKVICKEIGIFIQD--GVRRLAL 164
Nt-T     ITIGGNEGNPGRETISN-WFKIEKFERD----------YKLVYCPT-VCN-FCKVICKDVGIFIQD--GIRRLAL 164
Nt       IVTGGVEGNPGPQTLSS-WFKLEKLGTD----------YKFVFCPS-VCK-ICKVICKDVGIYTKD--GVRFLAL 165
Tb       VTTDGVRGQPGPNTLTS-WFKIERAG----------VLGYKFRFCPS-VCD-SCKTLCSDIGRHSDD-DGQIRLAL 169
Talisin  ITTNGVIGNPGPRTLMN-WFKLEKFSGG----------LNAYKIVHCPS-VCE-SCVTLCSDLGINY---DSKRRLVL 172
Lemir    ITLGGDQGNPGVETISN-WFKIEKYDRD----------YKLLYCPT-VCD-FCKVICRDIGIFIQD--GVRRLAL 165
CoMir    VLTDGVEGNSGCGTTAN-WFKIEAVGDR----------GYKLVFCPT-VCDSSSEAICKYVGIYHDD-DGTRRLAL 167
Sti      LPEGPAVKIGENKDAMDGWFRLERVSDDEFN---------NYKLVFCPQQAED----DKCGDIGISIDHDDGTRRLVV 163
             :           .  *::.          *::  .*.  .      *  :*      *  :*         *  :
```

Fig. 3 Multiple sequence alignments of MLP sequences from NCBI database and Rutaceae MLPs were done by ClustalW, deleting the signal sequences and few C-terminal residues. The predicted glycosylation site (*bold*), STI (Kunitz) protease inhibitors family signature (*gray shade*) and phosphorylation site (*underlined*) are indicated. Abbreviations used: MkI, *M. koenigii* MLP-I; MkII, *M. koenigii* MLP-II; CrI, *C. reticulata* MLP-I; CrII, *C. reticulata* MLP-II; Ccvs, *Citrus* cv *shiranuhi* MLP-2; Cxp2, *Citrus* × *paradisi* MLP-2; Cxp3, *Citrus* × *paradisi* MLP-3; Cxpm, *Citrus* × *paradisi* mRNA; Cu2, *C. unshiu* putative MLP-2; Cum, *C. unshiu* mRNA; Cj1, *C. jambhiri* Rlem MLP-1; Cj2, *C. jambhiri* Rlem MLP-2; Pt, *P. trichocarpa* Kunitz trypsin inhibitor; All, *A. lyrata* trypsin inhibitor; Sp, *S. palustre* MLP; Nt-T, *N. tabacum* tumor related protein; Nt, *N. tabacum* Kunitz trypsin inhibitor; Tb, *T. bicolor* trypsin inhibitor; and STI, *G. max* soybean trypsin inhibitor

under purifying selection (Table 3). To determine site-specific positive Darwinian selection and purifying selection, a web-server Selecton version 2.4 was used. The ratio between non-synonymous (K_a) and synonymous (K_s) substitutions at each site of the protein is graphically displayed on each site using a color-coding scheme indicating either positive or purifying selection (online resource 1). For type 1, based on K_a/K_s ratios, a total of 11 residues (Asp16, Leu33, Leu40, Leu52, Glu82, Trp103, Thr130, Gln132, Gly133, Thr134, and Phe149) were identified to be undergoing positive selection (K_a/K_s ratio >1) in MEC model. The residues Glu82, Gly133, and Thr134 had high K_a/K_s ratios of 1.8, 1.5, and 1.8, respectively. For type 2 MLPs, a total of 15 residues (Gln16, Leu33, Tyr40, Thr52, Leu82, Gly85, Arg86, Asp87, Tyr88, Trp107, Asn134, Pro136, Gly137, Thr138, and Lys153) were identified to be

undergoing positive selection (K_a/K_s ratio >1) in MEC model in Selecton server. The residues Leu82, Gly85, Arg86, Asp87, Pro136, Gly137, and Thr138 had high K_a/K_s ratios of 1.8, 1.5, 1.5, 2, 1.9, 1.5, and 1.8, respectively. For group III MLPs, a total of nine residues (Thr36, Val42, Val54, Phe85, Tyr108, Gly135, Ser136, Phe138, and Val153) were identified to be undergoing positive selection (K_a/K_s ratio >1) in MEC model in Selecton server. The residues Phe85, Gly135, and Phe138 had high K_a/K_s ratios of 1.8, 1.6, and 1.8, respectively. Significance test for positive selection was obtained by likelihood ratio test (LRT) and Akaike Information Criterion (AIC_c) (online resource 2). All the residues, except residues Leu33 in both type 1 and 2 MLPs, undergoing positive selection were located in the loop regions in crystal structures of MKMLP and Cj2 model, representatives of type 1 and 2 MLPs,

Table 3 Observed evolutionary properties of miraculin-like and its related proteins

Test	Parameters					
Tajima	m	S	p_s	Θ	π	D
	34	321	0.877049	0.214500	0.362282	2.630494
Z-test	Neutral/stat	Positive/stat	Purifying/stat			
	0.0000/−8.4051	1.0000/−8.8228	0.0000/8.3697			
SNAP	dS	dN	dS/dN	ps/pn		
	1.5059	0.4266	3.4243	1.9136		

The Tajima test statistic was estimated using MEGA 5. The analysis involved 34 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 366 positions in the final dataset

The Z-test was estimated in the overall population using MEGA 5. All positions containing gaps and missing data were eliminated from the dataset

m = number of sequences, S = number of segregating sites, $p_s = S/m$, $\Theta = p_s/a_1$, π = nucleotide diversity, and D is the Tajima test statistic

dS/dN, the ratio of synonymous to non-synonymous substitutions; ps, the proportion of observed synonymous substitutions; pn, the proportion of observed non-synonymous substitution

Fig. 4 Secondary structure analysis of MLPs was made by ESPRIPT server taking MKMLP structure (PDB id: 3IIR) as template. MkI (*M. koenigii* MLP), Cj2 (*C. jambhiri* Rlem MLP-2), Mir (miraculin) each representing group I, II, and III, respectively, and STI. β-Sheets, α helices, and disulfide linkages are shown as *arrows*, *spirals*, and *numbers* (1, 2, and 3), respectively. *PS* phosphorylation motif
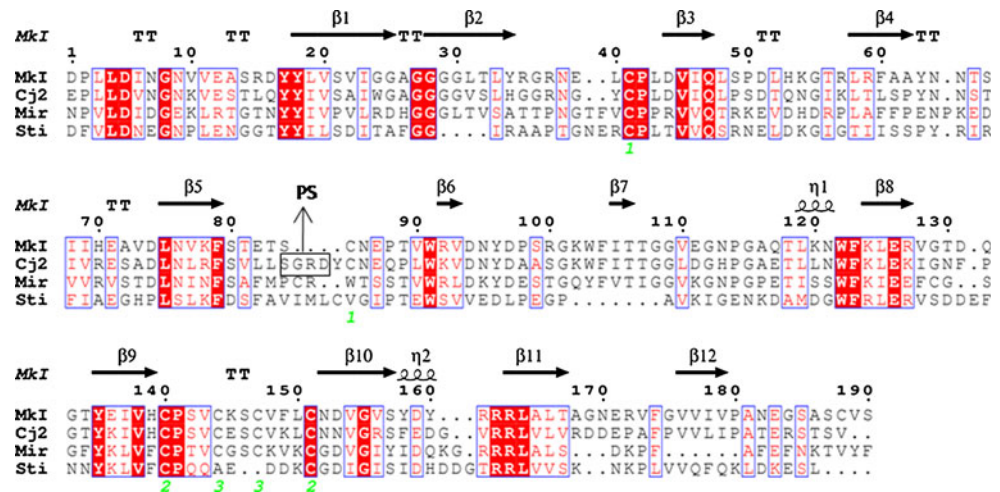
respectively. Residue Leu33 was located in β-strand. It is interesting to note that there are four additional residues in type 2 as compared to type 1, which are undergoing positive selection. These residues (85–88) are particularly noteworthy as they constitute phosphorylation motif created by insertion of four amino acids and are absent in type 1 MLPs.

Secondary structure analyses of MLPs and STI were done by ESPRIPT server (http://espript.ibcp.fr/ESPript/ESPript/) taking one member from each group. It showed that overall structures are conserved with the presence of three disulfide linkages in MLPs but only two disulfide linkages are present in STI (Fig. 4). The presence of 12 β-sheets and two short helices are predicted in all the three group members.

Overall model of Cj2 protein is similar to that of MKMLP consisting of β-trefoil fold made up of 12 antiparallel β-sheets connected by coils and two short stretches of α-helices (Fig. 5a). Procheck analysis shows only three residues (Ala120, His117, and Ala186) in disallowed

regions in the model. The Cj2 protein possesses six cysteine residues leading to three disulfide bridges. The central core superimposes well with the MKMLP structure with RMSD of 0.178, but subtle deviations are seen in conformations of connecting loops (Fig. 5b). A loop having four residues (85–88) insertion possessing phosphorylation motif is observed.

The MKMLP structure (PDB code: 3IIR Chain B) was docked on to the porcine trypsin (PDB code: 1AVW) by ZDOCK server to analyze the intermolecular contacts between MKMLP reactive loop and trypsin (Fig. 6a). The reactive loop interacting residues and types of interactions are shown (online resource 3). The mode of interactions observed in the MKMLP–trypsin complex at both primary and secondary specificity sites are significantly different from that of interactions in STI–trypsin complex. Asn65 (P1) in MKMLP is not able to interact extensively like Arg63 (P1) in STI with the residues forming S1 subsite of trypsin. It mainly interacts with Cys191, Gln192, Ser195, and Cys220 of trypsin (Fig. 6b) whereas Arginine in STI

Fig. 5 a The three-dimensional model of Cj2 was created by comparative modeling using the template model of MKMLP [PDBcode: 3IIR] from M. koenigii by MODELLER9v8. Central β-sheets, two short helices, three cysteines, and putative inhibitory loop are shown. b Superimposition of C-α atoms of MKMLP 3D-structure and Cj2 model
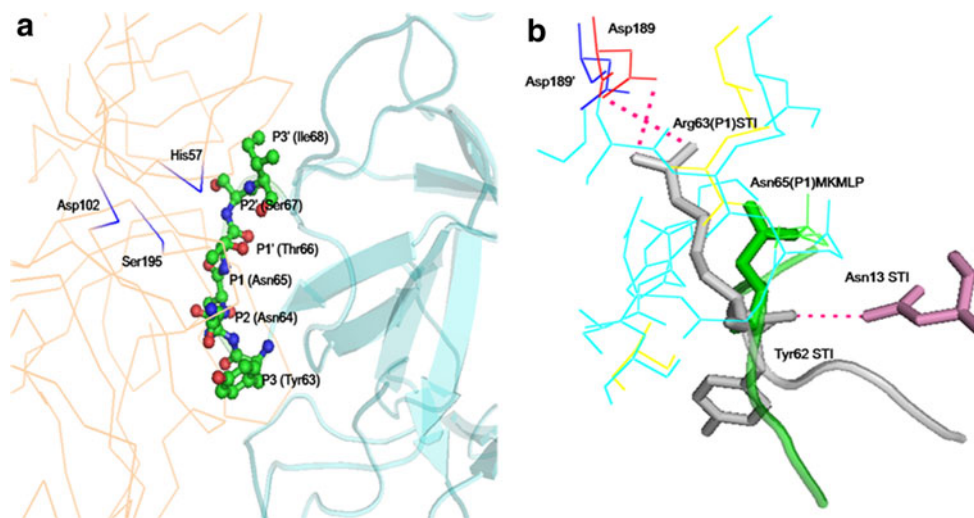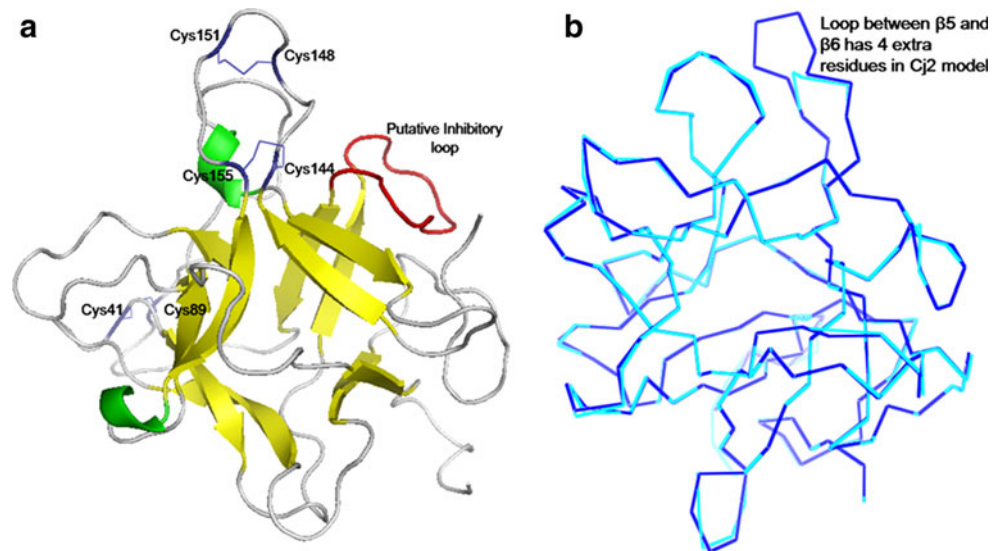
Fig. 6 a MKMLP–trypsin complex model: The MKMLP backbone is shown as cartoon and trypsin backbone is shown as C-α ribbon. Residues present at different positions (P1, P2, P3, P1′, P2′, and P3′) of the putative reactive loop of MKMLP are depicted as ball and stick representation. b Reactive loop of STI–trypsin complex superimposed on putative reactive loop of MKMLP–trypsin complex model. Arg63 of STI interacts with Asp189 of trypsin (lines), Asn65 of MKMLP cannot interact with Asp189′ of trypsin. Asn13 of STI interacts with Tyr62 and leads to stabilization of reactive loop. Interacting residues with P1 residues of STI and MKMLP are shown as lines

interacts additionally with Asp189, Ser190, Gly193, Asp194, Ser214, Trp215, Gly216, Gly219, and Gly226 of trypsin. It occupies the same position as Arginine or Lysine at P1 position in S1 subsite of trypsin, but it is too short to interact with Asp189, a key residue in binding pocket of trypsin. The geometry of the carbonyl group at the P1 position, important for the interaction between inhibitor and proteinase during catalysis, showed that carbonyl carbon atom is within van der Waals distance from Ser195 of trypsin. Also, carbonyl group forms a hydrogen bond with NH of Ser195 of trypsin. Asn64 (P2) does not interact with any of the residues of trypsin, whereas Tyr62 (P2) in STI interacts with His57, Phe94, Gly96, Leu99, Asp102,

Gln192, Ser214, and Trp215 of trypsin. Tyr63 (P3) in MKMLP makes hydrogen bond with Gln192 of trypsin and interacts with Leu99, Trp215, and Tyr217 of trypsin, whereas Pro61 (P3) in STI interacts only with Trp215 and Gly216 of trypsin. Thr66 (P1′) in MKMLP interacts only with Ser195 of trypsin, whereas Ile64 (P1′) in STI also makes contacts with Phe41, Cys42, His57, Gln192, and Gly193 of trypsin. In P2′ position, Ser67 of MKMLP makes one hydrogen bond with His40 and interacts with Phe41, Cys42, and Cys58, whereas Arg65 in STI interacts with His40, Phe41, Gly193, and Tyr151 of trypsin. A total of 12 hydrogen bonds present between MKMLP and trypsin involve residues from reactive site loop.

Discussion

This study focuses on sequence and thereby structural and functional evolution of MLPs within Kunitz super-family. MLPs constitute an important group of plant defense proteins within Kunitz super-family. It is to be noted that most MLPs have been characterized from Rutaceae family plants (Tsukuda et al. 2006; Gahloth et al. 2010), whereas most typical Kunitz inhibitors including STI have been characterized from Leguminosae family plants (Oliva et al. 2010). This shows that proteins of Kunitz super-family have evolved with refined specificities in plants of each family to serve as defense agents. MLPs show $\sim 30\%$ identity with classical soybean Kunitz inhibitor (STI) and $\sim 35\%$ identity to native miraculin, also a member of Kunitz family. Most MLPs possess conserved Kunitz signature motif (Fig. 2) and many characterized MLPs (Tsukuda et al. 2006; Shee and Sharma 2008) have shown trypsin inhibitory property and therefore, MLPs can be appropriately categorized under Kunitz family.

Phylogenetic analysis of MLPs along with STI demonstrated that MLPs and native miraculin clustered separately from STI implying that they have diverged long back from common ancestor. MLPs and native miraculin are broadly divided into three groups in phylogenetic analysis (Fig. 1). All the new MLPs cluster together in two groups, Groups I and II, in which group II can be further subdivided into two minor groups, whereas native miraculin and related MLPs cluster in a distinct group. There are orthologous and paralogous copies of type 1 and type 2 MLPs present in different members of Rutaceae family. Orthologous sequences across species are highly conserved implying functional maintenance. Multiple sequence alignments showed 98 and 75–95% identities among type 1 and 2 MLPs, respectively, while 49–58% identity was observed between the two groups. Group III MLPs showed 42–95% identity among themselves. Comparison with groups I and II showed ~ 39–45 and ~ 30–43% identities, respectively.

The data demonstrate that MLPs constitute a multigene family and have evolved from common ancestor by gene duplication, and later, functional diversification occurred because of speciation and specific local environments. Apart from trypsin inhibitory activity, new functions have been attributed to these proteins. They include taste-modifying property demonstrated only in native miraculin and antifungal property demonstrated in type 2 MLPs (Hirai et al. 2010; Tsukuda et al. 2006; Shee and Sharma 2007). The taste-modifying property has not been reported in MLPs. As demonstrated by mutagenesis studies, two histidine residues (His30 and 60 miraculin numbering), located in exposed regions, are considered responsible for miraculin taste-modifying activity (Ito et al. 2007). In type 1, only one of the Histidines is present (corresponding to His60) in exposed region, and the other is absent. Also, there are differences in gene expression patterns among these proteins. MKMLP, a type 1 MLP, is expressed constitutively in seeds and has been shown to be a seed storage protein (Shee and Sharma 2008), while type 2 MLPs characterized from *C. jambhiri* Lush, and tomato and coffee protein (*LeMir* and *CoMir*, respectively), both clustering with native miraculin in phylogenetic tree, are induced because of biotic and abiotic stress conditions like fungal, nematode, insect infestations ,and chemical treatments (Tsukuda et al. 2006; Brenner et al. 1998; Mondego et al. 2010). These observations suggest that regulatory differentiation might also take place allowing differential, spatial, and temporal gene regulations. These novel properties may have resulted from neofunctionalization.

In order to understand the selective pressure operating on the members of MLPs due to adaptive protein evolution, the average rates of non-synonymous and synonymous substitutions were calculated within and between phylogenetic groups. Surprisingly, the average dN/dS values for all pairwise comparisons were found to be less than 1, and other tests showed that sequences are under purifying selection (Table 3). It has been suggested that, when average rates are assessed, the positive selection gets masked, as only few residues in protein sequences are positively selected. Therefore, the site-specific positive Darwinian selection and purifying selection were estimated using a web server Selecton version 2.4. This server enables detecting the selective forces at a single amino-acid site by calculating the ratio of non-synonymous to synonymous substitutions (K_a/K_s ratio). Analysis of selection operating on type 1, type 2, and group III MLPs revealed a total of 11, 15, and 9 residues having K_a/K_s ratios >1, respectively, under MEC model and undergoing positive selection. All the residues were located in loop regions in all the three group members, except Leu33 (in type 1 and 2 members). Previous crystallographic analyses of the Kunitz inhibitors, interleukins-1β and 1α, and the acidic and basic fibroblast growth factors have shown that they contain β-trefoil fold (Murzin et al. 1992). Although these different proteins have very similar structures, many of their sequences have no significant similarities overall, therefore differences in function. Recent studies suggest that these proteins have similar key structural residues that are distributed symmetrically in their structures (Feng et al. 2010). MLPs are also characterized by β-trefoil fold with pseudo-threefold symmetry and consist of a six stranded-barrel capped by a triangular hairpin triplet. The loops connecting the β-strands vary in length and structure. The loops give the fold its varied binding capability, and the binding sites lie in different parts of the fold. In our results, we found that the residues undergoing positive Darwinian selection were in loop regions, suggesting evolution of

protein for novel interactions. We could recognize one such site at residue 40 in type 1 protein sequences undergoing positive selection where a change to amino acid proline at this site might render chymotrypsin inhibitory activity to the protein, as rest of the loop residues match exactly with double-headed, winged bean chymotrypsin inhibitors second reactive site loop (Dattagupta et al. 1999). In many cases, reactive loop residues have been observed to be likely targets for positive selection (Chakrabarty et al. 2006), but in some cases, reactive loop residues are not involved (Talyzina and Ingvarsson 2006). Amino acid divergences in the reactive loop are known to result in differences in binding affinity of inhibitors to various proteases. Although most residues undergoing positive selection in MLPs were located in loop regions, none of them was present in reactive loop. When type 1 and 2 MLPs are compared, the position of P1 residue asparagine is well conserved, but there are changes around the P1 residue in paralogous MLPs. We found that a few putative active site loop residues in MLPs are undergoing purifying selection perhaps to maintain reactive loop conformation. This suggests convergent evolution among paralogous MLPs and advocates that MLPs putative active site loop residues might have optimized for specific interactions, and purifying selection is acting to maintain it. Positive selection was also detected in predicted phosphorylation sites, and its implications are difficult to predict.

The comparison of amino acid sequences of MLPs among themselves, with miraculin and related proteins, and STI demonstrates notable features which points toward adaptive evolution and acquiring of new functions in this group of proteins. Many of these MLPs from both groups have been characterized at biochemical and structural levels. It has been shown that MLPs, like STI, possess trypsin inhibitory activity (Shee and Sharma 2008). However, drastic alterations in primary and secondary specificity sites have been observed when MLPs were compared with STI and native miraculin (Fig. 6). Compared to classical Kunitz inhibitors like STI and native miraculin, the conventional arginine/lysine at P1 position has been replaced by asparagine residue in type 1 and 2 MLPs, which is too short to interact with Asp189, a key residue in binding pocket of trypsin. This suggests that MLPs may not act as substrate-like inhibitors, as they lack arginine/lysine as active site residue, essential for trypsin specificity. The alterations in secondary specificity residues of putative reactive loop of MLPs as compared with STI will certainly result in differences in interactions with proteases important for activity. Also, subtle differences have been observed at secondary specificity sites among different MLPs, which may have resulted in response to specific needs. This clearly indicates alterations in specificities of Kunitz super-family inhibitors to counter digestive proteases of local pests and predators. Crystal structure analysis and biochemical characterization of one of the type 1 MLPs, MKMLP, helped in understanding some of the properties of this group of proteins (Gahloth et al. 2010; Shee and Sharma 2007). MKMLP in both native and heat-treated forms is remarkably stable against proteolysis, but its functional stability reduces with increase in temperature (Shee et al. 2007) as compared with soybean Kunitz trypsin inhibitors which are functionally stable even at high temperature. The remarkable structural stability of native and heat-treated proteins against proteolysis could be attributed to the presence of three disulfide bridges in the structure. The functional stability of STI has been attributed to the stabilization of reactive loop by a conserved Asn13 which forms a network of hydrogen bonds with reactive loop (Iwanaga et al. 2005). In type 1 and type 2 MLPs, the corresponding asparagine residue in STI is replaced by alanine and serine, respectively. Some of these type 2 MLP proteins have been shown to possess protease inhibitory as well as anti-fungal activities. The anti-fungal activity of type 2 MLPs has been attributed to the presence of thaumatin motif (Tsukuda et al. 2006). It has been suggested that the mechanism of anti-fungal activity may have nothing to do with trypsin inhibitory activity. Thaumatin and thaumatin-like proteins have pI > 8 and are proposed to bind to fungal cell wall components through an acidic cleft between their protein structures (Ghosh and Chakrabarti 2008) (online resource 4). Type 2 MLP sequences, however, have pI < 5.5, and the acidic residues distributed on surface are slightly clustered, but not present in cleft, which suggests a different mechanism of action. This clearly shows that the anti-fungal properties in MLP type 2 proteins have been acquired during the course of evolution to develop new specificities to counter local environmental stresses. A notable modification found only in type 2 MLPs is the presence of short phosphorylation motif created due to insertion of four residues after position 84, and all these four residues are undergoing strong positive selection. These features are absent in type 1 MLPs, native miraculin, and STI. Certainly, this will add to new unknown function in type 2 MLPs, which will be absent in other Kunitz super-family members.

From our data, we conclude that MLPs belongs to Kunitz super-family and represent a rapidly evolving gene family. Driven by gene duplication, neofunctionalization, and later by positive Darwinian selection, MLPs are being optimized for novel functions to counter local biotic and abiotic stress conditions.

References

Brenner ED, Lambert KN, Kaloshian I, Williamson VM (1998) Characterization of LeMir, a root-knot nematode-induced gene in tomato with an encoded product secreted from the root. Plant Physiol 118(1):237–247

Callaway E (2008) Bioterror: the green menace. Nature 452:148–150

Cavalcanti MDM, Oliva MLV, Fritz H, Jochum M, Mentele R, Sampaio M, Coelho LCBB, Batista IFC, Sampaio CAM (2002) Characterization of a Kunitz trypsin inhibitor with one disulfide bridge purified from *Swartzia pickellii*. Biochem Biophys Res Commun 291(3):635–639

Chakrabarty A, MacLean JA II, Hughes AL, Roberts RM, Green JA (2006) Rapid evolution of the trophoblast Kunitz domain proteins (TKDPs)—a multigene family in ruminant ungulates. J Mol Evol 63(2):274–282

Chen R, Li L, Weng Z (2003) ZDOCK: an initial stage protein docking algorithm. Proteins 52(1):80–87

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9(12):938–950

Dattagupta JK, Podder A, Chakrabarti C, Sen U, Mukhopadhyay D, Dutta SK, Singh M (1999) Refined crystal structure (2.3 Å) of a double headed winged bean chymotrypsin inhibitor and location of its second reactive site. Proteins 35(3):321–331

Fan J, Chen C, Yu Q, Brlansky RH, Li ZG, Gmitter FG Jr (2011) Comparative iTRAQ proteome and transcriptome analyses of sweet orange infected by "*Candidatus* Liberibacter asiaticus". Physiol Plant 143(3):235–245

Feng J, Li M, Huang Y, Xiao Y (2010) Symmetric key structural residues in symmetric proteins with beta-trefoil fold. PLoS ONE 5(11):e14138

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol 183(3):557–564

Gahloth D, Selvakumar P, Shee C, Kumar P, Sharma AK (2010) Cloning, sequence analysis and crystal structure determination of a miraculin-like protein from *Murraya koenigii*. Arch Biochem Biophys 494(1):15–22

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31(13):3784–3788

Ghosh R, Chakrabarti C (2008) Crystal structure analysis of NP24-I: a thaumatin-like protein. Planta 228(5):883–890

Gottwald TR (2010) Current epidemiological understanding of citrus huanglongbing. Annu Rev Phytopathol 48:119–139

Hansen D, Macedo-Ribeiro S, Veríssimo P, Yoo Im S, Sampaio MU, Oliva MLV (2007) Crystal structure of a novel cysteinless plant Kunitz-type protease inhibitor. Biochem Biophys Res Commun 360(4):735–740

Hirai T, Sato M, Toyooka K, Sun HJ, Yano M, Ezura H (2010) Miraculin, a taste-modifying protein is secreted into intercellular spaces in plant cells. J Plant Physiol 167(3):209–215

Ito K, Asakura T, Morita Y, Nakajima K, Koizumi A, Shimizu-Ibuka A, Masuda K, Ishiguro M, Terada T, Maruyama J, Kitamoto K, Misaka T, Abe K (2007) Microbial production of sensory-active miraculin. Biochem Biophys Res Commun 360(2):407–411

Iwanaga S, Yamasaki N, Kimura M, Kouzuma Y (2005) Contribution of conserved Asn residues to the inhibitory activities of Kunitz-type protease inhibitors from plants. Biosci Biotechnol Biochem 69(1):220–223

Kennedy AR (1998) Chemopreventive agents: protease inhibitors. Pharmacol Ther 78(3):167–209

Korber B (2000) HIV signature and sequence variation analysis. In: Computational analysis of HIV molecular sequences, Chap 4, pp 55–72

Laskowski M Jr, Kato I (1980) Protein inhibitors of proteinases. Annu Rev Biochem 49(1):593–626

Louis EJ (2007) Evolutionary genetics: making the most of redundancy. Nature 449(7163):673–674

Mondego JMC, Duarte MP, Kiyota E, Martínez L, de Camargo SR, De Caroli FP, Alves BSC, Guerreiro SMC, Oliva MLV, Guerreiro-Filho O, Menossi M (2010) Molecular characterization of a miraculin-like gene differentially expressed during coffee development and coffee leaf miner infestation. Planta 233(1):1–15

Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol 8(2):122–128

Murzin AG, Lesk AM, Chothia C (1992) β-Trefoil fold: patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. J Mol Biol 223(2):531–543

Oliva MLV, Souza-Pinto JC, Batista IFC, Araujo MS, Silveira VF, Auerswald EA, Mentele R, Eckerskorn C, Sampaio MU, Sampaio CAM (2000) *Leucaena leucocephala* serine proteinase inhibitor: primary structure and action on blood coagulation, kinin release and rat paw edema1. Biochim Biophys Acta 1477(1–2):64–74

Oliva MLV, Silva MCC, Sallai RC, Brito MV, Sampaio MU (2010) A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. Biochimie 92(11):1667–1673

Richardson M (1991) Seed storage proteins: the enzyme inhibitors. Methods Plant Biochem 5:259–305

Ryan C (1978) Proteinase inhibitors in plant leaves: a biochemical model for pest-induced natural plant protection. Trends Biochem Sci 3(3):148–150

Ryan CA (1990) Protease inhibitors in plants: genes for improving defenses against insects and pathogens. Annu Rev Phytopathol 28(1):425–449

Sali A, Blundell T (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815

Shee C, Sharma AK (2007) Purification and characterization of a trypsin inhibitor from seeds of *Murraya koenigii*. J Enzyme Inhib Med Chem 22(1):115–120

Shee C, Sharma AK (2008) Storage and affinity properties of *Murraya koenigii* trypsin inhibitor. Food Chem 107(1):312–319

Shee C, Islam A, Ahmad F, Sharma AK (2007) Structure–function studies of *Murraya koenigii* trypsin inhibitor revealed a stable core beta sheet structure surrounded by α-helices with a possible role for α-helix in inhibitory function. Int J Biol Macromol 41(4):410–414

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res 35(suppl 2):W506–W511

Talyzina NM, Ingvarsson PK (2006) Molecular evolution of a small gene family of wound inducible Kunitz trypsin inhibitors in *Populus*. J Mol Evol 63(1):108–119

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

Theerasilp S, Kurihara Y (1988) Complete purification and characterization of the taste-modifying protein, miraculin, from miracle fruit. J Biol Chem 263(23):11536–11539

Tina KG, Bhadra R, Srinivasan N (2007) PIC: protein interactions calculator. Nucleic Acids Res 35(Web Server issue):W473–W476

Tsukuda S, Gomi K, Yamamoto H, Akimitsu K (2006) Characterization of cDNAs encoding two distinct miraculin-like proteins and stress-related modulation of the corresponding mRNAs in *Citrus jambhiri* Lush. Plant Mol Biol 60(1):125–136