Undergraduate Theses   Undergraduate Works

5-13-2016

Sequencing and Analysis of a Miraculin Homolog from Two Ragweed Species

Kyj R. Mandzy
Bellarmine University, kmandzy01@bellarmine.edu

Follow this and additional works at: https://scholarworks.bellarmine.edu/ugrad_theses

Part of the Biochemistry Commons

Sequencing and Analysis of a Miraculin Homolog from Two Ragweed Species

April 21
2016

Our lab has previously isolated the gene for miraculin, a taste-modifying protein, from the pollen of *Ambrosia trifida*. This weedy plant, commonly known as Giant Ragweed, is a major producer of pollen that causes hay fever. The cDNA sequence of this gene was used to design PCR primers for the amplification and sequencing of genomic DNA from this species, as well as a related species, common ragweed (*Ambrosia artemisiifolia*). These genomic sequences were analyzed in terms of intron structure, phylogenetics, and protein structure and function. These gene sequences provide novel insights into the possible roles of proteases in plants.

Kyj Mandzy
Biochemistry
Honors Thesis

TABLE OF CONTENTS

Introduction

The average person in an industrialized country consumes 33.1 kg of sugar a year[3]. This accounts for a daily intake of 260 calories from sugar alone which is roughly one-eighth of the entire recommended daily caloric intake (2000 calories) the average individual needs for a day. This high intake of sugars has been implicated as one of the primary causes of obesity in the United States. Thirty-six of individuals in the United Sates are considered obese and a further 65% are considered overweight [1]. These individuals face a high number of health challenges and put a large strain on healthcare service costing $117 million annually[6]. Neuroimaging of the brain reveals that sugar consumption results in a dopamine release in the nucleus accumbens, an area associated with motivation, novelty, and reward[4]. Similar responses are elicited from the consumption of cocaine and heroin. Studies indicate that the responsiveness to sugars and sweetness has very ancient evolutionary beginnings[5]. One of the proposed methods of reducing the number of individuals suffering from obesity is to exchange the amount of real sugar in foods with alternative sweeteners which would not be as calorically dense but still taste similarly.

People taste sugar with their tongues, where the molecule binds to specific receptors and triggers a sensation of sweetness. Receptors for each of the basic tastes (sweet, sour, umami, spicy, and bitter) are found in each taste bud. Sweetness is detected by a variety of G protein gustducin coupled receptors found on the taste buds[7]. Incoming sweet molecules bind to their receptor and cause a conformational change in the molecule. The change activates the gustducin, which in turn, activates adenylate cyclase, catalyzing the conversion of ATP to cAMP. The cAMP molecule then

ultimately closes a potassium ion channel through an enzymatic cascade [7]. The concentration of excess potassium ions increases inside the cytoplasm and depolarizes the cell, which causes a release of dopamine, which is then received by a primary afferent neuron[7] . The individual then perceives a sweet taste. Measurement of this perceived sweetness is based on sucrose, with all sweeteners rated relative to that sugar. Sucrose has a sweetness index of 1 while most chemical sweeteners such as Stevia or Aspartame have sweetness indexes above two hundred[8]. Sweet-tasting proteins like saccharin are one of the many methods of altering the perception of a substance[2]. Taste-modifying proteins like curculin bind to the taste receptor itself and modify the taste of substances rather than simply amplifying the sensation of sweetness. These proteins would result in a different sensation of taste due to the different receptor activation.

Miraculin is a taste-modifying enzyme that is extracted from the fruit of *Synsepalum dulcifiurum* (Miracle fruit), a berry from West Africa. It is one of the more well-known taste-modifying proteins and is readily available for consumers to purchase. Ingestion of miraculin results in a perception of food being sweetened with a 17% sucrose solution. Miraculin induces a perceptional change which lasts an average of 20 minutes[2]. Native miraculin appears both as a dimer and a tetramer but has only been expressed in recombinant species as a dimer. The glycosylation of the miraculin structure is essential for its taste modifying mechanism and serves to support its 3D conformation[9] . Miraculin is classified as a Kunitz-type soybean trypsin inhibitor (STI) based on its primary protein structure. Similar to other members of this protein family, miraculin has a characteristic N-terminus, several cysteine residue disulfide bridges, and glycosylation

sites. How miraculin acts as a taste-modifying enzyme is not fully understood but the current model is based on two of its Histidines.

Miraculin is thought to have an invasive mechanism which changes the structure of taste receptors on the cells of the tongue. It is believed that one active site of Miraculin strongly binds to the taste receptor membrane, while the second active site fits into the sweet receptor site and changes its conformational shape[10]. Molecular modeling of miraculin's interaction with the human taste bud has focused on the interaction of the T1R2-T1R3 sweetness receptor. Histidine residues 29 and 59 on miraculin are believed to be responsible for the binding of the receptor to the negatively-charged cavity of the T1R2-T1R3 receptor. Many charged residues (four arginine, three aspartic acids, two glutamic acids, and one lysine) are in proximity to the histidine 29 residue which results in such a physioelectirc environment. Mutation studies suggest that the conformation of the two histidine residues (29 and 59) are responsible for this taste-modifying behavior[10]. Mutations of the histidine by replacing it with alanine residues, making them non-polar, generated miraculin with little to no sweetness-inducing ability [10]. The histidine residue 29 is theorized to act as the binding agent to the taste receptor at the intersection of the two monomer subunits. Figures 1 and 2 depict the structure of miraculin as it aligns to the T1R2-T1R3 receptor and the mechanism of perceived sweetness, respectively.
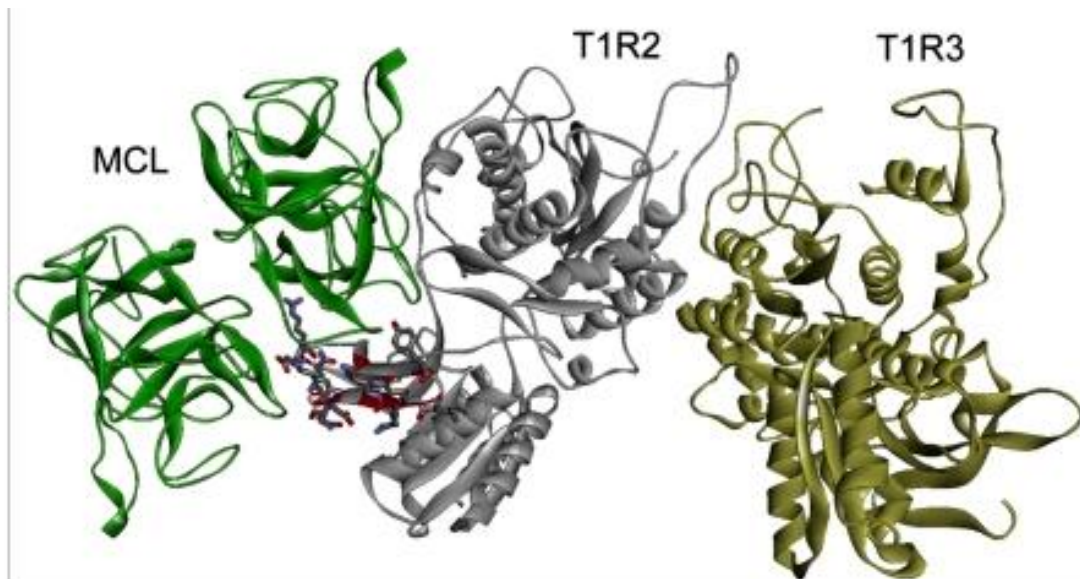
Figure 1.
Configuration of miraculin taste-modifying protein to the T1T2-T1T3 sweet-taste receptor. Koizumi, Ayako, *et al.* 2011
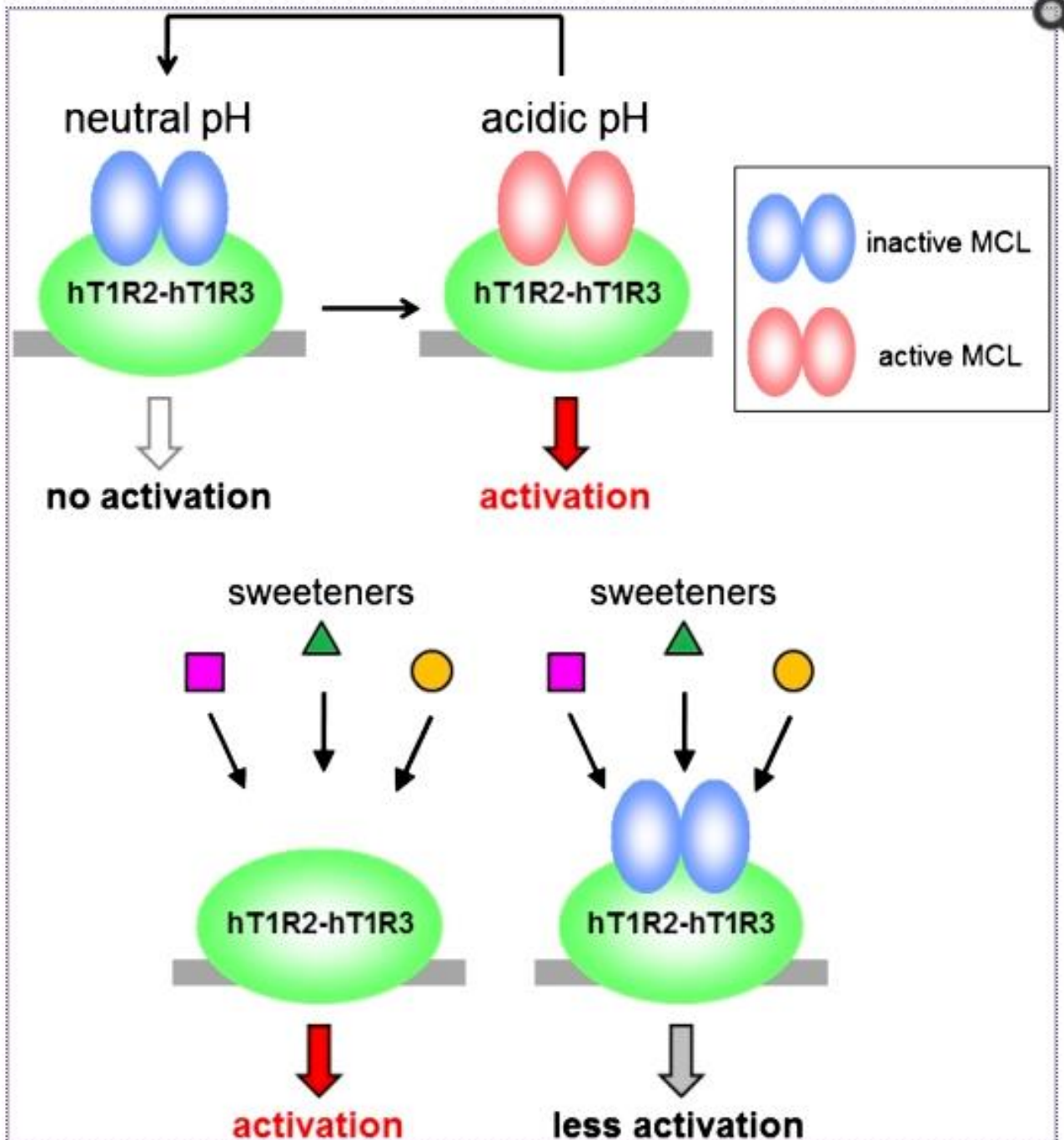
Figure 2.
Graphical representation of the interaction of the Miraculin taste-modifying protein to the T1T2-T1T3 sweet- taste receptor. Koizumi, Ayako, *et al*. 2011

There currently exists no identified genomic DNA sequence for miraculin, neither from *Synsepalum dulcifiurum* or any other plants. Genomic DNA refers to the sequence of nucleotides that resides within the nucleus. All published miraculin genes have been from cDNA. cDNA refers to the DNA sequence of the mRNA which makes up the codons for the amino acids in the expressed protein. cDNA lacks introns and is mostly the exons (protein coding region.) The full genomic sequencing of the DNA sequence of miraculin from ragweed would grant several insights. Currently, there exists a cDNA sequence of a giant ragweed miraculin homolog (clone # 153) collected from pollen that was isolated in the Robinson lab at Bellarmine University in the summer of 2014. That cDNA sequence presents a template for primer design with which to amplify genomic ragweed miraculin DNA. The sequencing of a variety of miraculin DNA sequences would allow analysis of gene structure, phylogenetics as well as a novel source of insight into the evolutionary development of both taste-modifying proteins and Soybean Kunitz-type trypsin inhibitor proteins[18]. Soybean Kunitz-type trypsin inhibitor proteins function primarily as protease inhibitors, many function as either a form of protection from predators by their anti-nutritive effects, as anti-carcinogens, or simply as a component of homoeostasis. However, a variety of other proteins fall into the same family based on their protein structure. Two atypical proteins of note in the Soybean Kunitz-Type family are miraculin, the taste modifying protein previously mentioned, and Lol p XI, a major grass pollen allergen[18,19]. This implies that the research may yield information which identifies the miraculin homolog as either a potential allergen or perhaps even as a novel anti-carcinogen[17].

Kunitz-Type Soybean trypsin inhibitors refer to a family of peptidase inhibitors. Exclusively found within Eukarya, two subfamilies exist: I3a and I3b. The I3a nomenclature refers to soybean trypsin inhibitors while I3b nomenclature refers to proteinase B type inhibitors. Both subfamilies are highly similar and are simply historically different; contemporary BLAST analysis has shown them to be 98% identical. The peptidase family I3 falls within the IC peptidase clan and is characterized by a specific Beta-trefoil [24] fold found within all of the species of inhibitors. The two subfamilies have differing disulfide bond patterns; however, both share the same reactive sites bound in a singular disulfide bond. The ability of the I3 protein family to inhibit multiple peptidases implies that structurally dissimilar peptidases can be inhibited by a single mechanism.

In regards to inhibitory protein families, there are two major types of interactions: 1) trapping reactions, which upon the cleavage of an internal peptide bond in the inhibitor triggers a conformational change that alters the ability of the inhibitor to function by keeping it bound to the inhibitor, and 2) reversible tight-binding interactions, when the inhibitor has a high-affinity for the active site of the target enzyme[24]. The tight-binding interaction is the more well-known mechanism and is referred to in literature as the standard mechanism. The inhibitory unit of a standard mechanism inhibitor has a single reactive-site peptide bond, and inhibition is caused by the binding of the inhibitor to the enzyme in a substrate-like fashion. Crystallographic structures of enzyme–inhibitor complexes show that the I3 family of proteins employs the standard mechanism of inhibitors [24].

The Soybean Kunitz-type inhibitors (peptidase class inhibitor I3) has been reported to have overlapping inhibitory sites for trypsin and papain, which imply a modular means of creating an inhibitory effect to the target proteinase. The reactive-site loops, resulting from the disulfide bridges found within the reactive site, are able to adopt the canonical conformation, which inhibit serine endopeptidases by the standard mechanism, but can be altered to match a wide variety of target ligands by the supporting scaffold structures of Kunitz-Type STI's. Figure 3, depicts the reactive loops found in I3a and I3b.
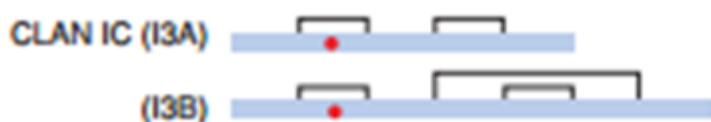

Figure 3.
Kunitz-Type STI conserved domain structures for both the I3a and I3b subfamilies. The disulfide bridges appear twice in the I3a subfamily and thrice in the I3b subfamily. The reactive loop site is represented by the red dot.
Rawlings, Neil D., Dominic P. Tolle, and Alan J. Barrett 2004

The Beta-Trefoil structure, by which proteins are placed into the Kunitz-STI superfamily, was first observed in Soybean trypsin inhibitors[24]. The trefoil consists of 12 strands of Beta sheets forming six hairpin structures, three of the hairpins then form a central barrel structure and three other hairpins form a triangular arrangement on one side. The resulting Y-shape shows an emerging Pseudo-3 fold axis which suggests that the gene emerged from the triplication of a singular trefoil unit. The trefoil units are then able to align into greater complex which allows for inhibition due to tight-binding to the substrate.

The triangular arrangement of the three hairpins is formed by hydrogen bonding. The coiling of the Beta sheets results in the inner strands of the trefoil closest to the 3-

fold axis to form hydrogen bonds with each other. The outermost Beta-sheet coils are only able to form hydrogen bonds with the inner triangle and not with each other. This symmetrical arrangement is made possible by two hydrogen bond pairs at each level, resulting in a total of four pairs of hydrogen bonds on each inner strand. The loops forming the triangular alignment correspond with 70% of the total surface of the protein with each loop varying wildly in size and sequence. While the Kunitz STI superfamily greatly shares the tertiary structure of the Beta-Trefoil, the overall primary and secondary structure differs greatly between individual inhibitors.

The side chains of the Beta-sheet amino acids are grouped in the alignment of the layers of the Beta-Trefold, the inner Beta-sheet alignment within the Beta-trefoil the outer two layers being generally composed of a variety of side chains, each being buried in the overall structure. The innermost layer is primarily composed of hydrophobic residues with a great deal of larger residues such as Tryptophan and Phenylalanine resulting in steric conditions which most likely benefit the interlocking of the Beta-Trefold by allowing the formation of the hydrogen bonds on the outside of each inner Beta-sheet coil. Figures 4 and 5 show the structure of the Beta-Trefoil and the hydrogen bonds necessary for its construction.
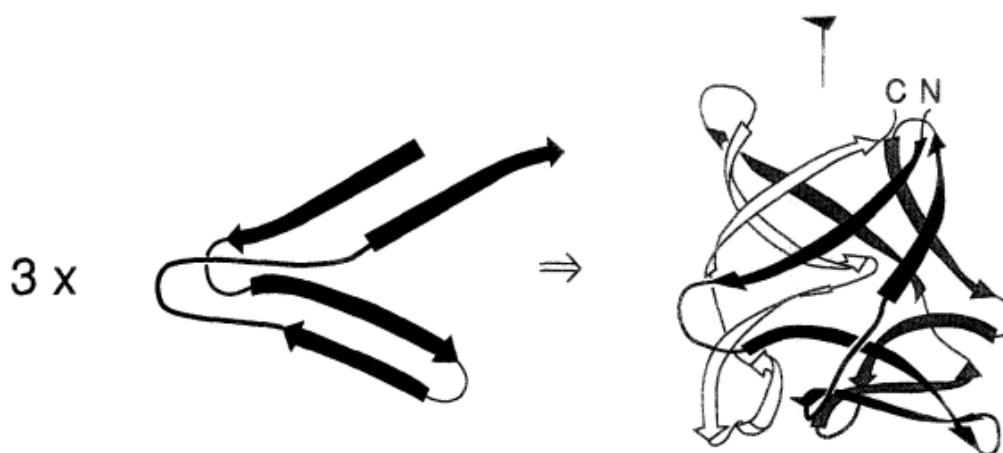
Figure 4. Lebeda, Frank J., *et al.* 1998
Schematic representation of the Beta-Trefoil found within the Kunitz-type STI family.
The individual trefoil unit (left) combine into the compound Beta-Trefoil (right) resulting
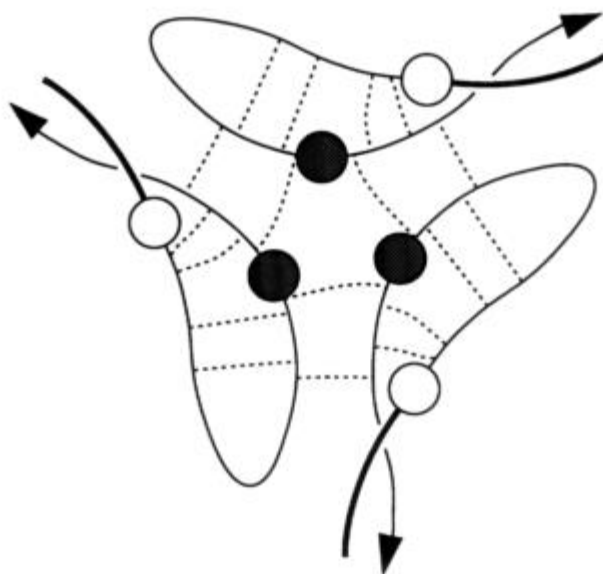in the barrel shape arrangement.

Figure 5. Lebeda, Frank J., *et al.* 1998
Schematic representation of the hydrogen bonds found within Beta-Trefoil. Four pairs
of hydrogen bonds branch from each inner Beta-sheet. Two pairs align to the other
two inner Beta-sheets while two simply extend to the outer Beta-sheet coils.

Allergen Ra3, the major allergen of Ragweed, is quite similar to miraculin and other STI proteins[20]. Like Lol p XI (the most common form of grass allergen), it also binds to IgE and shares both disulfide bridges and Glycosylation[19]. Ra3 binds to IgE through several disulfide bridges between its Cysteine amino acids and has a similarly active basic tail[21] .

The aim of this research in this study is to sequence and characterize the genomic DNA sequence of Miraculin homolog from *Ambrosia trifida* and the other plant species. The known cDNA sequence was used to design PCR primers to obtain the genomic DNA sequence of Miraculin. Amplified DNA was then be sequenced and analyzed. The multiple sequences were used for intron analysis, phylogenetics, and protein structure studies.

Experimental Methods

The cDNA sequence of Miraculin from *A. trifida* was used to provide a starting point for the designing of PCR primers in order to amplify the genomic DNA sequence. The sequence was utilized in a BLASTn search to find similar sequences and a consensus sequence was designed. The consensus sequence was the basis from which primers would be designed for performing PCR. Analogous annealing temperatures and purine/pyrimidine ratios were also considered in primer creation. Two pairs of primers were created, one utilizing 10 homologous plant sequences the other utilizing 12 plant sequences. Once obtained, an optimization PCR (Figure 6) was run to determine the ideal means of amplification of the genomic sequence of miraculin. Of the four different primer combinations, Primers F1 and R2 were ultimately deemed the most efficient at a 10 ng concentration of DNA and an annealing temperature of 42.5^{o}C. The

two primers were then used to successfully run PCR for the remaining three other plants of interest *Ambrosia artemisiifolia*, *Solidago rigida*, and *Aristolochia serpentaria*. DNA was extracted from the leaves of two plant species by means of a BIORAD Nucleic Acid Extraction kit (Hercules, CA). The DNA was then amplified by PCR and purified by means of BIORAD Quantum Prep PCR Kleen kit (Hercules, CA). The DNA was then sent to the University of Louisville Biochemistry Department DNA-Core for sequencing. Once the sequences were obtained, contigs were constructed from sequences by CAP3 sequence assembly program (doua.prabi.fr/software/cap3) and trimmed for extraneous sequences. The contig nucleotide sequences were then run through BLAST searches of both the DNA sequence as well as the projected protein sequence in order to be compared to other known sequences. The NetPlantGene Server was sued for intron analysis. Several other online resources were utilized: CFSSP (cho-fas.sourceforge.net/index.php) and Phyre 2.0 (http://www.sbg.bio.ic.ac.uk/phyre2/) to construct accurate models for the primary, secondary, and tertiary models of protein translated from the ragweed contigs.
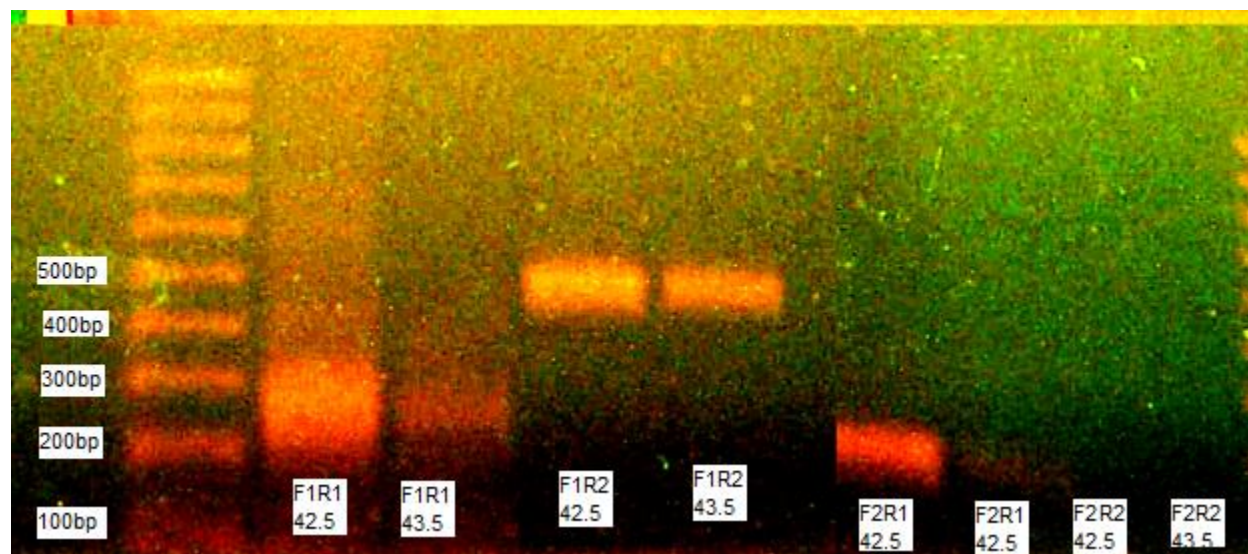
Primers utiized for PCR amplification

Primer F1
5' CACACAACATCATGAAAAC 3'

Primer R2
5' TCCTTGCCTATTTATTATCG 3'

Figure 6.

Optimization of PCR primers with various combinations of primers shown. The combination of F1R2 at the annealing temperature of 42.5 °C resulted in the brightest bands at the requisite size.

Results

The following contigs are consensus sequences crafted from a minimum of 4

sequencing DNA reactions. The consensus sequences were constructed by CLUSTAL

W alignment.

DNA_Contig_for Giant_Ragweed

```
AATATTCTCATTTTCTTTAGCCCAACCCTCACCCGATGCGGTTCGTGACATCAATGGCAACTTA
CTCCGATCGGGCACAAACTACTACATCCTCCCCGTCTTCAGAGGACGGGGCGGAGGCGTAACAC
TAGCCCCCACACGAAACGAGTCATGCCCACTTGATGTAGTCCAAGAAGGCTTCGAGCTACAGAA
TGGCCTCCCCTTAACATTCACACCTGTGGATCCCAAGAAAGGCGTTATACGCGAATCCACGGAT
CTTAACATTATTTTTTCGGCTTCGTCAATCTGTATTCAGTCGAACGTTTGGATGGTTGAAGAGT
ACCAAGGGCAACTTATTGTTACGGGTCATGGAACGGCTGGAAACCCTGGTCAGGGAACATTAAG
CAACTGGTTCAAGATAGAGAAGTATGAAGATGATTATAAGCTTGTGTTCTGTCCTACAGT
```

Final protein for the Giant Ragweed Contig (+2 reading frame):

```
TWIFSFSLAQPSPDAVRDINGNLLRSGTNYYILPVFRGRGGGVTLAPTRNESCPLDV
VQEGFELQNGLPLTFTPVDPKKGVIRESTDLNIIFSASSICIQSNVWMVEEYQGQLIVTG
HGTAGNPGQGTLSNWFKIEKYEDDYKLVFCPT
```

DNA_Contig_for Common_Ragweed

```
TTTTCTTTAGCCCAACCCTCACCAGATGCGGTTCGTGACATCAATGGCAACTTACTCCGGTCGG
GCACAAACTACTACATCCTCCCCGTCTTCAGAGGACGGGGCGGAGGCGTAACACTAGCCCCCAC
ACGAAACGAGTCATGCCCACTTGATGTAGTCCAAGAAGGCTTCGAGCTACAGAATGGCCTCCCC
TTAACATTCACACCTGTGGATCCCAAGAAAGGCGTTATACGCGAATCCACGGATCTTAACATTG
TTTTTTCGGCTTCGTCAATCTGTATTCAGTCGAACGTTTGGATGATTGAAGAGTACCAAGGGCA
ACTTATTGTTACGGGTCATGGAACGGCTGGAAACCCTGGTCAGGAAACATTAAGCAACTGGTTC
AAGATAGAGAAGTATGAAGATGATTATAAGCTTGTGTTCTGTCCTACA
```

Final protein for the Common Ragweed Contig (+1 reading frame):

```
FSLAQPSPDAVRDINGNLLRSGTNYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELQNGLP
LTFTPVDPKKGVIRESTDLNIVFSASSICIQSNVWMIEEYQGQLIVTGHGTAGNPGQETLSNWF
KIEKYEDDYKLVFCPT
```

Figure 7.
Consensus sequences formed from the sequenced PCR amplified DNA sent to the
UofL DNA-core and their translated protein products.

NCBI BLAST

Once the contigs were assembled, both the nucleotide contigs and the translated

protein products were run through the NCBI database by BLASTn, BLASTx, and

BLASTp searches. The results of which are presented below.

Giant Ragweed

BLASTn Results and Identity Percentages:

Helianthus annuus genotype RHA276 Kunitz-like protease inhibitor 86%
Youngia japonica gene of Miraculin homolog 79%
Taraxacum offcinale gene of Miraculin homolog 79%

BLASTp Results and Identity Percentages:

Kunitz-like Protease Inhibitor *Helianthus annuus* 88%
Kunitz inhibitor St1-like protein *Cynara cardunculus* 81%
Miraculin homologue *Youngia japonica* 83%
Miraculin homologue *Traxacum offcinale* 84%

BLASTn Phylogeny:

BLASTp Phylogeny:

Common Ragweed

BLASTn Results and Identity Percentages:

Helianthus annuus genotype RHA276 Kunitz-like protease inhibitor 86%
Youngia japonica gene of Miraculin homolog 79%
Taraxacum offcinale gene of Miraculin homolog 79%

BLASTp Results and Identity Percentages:

Kunitz-like Protease Inhibitor *Helianthus annuus* 88%
Kunitz inhibitor St1-like protein *Cynara cardunculus* 81%
Miraculin homologue *Youngia japonica* 83%
Miraculin homologue *Traxacum offcinale* 85%

BLASTn Phylogeny:

BLASTp Phylogeny:

The BLAST searches of both of the contigs yielded very similar results, with four of the same species of plants making the top matches. All four of the plants are found within the *Asteraceae* family of plants, indicating a high degree of conservation of sequence between the two plants. The BLASTx and BLASTp searches revealed the Kunitz-type STI domain present on both the Common Ragweed and Giant Ragweed sequences. Further, they both display the same reactive sites loops of the Kunitz-type STI superfamily.

Phylogenetic analysis of the nucleotide sequences results in both the common ragweed and giant ragweed bearing highest similarity to the *Helianthus annuus* Kunitz-type STI, with the miraculin-like proteins being more distantly related. This trend continues in the phylogenetic analysis of the protein sequences. The two sequences bear the greatest relationship to the *Cynara cardunuculus* Kunitz-type STI sequence, despite displaying lower BLAST identity values for the sequence. While *Youngia japonica* bears highest relatedness to the Ragweed sequences, *Traxacum offcinale* was least related of the four highly similar groups. The four highly similar sequences, identified by BLAST search, were then used in the analysis of the structure of the ragweed miraculin homologues.

Intron Analysis of Contigs

During the construction of the contigs, the reading frame was found to be +1 and +2 (Giant and Common Ragweed, respectively), after translation of the contigs themselves into readable proteins. Common Ragweed contained no intron sites on either the direct or complement strands. Giant Ragweed was found to have a singular donor Intron site on the complement strand with a high confidence (0.65), but due to the nature of the

translated protein it was ruled out as a significant indicator of intron interaction. Genome sequence analysis and online prediction revealed that all KTI genes are intronless and encode proteins with a putative signal peptide for cell secretion[27].

Donor splice sites, complement strand

```
Acceptor splice sites, direct strand
-----------------------------------
                pos 5'->3'  phase strand  confidence  5'      intron exon      3'
                    295       0      +        0.65        CTGTATTCAG^TCGAACGTTT
```

Figure 8..
Itnron analysis of the Giant Ragweed contig by Net Plant Gene.

Comparison of the two Ragweed Contig Sequences

The two Ragweed contigs were aligned by Clustal W (Figure 9).


```
Common_Ragweed-----------TTTTCTTTAGCCCAA 25    Common_RagweedCAAGAAAGGCGTTATACGCGAATCC 250
Giant_Ragweed AATATTCTCATTTTCTTTAGCCCAA        Giant_Ragweed CAAGAAAGGCGTTATACGCGAATCC

Common_RagweedCCCTCACCAGATGCGGTTCGTGACA 50    Common_RagweedACGGATCTTAACATTGTTTTTCGG 275
Giant_Ragweed CCCTCACCCGATGCGGTTCGTGACA        Giant_Ragweed ACGGATCTTAACATTATTTTTCGG

Common_RagweedTCAATGGCAACTTACTCCGGTCGGG 75    Common_RagweedCTTCGTCAATCTGTATTCAGTCGAA 300
Giant_Ragweed TCAATGGCAACTTACTCCGATCGGG        Giant_Ragweed CTTCGTCAATCTGTATTCAGTCGAA

Common_RagweedCACAAACTACTACATCCTCCCCGTC 100   Common_RagweedCGTTTGGATGATTGAAGAGTACCAA 325
Giant_Ragweed CACAAACTACTACATCCTCCCCGTC        Giant_Ragweed CGTTTGGATGGTTGAAGAGTACCAA

Common_RagweedTTCAGAGGACGGGGCGGAGGCGTAA 125   Common_RagweedGGGCAACTTATTGTTACGGGTCATG 350
Giant_Ragweed TTCAGAGGACGGGGCGGAGGCGTAA        Giant_Ragweed GGGCAACTTATTGTTACGGGTCATG

Common_RagweedCACTAGCCCCCACACGAAACGAGTC 150   Common_RagweedGAACGGCTGGAAACCCTGGTCAGGA 375
Giant_Ragweed CACTAGCCCCCACACGAAACGAGTC        Giant_Ragweed GAACGGCTGGAAACCCTGGTCAGGG

Common_RagweedATGCCCACTTGATGTAGTCCAAGAA 175   Common_RagweedAACATTAAGCAACTGGTTCAAGATA 400
Giant_Ragweed ATGCCCACTTGATGTAGTCCAAGAA        Giant_Ragweed AACATTAAGCAACTGGTTCAAGATA

Common_RagweedGGCTTCGAGCTACAGAATGGCCTCC 200   Common_RagweedGAGAAGTATGAAGATGATTATAAGC 425
Giant_Ragweed GGCTTCGAGCTACAGAATGGCCTCC        Giant_Ragweed GAGAAGTATGAAGATGATTATAAGC

Common_RagweedCCTTAACATTCACACCTGTGGATCC 225   Common_RagweedTTGTGTTCTGTCCTACA-- 450
Giant_Ragweed CCTTAACATTCACACCTGTGGATCC        Giant_Ragweed TTGTGTTCTGTCCTACAGT
```

Figure 9.
CLUSTAL W alignment of the two Ragweed contigs. Differeneces between the two are highlighted in gray.

The two sequences have a total of eighteen different nucleotides that are

different. Four of the differences highlighted arise from actual structural difference rather

than a difference in sequence size. This indicates that the Kunitz-Type STI is highly

conserved between the two species of Ragweed and further maintains the ability to

utilize the modularity of the I3 superfamily in order to be able to inhibit a variety of

proteases.

NCBI BLASTp was then used to align the two translated protein products against

each other.


```
Score            Expect  Method                            Identities     Positives      Gaps
294 bits(752)  1e-107  Compositional matrix adjust.  141/144(98%)  143/144(99%)  0/144(0%)

Query  6    FSLAQPSPDAVRDINGNLLRSGTNYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELQ  65
            FSLAQPSPDAVRDINGNLLRSGTNYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELQ
Sbjct  1    FSLAQPSPDAVRDINGNLLRSGTNYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELQ  60

Query  66   NGLPLTFTPVDPKKGVIRESTDLNIIFSASSICIQSNVWMVEEYQGQLIVTGHGTAGNPG  125
            NGLPLTFTPVDPKKGVIRESTDLNI+FSASSICIQSNVWM+EEYQGQLIVTGHGTAGNPG
Sbjct  61   NGLPLTFTPVDPKKGVIRESTDLNIVFSASSICIQSNVWMIEEYQGQLIVTGHGTAGNPG  120

Query  126  QGTLSNWFKIEKYEDDYKLVFCPT  149
            Q TLSNWFKIEKYEDDYKLVFCPT
Sbjct  121  QETLSNWFKIEKYEDDYKLVFCPT  144
```

Figure 10.
NCBI BLASTp alignment of the Common and Giant ragweed (Subject and Query respectively) proteins and the consensus sequence formed between the two.

Similar to the nucleotide alignment, the two sequences bear a great deal of

similarity with only three amino acids differing between the two species of ragweed. The

conserved Kuntiz-type Soybean Trypsin Inhibitor domain is present in both of the

proteins with the reactive site loops at sites 74-81 preserved in both proteins. giant

ragweed and common ragweed differ at two sites with valine and isoleucine

conservative difference. Both amino acids are non-polar but their steric nature

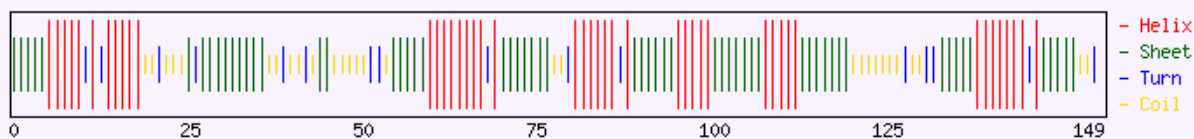structurally different. While Ile has an extra carbon, the two amino acids branch in the

same manner. As both are hydrophobic, they are likely found in the inner coils of the Beta-sheet triangle, where their similar branching is unlikely to cause steric change significant enough to alter either the Beta-Trefoil structure or the reactive site loop. The last difference is giant ragweed contains a glycine rather than the common ragweed glutamic acid at residue 127. The difference of the non-polar to neutral charge polar amino acid could have bearing on the protein conformation, but is apparently not significant enough to cause the deformation of the Beta-Trefoil structure.

Secondary Structure Analysis

The proteins translated from the Contigs were then run through the CFSSP: Chou and Fasman Secondary Structure Prediction Server. This server produces a two dimensional picture of the Helices and Sheets for the protein. Both the Ragweed Contig translated proteins, and the four high identity proteins were analyzed by CFSSP. The results of which are presented below in Figures 11-16.

Figure 11.
Secondary structure analysis of the Giant Ragweed translated protein.

Target Sequence:

```
FSLAQPSPDA VRDINGNLLR SGTNYYILPV FRGRGGGVTL APTRNESCPL DVVQEGFELQ NGLPLTFTPV
DPKKGVIRES TDLNIVFSAS SICIQSNVWM IEEYQGQLIV TGHGTAGNPG QETLSNWFKI EKYEDDYKLV
FCPT
```



```
                                                                    — Helix
                                                                    — Sheet
                                                                    — Turn
                                                                    — Coil
0           25          50          75          100         125        144
```
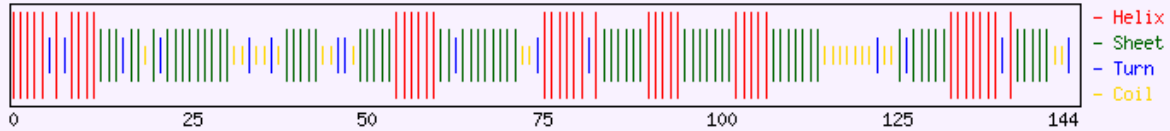
Secondary Structure:

```
                    *         *         *         *         *         *
Query 1    FSLAQPSPDAVRDINGNLLRSGTNYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELQNGLPLTFTPV 70
Helix 1    HHHHHHHHHHHHH               HH                    HHHHHHHHHHHHH           70
Sheet 1    EEEE        EEEEEE EEEEEEEEEEE        EEEEE      EEEEEEEEEEEEEEEEEEEEEEEEE 70
Turns 1        T T        T    T T        T T        TT        T       T           70
Struc 1    HHHHHTHTHHHHEEETEECETEEEEEEEEEECCTCCTCEEEEECCTTCEEEEEHHHHHHEETEEEEEEEEC 70

                    *         *         *         *         *         *
Query 71   DPKKGVIRESTDLNIVFSASSICIQSNVWMIEEYQGQLIVTGHGTAGNPGQETLSNWFKIEKYEDDYKLV 140
Helix 71     HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH          HHHHHHHHHHHHHHHHHH 140
Sheet 71         E EEEEEEEEE        EEEEEEEEEEEEEEEEEEEEE          EEEEEEEEEE      EEEE 140
Turns 71   T T     T                T               T           T  TT             T    140
Struc 71   CTHHHHHHTHEEEEEEHHHHHEEEEEEEEHHHHHEEEEEEEECCCCCCCTCCETEEEEEHHHHHHHTHEEEE 140


Query 141 FCPT 144
Helix 141      144
Sheet 141 E    144
Turns 141    T 144
Struc 141 ECCT 144


Total Residues: H: 82    E: 92    T: 21
        Percent: H: 56.9  E: 63.9  T: 14.6
```
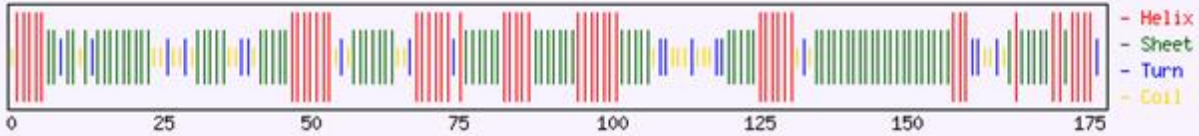
Figure 12.
Secondary structure analysis of the Common Ragweed translated protein.

Target Sequence:

```
PDAVRDIDGN LLRSGTEYYI LPVFRGRGGG VTLAPTRNES CPLDVVQEGF ELDNGLPLTF TPVDPKKGVI
RESTDLNIIF SASSICIQSN VWMLEEYEGQ RIISGRGTSG NPGGETISNW FKIEKYENGY KLVYCPTVCD
LCRPVCGDIG VVFAENGSRR LAISDVPFKI KFRKA
```


Secondary Structure:

```
                      *          *          *          *          *          *
Query  1    PDAVRDIDGNLLRSGTEYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELDNGLPLTFTPVDPKKGVI 70
Helix  1      HHHHHH              HH                  HHHHHHHHHH               HHHHH  70
Sheet  1           EEEEEE EEEEEEEEEEE        EEEEE      EEEEEEEEE      EEEEEEE       E 70
Turns  1      T        T    T T              T T        TT        T         T        T T  70
Struc  1    CHHHHHEETEECETEEEEEEEEEECCTCCTCEEEEECCTTCEEEEEHHHHHHHCTCEEEEEEEECCTHHHHH 70


                      *          *          *          *          *          *
Query  71   RESTDLNIIFSASSICIQSNVWMLEEYEGQRIISGRGTSGNPGGETISNWFKIEKYENGYKLVYCPTVCD 140
Helix  71   HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH                   HHHHHHHHHH          140
Sheet  71   EEEEEEEEE        EEEEEEEEEEEEEEEEEEEE             EEEEEEEE       EEEEEEEEEEE 140
Turns  71      T                 T              TT      TT    T    TT                 T    140
Struc  71   HTHEEEEEEEHHHHHEEEEEEEEHHHHHHHEEEEECTTCCCTCCCTTEEEEEHHHHHHCTCEEEEEEEEEEEE 140


                      *          *          *
Query  141  LCRPVCGDIGVVFAENGSRRLAISDVPFKIKFRKA 175
Helix  141          HHHHHHH       HHHHHHHHHHHHHHH   175
Sheet  141  EEEEEEEEEEEE         EEEEEEEEEEE        175
Turns  141                 TT  T             T  175
Struc  141  EEEEEEEEEEEEHHHTTCCTCEHEEEEEHHHEHHHHT 175


Total Residues: H: 86     E: 109    T: 26
       Percent: H: 49.1   E: 62.3   T: 14.9
```

Figure 13.
Secondary structure analysis of the *Helianthus annuus* translated protein.
Ascension number: AFL91226

Target Sequence:

```
MKTLSLLLLF STISLCLGQP SPDPVLDIDG NLVRSGTDYY ILPVFRGMGG GVTLASTRNE SCPLDVVQEP
QEVDNGLPLT FTPVDPKKGV IRESTDLNII FSALSICIQS NVWNLEEYEG QLIVSARGVS GNPGQETISN
WFKIEKYEDD YKIVFCPTVC DFCRPVCGDI GVIIGEDGSR RLAIRDVPFK MKTVLLFLLF STISLSFGQK
SPDQVRDTDG NLLRSGTDYY ILPVFRGRGG GVTLAPTRNE SCPLDVVQEG FELKFGLPLK FTPVNSKKGV
IRESTDLNIM FSASTSCVQS NVWMLEAYKG QLIVSGHGVW GNPGRETISN WFKIAKYEND YKLVFCPTVC
NYCKPVCGDI GVKIAENGSR HYYILPVFRG RGGGVTLAPT RNESCPLDVV QEGFELKFGL PLQFAPVNFK
KGVIRESTDL NIFSASTCQS NWMLEAYGQL IVSGHGVWGN PGRETISNWF KIAKYENDYK LVFCPTVCNY
CKPVCGDIGV KIAENGSRRL AISDVPFKVK FRKA
```


- Helix
- Sheet
- Turn
- Coil
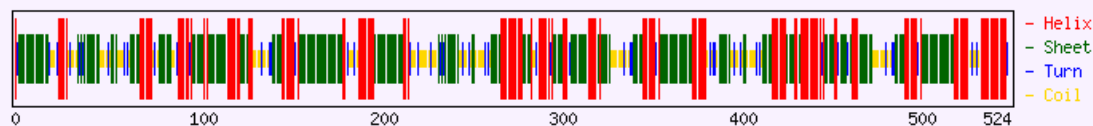
Secondary Structure:

```
                  *           *           *           *           *           *
Query   1  MKTLSLLLLFSTISLCLGQPSPDPVLDIDGNLVRSGTDYYILPVFRGMGGGVTLASTRNESCPLDVVQEP  70
Helix   1  HHHHHHHHHHHHHH        HHHHH               HH                  HHHHHHH      70
Sheet   1  EEEEEEEEEEEEEEEEEEE          EEEEEEEEEEEE      EEE       EEEEEE          70
Turns   1   T                T   T       T   T T              T            TT      T  70
Struc   1  HTEEEEEEEEEEEEEEEEEETCCCTHHHHHCTCCCETETEEEEEEEEECCCCTCEEECCCCTTCEEEEEHHHH  70


                  *           *           *           *           *           *
Query  71  QEVDNGLPLTFTPVDPKKGVIRESTDLNIIFSALSICIQSNVWNLEEYEGQLIVSARGVSGNPGQETISN  140
Helix  71  HHH              HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH        HHH    140
Sheet  71         EEEEEE         EEEEEEEEEEEEEEEEEEEEEE         EEEEEEE       EEEE   140
Turns  71  T   T          T T    T              T               T          T   TT    140
Struc  71  HHHCTCEEEEEEEECCTHHHHHHTHEEEEEEHHEEEEEEEEEEEHHHHHHHEEEEHHHCCCCTCCCTTEEEE   140


                  *           *           *           *           *           *
Query 141  WFKIEKYEDDYKIVFCPTVCDFCRPVCGDIGVIIGEDGSRRLAIRDVPFKMKTVLLFLLFSTISLSFGQK  210
Helix 141  HHHHHHHHHHHHH                  HHHHHHH      HHHHHHHHHHHHHHHHHHHHHHHHHHH    210
Sheet 141  EE          EEEEEEEEEEEEEEEEEEEEEEEEE                 EEEEEEEEEEEEEEEEEE   210
Turns 141           T                         T   T                             T   210
Struc 141  EHHHHHHHTHEEEEEEEEEEEEEEEEEEEEEEEEEEEHHTCCCTCHHHHHHHHHEEEEEEEEEEEEEEEHHHTC  210


                  *           *           *           *           *           *
Query 211  SPDQVRDTDGNLLRSGTDYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELKFGLPLKFTPVNSKKGV  280
Helix 211                 HH                          HHHHHHHHHHHHHHHHHHH      HHHH   280
Sheet 211             EEEEEEEEEEE      EE          EEEEEE         EEEEEE            280
Turns 211   T        T    T T        T T        TT          T              T T     280
Struc 211  CCTCCCCCCTCCCETETEEEEEEEECCTCCTCEECCCCCTTCEEEEEHHHHHHHHHHHHHHEEEEHCCTHHHH  280


                  *           *           *           *           *           *
Query 281  IRESTDLNIMFSASTSCVQSNVWMLEAYKGQLIVSGHGVWGNPGRETISNWFKIAKYENDYKLVFCPTVC  350
Helix 281  HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH           HHHHHHHHHHHHHHHHHH        350
Sheet 281   EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE        EEEEEE         EEEEEEEEEE  350
Turns 281   T          T       T          T         T   TT          T            350
Struc 281  HHTHEEEEEEHHHTEEEEEEEEEEHHHHHEHEEEEEECCCCCCCTCCTTEEEEEEHHHHHHHTHEEEEEEEEEE  350


                  *           *           *           *           *           *
Query 351  NYCKPVCGDIGVKIAENGSRHYYILPVFRGRGGGVTLAPTRNESCPLDVVQEGFELKFGLPLQFAPVNFK  420
Helix 351        HHHHHHHH         HH                    HHHHHHHHHHHHHHHHHHHHHHHHHH    420
Sheet 351  EEEEEEE             EEEEE        EE          EEEEEE         EEEEEEE      420
Turns 351                T TT         T T        TT          T                420
Struc 351  EEEEEEEHHHHHHHHCTCTTCEEEEEECCTCCTCEECCCCCTTCEEEEEHHHHHHHHEEEEHHEHHHHHH    420


                  *           *           *           *           *           *
Query 421  KGVIRESTDLNIFSASTCQSNWMLEAYGQLIVSGHGVWGNPGRETISNWFKIAKYENDYKLVFCPTVCNY  490
Helix 421  HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH              HHHHHHHHHHHHHHHH          490
Sheet 421     EEEEEEEEEEEEEEEEEEEEEEEEEEEEEE             EEEEEE        EEEEEEEEEEEE   490
Turns 421  T    T           T   T                 T  TT           T             490
Struc 421  HHHHHTHEEEEEEHHTEEEETEHHHHEEEEEEECCCCCCCTCCTTEEEEEHHHHHHHTHEEEEEEEEEEEEE   490
```
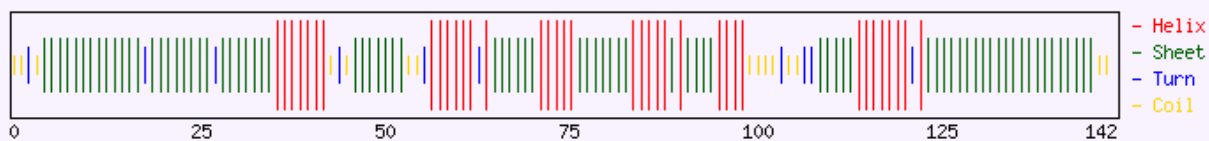
Figure 14.
Secondary structure analysis of the *Cynara caradunulus* translated protein. Ascension number: KVI09680

Target Sequence:

```
LRSGTEYYIL PVFRGMGGGL TLASTRNDTC PLDVVQADLE VDNGLPLTFT PVDPKKGVIR ESTDLNIIFS
ASSICIQSNV WNLEEYDGQL IVSAHGVAGN PGRETISNWF KIEKYEDDYK IVFCPTVCDF CRPVCGDIGV
LI
```


Secondary Structure:

```
                    *          *          *          *          *          *
Query 1    LRSGTEYYILPVFRGMGGGLTLASTRNDTCPLDVVQADLEVDNGLPLTFTPVDPKKGVIRESTDLNIIFS 70
Helix 1            HH                         HHHHHHHHH                HHHHHHHHHHHHHHHH 70
Sheet 1       EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE     EEEEEEE       E EEEEEEEEE    70
Turns 1     T T           T          T          T        T          T T    T        70
Struc 1    CCTCEEEEEEEEEEEEEEETEEEEEEEEETEEEEEEEEHHHHHHHCTCEEEEEEEECCTHHHHHHTHEEEEEEEHH 70

                    *          *          *          *          *          *
Query 71   ASSICIQSNVWNLEEYDGQLIVSAHGVAGNPGRETISNWFKIEKYEDDYKIVFCPTVCDFCRPVCGDIGV 140
Helix 71   HHHHHHHHHHHHHHHHHHHHHHHHH              HHHHHHHHHHHHHHHHH              HH   140
Sheet 71       EEEEEEEEE    EEEEEEE              EEEEEEEE      EEEEEEEEEEEEEEEEEEEEEEEE 140
Turns 71          T            T          T  TT          T                           140
Struc 71   HHHEEEEEEEEEHHHHHEHEEEEHHHHCCCCTCCTTEEEEEHHHHHHHTHEEEEEEEEEEEEEEEEEEEEEEEE 140


Query 141 LI 142
Helix 141    142
Sheet 141    142
Turns 141    142
Struc 141 CC 142


Total Residues: H: 72    E: 97    T: 15
       Percent: H: 50.7  E: 68.3  T: 10.6
```
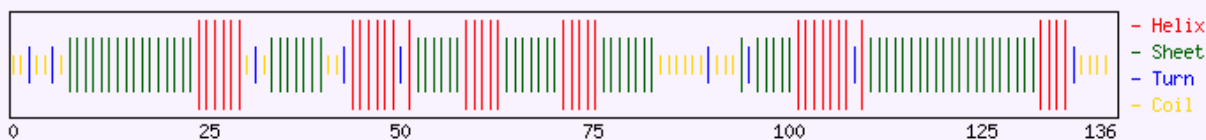
Figure 15.
Secondary structure analysis of the *Youngia japonica* translated protein.
Ascension number: BAA82840

Figure 16.
Secondary structure analysis of the *Traxacum Officinale* translated
protein. Ascension number: BAA82842

Each of the proteins contain the requisite amount of Beta-sheets necessary for
the construction of the Beta-Trefoil[25] indicative of the Kunitz-Type STI. Furthermore, all
of the proteins modeled contain the requisite amount of turns by which the triangular
inner Beta-Trefoil and outer barrel formation may be constructed. Despite the
differences in size found between the *Cynara caradunulus* and the other five protein
products, the Beta-Trefoil is conserved and the reactive site which allows the Kunitz-
type STI to function by tight-affinity binding. Three-dimensional modeling of the protein

sequences further confirms the maintenance of the Beta-Trefoil in each of the high-identity proteins.

Tertiary Structure Analysis and 3 Dimensional Modeling

The proteins translated from the contigs were then run through the Phyre[2]: Protein Homology/analogy Recognition Engine V 2.0 Server. This server produce three dimensional picture of the protein by aligning the sequence to the existing 3-D model based on identity. Both of the Ragweed Contigs translated proteins, and the four high identity proteins were analyzed by Phyre[2]: The results of which are presented in Figures 17-22.

Figure 17. Giant Ragweed 3D model drawn by usage of Phyre$^{2.0}$ Server
Modeling Identity: Fold:beta-Trefoil Superfamily:STI-like
Family:Kunitz (STI) inhibitors

Figure 18. Common Ragweed 3D model drawn by usage of Phyre$^{2.0}$ Server
Modeling Identity: Fold:beta-Trefoil Superfamily:STI-like
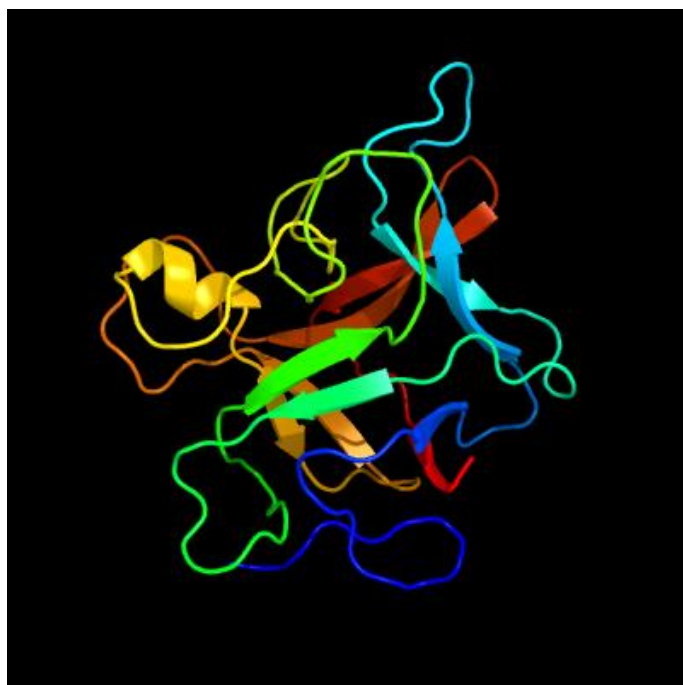Family:Kunitz (STI) inhibitors

Figure 19. *Helianthus annuus* 3D model drawn by usage of Phyre$^{2.0}$ Server
Modeling Identity: Fold:beta-Trefoil Superfamily:STI-like
Identity: crystal structure of a papaya latex serine protease inhibitor (ppi)
Ascension number: AFL91226

Figure 20. *Cynara cardunuclus* 3D model drawn by usage of Phyre[2.0] Server
Modeling Identity: Fold:beta-Trefoil Superfamily:STI-like
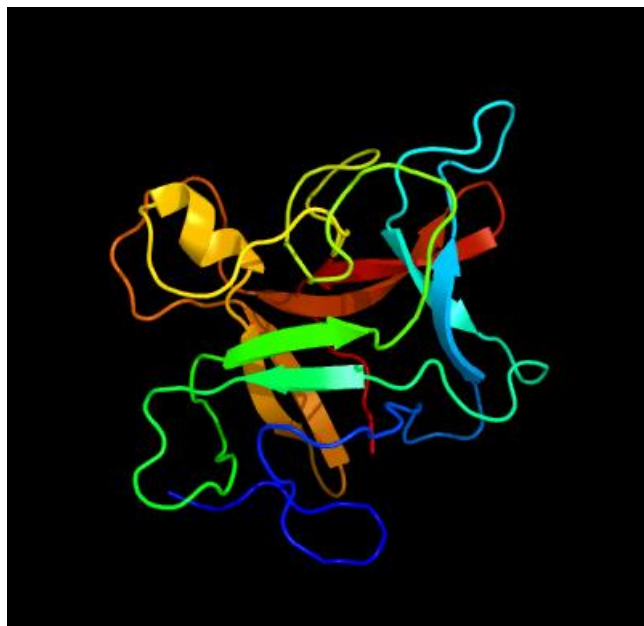Identity: crystal structure of a papaya latex serine protease inhibitor (ppi)
Ascension number: KVI09680

Figure 21. *Youngia japonica* 3D model drawn by usage of Phyre[2.0] Server
Modeling Identity: Fold:beta-Trefoil Superfamily:STI-like
Identity: crystal structure of Miraculin like protein from seeds of murraya2 koenigii
Ascension number: BAA82840

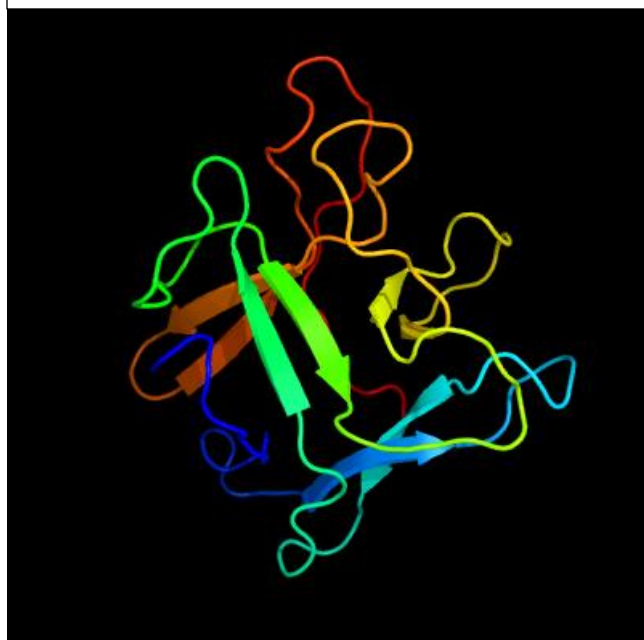**Figure 22.** *Traxacum offcinale* 3D model drawn by usage of Phyre[2.0] Server
Modeling Identity: Fold:beta-Trefoil Superfamily:STI-like
Identity: crystal structure of Miraculin like protein from seeds of murraya2 koenigii
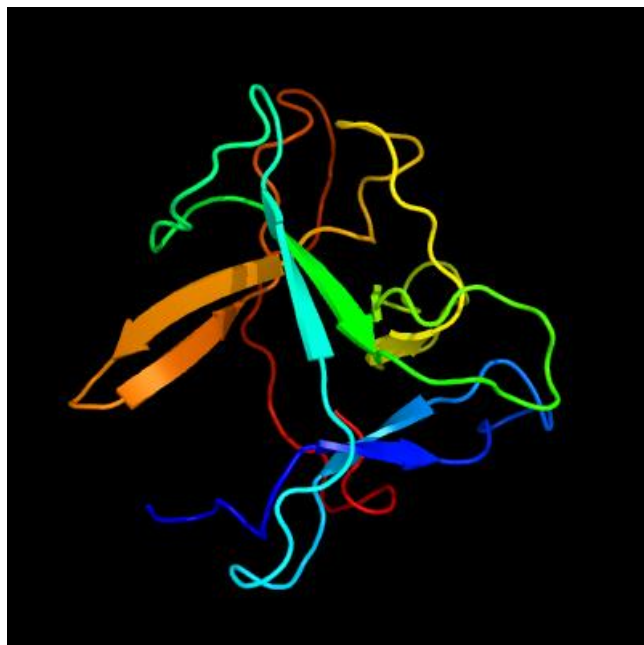Ascension number: BAA82842

As shown by three-dimensional modeling, both the Common and Giant Ragweed proteins show the presence of the Beta-Trefoil structure. As seen in the secondary structure analysis, the proteins showed a requisite amount of Beta sheet structures from which to create the barrel. The six analyzed proteins yielded three identity-based model structures. The two Ragweed proteins, which were translated from the constructed contigs, most closely matched Kunitz-type STI. The two most phylogenetically similar proteins *Youngia japonica* and *Cynara cardunuculus* were modeled on a Papaya latex serine-protease inhibitor. The Papaya latex serine protease itself is a member of the Kunitz-type Superfamily. The two are likely to be bear higher identities with the Papaya latex serine protease inhibitor due to the additional amino acid sequences outside of the Kunitz domain as the other two sequences are both longer. The other two highly proteins with high identity scores were modeled most similarly with a miraculin homologue, most likely due to the proteins themselves being miraculin homologs. Their protein models lack the post-translational modification of miraculin, the addition of four polysaccharide groups, which is likely why they still express the Beta-trefoil structure indicative of Kunitz-type STI proteins.

Ra3 Comparison

BLAST comparison of the known protein sequence of Ra3 (the major allergen of ragweed) [21] and the translated protein products of the ragweed contigs found no alignment between the two. It is unlikely that there is more to the similarity of the Ra3 allergen to Kunitz-Type STI's than superficial coincidence. The Ra3 protein was unable to become aligned to the translated protein products of the ragweed contigs.

Discussion

The analysis of the cDNA miraculin homolog from giant ragweed may present novel insight into the high degree of similarity between Kunitz-type Soybean Trypsin Inhibitor proteins and miraculin homologs. Miraculin falls into the Kunitz-type STI superfamiily by the crystalization of its protein, before it undergoes post-translational modification and is glycosylated. This miraculin protein, when crystalized, displays the classic Beta-Trefoil[23] of the I3 proteinase superfamily. Miraculin's structure, alongside its polysaccharide groups, has yet to be fully characterized. Existing characterization of the miraculin protein has only given the amounts of sugars attached to the miraculin protein, but their contributions to the overall structure of miraculin has yet to be determined. The amino acid sequence of miraculin contains the Kunitz domain with the reactive site loop. It is likely that the presence of polysaccharides in miraculin alters the Beta-Trefoil structure to where it is no longer able to engage in the tight-binding affinity of the I3 superfamily. Two possibilities may prevent the Beta-Trefoil structure. The first is the presence of the polysaccharide groups interfering with the steric alignment of hydrophobic phenylalanine and tryptophan within the inner coil. Both are large non-polar hydrophobic amino acids which, if displaced by the polysaccharides, would prevent the formation of the interior Beta-sheet coils neccesary for the Beta-Trefoil conformation. The second possibility is replacement of the original amino acid hydrogen bonds between the inner and outer Beta-sheet coils by a polysaccharide unit. There are four pairs hydrogen bonds between each pair of coils which results in the triangular alignment of the Beta-Trefoil. The presence of numerous hydroxyl groups on polysaccharides provide an environment where a diversity of hydrogen bonds may form

in the place of the hydrogen bonds between the amino acids. The inability of the Beta-Trefoil to form would similarly prevent the reactive site loop from forming and initating the tight-binding affinity typical to the I3 superfamily. The new conformation of miraculin may then have a conformation where the two histidines (His-29 and His-59) are able to bind to the taste-receptors in such a manner as to elicit the conformational change of the receptor resulting in the higher perceived sweetness.

In summary, the primers constructed for the amplification of both the common and giant ragweed sequences were successful. The contigs constructed from the amplified sequences had high fidelity which allowed for proper analysis of their protein products. Both the giant and common ragweed "Miraculin-homolog" fall within the I3A sub family of the Kunitz-type Soybean trypsin Inihibitors protein superfamily by the presence of their Beta-Trefoil. The conserved Kuntiz-type Soybean Trypsin Inhibitor domain is present in both of the proteins with the reactive site loops at sites 74-81 being preserved in both proteins, which makes it likely that that the better classification of the original cDNA sequence would be a Kunitz-Type STI, and both proteins at as protease inhibitors within ragweed, rather than taste-modifying proteins because one doesn't know if the protein would then be glycosylated in the post-translation process. The expression of these Kunitz-type STI's within ragweed is likely not limited to the pollen. Other Kunitz-type STI's in similar plants are expressed in all tissues of the plant (especially roots and stems) and that they might be helping the plant resist infection by fungi that cause disease similar plants.[27] The miraculin-homolog identified in the cDNA likely serves a similar function.

The identity of the miraculin homolog sequences as Knutiz-type STI serves to highlight the terminological differences used in the study of plant genomics and is important in broader

terms as it may serve as a basis for the reclassification of a great deal of published sequences. Instead of erroneously publishing sequences as miraculin-homologs, future studies may seek to assess whether the sequence has the two functional histidines, which trigger a conformational change in T1R2-T1R3 taste receptors. If the sequences then lacks this reactive site and does not undergo glycosylation as a part of its post-translational modification, then it is simply a Kunitz-type STI.

Literature Cited

1.   Flegal, K. M., Carroll, M. D., Kit, B. K., & Ogden, C. L. (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *Journal of The America Medical Association,* 307(5), 491-497.
2.   Kurihara, K., & Beidler, L. M. (1968). Taste-modifying protein from miracle fruit.*Science*, 161(3847), 1241-1243.
3.   Ismail, A. I., Tanzer, J. M., & Dingle, J. L. (1997). Current trends of sugar consumption in developing societies. *Community Dentistry and Oral Epidemiology*, 25(6), 438-443.
4.   Blass, E. M. (1987). Opioids, sweets and a mechanism for positive affect: Broad motivational implications. In *Sweetness* (pp. 115-126). Springer London.
5.   Adler, J., & Epstein, W. (1974). Phosphotransferase-system enzymes as chemoreceptors for certain sugars in Escherichia coli chemotaxis. *Proceedings of the National Academy of Sciences*, 71(7), 2895-2899.
6.   Thompson, D., & Wolf, A. M. (2001). The medical-care cost burden of besity. O*besity Reviews,* 2(3), 189-197.
7.   Kinnamon, S. C. (2012). Taste receptor signalling–from tongues to lungs. *Acta physiologica,* 204(2), 158-168.
8.   Joesten, Melvin D; Hogg, John L; Castellion, Mary E (2007). "Sweeteness Relative to Sucrose (table)". *The World of Chemistry: Essentials (4*th ed.). Belmont, California: Thomson Brooks/Cole. p. 359. ISBN 0-495-01213-0.
9.   Misaka, T. (2013, March). Molecular mechanisms of the action of Miraculin, a taste-modifying protein. In Seminars in cell & developmental biology (Vol. 24, No. 3, pp. 222-225). Academic Press
10.   Paladino, A., Costantini, S., Colonna, G., & Facchiano, A. M. (2008). Molecular modelling of Miraculin: structural analyses and functional hypotheses.*Biochemical and biophysical research communications, 3*67(1), 26-32.
11.   Duhita, N., Hiwasa-Tanase, K., Yoshida, S., & Ezura, H. (2009). Single-step purification of native Miraculin using immobilized metal-affinity chromatography.*Journal of agricultural and food chemistry,* 57(12), 5148-5151.
12.   Ito, K., Asakura, T., Morita, Y., Nakajima, K. I., Koizumi, A., Shimizu-Ibuka, A., ... & Abe, K. (2007). Microbial production of sensory-active Miraculin.Biochemical and biophysical research communications, 360(2), 407-411.
13.   Mabashi, Y., Kikuma, T., Maruyama, J. I., Arioka, M., & Kitamoto, K. (2006). Development of a versatile expression plasmid construction system for *Aspergillus oryzae* and its application to visualization of mitochondria.*Bioscience, biotechnology, and biochemistry,* 70(8), 1882-1889.
14.   Kanemori, Y., Gomi, K., Kitamoto, K., KUMAGAI, C., & TAMURA, G. (1999). Insertion analysis of putative functional elements in the promoter region of the *Aspergillus oryzae* Taka-amylase A gene (amyB) using a heterologous *Aspergillus nidulans* amdS-lacZ fusion gene system. *Bioscience, biotechnology, and biochemistry,* 63(1), 180-183.

15. Gomi, Kimura, Y., and Hara, S., Integrative transformation of *Aspergillus oryzae* with a plasmid containing the *Aspergillus nidulan*s argB gene. Agric. *Biol. Chem.*, 51, 2549–2555 (1987)

16. Theerasilp, S., & Kurihara, Y. (1988). Complete purification and characterization of the taste-modifying protein, Miraculin, from miracle fruit.J*ournal of Biological Chemistry,* 263(23), 11536-11539.

17. Kobayashi, H., Suzuki, M., Kanayama, N., & Terao, T. (2004). A soybean Kunitz trypsin inhibitor suppresses ovarian cancer cell invasion by blocking urokinase upregulation. *Clinical & experimental metastasis*, *21*(2), 159-166.

18. Lee, S. I., Lee, S. H., Koo, J. C., Chun, H. J., Lim, C. O., Mun, J. H., ... & Cho, M. J. (1999). Soybean Kunitz trypsin inhibitor (SKTI) confers resistance to the brown planthopper (Nilaparvata lugens Stål) in transgenic rice.*Molecular Breeding*, *5*(1), 1-9.

19. van Ree, R., Hoffman, D. R., van Dijk, W., Brodard, V., Mahieu, K., Koeleman, C. A., ... & Aalberse, R. C. (1995). Lol p XI, a new major grass pollen allergen, is a member of a family of soybean trypsin inhibitor-related proteins. *Journal of allergy and clinical immunology*, *95*(5), 970-978.

20. Goodfriend, L., Roebber, M., Lundkvist, U., & Choudhury, A. M. (1981). Two variants of ragweed allergen Ra3. *Journal of Allergy and Clinical Immunology*, *67*(4), 299-304.

21. Klapper, David G., L. Goodfriend, and J. D. Capra. "Amino acid sequence of ragweed allergen Ra3." *Biochemistry 19.25* (1980): 5729-5734.

22. Koizumi, A., Tsuchiya, A., Nakajima, K. I., Ito, K., Terada, T., Shimizu-Ibuka, A., & Abe, K. (2011). Human sweet taste receptor mediates acid-induced sweetness of miraculin. *Proceedings of the National Academy of Sciences*, *108*(40), 16819-16824.

23. Lebeda, F. J., Umland, T. C., Sax, M., & Olson, M. A. (1998). Accuracy of secondary structure and solvent accessibility predictions for a clostridial neurotoxin C-fragment. *Journal of protein chemistry*, *17*(4), 311-318.

24. Rawlings, Neil D., Dominic P. Tolle, and Alan J. Barrett. "Evolutionary families of peptidase inhibitors." *Biochemical Journal* 378.3 (2004): 705-716.

25. Xu, R., & Xiao, Y. (2005). A common sequence-associated physicochemical feature for proteins of beta-trefoil family. *Computational biology and chemistry*, *29*(1), 79-82.

26. Li, M., Huang, Y., Xu, R., & Xiao, Y. (2005). Nonlinear analysis of sequence symmetry of beta-trefoil family proteins. *Chaos, Solitons & Fractals*, *25*(2), 491-497.

27. Huang, H., Qi, S. D., Qi, F., Wu, C. A., Yang, G. D., & Zheng, C. C. (2010). NtKTI1, a Kunitz trypsin inhibitor with antifungal activity from Nicotiana tabacum, plays an important role in tobacco's defense response. *FEBS journal*, *277*(19), 4076-4088.