STUDIA UNIVERSITATIS

Revistă ştiinţifică a Universităţii de Stat din Moldova, 2009, nr.6(26)

PECULIARITY OF SUNFLOWER 11S SEED STORAGE GLOBULIN: TERTIARY STRUCTURE MODELING

Ellina MACAEVA, Andrei SHUTOV

Laboratory of Plant Biochemistry

Structura primară deosebită a subunităţii gb|AAA33374 a heliantininei 11S din floarea-soarelui se formează în urma deplasării cadrului de citire a secvenţei ce codifică β-structura în β-barelul N-terminal al genei M28832. Suprapunerea secvenţelor genei heliantininei cu ESTs din *Helianthus* demonstrează că deplasarea cadrului de citire poate fi provocată artificial în urma unei erori de secvenţare a genei. Modelarea structurii terţiare a heliantininei pe baza publicaţiei originale a consecutivităţii aminoacizilor AAA33374 şi secvenţei translate de pe cDNA heliantinina (EST gb|GE510121) confirmă această ipoteză. S-a demonstrat că in subunitatea AAA33374 lipseşte strandul C cu importanţă structurală, deoarece această subunitate poate fi instabilă. Se discută probabilitatea existenţei acestei subunităţi deosebite.

The deletion of a single nucleotide in helianthinin gene sequence M28832 inside exon 1 coding for N-terminal half of the N-terminal β-barrel, which is globally conserved in all other seed 11S storage globulins, led to a shift of the reading frame resulted in peculiar amino acid sequence in the region deduced from the exon sequence flanking to intron 1 [1]. Starting from the exon 2, the reading frame becomes renewed due to a downstream shift of the intron 3'-border resulted in deletion of five amino acid residues from respective coding region [1]. It remains unclear whether the unusual amino acid sequence of helianthinin subunit gb|AAA33374 reflects its real specificity or it is a result of gene sequencing error.

Aiming to choose between these possibilities, we constructed helianthinin structural models based on original AAA33374 amino acid sequence and that deduced from translated helianthinin cDNA sequence (EST gb|GE510121).

Materials and Methods

Amino acid and nucleotide sequences were obtained from http://www.ncbi.nlm.nih.gov/ and aligned using ClustalW http://www.ebi.ac.uk/. Helianthinin structure models were constructed and analyzed using programs http://swissmodel.expasy.org/ and Swiss-PdbViewer 3.7, respectively.

Results and Discussion

A single full-length helianthinin amino acid sequence AAA33374 is available. When this sequence was used as a query, the program BLAST (tblastn, search for translated nucleotide database using a protein query) revealed more than thirty highly similar translated EST sequences corresponding to helianthinin cDNAs from *H. annuus* and three other *Helianthus* species. All these sequences are highly similar to coding region of helianthinin gene M28832 at nucleotide level but dissimilar to AAA33374 amino acid sequence in the above described critical region underlined in Fig.1. Direct search for nucleotide sequence relatives of the gene M28832 revealed three *H. annuus* ESTs (GE510121, GE484425 and BU028043) probably coding for the same subunit (starting from signal, a single substitution was found among all 583 nucleotides common to these EST sequences). Remarkably, a nucleotide missing in M28832 sequence was found present in three ESTs mentioned above and three additional ESTs. A segment recognized as a 3'-edge of the intron 1 in the gene M28832 was present in all the EST sequences (Fig.1). As a result, translated sequences of all six ESTs reveal similarity to all ordinary 11S globulins in the critical region including five amino acid residues missing in AAA33374 sequence. Taken together, these circumstances allow suggesting that deletion of a single nucleotide in the gene M28832 (rather than in three identical ESTs) might be due to a sequencing error. The suggestion is supported by the unusual phase of the intron 1 in the gene M28832 (zero-phase) whereas globally conserved first-phase intron 1 is characteristic of all 11S globulin and homologous 7S globulin genes [2]. However, the distinction of M28832 and full-length EST sequences are much deeper than shown in Fig.1. Multiple technical errors in helianthinin gene sequence are hardly probable. Hence, two kinds of amino acid sequences shown in Fig.1 should be derived from two different genes. Thus, peculiar gene M28832 sequence might really exist.

```
 Q Q F Q C A W S I L F D T G F N L V A F S C L P T
M28832 935 caacagttccagtgtgcgtg-gtcgattttattcgacaccggattcaacctggtggccttctcttgccttcctac
GE510121 205 caacagttccagtgtgcgggtgtcgattttattcgacaccggattcaacctggtggccttctcttgccttcctac
 Q Q F Q C A G V D F I R H R I Q P G G L L L P S Y

 S T P L F W P S S R E <--intron 1------> G V I L P G C
M28832 gtcaacacccctattttggccttcgtcgagagagGT//GTAGGGGTATTCAGGGGGTATTATTGCCGGGATGC 1151
GE510121 gtcaacacccctattttggccttcgtcgagagag GTAGGGGTATTCAGGGGGTTATATTGCCGGGATGC 351
 V N T P I L A F V E R ⇑ G R G I Q G V I L P G C
```

Fig.1. Nucleotide and translated amino acid sequences of *H. annuus* 11S globulin gene M28832 and EST GE510121. Peculiar amino acid sequence is underlined. Non-identical nucleotides and 5'/3'-edges of zero-phase intron 1 sequence are printed in bold. Symbol ⇑ indicates position of first-phase intron 1 characteristic of all 11S and homologous 7S storage globulin genes. Sequence numbering corresponds to that of the gene and EST accessions.

Dramatic structural consequences are expected from peculiarity of helianthinin sequence in the critical region and especially from deletion of the five conserved amino acid residues (Fig.1). Therefore, we constructed two helianthinin structural models based on original AAA33374 sequence and that altered in accordance to Fig.1.

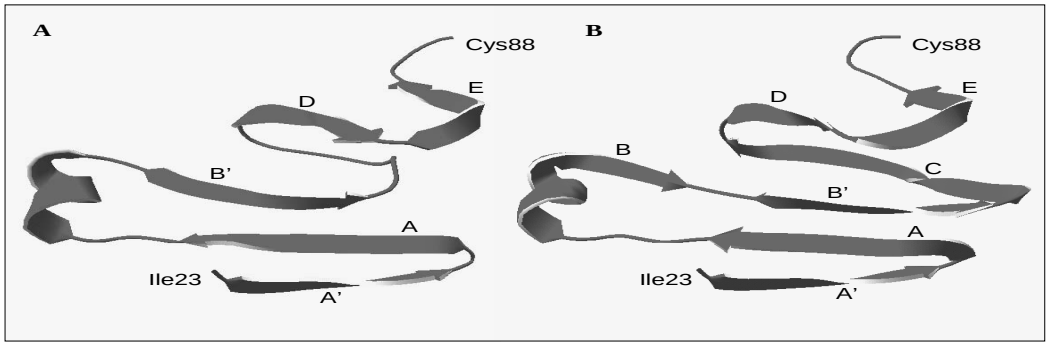

Fig.2. Ribbon diagrams of helianthinin structure models in the region covering the N-terminal half of the N-terminal domain (Fig. 3) encoded by exon 1 (Fig. 1). Target sequences and template structures, respectively: A, original subunit AAA33374 sequence and soybean 11S globulin pdb|1OD5; B, subunit sequence deduced from the EST GE510121 and pumpkin 11S globulin pdb|2E9Q. Sequence numbering corresponds to mature helianthinin subunit.

```
AAA33374 23 IEVIQAEAGVTEIWDAYDQQFQCAWSILFDTGFN-----LVAFSCLPTSTPLFWPSSREGVILPGC 88
1OD5 20 DHRVESEGGLIETWNSQHPELQCAGVTVSKRTLNRNGLHLPSYSPYPQMIIVVQGKGAIGFAFPGC 85
...*.*. * *.. ..*** . .* * ..* * . . * .***
GE510121 23 IEVIQAEAGVTEIWDAYDQQFQCAGVDFIRHRIQPGGLLLPSYVNTPILAFVERGRGIQGVILPGC 93
2E9Q 38 VRRAEAEAGFTEVWDQDNDEFQCAGVNMIRHTIRPKGLLLPGFSNAPKLIFVAQGFGIRGIAIPGC 103
. .**** **.** ...******. *** *.* *****.. *.* * ** .* **.*. .***
A' A helix B C D E
```

Fig.3. Pair-wise structural alignments of target helianthinin and respective template sequences used for structure modeling. Sequence region peculiar in the subunit AAA33374 is printed in bold. Secondary structures (see Fig.2) are shown in the lower row. Sequence numbering corresponds to mature subunits.

As expected, the model constructed on the basis of the amino acid sequence deduced from the EST GE510121 revealed a high degree of similarity to pumpkin (2E9Q and 2EVX) and soybean (1OD5 and 1FXZ) 11S globulin structures excluding interrupted strand B (Fig.2) and strand I. In contrast, the model constructed on the basis of the original AAA33374 sequence revealed missing of the strand C (Fig.2) and also both the strands J and J' (not shown) involved in formation of FEHCJ/J' sheet [3].

It should be taken into account that both the structures shown at Fig. 2 are models that have to be proved by X-ray diffraction analysis of respective crystals. However, presence of the entire FEHCJ/J' sheet in GE510121 model and disappearance of its half due to removal of the strand C in the AAA33374 model seems plausible. In any case, it seems very probable that the peculiar primary structure of helianthinin subunit AAA33374 should bring about peculiarity of its tertiary structure. Decreased stability and enhanced susceptibility of this peculiar structure to proteolytic attack can be suggested. Presence a fraction quickly removed by unlimited proteolysis in helianthinin preparation was reported [4]. It might be that this unstable fraction corresponds to helianthinin hexamers containing really existing subunit AAA33374.

Recently, we found presence of two helianthinin fractions different in their susceptibility to papain attack: an unstable fraction quickly removed by cooperative proteolysis and ordinary one slowly hydrolyzed [5]. One of subunits inside slowly hydrolyzed fraction turned out short and inaccessible for limited proteolysis. Amino acid sequence deduced from sunflower EST BU023319, which covers practically total α-chain region of 11S globulins, is extremely short and might correspond to this subunit. According to structural alignment (Fig. 4), the deduced sequence practically lacks extended disordered regions, which are targets for limited proteolysis in all ordinary 11S globulins [6].

```
Hel 678 NQCQIQRINALEPNERVEAEAGFTEFFDSGDQQFQCAGVEVIRHHIQPQGLMLPSYINTPLMVYILQGRGYQGI
Gly 1 itsskfNECQLNNLNALEPDHRVESEGGLIETWNSQHPELQCAGVTVSKRTLNRNGLHLPSYSPYPQMIIVVQGKGAIGF
 --1--- A' A h B C D E
Hel MLPGCPETFQSSqqaeer-------gfQDRHQKIRHFREGDMIIIPAGAAHWMYNDGQEEIVAIVMLDSTNEVNQLDQFHRRFF
Gly AFPGCPETFEKPqqqssrrgsrsqqqlQDSHQKIRHFNEGDVLVIPPGVPYWTYNTGDEPVVAISLLDTSNFNNQLDQNPRVFY
 E' -------2------- F' F G H I J
Hel LAGNPKRSGQETPEe--------------------GSGNIFRGFDLQMLSEVFHVDQETASKLQSPDDTRGHIVMVERGLQVVR
Gly LAGNPDIEHPETMQqqqqqkshggrkqgqhqqqeeEGGSVLSGFSKHFLAQSFNTNEDTAEKLRSPDDERKQIVTVEGGLSVIS
 h ----------3---------- h h h J'
Hel PPIriqgpa-----------------------------------------------------------n↓G 1
Gly PKWqeqedededdedeeyeqtpsypprrpshgkheddededeeedqprpdhppqrpsrpeqqeprgrgcqtrn↓G 493
 ---------------------------4----------------------------------------
```

Fig. 4. Structural alignment of amino acid sequences deduced from *H. annuus* EST BU023319 (Hel) and soybean glycinin 1OD5 (Gly) created from modeling of the EST-deduced sequence as a target and 1OD5 as a template. Identities and conserved substitutions are printed in bold. Low case letters, disordered regions 1-4 that are targets for limited proteolysis in 11S globulin structures. An arrow indicate processing site between 11S globulin α- and β-chains. Sequence numbering corresponds to EST (3'5', frame 1) and mature glycinin sequences.

In summary, modeling of α-chain structures of highly heterogeneous helianthinin [7] indicates real existence of two kinds of subunits, an ordinary one containing all structural elements (including extended disordered regions) characteristic of other 11S globulins, and a specifically short subunit that practically lacks usually extended EF loop, β-barrel/α-helix junction and hypervariable C-terminus. Although some circumstantial evidences support the existence of a third kind of subunit of peculiar primary and tertiary structure, it still remains unclear whether it is not a result of gene sequencing error.

References:

1. Zakharov A., Artykova G., Shutov A. Peculiarity of sunflower 11S seed storage globulin // Analele Ştiinţifice ale USM. Şeria „Ştiinţe chimico-biologice". - Chişinău: CEP USM, 2003, p.234-136.
2. Shutov A.D., Kakhovskaya I.A., Braun H., Bäumlein H., Müntz K. Legumin and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene // J. Mol. Evol. - 1995. - Vol.41. - P.1057-1069.
3. Adachi M., Takenaka Y., Gidamis A.B., Mikami B., Utsumi S. Crystal structure of soybean proglycinin A1aB1b homotrimer // J. Mol. Biol. - 2001. - Vol.305. - P.291-305.
4. Артыкова Г. Протеолиз запасного 11S глобулина подсолнечника // Analele Ştiinţifice ale USM. Seria „Ştiinţe chimico-biologice". - Chişinău: CEP USM, 2003, p.102-105.
5. Макаева Э., Лаптева Н., Рудакова А., Рудаков С., Каховская И., Шутов А. 11S глобулин подсолнечника: кинетика ограниченного и кооперативного протеолиза папаином // Studia Universitatis, Revista Ştiintifică a Universităţii de Stat din Moldova. - 2009 (в печати).
6. Shutov A.D., Blattner F.R., Bäumlein H., Müntz K. Storage and mobilization as antagonistic functional constraints of seed storage globulin evolution // J. Exp. Bot. - 2003. - Vol.54. - P.1645-1654.
7. Dalgalarrondo M., Raymond J., Azanza J.L. Sunflower seed proteins: characterization and subunit composition of the globulin fraction // J. Exp. Bot. - 1984. - Vol.35. - P.1618-1628.

Prezentat la 30.01.2009