Abstract: - For more than a century, natural oligopeptides have attracted scientific attention as important biochemical regulators. Since that time, thousands of natural oligopeptide regulators have been described, and now ~600 new natural oligopeptides emerge annually, out of a literature of >70 000 publications each year on oligopeptide chemistry and biology according to PubMed database. Their primary structure is determined either directly or by translation from nucleotide sequences. Both these ways are experimental and laborious. But there are a lot of unrecognized protein sequences in different public databases which can contain unknown oligopeptide structures. Thereupen we have carried out a theoretical structure–function analysis of known uncharacterized protein amino acid residue sequences in order to identify new oligopeptide primary structures. As an example, grape (Vitis vinifera) proteins were chosen. A special computer analysis was developed for such analysis. The data of GenBank and SwissProt databases containing primary structures of unrecognized grape proteins, EROP-Moscow database containing information on structure and functions of plant regulatory oligopeptides and specially created computer programs for the comparison of GenBank or SwissProt information with EROP-Moscow data were used. This method permitted to reveal new potentially active regulatory oligopeptide sequences after alignment procedure. It was been found 21 grape protein structure sites homologues to known regulatory oligopeptides elucidated from other plant species. Their similarity with other plant oligopeptide primary structures was from 54.4 to 95.7%. They can be characterized as putative antimicrobial oligopeptides and rapid alkalinization factors. The problem of existence of these oligopeptide structures in grape is discussed. It has been proposed that rapid alkalinization factors can also possess antimicrobial activity. This way of oligopeptide structure elucidation can be extended to oligopeptide structures of any functional class.

Key-Words: - Oligopeptide, Primary Structure, Antimicrobial Peptide, Rapid Alkalinization Factor, Protein Database, EROP-Moscow Database.

1 Introduction

For more than a century, natural oligopeptides have attracted scientific attention as important biochemical regulators. The first described oligopeptide, carnosine (β-ala-his), was discovered in Russia by V.S. Gulevitch and S. Amiradzhibi in 1900 [1] and its chemical structure was determined in 1918 [2].

Regulatory oligopeptides generally do not exceed ~50 amino acid residues [3], and they differ substantially from larger polypeptides (proteins) in their physicochemical and biological properties. It is known that these substances regulate nearly all vital processes [4, 5]. Primary structure of nearly 9000 oligopeptides from more than 1700 different living organisms representing all the biological kingdoms have been identified today [6]. They possess a wide spectrum of biological activity.

Most of the information on the structure and function of endogenous molecules of the chemical peptide class contained in different databases formed mainly on the basis of amino acid (e.g., SwissProt/TrEMBL database, http://ca.expasy.org) and/or nucleotide sequences (e.g., GenBank database, http://www.ncbi.nlm.nih.gov/Genbank) at present. There are a lot of other databases for different naturally occurring peptide structures, and
those are specialized for certain structural, functional and other features.

2 Problem Formulation

Since that time, thousands natural oligopeptide regulators have been described, and now ~600 new natural oligopeptides emerge annually, out of a literature of >70 000 publications each year on oligopeptide chemistry and biology according to PubMed database (http://www.ncbi.nlm.nih.gov). Their primary structure is determined either directly or by translation from nucleotide sequences. Both of them are experimental and laborious.

From the other hand there are a lot of unknown oligopeptide sequences. For example, one only grape (Vitis vinifera) oligopeptide primary structure was extracted and characterized before among more than 600 plant oligopeptide regulators [7] whereas information on numerous grape uncharacterized proteins has to be found in public protein databases.

Since antimicrobial peptides play important role in the innate defense system of plants [8] and represent a relatively unexplored source of antimicrobial peptides of biotechnological potential [9] it is necessary to obtain more information on their structures and functions.

Except for this, several years ago the oligopeptide possessing new functional properties was discovered in common tobacco (Nicotiana tabacum) leaves [10]. It consisted of 49 amino acid residues and induced a rapid alkalinization of the culture medium of tobacco suspension-cultured cells and a concomitant activation of an intracellular mitogen-activated protein kinase. It was named as rapid alkalinization factor (RALF). A family of oligopeptides inducing rapid pH alkalinization in hybrid poplar (Populus trichocarpa × Populus deltoides) cell culture medium was isolated later from hybrid poplar leaves [11]. RALF genes were elucidated in tomato (Lycopersicon esculentum), garden pea (Pisum sativum), barrel medic (Medicago truncatula), ice plant (Mesembryanthemum crystallinum), soybean (Glycine max), rice (Oryza sativa), wheat (Triticum aestivum), maize (Zea mays), sorghum (Sorghum vulgare), barley (Hordeum vulgare), cryptomeria (Cryptomeria japonica) [10], native tobacco Nicotiana attenuata [12], mouse-ear cress Arabidopsis thaliana [10, 13, 14], etc., but not in grape.

Thereupon the aim of the current study was an identification of unknown grape (Vitis vinifera) oligopeptide structures in uncharacterized proteins.

3 Problem Solution

We have carried out a theoretical structure–function analysis of all known uncharacterized grape protein amino acid residue sequences in order to identify new primary structures of oligopeptides possessing antimicrobial and RALF activities.

A special computer analysis was developed [15, 16]. The data of GenBank database (http://www.ncbi.nlm.nih.gov/sites/entrez) and SwissProt (http://www.expasy.org/sprot/) database containing primary structures of unrecognized grape proteins, EROP-Moscow (Endogenous Regulatory OligoPeptides) database (http://erop.inbi.ras.ru/) containing information on structure and functions of plant regulatory oligopeptides [3, 6] and specially created computer programs for the comparison of GenBank or SwissProt information with EROP-Moscow data were used. The EROP-Moscow database has been used because it contained data on many oligopeptides that were absent from other convenient databases. To date, protein databases contained more than 1 million plant (Viridiplantae) sequences and EROP-Moscow contained information on structure of 9003 natural oligopeptides. In total 91592 grape protein sequences were selected. Their primary structures were compared with all EROP-Moscow regulatory oligopeptides. The identity in the amino acid residue sequences was equal or more than 30%.

This method permitted to reveal several new potentially active antimicrobial oligopeptide sequences (Figure 1) after alignment procedure. 14 grape protein structure sites homologues to known regulatory oligopeptides elucidated from other plant species have been found. Some of them were identical and therefore 11 different primary structures were elucidated (I-XI). They belong to 5 structural families. Their similarity with other plant primary structures was from 65.9 to 95.7%.

It also appeared that all grape putative antimicrobial oligopeptide primary structures except one (IX) contained eight or four Cys residues forming potentially four or two disulfide bridges respectively. Their position in all grape oligopeptides was the same (without deletions or insertions) as in plant oligopeptides obtained from other sources. So, structures I-XIV were families of basic, cysteine-rich peptides of 41–50 amino acids in size. Overall homology at the amino acid residue level inside the group of natural oligopeptides might be high and low. However most groups contain several additional conserved residues other than Cys residues.
**Figure 1.** The result of alignment of putative grape antimicrobial oligopeptides. They are grouped in structural-homologous families. Accession number in EROP-Moscow (http://erop.inbi.ras.ru/) and Swiss-Prot/TrEMBL (http://www.expasy.org/sprot/) databases, name of the oligopeptide, isolation source, the amino acid sequence, homology level, and the reference are consecutively indicated in each line. The standard one-letter code is employed for amino acid residues. *translated from nucleotide sequence.

<table>
<thead>
<tr>
<th>RALF source</th>
<th>primary structure</th>
<th>identity, reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RALF (common tobacco)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>100.0 [6]</td>
</tr>
<tr>
<td>RALF (native tobacco)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>100.0 [6]</td>
</tr>
<tr>
<td>RALF (mouse-ear cress)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>100.0 [6]</td>
</tr>
<tr>
<td>RALF (tomato)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>98.0 [6]</td>
</tr>
<tr>
<td>RALF (garden pea)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>98.0 [6]</td>
</tr>
<tr>
<td>RALF1 (grape)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>89.8 this work</td>
</tr>
<tr>
<td>RALF (barrel medic)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>89.8 [6]</td>
</tr>
<tr>
<td>RALF1 (poplar hybrid)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>88.7 [7]</td>
</tr>
<tr>
<td>RALF2 (poplar hybrid)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>85.7 [7]</td>
</tr>
<tr>
<td>RALF3 (poplar)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>85.7 [6]</td>
</tr>
<tr>
<td>RALF (rice)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>85.7 [6]</td>
</tr>
<tr>
<td>RALF33 (mouse-ear cress)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>85.7 [10]</td>
</tr>
<tr>
<td>RALF22 (mouse-ear cress)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>83.5 [10]</td>
</tr>
<tr>
<td>RALF (soybean)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>83.7 [6]</td>
</tr>
<tr>
<td>RALF2 (grape)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>83.7 this work</td>
</tr>
<tr>
<td>RALF3 (grape)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>83.7 this work</td>
</tr>
<tr>
<td>RALF23 (mouse-ear cress)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>82.0 [10]</td>
</tr>
<tr>
<td>RALF3 (mouse-ear cress)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>81.6 [6]</td>
</tr>
<tr>
<td>RALF (cotton)</td>
<td>CT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>81.6 [6]</td>
</tr>
<tr>
<td>RALF (maize)</td>
<td>YGG-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>73.5 [6]</td>
</tr>
<tr>
<td>RALF (sorghum)</td>
<td>Y-NGS-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>73.5 [6]</td>
</tr>
<tr>
<td>RALF (barley)</td>
<td>Q-NGS-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>73.5 [6]</td>
</tr>
<tr>
<td>RALF (rice)</td>
<td>QQGSGY2YGDDLADDSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>70.0 [6]</td>
</tr>
<tr>
<td>RALF (wheat)</td>
<td>QQGSGY2YGDDLADDSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>68.0 [6]</td>
</tr>
<tr>
<td>RALF (cryptomeria)</td>
<td>AT-XXY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>65.4 [6]</td>
</tr>
<tr>
<td>RALF4 (grape)</td>
<td>AAGGKPSYGLKKNNPVRNSYRCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>64.1 this work</td>
</tr>
<tr>
<td>RALF6 (grape)</td>
<td>AAAHYRRLQKNNPVRNSYRCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>64.1 this work</td>
</tr>
<tr>
<td>RALF5 (grape)</td>
<td>AAHYRRLQKNNPVRNSYRCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>62.3 this work</td>
</tr>
<tr>
<td>RALF (pine)</td>
<td>AGY-RTY13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>59.6 [6]</td>
</tr>
<tr>
<td>RALF7 (grape)</td>
<td>VMKYKQXX13YGALQKNVSNPCRASGYYNCX-AGANPYNYSRGCSCAIRRC-RS</td>
<td>54.3 this work</td>
</tr>
</tbody>
</table>

**Figure 2.** The result of multisequence alignment of putative grape rapid alkalization factors. They are presented as structural-homologous family. Name of the oligopeptide, isolation source, the amino acid sequence, homology level, and the reference are consecutively indicated in each line. The standard one-letter code is employed for amino acid residues.
It has been shown that grape uncharacterized protein and chitinase sites considered as oligopeptides were plant oligopeptide structure homologs possessing different functions. But most of them were antimicrobial agents inhibiting the growth of numerous species of bacteria and fungi.

This method revealed also several new structures of potentially active rapid alkalinization factors in grape after alignment procedure (Figure 2). Nineteen (I-XIX) grape hypothetical proteins and unnamed protein products (potential RALF precursors) were found homologous to known regulatory oligopeptides (1-4 in Figure 2) isolated from common tobacco (Nicotiana tabacum) [10], native tobacco (Nicotiana attenuata) [12], and hybrid poplar (Populus trichocarpa × Populus deltoids) [11]. All of them were characterized by a conserved sequences containing four cysteine residues forming potentially two disulfide bridges. As a result primary structures (RALF1-RALF7) of grape unrecognized proteins formed a structural family containing seven new different putative oligopeptide structures not described before. They consisted of from 49 to 57 amino acid residues and their similarity with other plant RALF primary structures was from 54.4 to 90.0%.

Structures of RALF oligopeptide hormones were characterized as highly conserved homologs throughout the plant kingdom. In Figure 2 the amino acid residue sequences of seven found grape RALF oligopeptides compared with twenty three primary structures (determined directly or by translation from nucleotide sequences) were shown. These sequences have been reported from numerous plant species and extracted from the EROP-Moscow database. They exhibit a high percentage of similarity (from 54.4 to 89.8%). All RALFs contain two disulfide bridges formed by four cysteine residues and these cysteine residues are conserved in all of 19 grape (Vitis vinifera) RALF sequences. The conservation of these residues indicates that disulfide bonding and overall structure are similar among the RALFs. It has been shown that bridges were formed between Cys18 and Cys28 and between Cys41 and Cys47 in common tobacco [10].

The homology in primary structure found among widely divergent plant species suggests that the RALF primary sequences have been highly conserved over millions of years and that RALF have a role of fundamental importance in many plant families. The RALF in Figure 2 were derived from a variety of tissues including roots, shoots, leaves and flowers. Thus, the expression of the RALF precursor gene does not appear to be associated with any particular plant cell types, suggesting that it may be involved in some basic physiological process that is common to different aspects of growth and development [10].

It has been proposed that the mature tobacco RALF oligopeptide is released from its precursor by proteolysis [10, 14]. That appears likely because there is specific subtilase site [17] with a dibasic amino acid motif RRXL (did not shown in Figure 2) immediately upstream from the predicted N terminus of the mature tobacco and poplar RALF oligopeptides. It occurs in all the nineteen related grape structures. This site is identical to the protease recognition site in mouse-ea cress AtbZIP17, a membrane-associated transcription factor that is cleaved by mouse-ea cress AtS1P protease in response to salt stress [18]. AtS1P is Golgi-localized subtilisin-like serine protease (subtilase), similar to S1P in mammalian cells [19].

Existence of antimicrobial and RALF oligopeptide structures must be confirmed by their direct extraction from grape plants and subsequent sequencing. It has been performed already for three known plant oligopeptides listed in Figure 2: one for common tobacco [10] and two for hybrid poplar [11]. Twenty structures were found in the precursors of putative oligopeptides after translation of nucleotide sequences of EST to amino acid residue sequences and comparison of their precursors with precursors of known oligopeptides. Homologous precursor sites were named as predicted oligopeptide structures, but the exact place of oligopeptide excision from precursor was not detected experimentally.

The same way was chosen in our work. We did not know exactly where unrecognized plant proteins (precursors of putative oligopeptides) might be splitted. However, some evidences of predicted sequences existed. The same position of dibasic amino acid motif RRXL and oligopeptide in protein and in known precursors of plant oligopeptides at the C terminus was one of these proofs. It turned out that grape precursor sequence beyond a predicted oligopeptide might be homologous to such sequences in oligopeptide precursors of other plants. Except for this, exact number of amino acid residues in structure with absence of the beyond residues in many structures also testified that predicted sequence could exist really. Predicted oligopeptides of other subfamilies could be considered real only if high level of homology of precursor site with known plant RALFs was proved. However, the problem of existence of predicted oligopeptides might be finally solved after their direct extraction from the grape and following sequencing.
Functional type of the predicted oligopeptide is usually postulated to be the same type as of known homologs. However, Figure 2 demonstrates that RALF oligopeptides contain up to 14 positively charged residues Arg or Lys (grape RALF6), but negative charged residues Asp or Glu are found very rare. This feature is common to a wide variety of antimicrobial oligopeptides [15, 20, 21]. On the other hand, some antimicrobial oligopeptides have the same system of disulfide bridges (first-second and third-fourth cysteine pair linkages) as RALFs and similarity in primary structures. For example, this type of bridges was found in bacterial antimicrobial oligopeptides divercin V41 [22] and pediocin PA-1 [23]. They have low homology with RALF structures, but this is a distinctive property of plant antimicrobial oligopeptides [24, 25].

These similarities indicate that RALF molecules can potentially participate in other plant regulatory processes, i.e., to be the polyfunctional regulators. Unfortunately their functions were usually studied in one test only. For the development of our complete knowledge on regulatory role of both putative grape and other plant oligopeptides it is necessary to use different experimental models.

4 Conclusion

More than 20 new structures of potentially active oligopeptides of two types of activity were predicted. But this way of oligopeptide structure elucidation can be extended to structures of any functional class.

5 Acknowledgments

This work was supported by the Chilean National Science and Technology Research Fund FONDECYT (grant No. 1080504) and by the Program of the Russian Academy of Sciences Presidium “Molecular and Cell Biology”.

References:


