Sequence analysis in short functionally important peptides by combination of bioinformatics, molecular dynamics and testing of biological activity

N.T. Moldogazieva, D.D. Argun, A.A. Gerasim, R.A. Khanmuradov, I.M. Mokhosoev, O.S. Komarov, A.A. Terentiev

N.I. Pirogov Russian National Research Medical University,
Department of Biochemistry
117997, 1 Ostrovityanova street, Moscow, Russia
nmoldogazieva@mail.ru

K.V. Shaitan

M.V. Lomonosov Moscow State University
Department of Bioengineering,
119991, 1 Vorobiovy Gory, Moscow, Russia

Introduction: It has become recognized that most regulatory physiologically/biologically active proteins have poly-functional character that is provided by their multi-modular structure [1]. Moreover, it has been demonstrated that evolutionary unrelated and non-homologous proteins may possess similar structural modules that underlie their common functions [2]. This leads to their functioning in coordinated and cooperative manner in regulating cell proliferation, differentiation, migration and apoptosis. Among the protein structural/functional modules, short (of 3 to 10 amino acid residues) linear peptide motifs of proteins have been shown to participate in receptor binding (RGD), secretion and transportation of proteins from and to endoplasmic reticulum (KDEL and HDEL), SH3-domain-binding and cell signaling (PXXP motif) and so on [3, 4]. Previously, we discovered short heptapeptide segment with amino acid sequence LDSYQCT in polypeptide chain of alpha-fetoprotein (AFP), major human tumor-associated embryo-specific protein [5]. This segment was shown to encompass amino acid residues 14 to 20 in mature protein and was designated $AFP_{14–20}$. Afterward, $AFP_{14–20}$-like motifs were revealed in most regulatory protein including epidermal growth factor (EGF) family, transforming growth factor-beta (TGF-beta) superfamily, and pregnancy specific beta-1 glycoprotein (PSG) subfamily of carcino-embyonic antigen (CEA) family of Ig-superfamily and in CEA itself [6, 7]. In this work, we performed sequence analysis to discover influence of individual amino acid residue of conformational/dynamic properties and biological activity of $AFP_{14–20}$-like short peptide fragments of regulatory proteins mentioned above and correlation between these two groups of parameters. The novelty of this approach is in consideration of changes in conformational/dynamic behavior of amino acid residues upon making single amino acid substitution as a basis for change in biological activity of the peptides studied. Earlier, only physico-chemical properties of amino acids were considered to be responsible in change of

biological/physiological activities of peptides and proteins. In assessment of role of amino acid substitution, we took into account both naturally occurring and artificially performed replacements of a definite amino acid residue at the same position in the same peptide.

Materials and methods: Amino acid sequences of proteins were extracted from UniProtKB/Swiss-Prot database. FASTA algorithm was used to search for sites of local similarity and to perform local sequence alignment. In all similar peptide motifs revealed, quantitative and qualitative analysis of amino acid substitutions was performed. Short tetra- penta- and heptapeptide fragments were then computationally constructed. Point amino acid substitutions were made in each peptide to obtain their analogs. Molecular dynamics simulation method in implicit water model is used to study conformational and dynamic properties of short peptide fragments. For this purpose, analysis of 2D and 3D maps of free energy levels (based on Ramachandran maps) along with autocorrelation and cross-correlation functions are performed. Indirect immunofluorescence method was used to test ability of chemically synthesized peptides to regulate expression of CD antigens on the surface of T-lymphocytes.

Results: $AFP_{14–20}$-like heptapeptide motifs were revealed in human EGF and TGF-beta family growth factors along with CEA protein. They have the following sequences (Figure 1): LDKYACN in EGF (residues 26–32), LDTNYCF in TGF-beta (residues 2–8). Human PSGs and CEA penta- and tetrapeptide motifs were revealed (Table1). Among them consensus PxxP motifs with the following sequences: PETP (amino acid residues 171-174), PKLP (amino acid residues 237-240), PSVP (residues 281-284), and PDLP (residues 332-335). Also, they have tetrapeptide motifs of YxCx sequence represented by peptides YQCE (residues 215-218), YECE (residues 308-311) and YACS (residues 392-395) in PSGs. Additionally, all PSGs contain integrin-binding tripeptide RGD (residues 127-129). Molecular dynamics simulation study allowed revealing differences in conformational behavior and dynamic characteristics of each amino acid residue in the peptides studied. This gives rise to explanation of possible role of these residues in functioning of the peptide motifs in PSGs. Correlation between changing in conformational/dynamic properties and biological activity of the peptide studied was shown. Peptides LDSYQCT, PYECE and YECE were the most active in ability to elevate level of CD95 and its ligand CD95L on the surface of T-lymphocytes from patients with rheumatoid arthritis and infectious myocarditis. The same peptides were characterized by less amino acid residue conformational mobility.

Conclusion: Revealing role of individual amino acid residues in proteins and their biological; active segments allows studying molecular basis underlying their functioning. Our data may be used in constructing and designing new drugs based on the peptides studied with necessary properties.

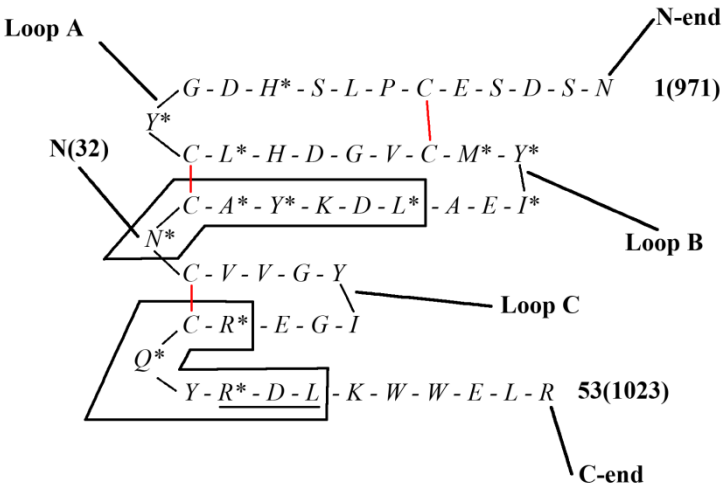

Figure 1. Primary structure of human EGF. Asterisks show amino acid residues that participate in receptor binding. $AFP_{14–20}$ –like peptide segments are shorn in frames.

Table 1. Short functionally active peptides of human PSGs

PSGs	Amino acid sequences of short peptides			
	RGD	PYECE	PYQCE	LYVCS/LYACS
PSG1	127–129 (GDD)	214–218	307–311	391–395 (LYVCS)
PSG2	127–129	214–218	——	298–302 (LYVCS)
PSG3	127–129	214–218	307–311	391–395 (LYACS)
PSG4	127–129 (RRD)	214–218	307–311	391–395 (LYACS)
PSG5	127–129	214–218	——	298–302 (LYTCS)
PSG6	126–128	213–217	306–310	390–394 (LYACS)

PSG7	127–129	214–218	307–311	391–395 (LYACS)
PSG8	127–129 (GGD)	214–218	307–311	391–395 (LYACS)
PSG9	127–129	214–218	307–311	391–395 (LYACS)
PSG11	127–129	214–218	——	298–302 (LYACS)

References

1. Zaretsky J.Z., Wreschner D.H. (2008) Protein multifunctionality: principles and mechanisms. *Translat. Oncogenomics*, 3: 99–136.

2. Neduva V., Russel R.B. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett.*, **579(15):**3342–3345.

3. Semenza J.C., Hardwick K.G., Dean N., Pelham H.R. ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. (1990) *Cell*, **61(7):**1349–1357.

4. Ruoslahti E. (2003) The RGD story: a personal account. *Matrix Biol.*, **22:**459–465.

5. Terentiev A. A., Moldogazieva N.T. Alpha-fetoprotein: a renaissance. (2013) *Tumor Biol.*, **34(2):**2075-2091.

6. Terentiev A.A., Moldogazieva N.T. Epidermal growth factor-like motifs in proteins associated with embryonic development and tumor growth: possible markers for common functions. (2007) *In: "Tumor Markers Research Focus." (Ed.: Deborah H. Chang). – NOVA Science Publishers (Chapter VII)*, p.163–176.

7. Moldogazieva N.T., Terentiev A. A., Antonov M. Yu., Kazimirsky A. N., Shaitan K.V. (2012) Correlation between biological activity and conformational and dynamic properties of tetra- and pentapeptide fragments of fetoplacental proteins. Biochemistry (Mosc.), **77 (5):**469–484.