Volume 239, number 1, 141–146 FEB 06437 October 1988

Thymopoietins and long postsynaptic neurotoxins share common information in their primary structures

Vesna Skerl and Mirjana Pavlović

The Boris Kidrič Institute, Vinča Laboratory for Multidisciplinary Research 180/02, 11001 Beograd, PO Box 522, Yugoslavia

Received 16 August 1988

The informational content of the primary structure of thymopoietin (TP) is investigated using the informational spectrum method (ISM). We show that the sequence of TP shares common information with the sequences of long postsynaptic snake neurotoxins, although no apparent similarity was found among their primary structures. The most sensitive point in the sequence of TP, concerning this information, is D-34, previously determined as being the residue responsible for TP's effect on neuromuscular transmission. Our results suggest that TP and long toxins recognize the neuromuscular nicotinic acetylcholine receptor (AChR) and/or bind to the AChR in a different mode than the short toxins do.

Sequence analysis; Thymopoietin; Postsynaptic neurotoxin; Nicotinic acetylcholine receptor

1. INTRODUCTION

Thymopoietin (TP) is a polypeptide immunoregulatory hormone isolated from the thymus, discovered by its effect in causing a delayed impairment of neuromuscular transmission [1]. The sequences of bovine and human TP (bTP and hTP) consist of 49/48 residues as are the sequences of bovine and human splenin (bSP and hSP), homologous proteins isolated from the spleen, but which do not affect neuromuscular transmission [2–6]. Synthetic pentapeptides called thymopentin (TP5) and splenopentin (SP5), corresponding to residues 32–36 of TP/SP, have been shown to reproduce the biological activities of the native proteins and probably comprise their active sites [2,3]. The only structural difference between TP5 and SP5 occurs at position 3, corresponding to the residue 34 of TP/SP. TP binds with high affinity to the nicotinic acetylcholine receptor (AChR) from the electric organ of *Torpedo californica*, competing with the snake venom α-bungarotoxin for the same binding region on AChR [4]. At lower concentrations and in the presence of Ca^{2+}, TP favors the AChR desensitization, displacing the conformational equilibrium of the AChR towards its desensitized state, and increasing the transition rate towards the same state [5].

Postsynaptic snake neurotoxins are protein components of snake venoms producing severe impairments of neuromuscular transmission by binding specifically to nicotinic AChR on the postsynaptic membrane of the motor endplate [7]. They can be classified into two classes usually called 'short' and 'long' toxins. Short neurotoxins contain 60–62 amino acids, while the long ones usually have 71–73 residues. Toxins from both classes produce a neuromuscular block upon binding to the AChR, but their binding modes seem to differ, since they produce different effects on AChR [8]. The binding of postsynaptic toxins to AChR probably involves a conformational change of the toxin and/or AChR, which may enhance their affinities for each other and/or allow them to produce a mutually locked complex [8].

Correspondence address: V. Skerl, The Boris Kidrič Institute, Vinča Laboratory for Multidisciplinary Research 180/02, 11001 Beograd, PO Box 522, Yugoslavia

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

Up to now, more than 60 postsynaptic neurotoxins from snake venoms have been isolated and sequenced, forming one of the largest groups of homologous protein sequences with determined structural characteristics [7,9]. It is intriguing to find out why TP, with no apparent similarity in its primary structure with snake toxins, affects the neuromuscular transmission, by binding to AChR in a competitive way with α-bungarotoxin. We show in this paper that common information exists in the primary structures of TPs and of long neurotoxins, undetectable by the sequence homology analysis but evident with ISM, which could result in their similar behavior concerning their ability to recognize the AChR and/or to bind to it.

2. THE INFORMATIONAL SPECTRUM METHOD

The informational spectrum method (ISM) is a theoretical method for analysis of the informational content of protein and nucleic acid sequences [10–12]. The physical basis of the ISM and its mathematical apparatus have been explained in detail elsewhere [12,13], and here we will give a description of the approach.

The basic physical parameter which influences the energy and distribution of valence electrons in an organic molecule is the potential in which they move. The electron-ion interaction potential (EIIP) determines energy states of valence electrons [14] and influences the physico-chemical properties of molecules such as their hydropathy, charge, dipole momentum, etc. [15]. The EIIP has been shown to correlate with the toxicity, carcinogenicity, antibiotic activity and other properties of small organic molecules [13,16,17]. In an attempt to determine the biological function of a macromolecule from its primary structure, the authors of the ISM consider the EIIP of amino acid or nucleotide residues to be the relevant characteristic [10–12].

The value of the EIIP corresponding to an organic molecule can be determined using the following expression derived from the general model pseudopotential [18,19]:

$$W = 0.25Z^* \sin(1.04\pi \cdot Z^*)/2\pi$$

where Z^* is the average quasi-valence number of the molecule determined by:

$$Z^* = \sum_{1}^{N} Z_i/N$$

where Z_i is the valence number of the i-th atom and N is the total number of atoms in the molecule.

Applying the given expression to 20 amino acids, the following values are obtained (in Ry): L 0.0000, I 0.0000, N 0.0036, G 0.0050, V 0.0057, E 0.0058, P 0.0198, H 0.0242, K 0.0371, A 0.0373, Y 0.0516, W 0.0548, Q 0.0761, M 0.0823, S 0.0829, C 0.0829, T 0.0941, F 0.0946, R 0.0959 and D 0.1263 [13].

DNA, RNA and proteins can be considered as informational macromolecules. The information they carry is stored in their primary structures and is coded by the distribution of their constitutive elements (nucleotides or amino acids). In analyses of the structure/function relationship, this information has to be decoded in terms of the biological/biochemical function.

The informational content of a primary structure can be presented in different forms. Let us consider three of them: the so-called 'symbolic', 'numerical' and 'spectral' presentation.

In the simplest and most commonly used symbolic presentation, each residue in the primary structure is represented by a symbol (usually a letter). The primary structures given as sequences of symbols can be compared among themselves through sequence alignments. However, a high sequence similarity score obtained through such analysis should not be regarded as a measure of functional relatedness of the analyzed sequences (in some cases even a single mutation in a sequence can destroy/alter its biological function), but only as an indicator of their common ancestry [20]. This implies that the symbolic presentation of the genetic information is not convenient for the analyses of the structure/function relationship and of the functional relatedness among different primary structures.

In the numerical presentation, the primary structure is given in the form of the numerical series, where each number in the series represents the value of some chosen physico-chemical property of the corresponding side-group. This presentation enables the determination of the distribution of the followed property along the residues in the sequence of the analyzed macromolecule. The followed characteristic may be any physico-chemical parameter expected to influence some property of the whole macromolecule. The numerical presentation of the hydropathy values, corresponding to amino acids of some protein, is very propitious for the prediction of its transmembrane parts, antigenic segments and secondary structure [21]. The physical property expected in the ISM approach to influence the biological function of the macromolecule is the EIIP, so the first step in the ISM analysis of the primary structure is the substitution of each residue symbol in the sequence by the corresponding EIIP value. The obtained EIIP series represents the side-group-influenced changes of local environments along the backbone of the macromolecule, through which a hypothetical electrical impulse propagates [15,22]. The putative high-temperature superconductivity in proteins and DNA (proposed by Little [23]) could influence the long-distance intermolecular recognitions (proposed by Frölich [24]) and preselections, based on the information carried in their primary structures (proposed by Veljković [15,22]).

As the numerical presentation of primary structures is also inconvenient for comparative analyses of functional relatedness among sequences, it is possible to transform it into the form of a spectrum which is much more convenient for further analyses. The informational spectrum (IS) of a sequence in the ISM is defined as the energy density spectrum of a discrete Fourier transform applied on the series of the EIIP values corresponding to the analyzed primary structure [10–12]. The series of EIIP values is treated in the ISM as a discrete signal, and it is assumed that the points in the signal are equidistant with the distance $d = 1$. The Fourier transform decomposes this signal into a sum of sine waves, and the obtained Fourier coefficients describe their frequencies, amplitudes and phases. Each point in

the IS, which defines the (square of the) amplitude and the frequency of one of these sine waves, depends on the collective effect of all of the constitutive elements of the sequence [12]. The maximal frequency in the IS is $F = 1/2d = 0.5$, so the IS-frequency range 0.0–0.5 is independent of the length of the analyzed sequence [10–12]. The total number of points in a sequence influences only the resolution of the IS and the accuracy of the transform.

The cross spectrum (CS) of two or more ISs calculated with the same resolution is used to extract their common information. It is defined in the following way: the intensity (amplitude) corresponding to a frequency in the CS is the product of the corresponding amplitudes from the ISs from which this CS originates. In general, the CS of several proteins contains only those peaks which appear in all their ISs. It was found that the CS of functionally unrelated sequences does not contain any significant peak, while the CS of biologically related sequences contains one or more significant peaks [10–12].

The consensus spectrum is defined as a CS of a large group of sequences which share a common biological function. It usually contains one (sometimes several) extremely significant peak(s). The frequency of this peak can be related to that particular biological function, and is considered to be 'characteristic' for that function [10–12]. The characteristic frequencies are, up to now, obtained for more than 20 groups of proteins (oncogenes, kinases, interferons, growth factors, haemoglobins, etc.) as well as for several types of DNA regulatory sequences (promotors, terminators, enhancers, SOS-operators) [10–12,22,25–30].

Some of the possibilities which the ISM offers can be summarized as follows: (i) it is possible to predict the biological function of a sequence with an unknown function; (ii) it is possible to compare the strength of the biological activity within a group of sequences with the same function; (iii) within the sequence with a known biological function, the mutations which could increase or decrease this function can be predicted; (iv) it is possible to predict which group of proteins would bind to a particular kind of DNA regulatory sequence; and (v) ISM also enables the design of completely artificial sequences which could be expected to have some desired biological function.

In this paper we use for the first time the 'minimal change mutation' (MCM) procedure, which we describe here in detail. The set of 20 amino acids is linearly arranged according to the corresponding values of the EIIP in the following way: D ≫ R > F > T > C > = S > M > Q > W > Y > A > K > H > P > E > V > G > N > L = I.

For each amino acid we define its MCMs as the two neighbouring amino acids in this order. For example the minimal 'up' mutation for W is Q and its minimal 'down' mutation is Y, so the MCMs for W are Q and Y. To determine which parts of the protein sequence are most sensitive to changes in respect to a fixed IS frequency f, we proceed as follows: (1) we determine the amplitude $a(f)$ in the IS of that protein and (2) repeat the 2a and 2b for every position in the analyzed sequence. (2a) We change the amino acid on that position by its MCMs (if both defined) and obtain two (or one) sequences which differ from the original one only in this position. (2b) We calculate $a(f)$ from each IS of these altered sequences and express it as a percentage of the same value from the IS of the original sequence. (3) Having these values for all the positions in the sequence we can compare them and conclude which positions are the most sensitive to minimal mutations in respect to frequency f. Other mutations (non-minimal) should have even more drastic influence on the change of $a(f)$ than the minimal ones. While the inverse Fourier transform enables the finding of the optimal changes in the primary structure which would decrease/increase $a(f)$ [28], the MCM procedure defines the positions in the primary structure where even minimal changes significantly affect $a(f)$.

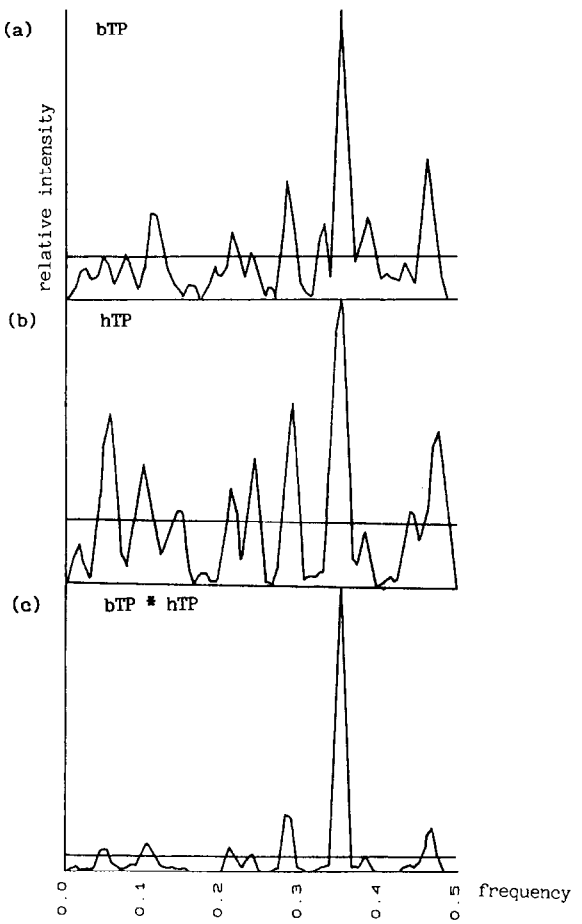

Fig.1. (a) The IS of bovine thymopoietin. (b) The IS of human thymopoietin. (c) The CS of bovine and human thymopoietins. The obtained dominant frequency in all of the three spectra is 0.35156. The EIIP signals were extended on 128 points, giving spectra with 64-point (frequencies) resolution. In each spectrum the abscissa presents the IS-frequency domain 0.0–0.5. The ordinate in each IS presents the intensity of the Fourier transform applied to the sequence of EIIP values corresponding to the amino acids in the analyzed primary structure. The ordinate in the CS presents the product of the intensities of the corresponding ISs. The horizontal line in each spectrum presents the noise level (mean amplitude value).

3. RESULTS AND DISCUSSION

The aim of our work was to study the primary structure of thymopoietin (TP) using the previously described ISM. The ISs of bovine and human TP (bTP and hTP) were calculated and their CS was determined, as presented in fig.1. The dominant peak in all three spectra appears on IS at frequency 0.35156, or, according to the accuracy of the method, in the frequency domain 0.34136–0.36176. Repeating the same procedure starting from the sequences of bovine and human splenins (bSP and hSP), we obtained similar spectra as in the case of TPs, with the same, but less significant dominant peak. The comparative results concerning IS frequency 0.35156 in the spectra of TPs and SPs are given in table 1. These results suggest that if TPs and SPs were involved in a biological process represented in the ISM model by the frequency 0.35156, then we would expect TPs to be more efficient in this process than the corresponding SPs. Which biological processes correlate with frequency 0.35156 in the spectral representation are still to be determined.

In our previous work we proposed the same frequency 0.35156 to be related to postsynaptic snake neurotoxins, as the peak on this frequency appears as the dominant one in the CS of five of them [29]. Our further studies showed that this spectral component appears as dominant in the ISs of some of the long toxins, but never in the ISs of short toxins (unpublished). In fig.2, we present the consensus spectra of two classes of postsynaptic snake neurotoxins in comparison with the CS of TPs. The obtained characteristic frequencies are: 0.35156 ± 0.01042 for TPs, 0.34375 ± 0.00758 for 22 long toxins and 0.48438 for 37 short toxins. These results reveal that a characteristic frequency domain 0.34136–0.35133 shared between the CSs of two different groups of proteins, TPs and long toxins exists, but is not present in the CS of short toxins which further means that the primary structures of TPs and of long toxins carry a common piece of information not present in the primary structures of short toxins. We concluded that this common information could be responsible for the

Table 1

The values of the amplitude and of the signal-to-noise ratio (S/N) corresponding to the IS frequency 0.35156 in the ISs and CSs of thymopoietins and splenins of bovine and human origin

	a(f)	*S/N*
bTP	0.494	6.78
bSP	0.304	4.82
hTP	0.249	4.61
hSP	0.177	3.72
bTP * hTP	0.123	18.33
bSP * hSP	0.054	11.98

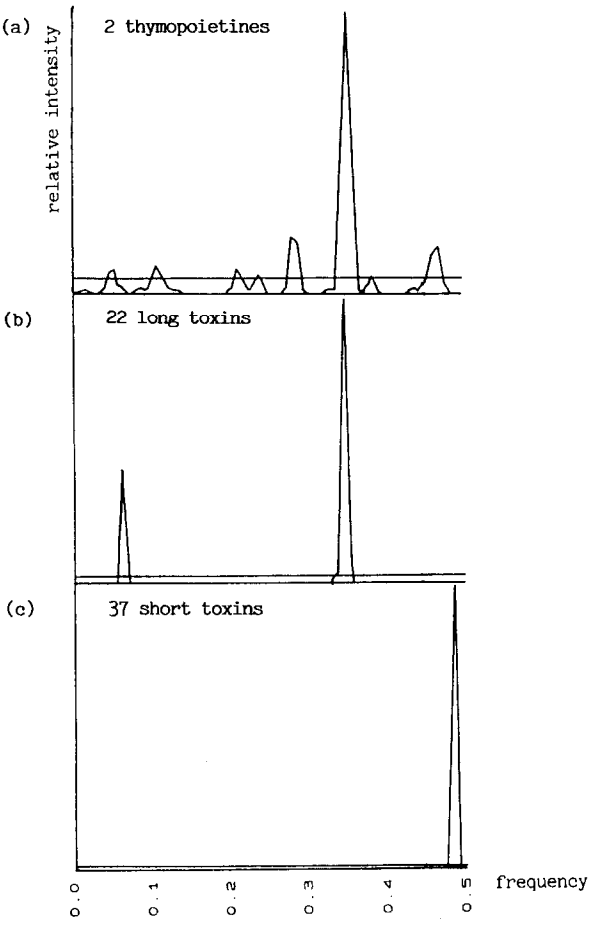

Fig.2. (a) The CS of thymopoietins of bovine and human origin with the dominant peak at frequency 0.35156. (b) The consensus spectrum of the group of 22 long postsynaptic snake neurotoxins with the dominant peak on frequency 0.34375. (c) The consensus spectrum of the group of 37 short postsynaptic snake neurotoxins with the dominant peak on frequency 0.48438. The primary structures of the toxins are taken from [7–9]. The resolution, abscissa and ordinate of each spectrum are as described for fig.1.

Table 2

The results of the MCM procedure applied to the sequence of bovine thymopoietin (bTP) and human thymopoietin (hTP)

bTP residue	MCM up	MCM down	hTP residue	MCM up	MCM down
1 P	99	103	1 G	100	101
2 E	100	100	2 L	101	–
3 F	100	100	3 P	101	98
4 L	99	–	4 K	100	105
5 E	101	100	5 E	104	100
6 D	–	95	6 V	100	100
7 P	99	104	7 P	98	105
8 S	102	100	8 A	105	100
9 V	100	100	9 V	100	100
10 L	99	–	10 L	99	–
11 T	100	97	11 T	100	96
12 K	100	100	12 K	100	103
13 E	97	100	13 Q	99	103
14 K	100	96	14 K	100	95
15 L	100	–	15 L	99	–
16 K	100	102	16 K	100	100
17 S	103	100	17 S	104	100
18 E	97	100	18 E	95	100
19 L	100	–	19 L	100	–
20 V	100	100	20 V	100	100
21 A	96	100	21 A	94	100
22 N	100	100	22 N	100	99
23 N	100	99	23 G	100	100
24 V	100	100	24 V	100	100
25 T	100	98	25 T	100	96
26 L	100	–	26 L	100	–
27 P	99	104	27 P	99	104
28 A	103	100	28 A	106	100
29 G	100	100	29 G	100	100
30 E	97	100	30 E	97	100
31 Q	102	94	31 M	100	98
32 R	98	100	32 R	92	100
33 K	100	102	33 K	100	101
34 D	–	*91*	*34 D*	–	*90*
35 V	100	100	35 V	100	100
36 Y	100	101	36 Y	100	99
37 V	100	100	37 V	100	100
38 E	97	100	38 E	95	100
39 L	100	–	39 L	101	–
40 Y	101	97	40 Y	101	98
41 L	99	–	41 L	99	–
42 Q	101	98	42 Q	102	94
43 S	102	100	43 H	100	100
44 L	99	–	44 L	99	–
45 T	100	98	45 T	100	96
46 A	101	100	46 A	98	100
47 L	99	–	47 L	99	–
48 K	100-	97	48 H	105	98
49 R	99	100			

The MCM parameter is a value of the amplitude on IS frequency 0.35156 in the IS of the altered sequence, expressed as a percentage of the same value in the IS of the native protein bTP

ability of TP to recognize AChR in neuromuscular synapses and/or bind to it. Our further investigation was directed to test this hypothesis.

This common information shared among the sequences of TPs and of long toxins is represented in the IS of TP with the spectral component 0.35156. To test whether this frequency correlates with the ability of TP to influence the neuromuscular transmission, we determined which parts of the primary structure of TP are the most sensitive, concerning the influence of mutations in them on frequency 0.35156. We used the previously described MCM procedure on the sequences of bTP and hTP, and followed the changes of amplitude on frequency 0.35156. The results are presented in table 2. In both cases, the maximal absolute value of the change of the amplitude on the followed frequency (9% decrease for bTP and 10% decrease for hTP) occurs when D-34 is substituted by its MCM (R-34). This means that residue D-34 is the most sensitive point in the sequence of TP, according to the influence of its change on the IS frequency 0.35156. As the same residue was previously determined as the main one in the active site of TP, responsible for its ability to affect the neuromuscular transmission [2,3], we concluded that the common spectral component (frequency domain) in the ISs of TPs and of long neurotoxins correlates with their ability to recognize the AChR and/or to bind to it. Further, as this spectral component is not significant in the ISs of short neurotoxins, our results are in agreement with the assumption that the modes of action of short and long toxins differ among themselves.

REFERENCES

[1] Goldstein, G. (1974) Nature 247, 11–14.
[2] Audhya, T., Schlesinger, D.H. and Goldstein, G. (1981) Biochemistry 20, 6195–6200.
[3] Audhya, T., Scheid, M.P. and Goldstein, G. (1984) Proc. Natl. Acad. Sci. USA 81, 2847–2849.
[4] Venkatasubramanian, K., Audhya, T. and Goldstein, G. (1986) Proc. Natl. Acad. Sci. USA 83, 3171–3174.
[5] Revah, F., Mulle, C., Pinset, C., Audhya, T., Goldstein, G. and Changeux, J.-P. (1987) Proc. Natl. Acad. Sci. USA 84, 3477–3481.
[6] Audhya, T., Schlesinger, D.H. and Goldstein, G. (1987) Proc. Natl. Acad. Sci. USA 84, 3545–3549.
[7] Lee, C.Y. (1979) in: Adv. in Cytopharmacology vol.3 (Ceccarelli, B. and Clementi, F. eds) pp.1–16, Raven, New York.

[8] Endo, T., Nakanishi, M., Furukawa, S., Joubert, F.J., Tamiya, N. and Hayashi, K. (1986) Biochemistry 25, 395–404.
[9] Karlsson, E. (1978) Handb. Exp. Pharmacol. 52.
[10] Veljković, V., Ćosić, I., Lalović, D. and Dimitrijević, B. (1985) IEEE Biomed. Eng. 32, 337–341.
[11] Ćosić, I. (1985) Digital Signal Analysis Applied to Protein and Nucleotide Sequences, PhD Thesis, University of Beograd.
[12] Veljković, V. and Ćosić, I. (1987) Cancer Biochem. Biophys. 9, 139–148.
[13] Veljković, V. (1980) A Theoretical Approach to the Preselection of Carcinogens and Chemical Carcinogenesis, Gordon and Breach, New York, USA.
[14] Heine, V., Koen, M. and Vir, D. (1973) Theory of Pseudopotentials, MIR, Moscow.
[15] Veljković, V., personal communication.
[16] Veljković, V. and Lalović, D. (1976) Cancer Biochem. Biophys. 1, 295.
[17] Ajdačić, V. and Veljković, V. (1978) Experientia 34, 633.
[18] Veljković, V. and Slavić, I. (1972) Phys. Rev. Lett. 29, 105–109.
[19] Veljković, V. (1973) Phys. Lett. 45A, 41.
[20] Doolittle, R.F. (1981) Science 214, 149–159.
[21] Kyte, J. and Doolittle, R.F. (1982) S. Molec. Biol. 157, 105.
[22] Vojvodić, V., Veljković, V. and Skerl, V. (1986) Proc. II Int. Symp. Protection Against Chemical Warfare Agents, pp.347–358, Stockholm, Sweden.
[23] Little, W.A. (1964) Phys. Rev. 134 (6A), A1416–A1424.
[24] Frölich, H. (1970) Nature 228, 1093.
[25] Ćosić, I., Nešić, D., Pavlović, M. and Williams, R. (1986) Biochem. Biophys. Res. Commun. 141, 831–839.
[26] Raspopović, J., Škrbić, N. and Skerl, V. (1988) Proc. III Int. Conf. MATH/CHEM/COMP, Elsevier, Amsterdam, in press.
[27] Veljković, V. and Metlaš, R. (1987) Proc. Protein Engineering 87, 102.
[28] Ćosić, I. and Nešic, D. (1987) Eur. J. Biochem. 170, 247–252.
[29] Vojvodić, V., Veljković, V. and Skerl, V. (1986) Proc. II Int. Symp. Protection Against Chemical Warfare Agents, pp.303–309, Stockholm, Sweden.
[30] Veljković, V. and Metlaš, R. (1988) Cancer Biochem. Biophys., in press.