J Mol Model (2001) 7:318–323
DOI 10.1007/s008940100047

ORIGINAL PAPER

Guang Wu · Shaomin Yan

Analysis of distributions of amino acids and amino acid pairs in human tumour necrosis factor precursor and its eight mutations according to a random mechanism

Received: 14 May 2001 / Accepted: 16 July 2001 / Published online: 5 September 2001

Abstract In this study, we use the random principle to analyse the distributions of amino acids and amino acid pairs in human tumour necrosis factor precursor (TNF-α) and its eight mutations, to compare the measured distribution probability with the theoretical distribution probability and to rank the measured distribution probability against the theoretical distribution probability. In this way, we can suggest that distributions with a high random rank should not be deliberately evolved and conserved and those with a low random rank should be deliberately evolved and conserved in human TNF-α. An increased distribution probability in a mutation means probabilistically that the mutation is more likely to occur spontaneously, whereas a decreased distribution probability in a mutation means probabilistically that the mutation is less likely to occur spontaneously and perhaps is more related to a certain cause. The results, for example, show that the distributions of 30% of the amino acids are identical with their probabilistic simplest distributions, and the distributions of some of the remaining amino acids are very close to their probabilistic simplest distributions. With respect to probabilities of distributions of amino acids in mutations, the results show that mutations lead to an increase in eight probabilities, which are thus more likely to occur. Eight probabilities decrease and are thus less likely to occur. With respect to the random ranks against the theoretical probabilities of distributions of amino acids, the results show that mutations lead to an increase in seven and a decrease in seven probabilities, with two probabilities unchanged.

Keywords Amino acids · Tumour necrosis factor precursor · Probability

G. Wu (✉)
Laboratoire de Toxicocinétique et Pharmacocinétique,
Faculté de Pharmacie,
Université de la Méditerranée Aix-Marseille II, Marseille, France

S. Yan
Cattedra di Anatomia Patologica,
Dipartimento di Ricerche Mediche e Morfologiche,
Facoltà di Medicina e Chirurgia, Università degli Studi di Udine,
Udine, Italy

Present address:
G. Wu, DMPK, WKL-135.1.16, Novartis Pharma AG,
4002 Basel, Switzerland
e-mail: guang.wu@pharma.novartis.com
Tel.: +41 61 696 7746, Fax: +41 61 696 6992

Introduction

The primary structure of proteins is the objective of numerous studies because it is the basis for the higher level structures and protein functions. However, the primary structure of proteins also provides the basis for studies and modelling of (i) the patterns of amino acid composition, (ii) the patterns of natural and artificial mutations, (iii) the similarity within a protein family, (iv) the similarity between protein families, (v) the mechanism for construction of higher level structures, (vi) the topological base for higher level structures, etc.

The patterns of amino acid composition and mutations are archived via experimental methods and annotation. [1] The most popular analysis of the similarity within a protein family and between protein families is archived via multiple sequence comparisons and alignments using standard software, for example, BlastP. [2] There are several other approaches for the similarity analysis, such as fast Fourier transform, [3] pattern graph, [4] linguistic approaches, [5, 6] the statistical approach, [7] etc.

These approaches provide a way to identify the signature pattern in a protein, because a function is assigned to a protein based on its sequence similarity to other sequences within a protein family and between protein families. However, the underlying reasoning for the patterns of amino acid composition and the patterns of natural mutations cannot be revealed by these approaches, i.e. these approaches can describe and find the patterns but cannot answer why there are such patterns, although artificial mutation can directly be related to the experimental condition.

Table 1 Distributions of four "H"s in four parts of human TNF-α, distribution pattern, calculations and distribution probabilities

Part 1	Part 2	Part 3	Part 4	Distribution pattern	Probability=4!×4!×4^{-4} divided by	Probability
H	H	H	H	1, 1, 1, 1	(0!×4!×0!×0!×0!)×(1!×1!×1!×1!)	0.09375
	H	H	HH	0, 1, 1, 2	(1!×2!×1!×0!×0!)×(0!×1!×1!×2!)	0.56250
		HH	HH	0, 0, 2, 2	(2!×0!×2!×0!×0!)×(0!×0!×2!×2!)	0.14063
		H	HHH	0, 0, 1, 3	(2!×1!×0!×1!× 0!)×(0!×0!×1!×3!)	0.18750
			HHHH	0, 0, 0, 4	(3!×0!×0!×0!×1!)×(0!×0!×0!×4!)	0.01563

Perhaps the random analysis may throw light on the underlying reasoning, not only because pure chance is now considered to lie at the very heart of nature, [8] but also because the sequence of proteins with a high probability of occurrence would be less costly and less time-consuming during its synthesis. Moreover, the random analysis may provide a clue for evolution, because it is still not clear whether or not nature selects the sequence with a high probability of occurring for protein functions.

In order to apply the random analysis to the primary structure of a protein, we choose human tumour necrosis factor precursor (TNF-α) for our analysis. Human tumour necrosis factor is mainly secreted by macrophages and belongs to the tumor necrosis factor family. It is a cytokine with a wide variety of functions: it can cause cytolysis of certain tumor cell lines, it is implicated in the induction of cachexia, it is a potent pyrogen causing fever by direct action or by stimulation of interleukin 1 secretion, it can stimulate cell proliferation and induce cell differentiation under certain conditions.

In this study, we use the random principle to analyse the possible distributions of amino acid and amino acid pairs in human TNF-α, to compare the measured distribution probability with the theoretical distribution probability and to rank the measured distribution probability against the theoretical distribution probability. In this way, we can suggest that the distributions with a high rank of randomness should not be deliberately evolved and conserved and the distributions with a low rank of randomness should be deliberately evolved and conserved.

For example, there are two tryptophans (W) among 233 amino acids in human TNF-α. Our intuition with regard to their randomness may suggest that there would be one "W'' in the first half of human TNF-α and another "W'' in the second half of human TNF-α, which is true, i.e. one "W'' is at position 104 and another "W'' is at position 190. In fact there are only three distributions of "W'' in human TNF-α, i.e. (i) both "W''s are in the first half, (ii) each "W'' is in each half, and (iii) both "W''s are in the second half, thus each distribution of "W'' has the probability of 1/3. This results in a predictable distribution in agreement with actual distributions of "W'', because any distribution has the same probability to occur, although we are still not able to define their exact positions.

To take this one step further, if we do not distinguish either the first or second half and we are simply interested in whether both "W''s are in both halves or in any half, we will have the probability of 1/2 for each situation, as if we toss two coins, either both have the same faces up or both have different faces up. This is also not surprising for the distributions of "W''s in human TNF-α because any distribution follows a random mechanism.

A little more complicated distribution of amino acids and amino acid pairs, for example, is the amino acid histidine (H), which is the second least abundant amino acid, and there are four "H''s in human TNF-α. As in the previous example, we can group the human TNF-α into four parts, each containing about 58 amino acids (233/4=58.25), and our intuition with regard to randomness may once again suggest that each part would contain an "H''. In order to find a probabilistic model for our analysis, we can relate it to statistical mechanics, in which the distribution of elementary particles among a set of energy states can be classified according to three assumptions with respect to whether or not to distinguish each particle and energy state, i.e. the Maxwell–Boltzmann, Fermi–Dirac and Bose–Einstein assumptions. [9] In our case, four "H''s and four parts are analogous to four particles and four states, respectively. If we do not distinguish four parts and four "H''s because we are only interested in how many "H''s occur in a part, then there is no difference between this "H'' and that "H'' and between this part and that part, we only have five distributions of "H''.

Table 1 shows the probabilities of five distributions of "H'' in four parts. The distribution in which one part contains zero "H''s, two parts contain one "H'' and one part contains two "H"s has the highest probability of 0.56250. This is somewhat contrary to our intuition with regard to the random distribution of "H''s, because our intuition is that each part containing an "H'' has the highest probability. What happens in the distribution of "H''s is similar to what happens in our daily life, for example, we would not expect to receive one letter every day from Monday to Saturday if six letters randomly arrive from Monday to Saturday. Thus the randomness precludes an "even" distribution of human TNF-α.

If we look at the real "H'' distributions in human TNF-α, we find that they are the same as the highest probabilistic configuration in Table 1, i.e. one part contains zero "H''s, two parts contain one "H'' and one part contains two "H''s. Here is something surprising: it is likely that nature has already chosen the most-likely-to-occur distribution of "H'' in human TNF-α or that nature is very economic and the evolution of human TNF-α had

proceeded along the probabilistically simplest and most economic way.

Furthermore, it would be of interest to know whether a mutation of human TNF-α leads to the distribution probabilities of amino acids and amino acid pairs increasing or decreasing compared with the human TNF-α, because the ultimate origin of all evolutionary change is rooted in mutations. [10] Increasing distribution probabilities would suggest that the mutation develops along the probabilistically simplest pathway and thus more easily occurs spontaneously, whereas along the probabilistically difficult pathway it is more difficult to occur spontaneously.

Materials and methods

The human TNF-α sequence was obtained from the SWISS-PROT Protein Sequence Database (access number, P01375, http://srs.ebi.ac.uk or http://www4.ncbi.nlm.nih.gov/PubMed). [11] The eight mutations causing biological lower activity or inactivity in human TNF-α are (i) L→S at position 105, (ii) R→W at position 108, (iii) L→F at position 112, (iv) A→V at position 160, (v) S→F at position 162, (vi) V→A at position 167, (vii) V→D at position 167, and (viii) E→K at position 222. [12, 13, 14, 15, 16]

The calculation of distributions of amino acids and amino acid pairs is according to the calculation of occupancy problems of subpopulations and partitions. [9] For each of distributions of amino acids and amino acid pairs, the probability is $r!/(q_0!\times q_1!\times...\times q_n!)\times r!/(r_1!\times r_2!\times...\times r_n!)\times n^{-r}$.

In the equation,! is the factorial function. r is the number of a given kind of amino acid or amino acid pair, for example, we have r=4 for "H'' and r=19 for alanine (A) because there are four "H''s and 19 "A''s in human TNF-α. n is the number of grouped parts in human TNF-α for a given kind of amino acid or amino acid pair, for example, n=4 for "H'', in fact we have r=n in this study. r_1, r_2,... r_n are the number of a given kind of amino acid or amino acid pair in part 1, 2,... n, for example, when four "H"s appear in each of four parts, we have r_1=1, r_2=1, r_3=1 and r_4=1. q is the number of parts with the same number of amino acid or amino acid pair, for example, when four "H''s appear in each of four parts, we have q_0=0, q_1=4, q_2=0, q_3=0 and q_4=0, i.e. zero parts have zero "H''s, four parts have one "H'', zero parts have two "H''s, and zero parts have three "H''s, zero parts have four "H''s.

Table 1 shows the calculation using this equation with respect to the distributions of four "H''s in human TNF-α. From Table 1, we can understand not only which distribution is more likely to occur, but also the comparison between them, for example, the distribution of "H'', "H'', and "HH" is 36 times (0.56250/0.01563) more likely to occur than the distribution of "HHHH".

Results

Table 2 shows amino acids, their numbers, their measured distribution probabilities, their maximum theoretical distribution probabilities, their random ranks against theoretical distribution probabilities and their "even" distribution probabilities in human TNF-α. When ranking a distribution to be 1, we mean that the distribution has the highest probability to occur, for example "Q''. It can be seen that the distributions of "F'' and "T'' have the lowest random ranks among 20 kinds of amino acids, thus these distributions are highly unpredictable. However 30% (6/20) of distributions of amino acids are identical with their probabilistically simplest distributions, and the distributions of the remaining amino acids are very near to their probabilistically simplest distributions because their random ranks are near to 1. Table 2 also shows that any measured distribution probability is far away from the "even" distribution probability, i.e. none of the amino acids is distributed homogeneously along human TNF-α.

Table 3 shows amino acids, their numbers, measured distribution probability, random ranks against the theoretical distribution probability and their tendencies in eight mutations compared with those in human TNF-α. With respect to probabilities of distributions of amino acids in mutations, the mutations lead eight probabilities to increase and thus to occur more easily, and eight probabilities to decrease and thus to occur with more difficulty. With respect to the random ranks against the theoretical probabilities of distributions of amino acids, the mu-

Table 2 Amino acids, their numbers, measured distribution probability (MDP), maximum theoretical distribution probability (MTDP), random ranks against theoretical distribution probability (RRTDP) and "even" distribution probability (EDP) in human TNF-α

Amino acid	Number	MDP	MTDP	RRTDP	EDP
A[a]	19	0.0895	0.1118	2	6.1486×10^{-8}
R[b]	14	0.0618	0.1649	6	7.8454×10^{-6}
N[c]	7	0.0357	0.3213	6	6.1199×10^{-3}
D[d]	7	0.1071	0.3213	4	6.1199×10^{-3}
C[e]	4	0.1406	0.5625	3	9.3750×10^{-2}
E[f]	16	0.0341	0.1362	9	1.1342×10^{-6}
Q[g]	13	0.1544	0.1544	1	2.0560×10^{-5}
G[h]	17	0.0732	0.1280	3	4.2997×10^{-7}
H[i]	4	0.5625	0.5625	1	9.3750×10^{-2}
I[j]	12	0.0399	0.1862	8	5.3723×10^{-5}
L[k]	30	0.0381	0.0593	6	1.2883×10^{-12}
K[l]	8	0.1682	0.2523	3	2.4033×10^{-3}
M[m]	2	0.5000	0.5000	1	5.0000×10^{-1}
F[n]	10	0.0286	0.1905	10	3.6288×10^{-4}
P[o]	15	0.1569	0.1569	1	2.9863×10^{-6}
S[p]	20	0.0965	0.0965	1	2.3202×10^{-8}
T[q]	10	0.0286	0.1905	10	3.6288×10^{-4}
W[r]	2	0.5000	0.5000	1	5.0000×10^{-1}
Y[s]	7	0.1071	0.3213	4	6.1199×10^{-3}
V[t]	16	0.0639	0.1362	5	1.1342×10^{-6}

[a] Alanine; [b] Arginine; [c] Asparagine; [d] aspartic acid; [e] Cysteine; [f] Glutamic acid; [g] Glutamine; [h] Glycine; [i] Histidine; [j] Isoleucine; [k] Leucine; [l] Lysine; [m] Methionine; [n] Phenylalanine; [o] Proline; [p] Serine; [q] Threonine; [r] Tryptophan; [s] Tyrosine; [t] Valine

Table 3 Amino acids, their numbers, MDP, RRTDP and their tendencies in eight mutations compared with those in human TNF-α (parentheses)

	Amino acid	Number	MDP	Tendency[a]	RRTDP	Tendency[a]
Mutation 1	L	29 (30)	0.0453 (0.0381)	↑	3 (6)	↑
L→S in 105	S	21 (20)	0.0954 (0.0965)	↓	1 (1)	–
Mutation 2	R	13 (14)	0.1158 (0.0618)	↑	3 (6)	↑
R→W in 108	W	3 (2)	0.6667 (0.5000)	↑	1 (1)	–
Mutation 3	L	29 (30)	0.0453 (0.0381)	↑	3 (6)	↑
L→F in 112	F	11 (10)	0.0539 (0.0286)	↑	7 (10)	↑
Mutation 4	A	18 (19)	0.0748 (0.0895)	↓	4 (2)	↓
A→V in 160	V	17 (16)	0.0041 (0.0639)	↓	31 (5)	↓
Mutation 5	S	19 (20)	0.0373 (0.0965)	↓	9 (1)	↓
S→F in 162	F	11 (10)	0.0539 (0.0286)	↑	7 (10)	↑
Mutation 6	V	15 (16)	0.0062 (0.0639)	↓	22 (5)	↓
V→A in 167	A	20 (19)	0.0422 (0.0895)	↓	5 (2)	↓
Mutation 7	V	15 (16)	0.0062 (0.0639)	↓	22 (5)	↓
V→D in 167	D	8 (7)	0.2523 (0.1071)	↑	1 (4)	↑
Mutation 8	E	15 (16)	0.0028 (0.0341)	↓	33 (9)	↓
E→K in 222	K	9 (8)	0.1967 (0.1682)	↑	1 (3)	↑

[a] ↑, ↓ and – are the random rank to increase, to decrease and unchanged, respectively

Table 4 Amino acid pairs that appeared more than twice, their numbers, MDP, MTDP and RRTDP in human TNF-α

Amino acid pair	Number	MDP	MTDP	RRTDP
AN	3	0.1111	0.6667	3
AE	5	0.3840	0.3840	1
AL	3	0.2222	0.6667	2
EG	3	0.6667	0.6667	1
GQ	3	0.6667	0.6667	1
GV	3	0.2222	0.6667	2
LA	3	0.6667	0.6667	1
LI	3	0.6667	0.6667	1
LL	4	0.5625	0.5625	1
LF	4	0.1875	0.5625	2
LS	4	0.0938	0.5625	4
PQ	3	0.6667	0.6667	1
PS	3	0.1111	0.6667	3
QL	3	0.6667	0.6667	1
RD	3	0.6667	0.6667	1
SR	3	0.2222	0.6667	2
VA	3	0.6667	0.6667	1
YL	3	0.6667	0.6667	1

tations lead seven probabilities to increase, seven probabilities to decrease and two probabilities to be unchanged. It can also be seen that the deceased probabilities in mutations are most related to "V'', which leads to a dramatic decrease in the random ranks against the theoretical probabilities.

Table 4 shows amino acid pairs that appeared more than twice, their numbers, measured distribution probability, maximum theoretical distribution probability and random ranks against the theoretical distribution probability in human TNF-α, because the distribution probabilities of amino acid pairs which appear once or twice are predictable as described in the introduction and we did not show them. It can be seen that 61% (11/18) of distributions of amino acid pairs follow the probabilistically simplest pathway. Moreover, the remaining distributions of amino acid pairs are very close to the probabilistically simplest pathway. Although we would like to know the probabilities of distributions of amino acid triplets, no amino acid triplet appears more than twice.

Table 5 shows amino acid pairs, their appearances, and their random ranks related to the changes between human TNF-α and its mutations. A mutation leads to changes in four amino acid pairs, for example, mutation 1 leads to the change from "L'' to "S'' at position 105 in human TNF-α, and "W'' and "N'' are at positions 104 and 106, thus mutation 1 leads to the changes from "WL" to "WS" and "LN" to "SN". Thus we have a total of 32 changes with respect to these mutations. It can be seen that the mutations cause the random rank of one amino acid pair to decrease, the random ranks of two amino acid pairs to increase, and the random ranks of 11 amino acid pairs to be unchanged.

Discussion

This is the second approach we are developing to use the random principle to analyse protein primary structure. What we had done using our first approach is (i) to use the random principle to predict the presence and absence of amino acid pairs and triplets in different proteins from their amino acid compositions, (ii) to compare the predicted frequencies with the counted frequencies and (iii) to suggest that the sequences being predictable should not be deliberately evolved and conserved and the sequences being unpredictable should be deliberately evolved and conserved. [17, 18, 19, 20, 21, 22] Using this approach, we may suggest that the distributions with a high random rank should not be deliberately evolved and conserved and the distributions with a low random rank should be deliberately evolved and conserved.

In our approaches we concentrate on the random analysis of short sequences; the probabilistic consideration is

Table 5 Amino acid pairs, their appearances and their random ranks (in parentheses) related to the changes between human TNF-α and its mutations

Amino acid pair	Normal[a]	I	II	III	IV	V	VI	VII	VIII	Tendency[b]
LN	1 (1)	0								
SN	0	1 (1)								
WL	1 (1)	0								
WS	0	1 (1)								
RA	1 (1)		0							
RR	2 (1)		1 (1)							–
RW	0		1 (1)							
WA	0		1 (1)							
AL	3 (2)			2 (1)						↑
AF	0			1 (1)						
LL	4 (1)			3 (2)						↓
FL	2 (1)			3 (1)						–
AV	2 (1)				1 (1)					–
IA	2 (1)				1 (1)					–
IV	1 (1)				2 (1)					–
VV	2 (1)				3 (1)					–
FY	0					1 (1)				
SY	1 (1)					0				
VF	1 (1)					2 (1)				–
VS	1 (1)					0				
AN	3 (3)						4 (2)			↑
KA	0						1 (1)			
KV	1 (1)						0			
VN	1 (1)						0			
DN	1 (1)							2 (1)		–
KD	0							1 (1)		
KV	1 (1)							0		
VN	1 (1)							0		
AE	5 (1)								4 (1)	–
AK	1 (1)								2 (1)	–
ES	2 (1)								1 (1)	–
KS	1 (1)								2 (1)	–

[a] If the random ranks are different from the human TNF-α, they are presented in parentheses in related mutations

[b] ↑, ↓ and – are the random rank to increase, to decrease and unchanged, respectively. I, II, III, IV, V, VI, VII and VIII are mutations 1, 2, 3, 4, 5, 6, 7 and 8, respectively

that the long sequences are predictable and the biological consideration is that a good signature pattern of a protein must be as short as possible and many short sequences (not more than four or five residues long conserved sequence) are often diagnostic of certain binding properties or active sites. [23] It is therefore considered important to analyse the short sequences by our approaches.

We could not use the approach used in this study to correlate the protein mutations with their biological implications although our other approach can do so, [17, 18, 19, 20, 21, 22] because the biological implications of mutations are the consequence of mutations while our approach is concerned with the probabilistic cause of mutations.

Our random analysis is different from random genetic drift, which is related to the evolution in a population and determined by three main parameters: the size of the population, the underlying substructure of the population and the mutation rate and is calculated using the Hardy–Weinberg law, whereas our approach provides a quantitative estimation on single protein.

In this study, we notice that the distribution probabilities for most amino acids and amino acid pairs are at or very near to the probabilistically simplest distribution, which is not a homogenous distribution along a protein, but rather a heterogeneous distribution along a protein. Such a heterogeneous distribution with a high probability may provide the base for the protein function.

The results show that mutations can lead to the increased and decreased distribution probabilities of amino acids and amino acid pairs. From a probabilistic viewpoint, an increased distribution probability means that the mutation is more likely to occur spontaneously, whereas a decreased distribution probability means that the mutation is less likely to occur spontaneously and perhaps is more related to a certain cause. In fact, a mutation often leads to both distribution probabilities of replaced amino acid and replacing amino acid to increasing, for example, mutation 2 in Table 3. In such a case, the mutation causes both distribution probabilities to increase along the probabilistically simplest pathway, although their random ranks against the theoretical distribution probabilities are unchanged.

We can know the magnitude in distribution probability between human TNF-α and its mutations as stated in the Materials and methods section, for example, the distribution of "L"s in mutation 1 has about 1.1890

(0.0453/0.0381) times higher chance to occur than that in human TNF-α (Table 3).

Although it is generally considered that proteins do not have random distributions of amino acids, our approach does show such random distributions. Perhaps natural selection goes along the probabilistically simplest line, but much work is needed to test various hypotheses.

References

1. Barker WC, Garavelli JS, Hou Z, Huang H, Ledley RS, McGarvey PB, Mewes HW, Orcutt BC, Pfeiffer F, Tsugita A, Vinayaka CR, Xiao C, Yeh LS, Wu C (2001) Nucleic Acids Res 29:29–32
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) J Mol Biol 215:403–410
3. Benson DC (1990) Nucleic Acids Res 18:6305–6310
4. Jonassen I (1997) Comput Appl Biosci 13:509–522
5. Searls DB (1997) Comput Appl Biosci 13:333–344
6. Popov O, Segal DM, Trifonov EN (1996) Biosystems 38:65–74
7. Karlin S, Bucher P, Brendel V, Altschul SF (1991) Annu Rev Biophys Biophys Chem 20:175–203
8. Everitt BS (1999) Chance rules: an informal guide to probability, risk, and statistics. Springer, New York
9. Feller W (1968) An introduction to probability theory and its applications, 3rd edn, vol I. Wiley, New York, pp 38–40
10. Majerus M, Amos W, Hurst G (1996) Evolution: the four billion year war. Longman, Essex
11. Bairoch A, Apweiler R (1999) Nucleic Acids Res 27:49–54
12. Pennica D, Nedwin GE, Hayflick JS, Seeburg PH, Derynck R, Palladino MA, Kohr WJ, Aggarwal BB, Goeddel DV (1984) Nature 312:724–729
13. Ostade XV, Tavernier J, Prange T, Fiers W (1991) EMBO J 10:827–836
14. Stevenson FT, Bursten SL, Locksley RM, Lovett DH (1992) J Exp Med 176:1053–1062
15. Cha SS, Kim JS, Cho HS, Shin NK, Jeong W, Shin HC, Kim YJ, Hahn JH, Oh BH (1998) J Biol Chem 273:2153–2160
16. Reed C, Fu ZQ, Wu J, Xue YN, Harrison RW, Chen MJ, Weber IT (1997) Protein Engineer 10:1101–1107
17. Wu G (2000) Biochem Biophys Res Commun 268:823–826
18. Wu G (2000) Cancer Lett 153:145–150
19. Wu G (2000) J Biochem Mol Biol Biophys 4:179–185
20. Wu G (2000) Alcohol 35:302–306
21. Wu G, Yan SM (2000) J Mol Microbiol Biotechnol 2:277–281
22. Wu G, Yan SM (2001) J Mol Mod 5:120–124
23. PROSITE (2001) A dictionary of protein sites and patterns user manual. http://www.expasy.ch/prosite/