Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 319 (2004) 27–31
BBRC
www.elsevier.com/locate/ybbrc

Determination of amino acid pairs sensitive to variants in human copper-transporting ATPase 2

Guang Wu* and Shaomin Yan

DreamSciTech Consulting Co. Ltd., 301, Building 12, Nanyou A-zone, Jiannan Road, Shenzhen, Guangdong Province CN-518054, China

Received 4 April 2004
Available online 10 May 2004

Abstract

In this study, we use our probabilistic approach to analyze the amino acid pairs in human copper-transporting ATPase 2 (ATP7B) in order to determine which amino acid pairs are more sensitive to 125 variants from missense mutant human ATP7B. The results show 97.6% of 125 variants occur at randomly unpredictable amino acid pairs, which account for 80.9% of amino acid pairs in ATP7B, and the chance of occurring of variant is about 9 times higher in randomly unpredictable amino acid pairs than in predictable pairs. Thus, the randomly unpredictable amino acid pairs are more sensitive to variants in human ATP7B.

Keywords: Amino acid; ATPase; Genetic Stability; Modeling; Protein; Wilson disease

Copper is an essential element for the activity of a number of physiologically important enzymes [1]. The Wilson disease is an autosomal recessive disorder of copper transport, with a prevalence of approximately one in 35,000 [2,3]. The disease results from the absence or dysfunction of a copper-transporting P-type ATPase (ATP7B) encoded on chromosome 13 [4]. This ATP7B is expressed in hepatocytes and transports copper into the secretory pathway for incorporation into ceruloplasmin and excretion into the bile [5]. Affected individuals exhibit excessive copper accumulation in the liver and brain, deficient holoceruloplasmin biosynthesis, and a marked impairment in biliary copper excretion [6,7]. Clinically, the disorder leads to progressive liver dysfunction, cirrhosis, and/or extrapyramidal features such as parkinsonism and dystonia [8].

To date, about 125 variants from missense mutation have been found in the ATP7B [9]. Generally a single amino acid is related to a variant, however this amino acid (except for the one at terminal) has a connection with two neighboring amino acids, whose connection constructs two amino acid pairs. We, therefore, regard the amino acid pairs as the basic unit for analysis following the concept that a good signature pattern of a protein must be as short as possible, but the conserved sequence is not longer than four or five residues [10]. In the case of 125 variants in ATP7B, little is known about which amino acid pairs are more sensitive to variants, thus it is difficult to draw a general rule to judge the sensitivity of amino acid pair to variants. If such a general rule could be drawn, then we could gain not only more insight into the relationship between ATP7B variants and Wilson disease, but more importantly we can pay much attention to these amino acid pairs in order to prevent them from variants. Moreover, we even can principally predict the possible pairs sensitive to the currently unknown variants.

This problem can be assessed using different approaches such as empirical (regression analysis), experimental (artificial and natural mutations), computation (multiple sequence comparisons and alignments), etc. Currently two explanations are commonly proposed to explain why some amino acids are mutated more frequently than others. The first is targeted mutagenesis, which defines the "hotspot" sites sensitive to endogenous and exogenous mutagens [11–13]. The second is the function selection, which indicates that the disruption of

* Corresponding author. Fax: +86-755-2520-8256.
E-mail address: hongguanglishibahao@yahoo.com (G. Wu).

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.04.143

protein functions may depend upon the position of the variant in a protein [14–16].

Over the last several years we have developed three models using random approaches to analyze primary structure of proteins (for review, see [17]). Generally, in terms of actual and predicted frequencies, our first model classifies the amino acid pairs in a protein into two categories: the randomly predictable and the unpredictable. We suggest that the randomly predictable amino acid pairs should not be deliberately evolved, whereas the randomly unpredictable amino acid pairs should be deliberately evolved. Accordingly the randomly unpredictable amino acid pairs are more related to protein function, and the variants in these amino acid pairs may lead to the dysfunction of protein.

More recently we found that a mutation, which leads to the dysfunction of rat monoamine oxidase B, is located in a randomly unpredictable amino acid pair. On the contrast, another mutation, which does not affect rat monoamine oxidase B function, is located in randomly predictable amino acid pairs [18]. This implies that a harmful variant is likely to be located at the randomly unpredictable amino acid pair, and a harmless variant is likely to be located at the randomly predictable amino acid pair. As a harmful variant is easy to draw our attention, we would expect that a harmful variant could be easily documented in the literature, hence we would expect to see more variants to be located at the randomly unpredictable amino acid pairs. On the other hand, we would also expect that the randomly unpredictable amino acid pair is not as stable as the randomly predictable amino acid pair because randomness governs the highest probability of occurrence. As a result, a variant is readily to occur at the randomly unpredictable amino acid pair, and the randomly unpredictable amino acid pair is more sensitive to variant.

In this study, we analyze 125 variants from human ATP7B in order to determine whether or not the randomly unpredictable amino acid pairs are more sensitive to variants.

Materials and methods

The amino acid sequences of the human ATP7B and its 125 variants are obtained from the SWISS-PROT data bank [16] (Accession No. P35670). The detailed calculations and its rationales have already been published in a number of our previous studies (for the details, see [17]). The calculation procedure with its examples is as follows.

Amino acid pairs in human ATP7B. The human ATP7B is composed of 1465 amino acids. We count the first and second amino acids as an amino acid pair, the second and third as another amino acid pair, the third and fourth, until the 1464th and 1465th, thus there are 1464 amino acid pairs. As there are 20 types of amino acids, so there are 400 theoretically possible kinds of amino acid pairs. Again there are 1464 amino acid pairs in human ATP7B, which are more than 400 kinds of theoretically possible amino acid pairs, clearly some of 400 kinds of theoretically possible amino acid pairs should appear more than once. Meanwhile, we may expect that some of 400 kinds of theoretically possible amino acid pairs are absent from human ATP7B.

Randomly predicted frequency and actual frequency. The randomly predicted frequency is calculated according to the simple permutation principle [19]. For example, there are 134 alanines (A) and 62 aspartic acids (D) in human ATP7B, and the predicted frequency of amino acid pair "AD" is 6 ($134/1465 \times 62/1464 \times 1464 = 5.671$). Actually we find 6 "AD"s in human ATP7B, so the actual frequency of "AD" is 6. Hence, we have three relationships between actual and predicted frequencies, i.e., the actual frequency is smaller than, equal to, and larger than the predicted frequency, respectively.

Randomly predictable present amino acid pairs. As described above, the frequency of randomly presence of amino acid pair "AD" is 6 and "AD" really appears 6 times in human ATP7B, so the presence of "AD" is randomly predictable.

Randomly unpredictable present amino acid pairs. There are 69 glutamines (Q) in human ATP7B, and the frequency of random presence of amino acid pair "AQ" is 6 ($134/1465 \times 69/1464 \times 1464 = 6.311$). But actually the "AQ" appears 9 times, so the presence of "AQ" is randomly unpredictable. This is the case that the actual frequency is larger than the predicted frequency. Another case is that the actual frequency is smaller than the predicted frequency. For example, there are 53 arginines (R) and 104 glycines (G) in human ATP7B and the predicted frequency of "RG" is 4 ($53/1465 \times 104/1464 \times 1464 = 3.762$), whereas the actual frequency of "RG" is only one.

Randomly predictable absent amino acid pairs. There are 29 cysteines (C) and 11 tryptophans (W) in human ATP7B, and the frequency of random presence of "CW" is 0 ($29/1465 \times 11/1464 \times 1464 = 0.218$), i.e., the "CW" would not appear in human ATP7B, which is true in the real situation. Thus, the absence of "CW" is randomly predictable.

Randomly unpredictable absent amino acid pairs. There are 80 glutamic acids (E) in human ATP7B, and the frequency of random presence of "CE" is 2 ($29/1465 \times 80/1464 \times 1464 = 1.584$), i.e., there would be two "CE"s in human ATP7B. However no "CE" appears in human ATP7B, therefore the absence of "CE" from human ATP7B is randomly unpredictable.

Difference between actual and randomly predicted frequencies. We calculate the difference between actual frequency (AF) and predicted frequency (PF) of amino acid pairs in variant ATP7B, i.e., $\sum(\mathrm{AF} - \mathrm{PF})$. For instance, the variant at position 642 substitutes "D" to "H" which leads to the amino acid pairs "LD" and "DH" mutate to "LH" and "HH," because the amino acids are "L" and "H" at positions 641 and 643. The actual frequency and predicted frequency are 4 and 6 for "LD," 0 and 3 for "LH," 5 and 2 for "DH," and 1 and 1 for "HH," respectively. With respect to these amino acid pairs, the difference between actual and predicted frequencies before mutation is 1, i.e., $(4 - 6) + (5 - 2)$, while the difference between actual and predicted after mutation is −3, i.e., $(0 - 3) + (1 - 1)$. In this manner, we can compare the difference in the amino acid pairs before and after mutations.

Statistics. The difference between actual and predicted frequencies before and after mutations is compared using the Mann–Whitney U test with SigmaStat for Windows (SPSS, 1992–2003), and the $p < 0.05$ is considered statistically significant.

Results and discussion

General information on amino acid pairs in human ATP7B

Table 1 details the general information on amino acid pairs in human ATP7B, for example, 56 of 400 kinds of theoretically possible amino acid pairs are absent from

Table 1
Appearance of theoretical kinds of amino acid pairs in human ATP7B protein

Appearances	Number of theoretically possible kinds of amino acid pairs
0	56
1	68
2	71
3	50
4	36
5	32
6	19
7	16
8	12
9	9
10	7
11	5
12	7
13	3
14	1
15	4
17	1
18	1
20	2

human ATP7B including 13 randomly predictable and 43 randomly unpredictable. Consequently, 1464 amino acid pairs in human ATP7B include only 344 kinds of theoretically possible amino acid pairs (400 − 56 = 344), i.e., some amino acid pairs should appear more than once (Table 1).

Of 344 kinds of theoretically possible amino acid pairs in human ATP7B, 91 kinds are randomly predictable and 253 kinds are randomly unpredictable. As mentioned above, some kinds of amino acid pairs appear more than once, thus of 1464 amino acid pairs in human ATP7B, 280 pairs are randomly predictable and 1184 pairs are randomly unpredictable. We therefore can find how many variants occur with respect to these present amino acid pairs in human ATP7B (Table 2).

Variants of human ATP7B in randomly predictable and unpredictable present amino acid pairs

As mentioned under Materials and methods, a missense mutant protein leads to two amino acid pairs to be replaced by another two, and their actual frequency can be smaller than, equal to, and larger than the randomly predictable frequency. Tables 3 and 4 detail the situations related to the amino acid pairs before and after mutations, and the relationship between their actual and predicted frequencies.

Table 3 can be read as follows. The first column classifies the amino acid pairs into randomly predictable and unpredictable. The second and third columns show the variant occurs in which type of amino acid pairs, for example, the first two cells in columns 2 and 3 indicate that the actual frequencies are equal to the predicted frequencies in amino acid pairs I and II. The fourth and five columns indicate the number of variants occur in amino acid pairs I and II, for instance, 3 of 125 variants (2.4%) occur at amino acid pairs whose actual frequencies are equal to predicted frequencies in both pairs. The sixth column indicates the percentage of variants occurring at predictable and unpredictable amino acid pairs.

Tables 2 and 3 indicate that 97.6% of variants occur at randomly unpredictable present amino acid pairs and 2.4% of variants occur in randomly predictable amino acid pairs. These results imply that 253 kinds of randomly unpredictable present amino acid pairs account for 97.6% variants in human ATP7B, whereas 91 kinds of randomly predictable present amino acid pairs

Table 2
Occurrences of variants with respect to randomly predictable and unpredictable amino acid pairs in human ATP7B protein

ATP7B	Kinds		Pairs		Variants		Ratio	
	Number	Percentage	Number	Percentage	Number	Percentage	Variants/kinds	Variants/pairs
Predictable	91	26.45	280	19.13	3	2.40	3/91 = 0.03	3/280 = 0.01
Unpredictable	253	73.55	1184	80.87	122	97.60	122/253 = 0.48	122/1184 = 0.10
Total	344	100	1464	100	125	100	125/344 = 0.36	125/1464 = 0.09

Table 3
Classification of original amino acid pairs induced by variants in human ATP7B

	Amino acid pair		Variants		Total (%)
	I	II	Number	Percentage	
Predictable	AF = PF	AF = PF	3	2.40	2.40
Unpredictable	AF > PF	AF > PF	37	29.60	97.60
	AF > PF	AF = PF	25	20.00	
	AF > PF	AF < PF	34	27.20	
	AF < PF	AF = PF	17	13.60	
	AF < PF	AF < PF	9	7.20	

AF, actual frequency; PF, predicted frequency.

Table 4
Classification of mutant amino acid pairs induced by variants in human ATP7B

Amino acid pair		Variants		Total (%)
I	II	Number	Percentage	
AF = 0, PF > 0	AF = 0, PF > 0	0[a]	0	12.00
AF = 0, PF > 0	AF = PF = 0	1[a]	0.80	
AF = 0, PF > 0	AF = PF > 0	1[a]	0.80	
AF = 0, PF > 0	AF < PF, AF ≠ 0	4[a]	3.20	
AF = 0, PF > 0	AF > PF	5[a]	4.00	
AF = PF = 0	AF = PF = 0	1	0.80	
AF = PF = 0	AF = PF > 0	3	2.40	
AF = PF = 0	AF < PF, AF ≠ 0	0	0	
AF = PF = 0	AF > PF	0	0	
AF < PF, AF ≠ 0	AF < PF, AF ≠ 0	23[a]	18.40	88.00
AF < PF, AF ≠ 0	AF = PF > 0	21[a]	16.80	
AF < PF, AF ≠ 0	AF > PF	37[a]	19.60	
AF = PF > 0	AF = PF > 0	6	6.80	
AF > PF	AF > PF	11	8.80	
AF = PF > 0	AF > PF	12	9.60	

[a] Indicates the variants which target one or both mutant amino acid pairs with their actual frequency smaller than predicted one (totally 73.60%).

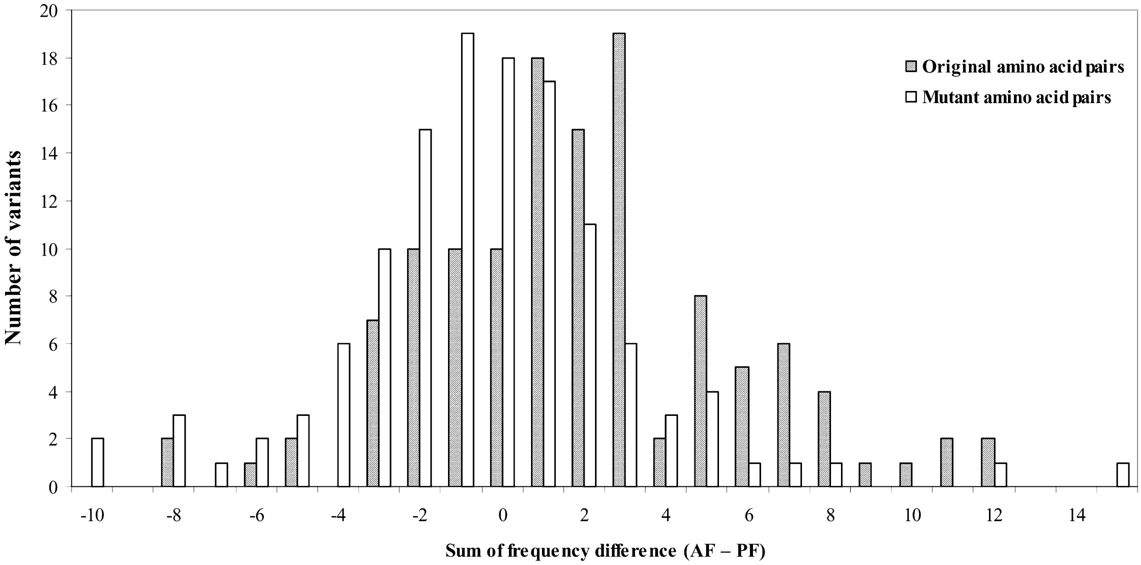

Fig. 1. Frequency difference between original (filled) and mutant (unfilled) amino acid pairs induced by variants from human ATP7B.

account for only 2.4% of variants. Still we can see the ratio in Table 2 that the chance of occurring of variants in unpredictable amino acid pairs is larger than in predictable amino acid pairs. For example, the chance of occurring of variant is 15-fold higher in unpredictable kind than in predictable kind (0.48 vs. 0.03). These results strongly support our rationale that the harmful variants are more likely to occur at randomly unpredictable present amino acid pairs, which therefore are more sensitive to the variants.

When looking at the unpredictable pairs in Table 3, we find that the majority of these pairs are characterized by one or both original pairs whose actual frequency is larger than predicted frequency (the first three rows in unpredictable pairs). Comparing each variant, we find that the impact of variants is to diminish the difference between actual and predicted frequencies by means of reducing the actual frequency, which indicates that the variants lead to the construction of amino acid pairs more randomly. In other words, the variants result in the construction of amino acid pairs to be more naturally easy to occur.

Table 4 can be read as follows. The first and second columns indicate the actual and predicted situations in amino acid pairs I and II. The third and fourth columns indicate the number of variants occurs at amino acid pairs I and II and their percents. The fifth column is the total percentage of our classifications. Table 4 shows that 12% of variants bring about one or both mutant amino acid pairs which are absent in normal human ATP7B (AF = 0). Also 73.60% of variants target one or both mutant amino acid pairs with their actual frequency smaller than predicted frequency (a). These phenomena indicate that the amino acid pairs in mutant proteins are more randomly constructed.

Frequency difference of amino acid pairs affected by variants

The difference between actual and predicted frequencies represents a measure of randomness of construction of amino acid pairs, i.e., the smaller the difference, the more random the construction of amino acid pairs. In particular, (i) the larger is the positive difference, the more randomly unpredictable is the presence of amino acid pairs; and (ii) the larger is the negative difference, the more randomly unpredictable is the absence of amino acid pairs.

Considering all 125 variants, the difference between actual and predicted frequencies is 1.91 ± 0.34 (means ± SE, ranging from −8 to 12) for original amino acid pairs. This means that the variants occur in the amino acid pairs that appear more frequently than their predicted frequency. Meanwhile, the difference between actual and predicted frequencies is -0.35 ± 0.32 (means ± SE, ranging from −10 to 15) for mutant amino acid pairs, which implies that the mutant amino acid pairs are randomly constructed in the mutant ATP7B, as their actual and predicted frequencies are about the same. Striking statistical difference is found between the original and mutant amino acid pairs ($p < 0.0001$). Fig. 1 shows the distribution of difference between actual and predicted frequencies. As the predicted frequency is the highest chance for construction of amino acid pairs, it is important to find out whether the variants lead to the actual frequency to approach the predicted frequency. If so, we can understand that the protein has a natural trend to variants; if not, the protein does not have a natural trend to variants. The present study reveals that the human ATP7B has a natural trend to variants.

In conclusion, the results in this study suggest that the randomly unpredictable amino acid pairs are more sensitive to the variants. Diminishing of difference between actual and predicted frequencies has been shown in this study, thus the variants in fact are a degeneration process inducing Wilson disease related to ATP7B variants.

References

[1] M.C. Linder, Biochemistry of copper, in: E. Frieden (Ed.), Biochemistry of the Elements, Plenum, New York, 1991.

[2] M. Orth, A.H.V. Schapira, Mitochondria and degenerative disorders, Am. J. Med. Genet. 106 (2001) 27–36.

[3] J.F.B. Mercer, The molecular basis of copper-transport diseases, Trends Mol. Med. 7 (2001) 64–69.

[4] Y. Yamaguchi, M.E. Heiny, J.D. Gitlin, Isolation and characterisation of a human liver cDNA as a candidate gene for Wilson disease, Biochem. Biophys. Res. Commun. 197 (1993) 271–277.

[5] Y. Murata, E. Yamakawa, T. Iizuka, H. Kodama, T. Abe, Y. Seki, M. Kodama, Failure of copper incorporation into ceruloplasmin in the Golgi apparatus of LEC rat hepatocytes, Biochem. Biophys. Res. Commun. 209 (1995) 349–355.

[6] D.M. Danks, Copper deficiency in humans, Annu. Rev. Nutr. 8 (1988) 235–257.

[7] D. Strausak, J.F.B. Mercer, H.H. Dieter, W. Stremmel, G. Multhaup, Copper in disorders with neurological symptoms: Alzheimer's, Menkes, and Wilson diseases, Brain Res. Bull. 55 (2001) 175–185.

[8] G. Loudianos, J.D. Gitlin, Wilson's disease, Semin. Liver Dis. 20 (2000) 353–364.

[9] W.M. Rideout, G.A. Coetzee, A.F. Olumi, P.A. Jones, 5-Methylcytosine as an endogenous mutagen in human LL receptor and p53 genes, Science 249 (1990) 1288–1290.

[10] PROSITE: a dictionary of protein sites and patterns user manual. Available from <http://www.expasy.ch/prosite/>.

[11] R. Montesano, P. Hainaut, C.P. Wild, Hepatocellular carcinoma: from gene to public health, J. Natl. Cancer Inst. 89 (1997) 1844–1851.

[12] P. Hainaut, G.P. Pfeifer, Patterns of p53 G → T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke, Carcinogenesis 22 (2001) 367–374.

[13] K. Ory, Y. Legros, C. Auguin, T. Soussi, Analysis of the most representative tumour-derived p53 mutants reveals that changes in protein conformation are not correlated with loss of transactivation or inhibition of cell proliferation, EMBO J. 13 (1994) 3496–3504.

[14] K. Forrester, S.E. Lupold, V.L. Ott, C.H. Chay, V. Band, X.W. Wang, C.C. Harris, Effects of p53 mutants on wild-type p53-mediated transactivation are cell type dependent, Oncogene 10 (1995) 2103–2111.

[15] T. Aas, A.L. Borresen, S. Geisler, B. Smith-Sorensen, H. Johnsen, J.E. Varhaug, L.A. Akslen, P.E. Lonning, Specific p53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients, Nat. Med. 2 (1996) 811–814.

[16] A. Bairoch, R. Apweiler, The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 2000, Nucleic Acids Res. 28 (2000) 45–48.

[17] G. Wu, S.M. Yan, Randomness in the primary structure of protein: methods and implications, Mol. Biol. Today 3 (2002) 55–69.

[18] G. Wu, S.M. Yan, Prediction of presence and absence of two- and three-amino-acid sequence of human monoamine oxidase B from its amino acid composition according to the random mechanism, Biomol. Eng. 18 (2001) 23–27.

[19] W. Feller, An Introduction to Probability Theory and its Applications, vol. I, third ed., Wiley, New York, 1968.