Mol Divers (2008) 12:119–129
DOI 10.1007/s11030-008-9084-x

FULL LENGTH PAPER

Quantitative relationship between mutated amino-acid sequence of human copper-transporting ATPases and their related diseases

Shaomin Yan · Guang Wu

Received: 29 January 2008 / Accepted: 19 July 2008 / Published online: 8 August 2008
© Springer Science+Business Media B.V. 2008

Abstract Copper-transporting ATPase 1 and 2 (ATP7A and ATP7B) are two highly homologous P-type copper ATPase exporters. Mutations in ATP7A can lead to Menkes disease which is an X-linked disorder of copper deficiency. Mutations in ATP7B can cause Wilson disease which is an autosomal recessive disorder of copper toxicity. In this study, we attempt to build a quantitative relationship between mutated ATPase and Menkes/Wilson disease. First, we use the amino-acid distribution probability as a measure to quantify the difference in ATPase before and after mutation. Second, we use the cross-impact analysis to define the quantitative relationship between mutant ATPase protein and Menkes/Wilson disease, and compute various probabilities. Finally, we use the Bayesian equation to determine the probability that Menkes/Wilson disease is diagnosed under a mutation. The results show (i) the vast majority of mutations lead to the amino-acid distribution probability increase in mutant ATP7As and decrease in ATP7Bs, and (ii) the probability that a mutation causes Menkes/Wilson disease is about nine tenth. Thus we provide a way to use the descriptively probabilistic method to couple the mutation with its clinical outcome after quantifying mutations in proteins.

Keywords Amino acid · ATP7A · ATP7B · Bayes' law · Computational · Cross-impact analysis · Distribution probability · Menkes disease · Prediction · Wilson disease

S. Yan
Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi 530007, China

G. Wu (✉)
Computational Mutation Project, DreamSciTech Consulting, 301, Building 12, Nanyou A-zone, Jiannan Road, Shenzhen, Guangdong Province 518054, China
e-mail: hongguanglishibahao@yahoo.com

Introduction

With worldwide efforts on correlating each disease to genetic level, it is very meaningful to build a quantitative relationship between a mutated protein and its phenotype outcome. For this purpose, we must quantify a protein sequence in order to use a numeric sequence to represent the protein sequence no matter what mechanism we use for quantification. Currently, there are many ways to quantify a protein sequence such as physicochemical property, electrostatic property, steric property, hydrophobic property, hydrogen bond property, etc. [1–10]. These quantifications were developed by physicists and chemists for their own purpose, but not for the study of mutations. When applying these quantifications for mutation studies, it is possible that these quantifications will not change before and after a mutation. For example, pI (isoelectric point) would not change before and after mutation. Furthermore, these borrowed quantification descriptors have their own special dimensions (units), which are not associated with mutations. Although a relationship between structure and mutation is generally modeled using a regression, this system is associated to a phenomenological model rather than to a kinetic/dynamic model. Thus, these borrowed quantifications could be more suitable for studying static state.

In this study, we apply our own quantifications which are subject to mutation and have no problems with dimensions. Since 1999, we have developed three approaches to quantify each amino acid in a protein and a whole protein (for review, see [11–13]). Our three quantifications are not only sensitive to mutations and dimensionless, but also sensitive to the length of protein, composition of amino acids, amino acid positions and neighboring amino acids.

We have developed a kinetic model to predict mutations in proteins from influenza A virus [14–23]. With our approach

we can look at it from four aspects: (i) prediction of mutation position, (ii) prediction of would-be-mutated amino acids at the predicted positions, (iii) timing of mutation [24,25], and (iv) prediction of new function led by mutation. To accomplish the last prediction, the first step is to develop a quantitative model to couple the changed protein sequence with its outcome, which at this stage would be descriptive. In this study, we attempt to use the amino-acid distribution probability as a measure to build a quantitative relationship between mutated amino-acid sequence of human copper-transporting ATPases (ATP7A and ATP7B) and their related diseases (Menkes and Wilson diseases).

As a critical catalyst, copper is an essential transition metal required for the activity of multiple mammalian enzymes, including cytochrome c oxidase, superoxide dismutase and lysyl oxidase [26,27]. However, copper ions are efficient generators of free radicals, which accounts for the toxicity of copper when homeostatic mechanisms are disrupted under certain conditions [26]. The Menkes and Wilson's disease proteins are two highly homologous P-type copper ATPase exporters [28,29]. They supply copper to the secreted cuproenzyme and ceruloplasmin, which has a role in iron metabolism [30–34].

The Menkes disease gene ATP7A was the first mammalian heavy metal ion transporter cDNA to be cloned, which locates at Xql3.3 and encodes the copper-transporting ATPase 1 consisting of 1,500 amino acids (ATP7A) [35–37]. This gene is expressed in most tissues, except in the liver. Mutations in the gene cause Menkes disease [38–41], which is an X-linked disorder of copper transport leading to Cu^{2+} accumulation predominantly in the intestine and kidney causing damage to certain tissues, neurodegeneration, and death in early childhood [42].

The Wilson disease gene ATP7B was localized to the q14.3 band of chromosome 13 and cloned in 1993 [43–46]. This gene encodes the copper transporting ATPase 2 with 1,465 amino acids (ATP7B) expressed mainly in the liver [47], where it is responsible for biliary excretion of copper [48]. Thus, ATP7B is the copper pump regulating the rate of biliary copper excretion when copper levels start to rise in liver [49–51]. Mutations in the ATP7B gene result in Wilson disease [52–55], which is an autosomal recessive copper toxicosis described firstly by Kinnear Wilson in 1912 [56]. This disease is characterized by a massive accumulation of copper in the liver, with subsequent deposition of copper in other tissues, such as the central nervous system [57–60]. Patients with Wilson disease manifest a spectrum of liver pathologies ranging from hepatitis and cirrosis to liver failure [61]. About half of patients with Wilson disease develop central nervous system toxicity as the initial clinical manifestation of the disease [62].

Approximately 200 mutations have been identified in unrelated Menkes patients [63], and over 150 in Wilson patients [64–70]. All types of mutations have been found the copper-transporting ATPases. For their outcome, splice site, nonsense, duplication and deletion mutations would be predicted to truncate the mRNA or the protein and severely affect function [71].

Materials and methods

Data

Both human copper-transporting ATPases with their documented mutations were obtained from UniProtKB/S wiss-Prot entry: ATP7A with its 37 mutations (accession number Q04656; December 4, 2007; Entry version 99, http://expasy.org/uniprot/Q04656) and ATP7B with its 165 mutations (accession number P35670; Novmeber 13, 2007; Entry version 100, http://expasy.org/uniprot/P35670).

Amino-acid distribution probability

Among three approaches developed by us, the amino-acid distribution probability is mainly related to the positions of amino acids along the protein, which is suitable for mutation analysis, and we have used this approach in a number of our previous studies [11–23,25,72–81]. The quantification was developed along such a thought, for example, there are two methionines (M) among 141 amino acids in human hemoglobin α-chain [72]. With regard to their random distribution, our intuition might suggest that there would be one M in the first half of the chain and another M in the second half, which is true in real-life case. In fact, there are only three possible distributions of Ms in human hemoglobin α-chain, i.e. (i) both Ms are in the first half, (ii) one M is in each half and (iii) both Ms are in the second half. Thus each distribution of Ms has the probability of 1/3. If we do not distinguish either the first half or second half but are simply interested in whether both Ms are in both halves or in any half, we have the probability of 1/2 for each distribution.

If we are interested in the distribution probability of three amino acids in a protein, we naturally imagine to grouping the protein into three partitions, and our intuition may suggest that each partition contains an amino acid. If we do not distinguish the first, second and third partition, actually there are totally three types of distributions, i.e. (i) each amino acid is in each partition, (ii) two amino acids are in a partition and an amino acid in another partition, and (iii) three amino acids are in a partition.

At this moment, the distribution probability can be computed according to the statistical mechanics, which classifies the distribution of elementary particles in energy states according to three assumptions of whether distinguishing each particle and energy state, i.e. Maxwell-Boltzmann,

Fermi-Dirac and Bose-Einstein assumptions [82]. We actually used the Maxwell-Boltzmann assumption for computing amino-acid distribution probability, which is [82]

$$\frac{r!}{q_0! \times q_1! \times \ldots \times q_n!} \times \frac{r!}{r_1! \times r_2! \times \ldots \times r_n!} \times n^{-r},$$

where r is the number of amino acids, n is the number of partitions, r_n is the number of amino acids in the nth partition, q_n is the number of partitions with the same number of amino acids, and ! is the factorial function.

Thus, the distribution probabilities are different for these three types of distributions of three amino acids, say, 0.2222 for (i), 0.6667 for (ii) and 0.1111 for (iii). Clearly the protein can only adopt one type of distribution for these three amino acids, which is the actual distribution probability.

For four amino acids, we have five distribution probabilities, i.e. (i) each partition contains an amino acid, (ii) a partition contains two amino acids and two partitions contain an amino acid each, (iii) two partitions contain two amino acids each, (iv) a partition contains an amino acid and a partition contains three amino acids, and (v) a partition contains four amino acids. Their distribution probabilities are 0.0938 for (i), 0.5625 for (ii), 0.1406 for (iii), 0.1875 for (iv), and 0.0156 for (v). Furthermore, there are seven distributions for five amino acids, 11 distributions for 6 amino acids, 15 distributions for 7 amino acids, and so on.

Quantification of normal copper-transporting ATPases

For the human ATP7A and ATP7B before mutation, we used the equation in the above subsection to calculate the amino-acid distribution probability for amino acids listed in Table 1. We showed the simplest example for computation of amino-acid distribution probability in Appendix.

In this way, we quantified each amino acid in normal human ATP7A and 7B although the calculation became more complicated with respect to other types of amino acids in Table 1 because the q, r, n and factorial function increased. Thereafter, we assigned the calculated distribution probability to each amino acid along the sequence in order to have the visualized concept (Fig. 1).

Quantification of mutated copper-transporting ATPases

The amino-acid distribution probability calculated in the above subsection serves as a reference because the mutation does not occur. Obviously any mutation changes the amino-acid composition of copper-transporting ATPases, which certainly changes the distribution pattern of both original and mutated amino acids, thus the amino-acid distribution probability differs for both original and mutated amino acids.

For example, the mutation of the conserved histidine residue is one of the most common mutations found in Wilson disease [52,83], which actually changes histidine (H) to glutamine (Q) at position 1069 leading to the Wilson disease. There are 69 Qs and 36 Hs before mutation, but there are 70 Qs and 35 Hs after mutation. For the sake of simplicity, we used Hs as an example in Appendix.

Table 1 Amino acids, their compositions and distribution probability in normal human copper-transporting ATPases

Amino acid	ATP7A		ATP7B	
	Number	Distribution probability	Number	Distribution probability
A	106	0.004930	134	0.000279
R	48	0.000930	53	0.018279
N	67	0.000099	49	0.022866
D	66	0.003838	62	0.003048
C	26	0.040315	29	0.000677
E	92	0.003302	80	0.002652
Q	54	0.000865	69	0.003584
G	92	0.003350	104	0.000924
H	35	0.000143	36	0.013806
I	129	0.000027	106	0.002020
L	135	0.000670	132	0.000038
K	88	0.001071	75	0.004623
M	44	0.000110	44	0.008612
F	43	0.001010	36	0.000407
P	62	0.000964	72	0.004417
S	133	0.000000	126	0.000032
T	100	0.000341	82	0.000460
W	9	0.177028	11	0.040398
Y	30	0.043096	24	0.006342
V	141	0.002731	141	0.003840

A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; E, glutamic acid; Q, glutamine; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; V, valine

In this manner, we found that the distribution probabilities were 0.0138 and 0.0233 before and after mutation, which means that this mutation led the distribution probability increased in the original amino acid H. On the other hand, the distribution probability was different in the mutated amino acid Q before and after mutation, which were 0.0036 and 0.0111. Therefore the overall effect for this mutation on ATP7B was $(0.0233-0.0138)+(0.0111-0.0036) = 0.017$, that is, the mutation increased the distribution probability of amino acids.

Hence, we had the quantitative measure for the changed amino-acid distribution probability of mutated copper-transporting ATPases and we also had their related diseases in documentation, thus we could build a quantitative relationship between mutated protein sequence and corresponding clinical outcome.

Probabilistic relationship

For building of quantitative relationship between mutation and clinical outcome, we used the descriptively probabilistic

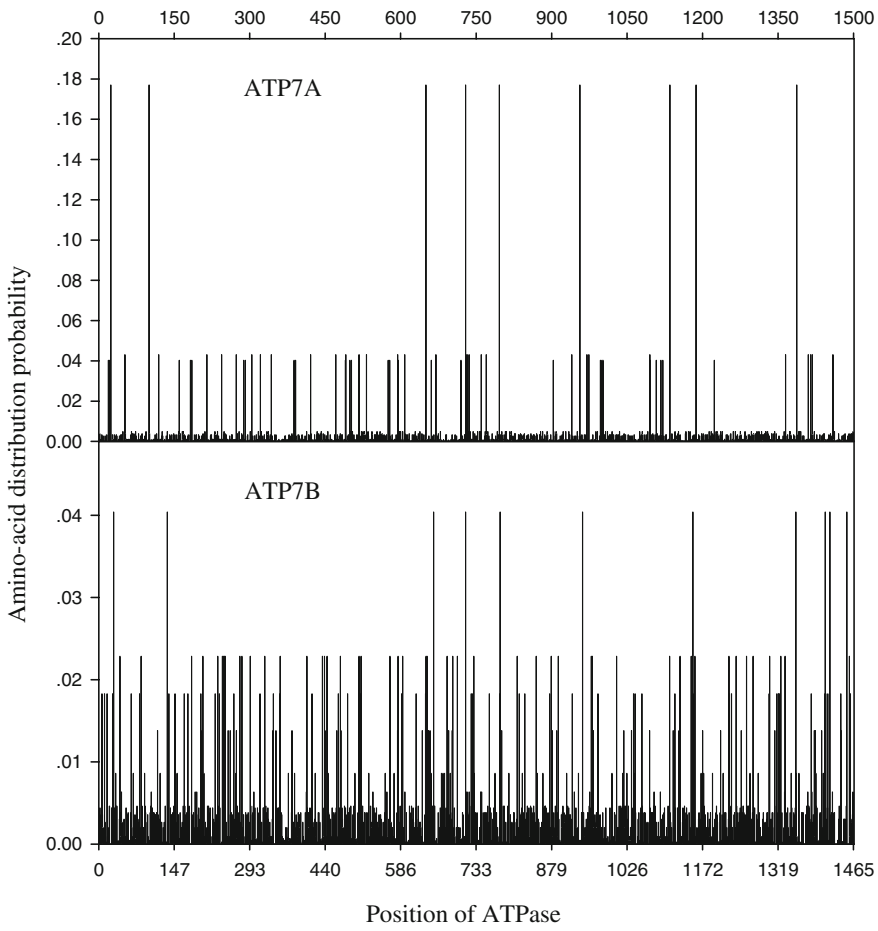

Fig. 1 Visualization of amino-acid distribution probability along ATP7A and ATP7B sequence

method, because our measure was amino-acid distribution probability and each individual mutation related to its clinical outcome was presented as frequency. Therefore, we used the cross-impact analysis to couple them, because the amino-acid distribution probability either increased or decreased after mutation, which was a 2-possibilty event, and the clinical outcome either occurred or did not occur after mutation, which was a yes-and-no event. Thereafter, we used the Bayesian equation to calculate the probability of occurrence of clinical outcome under mutation.

Statistical validation

The descriptive statistics or probability produces the general trends or estimates in a population, such as population weight, in our case the probability of occurrence of clinical outcome under mutation. Therefore, the validation is directly related to the increase in sampling size of population of interest, which could lead the population estimates to move from the current values, such as survey on different years. Therefore we will closely monitor the change in documented mutation and clinical outcome to re-determine the probability, which is the way to validate the descriptive analysis.

Results and discussion

In principle, the descriptive method was the first step for the development modeling, we therefore firstly dealt with the 37 ATP7A mutations, among which 33 mutations were documented as the Menkes disease, one as the occipital horn syndrome and the rest as polymorphisms. As the cross-impact analysis was particularly suited for two relevant events coupled together [80,84–90], we used this analysis to build a quantitative relationship between the increase/decrease of distribution probability after mutation and the diagnosis of Menkes disease.

Figure 2 showed the cross-impact analysis on the relationship between changed primary structure and diagnosis of Menkes disease. At the level of amino-acid distribution probability, $P(2)$ and $P\left(\bar{2}\right)$ were the decreased and increased probabilities led by mutations, and 12 point mutations resulted in the probability decreased whereas 25 point mutations led to the probability increased. At the level of diag-

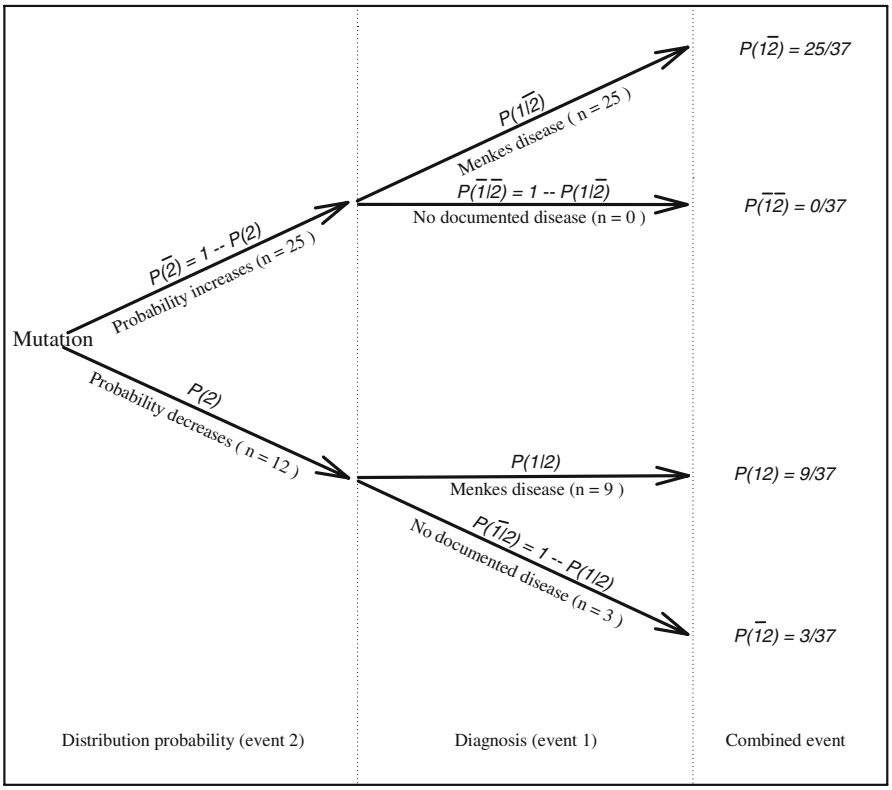

Fig. 2 Cross-impact relationship among mutant human ATP7A, changed amino-acid distribution probability, and diagnosis of Menkes disease

Table 2 Computation on cross-impact analysis in Fig. 2

$P(2) = 12/37 = 0.3243$
$P\left(\bar{2}\right) = 1 - P(2) = 1 - 0.3243 = 0.6757 = 25/37$
$P\left(1|\bar{2}\right) = 25/25 = 1$
$P\left(\bar{1}|\bar{2}\right) = 1 - P\left(1|\bar{2}\right) = 1 - 1 = 0 = 0/25$
$P(1|2) = 9/12 = 0.75$
$P\left(\bar{1}|2\right) = 1 - P(1|2) = 1 - 0.75 = 0.25 = 3/12$
$P\left(1\bar{2}\right) = P\left(1|\bar{2}\right) \times P\left(\bar{2}\right) = 25/25 \times 25/37 = 0.6757 = 25/37$
$P\left(\bar{1}\bar{2}\right) = P\left(\bar{1}|\bar{2}\right) \times P\left(\bar{2}\right) = 0/25 \times 25/37 = 0 = 0/37$
$P(12) = P(1|2) \times P(2) = 9/12 \times 12/37 = 0.2432 = 9/37$
$P\left(\bar{1}2\right) = P\left(\bar{1}|2\right) \times P(2) = 3/12 \times 12/37 = 0.0811 = 3/37$

nosis of Menkes disease under mutation: (i) $P\left(1|\bar{2}\right)$ was the impact probability (conditional probability) that Menkes disease could be diagnosed under the condition of increased distribution probability, and 25 mutations had such results. (ii) $P\left(\bar{1}|\bar{2}\right)$ was the impact probability that no disease was documented under the condition of increased distribution probability, and none of mutations worked in such a manner. (iii) $P(1|2)$ was the impact probability that Menkes disease could be diagnosed under the condition of decreased distribution probability, and nine mutations played such a role. (iv) $P\left(\bar{1}|2\right)$ was the impact probability that no disease was documented under the condition of decreased distribution probability, and three mutations fell into this category. At the level of combined events, we could see the combined results of changed primary structure and diagnosis of Menkes disease.

Table 2 listed the computed probabilities with respect to Fig. 2, from which we could see several interesting points. (i) As $P\left(\bar{2}\right)$ was larger than $P(2)$, a mutation had two third chance of increasing the distribution probability in mutant human ATP7A. (ii) As $P\left(\bar{1}|\bar{2}\right)$ was equal to zero, a mutation that increased the distribution probability had 100% chance of excluding Menkes disease. (iii) As $P(1|2)$ was much larger than $P\left(\bar{1}|2\right)$, a mutation that decreased the distribution probability had three fourth chance of bringing out Menkes disease.

Here, it was very meaningful to use the Bayes' law [91],

$$P\left(1|2\right) = P\left(2|1\right)\frac{P\left(1\right)}{P\left(2\right)},$$

which indicates the probabilities of occurrences of two events, to determine the probability, $P(1)$, the diagnosis of Menkes disease under a mutation, because $P(2)$ and $P\left(1|2\right)$ had already been defined in cross-impact analysis, while $P\left(2|1\right)$ was the probability that the distribution probability decreased under the condition that the Menkes disease was diagnosed.

Thus, we had $P(1|2) = 9/12 = 0.75$ (Table 2), and $P\left(2|1\right) = 9/(9+25) = 0.2647$, then we had

$$P\left(1\right) = \frac{P\left(1|2\right)}{P\left(2|1\right)}P\left(2\right) = \frac{0.75 \times 0.3243}{0.2647} = 0.9189,$$

namely, the chance that Menkes disease was diagnosed under ATP7A mutation was larger than nine tenth.

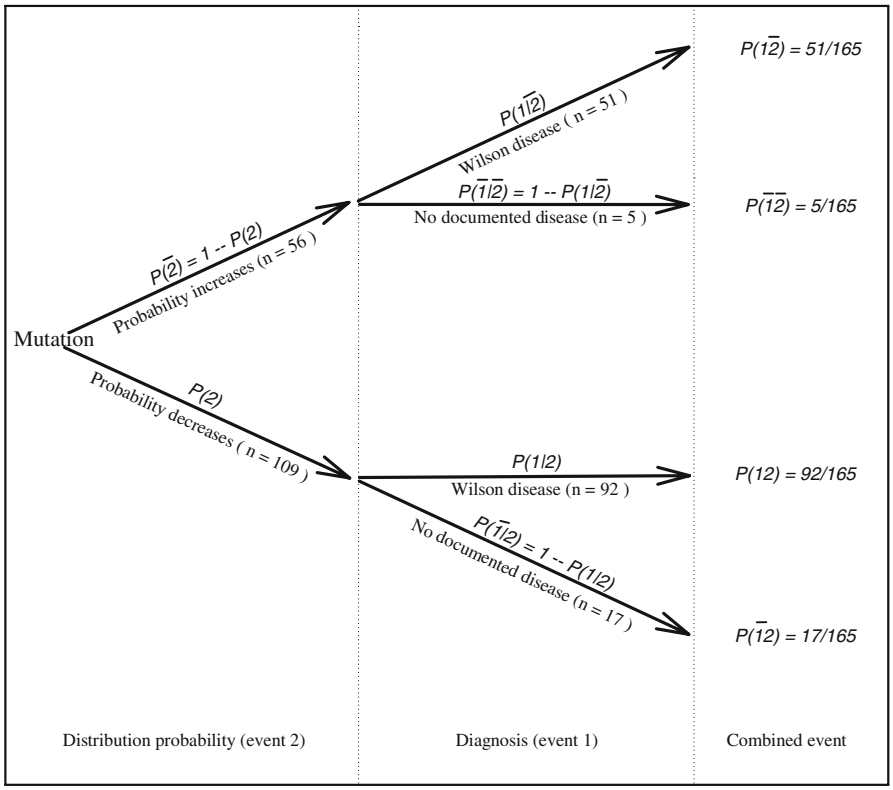

Fig. 3 Cross-impact relationship among mutant human ATP7B, changed amino-acid distribution probability, and diagnosis of Wilson disease

Secondly we turned our focus on the human ATP7B with its 165 mutations, of which 143 mutations were documented with Wilson disease. Figure 3 illustrated the cross-impact analysis on the relationship between changed primary structure of ATP7B and diagnosis of Wilson disease.

Similarly, at the level of amino-acid distribution probability, $P(2)$ and $P\left(\bar{2}\right)$ were the decreased and increased probabilities led by mutations, and 109 and 56 mutations resulted in the probability decreased and increased, respectively. At the level of diagnosis of Wilson disease under a mutation: (i) $P\left(1|\bar{2}\right)$ was the impact probability that Wilson disease could be diagnosed under the condition of increased distribution probability, and 51 mutations had such results. (ii) $P\left(\bar{1}|\bar{2}\right)$ was the impact probability that no disease was documented under the condition of increased distribution probability, and five mutations worked in such a manner. (iii) $P(1|2)$ was the impact probability that Wilson disease could be diagnosed under the condition of decreased distribution probability, and 92 mutations played such a role. (iv) $P\left(\bar{1}|2\right)$ was the impact probability that no disease was documented under the condition of decreased distribution probability, and 17 mutations fell into this category. At the level of combined events, we could see the combined results of changed primary structure of ATP7B and diagnosis of Wilson disease.

Table 3 displayed the computed probabilities with respect to Fig. 3, from which we could draw some features. (i) As $P\left(\bar{2}\right)$ was smaller than $P(2)$, a mutation had one third chance of increasing the distribution probability in mutant human ATP7B. (ii) As $P\left(\bar{1}|\bar{2}\right)$ was less than 0.1, a mutation that increased the distribution probability had more than nine tenth chance of leading to Wilson disease. (iii) As $P(1|2)$ was much larger than $P\left(\bar{1}|2\right)$, a mutation that decreased the distribution probability had more than eight tenth chance of resulting in Wilson disease.

Table 3 Computation on cross-impact analysis in Fig. 3

$P(2) = 109/165 = 0.6606$
$P\left(\bar{2}\right) = 1 - P(2) = 1 - 0.6606 = 0.3394 = 56/165$
$P\left(1|\bar{2}\right) = 51/56 = 0.9107$
$P\left(\bar{1}|\bar{2}\right) = 1 - P\left(1|\bar{2}\right) = 1 - 0.9107 = 0.0893 = 5/56$
$P(1|2) = 92/109 = 0.8440$
$\left(\bar{1}|2\right) = 1 - P(1|2) = 1 - 0.8440 = 0.1560 = 17/109$
$P\left(1\bar{2}\right) = P\left(1|\bar{2}\right) \times P\left(\bar{2}\right) = 51/56 \times 56/165 = 0.3091 = 51/165$
$P\left(\bar{1}\bar{2}\right) = P\left(\bar{1}|\bar{2}\right) \times P\left(\bar{2}\right) = 5/56 \times 56/165 = 0.0303 = 5/165$
$P(12) = P(1|2) \times P(2) = 92/109 \times 109/165 = 0.5576 = 92/165$
$P\left(\bar{1}2\right) = P\left(\bar{1}|2\right) \times P(2) = 17/109 \times 109/165 = 0.1030 = 17/165$

Also, we could use the Bayesian equation to determine the probability, $P(1)$, that Wilson disease was diagnosed under ATP7B mutations. As we had $P(1|2) = 92/109 = 0.8440$ (Table 3), and $P\left(2|1\right) = 92/(92 + 51) = 0.6434$, so

$$P\left(1\right) = \frac{P\left(1|2\right)}{P\left(2|1\right)} P\left(2\right) = \frac{0.8440 \times 0.6606}{0.6434} = 0.8666,$$

say, a ATP7B mutation had about nine tenth chance of causing Wilson disease.

Although both ATP7A and ATP7B genes and their encoding proteins have similar structure [92], the ATP7A mutations in humans cause the Menkes disease characterized by a copper deficiency disorder, while the mutations in ATP7B result in the Wilson disease characterized by a copper toxicity condition. The reason for the very different diseases caused by mutations is in part due to their different pattern of cellular trafficking [93–96] and distinct pattern of tissue expression [97].

Also, different mutation patterns are found in both genes. For example, deletions of varying sizes within the ATP7A gene have been identified in 15–20% of patients with classical Menkes disease [42]. Unlike the situation in Menkes disease, large gene deletions have not been reported for patients with Wilson disease [52]. Thus, less point mutations were documented in ATP7A but much more mutations in ATP7B.

Conclusion

From the probabilistic viewpoint, vast majority of mutations lead to the amino-acid distribution probability increase in mutant ATP7As and decrease in ATP7Bs, which may throw lights on the difference between Menkes and Wilson disease.

The genotype–phenotype relationship is a crucial issue for diagnosing genetic defective diseases, such as Menkes and Wilson disease [83,97]. There is, in part, a correlation between the severity of the mutation and the disease severity, but other factors are also involved [98]. Using the cross-impact analysis and the Bayesian equation, we can calculate the probability, $P(1)$, that Menkes/Wilson disease can be diagnosed under ATP7A/ATP7B mutations, which will benefit for early clinical diagnosis as treatment can prevent tissue damage.

Appendix

Example of computation of amino-acid distribution probability for normal ATP7A

For the simplest example, there are nine tryptophans (W) in ATP7A, which are the least abundant amino acids in this protein (Table 1). How do these nine Ws distribute along the

Appendix Table 1 All possible distributions of nine tryptophans in human ATP7A

Partition									Distribution probability
I	II	III	IV	V	VI	VII	VIII	IX	
1	1	1	1	1	1	1	1	1	9.3666e-4
	1	1	1	1	1	1	1	2	0.0337
		1	1	1	1	1	1	3	0.0393
			1	1	1	1	1	4	0.0197
				1	1	1	1	5	4.9174e-3
					1	1	1	6	6.5566e-4
						1	1	7	4.6833e-5
							1	8	1.6726e-6
								9	2.3231e-8
		1	***1***	***1***	***1***	***1***	***2***	***2***	***0.1770***
			1	1	1	1	2	3	0.1967
				1	1	1	2	4	0.0492
					1	1	2	5	5.9009e-3
						1	2	6	3.2783e-4
							2	7	6.6904e-6
			1	1	1	2	2	2	0.1967
				1	1	2	2	3	0.1475
					1	2	2	4	0.0148
						2	2	5	4.9174e-4
				1	2	2	2	2	0.0369
					2	2	2	3	9.8349e-3
				1	1	1	3	3	0.0328
					1	1	3	4	9.8349e-3
						1	3	5	6.5566e-4
							3	6	1.5611e-5
					1	2	3	3	0.0197
						2	3	4	1.6391e-3
						3	3	3	3.6426e-4
						1	4	4	4.0979e-4
							4	5	2.3416e-5

Bold and italic is the real distribution

ATP7A? Accordingly, we can imagine to dividing the ATP7A into nine equal partitions, and each partition has about 167 amino acids (1,500/9) because the human ATP7A is composed of 1,500 amino acids, then there would be 30 configurations for all the possible distributions of nine Ws (Appendix Table 1).

Here, we calculate two distribution probabilities in this Table as example according to the equation. For the nine Ws equally distribute in each partition (the third row in the Table), we have $r_1 = 1, r_2 = 1, \ldots r_9 = 1, q_0 = 0, q_1 = 9, \ldots q_9 = 0$. Thus, we have the distribution probability as

$$\frac{9!}{0! \times 9! \times 0! \times 0! \times 0! \times 0! \times 0! \times 0! \times 0! \times 0!} \times \frac{9!}{1! \times 1! \times 1! \times 1! \times 1! \times 1! \times 1! \times 1! \times 1!} \times 9^{-9} = \frac{362880}{1 \times 362880 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1} \times \frac{362880}{1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1} \times \frac{1}{387420489} = 9.3666\mathrm{e}^{-4}.$$

Clearly, the human ATP7A can adopt only one pattern, which is that two partitions contain zero W, five partitions contain one W and two partitions contain two Ws (row 12 in the Table). So we have $r_1 = 0, r_2 = 0, r_3 = 1, r_4 = 1, r_5 = 1, r_6 = 1, r_7 = 1, r_8 = 2, r_9 = 2, q_0 = 2, q_1 = 5, q_2 = 2, q_3 = 0, q_4 = 0, q_5 = 0, q_6 = 0, q_7 = 0, q_8 = 0, q_9 = 0$, that is,

$$\frac{9!}{2! \times 5! \times 2! \times 0! \times 0! \times 0! \times 0! \times 0! \times 0! \times 0!} \times \frac{9!}{0! \times 0! \times 1! \times 1! \times 1! \times 1! \times 1! \times 2! \times 2!} \times 9^{-9} = \frac{362880}{2 \times 120 \times 2 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1} \times \frac{362880}{1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 2 \times 2} \times \frac{1}{387420489} = 0.1770.$$

Example of computation of amino-acid distribution probability for ATP7A after mutation

Appendix Table 2 shows the distribution patterns for H before and after mutation at position 1069. Accordingly, we have $r_1 = 0, r_2 = 0, r_3 = 1, r_4 = 0, r_5 = 1, r_6 = 0, r_7 = 3, r_8 = 0, r_9 = 2, r_{10} = 1, r_{11} = 1, r_{12} = 2, r_{13} = 0, r_{14} = 0, r_{15} = 1, r_{16} = 4, r_{17} = 2, r_{18} = 1, r_{19} = 0, r_{20} = 1, r_{21} = 0, r_{22} = 2, r_{23} = 0, r_{24} = 1, r_{25} = 1, r_{26} = 1, r_{27} = 1, r_{28} = 2, r_{29} = 1, r_{30} = 1, r_{31} = 1, r_{32} = 1, r_{33} = 0, r_{34} = 2, r_{35} = 1, r_{36} = 1$; $q_0 = 11, q_1 = 17, q_2 = 6, q_3 = 1, q_4 = 1, q_5 = 0, q_6 = 0, q_7 = 0, q_8 = 0, q_9 = 0, q_{10} = 0, q_{11} = 0, q_{12} = 0, q_{13} = 0, q_{14} = 0, q_{15} = 0, q_{16} = 0, q_{17} = 0, q_{18} = 0, q_{19} = 0, q_{20} = 0, q_{21} = 0, q_{22} = 0, q_{23} = 0, q_{24} = 0, q_{25} = 0, q_{26} = 0, q_{27} = 0, q_{28} = 0, q_{29} = 0, q_{30} = 0, q_{31} = 0, q_{32} = 0, q_{33} = 0, q_{34} = 0, q_{35} = 0, q_{36} = 0$ and $q_{37} = 0$ before mutation, whose distribution probability is

$$\frac{36!}{11! \times 17! \times 6! \times 1! \times 1! \times 0! \ldots \times 0!} \times \frac{36!}{0! \times 0! \times 1! \times 0! \times 1! \times 0! \ldots \times 1!} \times 36^{-36} = 0.0138.$$

We also have $r_1 = 0, r_2 = 0, r_3 = 1, r_4 = 0, r_5 = 1, r_6 = 1, r_7 = 2, r_8 = 0, r_9 = 3, r_{10} = 0, r_{11} = 2, r_{12} = 1, r_{13} = 0, r_{14} = 1, r_{15} =, 1, r_{16} = 3, r_{17} = 2, r_{18} = 1, r_{19} = 1, r_{20} = 0, r_{21} =, 1, r_{22} = 1, r_{23} = 1, r_{24} = 0, r_{25} = 2, r_{26} = 0, r_{27} = 2, r_{28} = 1, r_{29} = 1, r_{30} = 1, r_{31} = 0, r_{32} = 1, r_{33} = 0, r_{34} = 3, r_{35} = 1$; $q_0 = 11, q_1 = 16, q_2 = 5, q_3 = 3, q_4 = 0, q_5 = 0, q_6 = 0, q_7 = 0, q_8 = 0, q_9 = 0, q_{10} = 0, q_{11} = 0, q_{12} = 0,$

Appendix Table 2 Distribution of histidines before and after mutation at position 1069 in human ATP7B

Partition	Before mutation	After mutation
I	0	0
II	0	0
III	1	1
IV	0	0
V	1	1
VI	0	1
VII	3	2
VIII	0	0
IX	2	3
X	1	0
XI	1	2
XII	2	1
XIII	0	0
XIV	0	1
XV	1	1
XVI	4	3
XVII	2	2
XVIII	1	1
XIX	0	1
XX	1	0
XXI	0	1
XXII	2	1
XXIII	0	1
XXIV	1	0
XXV	1	2
XXVI	1	0
XXVII	1	2
XXVIII	2	1
XXIX	1	1
XXX	1	1
XXXI	1	0
XXXII	1	1
XXXIII	0	0
XXXIV	2	3
XXXV	1	1
XXXVI	1	–

$q_{13} = 0, q_{14} = 0, q_{15} = 0, q_{16} = 0, q_{17} = 0, q_{18} = 0, q_{19} = 0, q_{20} = 0, q_{21} = 0, q_{22} = 0, q_{23} = 0, q_{24} = 0, q_{25} = 0, q_{26} = 0, q_{27} = 0, q_{28} = 0, q_{29} = 0, q_{30} = 0, q_{31} = 0, q_{32} = 0, q_{33} = 0, q_{34} = 0, q_{35} = 0$, and $q_{36} = 0$ after mutation, whose distribution probability is

$$\frac{35!}{11! \times 16! \times 5! \times 3! \times 0! \ldots \times 0!} \times \frac{35!}{0! \times 0! \times 1! \times 0! \times 1! \ldots \times 1!} \times 35^{-35} = 0.0233.$$

References

1. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E (2008) Proteomics, networks and connectivity indices. Proteomics 8:750–778. doi:10.1002/pmic.200700638
2. González-Díaz H, Vilar S, Santana L, Uriarte E (2007) Medicinal chemistry and bioinformatics–current trends in drugs discovery with networks topological indices. Curr Top Med Chem 7:1015–1029. doi:10.2174/156802607780906771
3. González-Díaz H, Pérez-Castillo Y, Podda G, Uriarte E (2007) Computational chemistry comparison of stable/nonstable protein mutants classification models based on 3D and topological indices. J Comput Chem 28:1990–1995. doi:10.1002/jcc.20700
4. González-Díaz H, Uriarte E (2005) Proteins QSAR with Markov average electrostatic potentials. Bioorg Med Chem Lett 15: 5088–5094. doi:10.1016/j.bmcl.2005.07.056
5. González-Díaz H, Molina R, Uriarte E (2005) Recognition of stable protein mutants with 3D stochastic average electrostaticpotentials. FEBS Lett 579:4297–4301. doi:10.1016/j.febslet.2005.06.065
6. González-Díaz H, Uriarte E, Ramos de Armas R (2005) Predicting stability of Arc repressor mutants with protein stochastic moments. Bioorg Med Chem 13:323–331. doi:10.1016/j.bmc.2004.10.024
7. Marrero-Ponce Y, Medina-Marrero R, Castillo-Garit JA, Romero-Zaldivar V, Torrens F, Castro EA (2005) Protein linear indices of the macromolecular pseudograph alpha-carbon atomadjacency matrix' in bioinformatics. Part 1: prediction of protein stability effects of a complete set of alanine substitutions in Arc repressor. Bioorg Med Chem 13:3003–3015. doi:10.1016/j.bmc.2005.01.062
8. Fernández M, Caballero J, Fernández L, Abreu JI, Acosta G (2008) Classification of conformational stability of protein mutants from 3Dpseudo-folding graph representation of protein sequences using support vector machines. Proteins 70:167–175. doi:10.1002/prot.21524
9. Fernández L, Caballero J, Abreu JI, Fernández M (2007) Amino acid sequence autocorrelation vectors and Bayesian-regularized geneticneural networks for modeling protein conformational stability: gene V protein mutants. Proteins 67:834–852. doi:10.1002/prot.21349
10. Caballero J, Fernández L, Garriga M, Abreu JI, Collina S, Fernández M (2007) Proteometric study of ghrelin receptor function variations upon mutations using amino acid sequence autocorrelation vectors and genetic algorithm-based least square support vector machines. J Mol Graph Model 26:166–178. doi:10.1016/j.jmgm.2006.11.002
11. Wu G, Yan S (2002) Randomness in the primary structure of protein: methods and implications. Mol Biol Today 3: 55–69
12. Wu G, Yan S (2006) Mutation trend of hemagglutinin of influenza A virus: a review from computational mutation viewpoint. Acta Pharmacol Sin 27:513–526. doi:10.1111/j.1745-7254.2006.00329.x
13. Wu G, Yan S (2008) Lecture notes on computational mutation. Nova Science Publishers, New York
14. Wu G, Yan S (2006) Prediction of possible mutations in H5N1 hemagglutinins of influenza A virus by means of logistic regression. Comp Clin Pathol 15:255–261. doi:10.1007/s00580-006-0638-y
15. Wu G, Yan S (2006) Prediction of mutations in H5N1 hemagglutinins from influenza A virus. Protein Pept Lett 13:971–976. doi:10.2174/092986606778777533
16. Wu G, Yan S (2007) Improvement of model for prediction of hemagglutinin mutations in H5N1 influenza viruses with distinguishing of arginine, leucine and serine. Protein Pept Lett 14:191–196. doi:10.2174/092986607779816032
17. Wu G, Yan S (2007) Improvement of prediction of mutation positions in H5N1 hemagglutinins of influenza A virus using neural network with distinguishing of arginine, leucine and serine. Protein Pept Lett 14:465–470. doi:10.2174/092986607780782713
18. Wu G, Yan S (2008) Prediction of mutations initiated by internal power in H3N2 hemagglutinins of influenza A virus from North America. Int J Pept Res Ther 14:41–51. doi:10.1007/s10989-007-9104-1
19. Wu G, Yan S (2008) Prediction of mutations in H1 neuraminidases from North America influenza A virus engineered by internal randomness. Mol Divers 11:131–140. doi:10.1007/s11030-008-9067-y
20. Wu G, Yan S (2008) Prediction of mutations engineered by randomness in H5N1 neuraminidases from influenza A virus. Amino Acids 34:81–90. doi:10.1007/s00726-007-0579-z
21. Wu G, Yan S (2008) Prediction of mutation in H3N2 hemagglutinins of influenza A virus from North America based on different datasets. Protein Pept Lett 15: 144–152
22. Wu G, Yan S (2008) Prediction of mutations engineered by randomness in H5N1 hemagglutinins of influenza A virus. Amino Acids. doi:10.1007/s00726-007-0602-4
23. Wu G, Yan S (2008) Three sampling strategies to predict mutations in H5N1 hemagglutinins from influenza A virus. Protein Pept Lett http://www.bentham.org/ppl/CurrentIssue.htm
24. Wu G, Yan S (2005) Timing of mutation in hemagglutinins from influenza A virus by means of unpredictable portion of amino-acid pair and fast Fourier transform. Biochem Biophys Res Commun 333:70–78. doi:10.1016/j.bbrc.2005.05.094
25. Wu G, Yan S (2006) Timing of mutation in hemagglutinins from influenza A virus by means of amino-acid distribution rank and fast Fourier transform. Protein Pept Lett 13:143–148. doi:10.2174/092986606775101616
26. Linder MC (1991) Biochemistry of copper. Plenum Press, New York
27. Pardo CA, Xu Z, Borchelt DR, Price DL, Sisodia SS, Cleveland DW (1995) Superoxide dismutase is an abundant component of cell bodies, dendrites and axons of motor neurons and in a subset of other neurons. Proc Natl Acad Sci USA 92:954–958. doi:10.1073/pnas.92.4.954
28. Rolfs A, Hediger MA (1999) Metal ion transporters in mammals: structure, function and pathological implications. J Physiol 518: 1–12. doi:10.1111/j.1469-7793.1999.0001r.x
29. Tanzi RE, Petrukhin K, Chernov I, Pellequer JL, Wasco W, Ross B et al (1993) The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. Nat Genet 5:344–350. doi:10.1038/ng1293-344
30. Terada K, Kawarada Y, Miura N, Yasui O, Koyama K, Sugiyama T (1995) Copper incorporation into ceruloplasmin in rat livers. Biochim Biophys Acta 1270: 58–62
31. Yoshida K, Furihata K, Takeda S, Nakamura A, Yamamoto K, Morita H et al (1995) A mutation in the ceruloplasmin gene is associated with systemic hemosiderosis in humans. Nat Genet 9:267–272. doi:10.1038/ng0395-267

32. Terada K, Nakako T, Yang XL, Iida M, Aiba N, Minamiya Y et al (1998) Restoration of holoceruloplasmin synthesis in LEC rat after infusion of recombinant adenovirus bearing WND cDNA. J Biol Chem 273:1815–1820. doi:10.1074/jbc.273.3.1815
33. Camakaris J, Voskoboinik I, Mercer JFB (1999) Molecular mechanisms of copper homeostasis. Biochem Biophys Res Commun 261:225–232. doi:10.1006/bbrc.1999.1073
34. Peña MMO, Lee J, Thiele DJ (1999) A delicate balance: homeostatic control of copper uptake and distribution. J Nutr 129: 1251–1260
35. Chelly J, Tumer Z, Tonnesen T, Petterson A, Ishikawa-Brush Y, Tommerup N et al (1993) Isolation of a candidate gene for Menkes disease that encodes a potential heavy metal binding protein. Nat Genet 3:14–19. doi:10.1038/ng0193-14
36. Mercer JF, Livingston J, Hall B, Paynter JA, Begy C, Chandrasekharappa S et al (1993) Isolation of a partial candidate gene for Menkes disease by positional cloning. Nat Genet 3:20–25. doi:10.1038/ng0193-20
37. Vulpe C, Levinson B, Whitney S, Packman S, Gitschier J (1993) Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. Nat Genet 3:7–13. doi:10.1038/ng0193-7
38. Tuemer Z, Moeller LB, Horn N (1999) Mutation spectrum of ATP7A, the gene defective in Menkes disease. Adv Exp Med Biol 448: 83–95
39. Ambrosini L, Mercer JFB (1999) Defective copper-induced trafficking and localization of the Menkes protein in patients with mild and copper-treated classical Menkes disease. Hum Mol Genet 8:1547–1555. doi:10.1093/hmg/8.8.1547
40. Ogawa A, Yamamoto S, Takayanagi M, Kogo T, Kanazawa M, Kohno Y (1999) Identification of three novel mutations in the MNK gene in three unrelated Japanese patients with classical Menkes disease. J Hum Genet 44:206–209. doi:10.1007/s100380050144
41. Gu Y-H, Kodama H, Murata Y, Mochizuki D, Yanagawa Y, Ushijima H et al (2001) ATP7A gene mutations in 16 patients with Menkes disease and a patient with occipital horn syndrome. Am J Med Genet 99:217–222. doi:10.1002/1096-8628(2001)9999:9999<::AID-AJMG1167>3.0.CO;2-R
42. Turner Z, Horn N (1997) Menkes disease: recent advances and new aspects. J Med Genet 34: 265–274
43. Bull PC, Thomas GR, Rommens JM, Forbes JR, Cox DW (1993) The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. Nat Genet 5:327–337. doi:10.1038/ng1293-327
44. Yamaguchi Y, Heiny ME, Gitlin JD (1993) Isolation and characterization of a human liver cDNA as a candidate gene for Wilson disease. Biochem Biophys Res Commun 197:271–277. doi:10.1006/bbrc.1993.2471
45. Petrukhin K, Lutsenko S, Chernov I, Ross BM, Kaplan JH, Gilliam TC (1994) Characterization of the Wilson disease gene encoding a P-type copper transporting ATPase: genomic organization, alternative splicing, and structureÏfunction predictions. Hum Mol Genet 3:1647–1656. doi:10.1093/hmg/3.9.1647
46. Oh WJ, Kim EK, Park KD, Hahn SH, Yoo OJ (1999) Cloning and characterization of the promoter region of the Wilson disease gene. Biochem Biophys Res Commun 259:206–211. doi:10.1006/bbrc.1999.0732
47. Tao TY, Gitlin JD (2003) Hepatic copper metabolism: insights from genetic disease. Hepatology 37:1241–1247. doi:10.1053/jhep.2003.50281
48. Linder MC, Wooten L, Cerveza P, Cotton S, Shulze R, Lomeli N (1998) Copper transport. Am J Clin Nutr 67(5 Suppl): 965S–971S
49. Schaefer M, Hopkins RG, Failla ML, Gitlin JD (1999) Hepatocyte-specific localization and copper-dependent trafficking of the Wilson's disease protein in the liver. Am J Physiol 276: G639–G646
50. Terada K, Aiba N, Yang XL, Iida M, Nakai M, Miura N et al (1999) Biliary excretion of copper in LEC rat after introduction of copper transporting P-type ATPase, ATP7B. FEBS Lett 448:53–56. doi:10.1016/S0014-5793(99)00319-1
51. Roelofsen H, Wolters H, Van Luyn MJA, Miura N, Kuipers F, Vonk RJ (2000) Copper-induced apical trafficking of ATP7B in polarized hepatoma cells provides a mechanism for biliary copper excretion. Gastroenterology 119:782–793. doi:10.1053/gast.2000.17834
52. Thomas GR, Forbes JR, Roberts EA, Walshe JM, Cox DW (1995) The Wilson disease gene: spectrum of mutations and their consequences. Nat Genet 9:210–216. doi:10.1038/ng0295-210
53. Loudianos G, Dessi V, Angius A, Lovicu M, Loi A, Deiana M et al (1996) Wilson disease mutations associated with uncommon haplotypes in Mediterranean patients. Hum Genet 98:640–642. doi:10.1007/s004390050275
54. Shah AB, Chernov I, Zhang HT, Ross BM, Das K, Lutsenko S et al (1997) Identification and analysis of mutations in the Wilson disease gene (ATP7B): population frequencies, genotype–phenotype correlation, and functional analyses. Am J Hum Genet 61:317–328. doi:10.1086/514864
55. Kim EK, Yoo OJ, Song KY, Yoo HW, Choi SY, Cho SW et al (1998) Identification of three novel mutations and a high frequency of the Arg778Leu mutation in Korean patients with Wilson disease. Hum Mutat 11:275–278. doi:10.1002/(SICI)1098-1004(1998)11:4<275::AID-HUMU4>3.0.CO;2-L
56. Wilson SAK (1912) Progressive lenticular degeneration: a familial nervous disease associated with cirrhosis of the liver. Brain 34:295–508. doi:10.1093/brain/34.4.295
57. Thomas GR, Forbes JR, Roberts EA, Walshe JM, Cox DW (1995) The Wilson disease gene: spectrum of mutations and their consequences. Nat Genet 9:210–217. Erratum in Nat Genet 1995; 9:451. doi:10.1038/ng0295-210
58. Schilsky ML (1996) Wilson disease: genetic basis of copper toxicity and natural history. Semin Liver Dis 16: 83–95
59. Mercer JFB (2001) The molecular basis of copper transport diseases. Trends Mol Med 7:64–69. doi:10.1016/S1471-4914(01)01920-7
60. Shim H, Harris ZL (2003) Genetic defects in copper metabolism. J Nutr 133: 1527S–1531S
61. Das SK, Ray K (2006) Wilson's disease: an update. Nat Clin Pract Neurol 2:482–493. doi:10.1038/ncpneuro0291
62. Brewer GJ, Yuzbasiyan-Gurkan V (1992) Wilson disease. Medicine 71:139–164. doi:10.1097/00005792-199205000-00004
63. Moeller LB, Bukrinsky JT, Moelgaard A, Paulsen M, Lund C, Tuemer Z et al (2005) Identification and analysis of 21 novel disease-causing amino acid substitutions in the conserved part of ATP7A. Hum Mutat 26:84–93. doi:10.1002/humu.20190
64. Loudianos G, Dessi V, Lovicu M, Angius A, Nurchi A, Sturniolo GC et al (1998) Further delineation of the molecular pathology of Wilson disease in the Mediterranean population. Hum Mutat 12:89–94. doi:10.1002/(SICI)1098-1004(1998)12:2<89::AID-HUMU3>3.0.CO;2-G
65. Loudianos G, Dessi V, Lovicu M, Angius A, Altuntas B, Giacchino R et al (1999) Mutation analysis in patients of Mediterranean descent with Wilson disease: identification of 19 novel mutations. J Med Genet 36: 833–836
66. Okada T, Shiono Y, Hayashi H, Satoh H, Sawada T, Suzuki A et al (2000) Mutational analysis of ATP7B and genotype–phenotype correlation in Japanese with Wilson's disease. Hum Mutat 15:454–462. doi:10.1002/(SICI)1098-1004(200005)15:5<454::AID-HUMU7>3.0.CO;2-J
67. Caca K, Ferenci P, Kuehn H-J, Polli C, Willgerodt H, Kunath B et al (2001) High prevalence of the H1069Q mutation in East German patients with Wilson disease: rapid detection of mutations by limited sequencing and phenotype–genotype analysis. J Hepatol 35:575–581. doi:10.1016/S0168-8278(01)00219-7

68. Gu Y-H, Kodama H, Du S-L, Gu Q-J, Sun H-J, Ushijima H (2003) Mutation spectrum and polymorphisms in ATP7B identified on direct sequencing of all exons in Chinese Han and Hui ethnic patients with Wilson's disease. Clin Genet 64:479–484. doi:10.1046/j.1399-0004.2003.00179.x
69. Margarit E, Bach V, Gomez D, Bruguera M, Jara P, Queralt R et al (2005) Mutation analysis of Wilson disease in the Spanish population-identification of a prevalent substitution and eight novel mutations in the ATP7B gene. Clin Genet 68: 61–68
70. Vrabelova S, Letocha O, Borsky M, Kozak L (2005) Mutation analysis of the ATP7B gene and genotype/phenotype correlation in 227 patients with Wilson disease. Mol Genet Metab 86:277–285. doi:10.1016/j.ymgme.2005.05.004
71. Cox DW (1999) Disorders of copper transport. Br Med Bull 55:544–555. doi:10.1258/0007142991902619
72. Wu G, Yan S (2000) Prediction of distributions of amino acids and amino acid pairs in human haemoglobin α-chain and its seven variants causing α-thalassemia from their occurrences according to the random mechanism. Comp Haematol Int 10:80–84. doi:10.1007/s005800070012
73. Wu G, Yan S (2001) Analysis of distributions of amino acids, amino acid pairs and triplets in human insulin precursor and four variants from their occurrences according to the random mechanism. J Biochem Mol Biol Biophys 5: 293–300
74. Wu G, Yan S (2001) Analysis of distributions of amino acids and amino acid pairs in human tumor necrosis factor precursor and its eight variants according to random mechanism. J Mol Model 7: 318–323
75. Wu G, Yan S (2002) Random analysis of presence and absence of two- and three-amino-acid sequences and distributions of amino acids, two- and three-amino-acid sequences in bovine p53 protein. Mol Biol Today 3: 31–37
76. Wu G, Yan S (2002) Analysis of distributions of amino acids in the primary structure of apoptosis regulator Bcl-2 family according to the random mechanism. J Biochem Mol Biol Biophys 6:407–414. doi:10.1080/1025814021000036106
77. Wu G, Yan S (2002) Analysis of distributions of amino acids in the primary structure of tumor suppressor p53 family according to the random mechanism. J Mol Model 8:191–198. doi:10.1007/s00894-002-0087-8
78. Wu G, Yan S (2004) Determination of sensitive positions to mutations in human p53 protein. Biochem Biophys Res Commun 321:313–319. doi:10.1016/j.bbrc.2004.06.117
79. Wu G, Yan S (2005) Searching of main cause leading to severe influenza A virus mutations and consequently to influenza pandemics/epidemics. Am J Infect Dis 1: 116–123
80. Wu G, Yan S (2005) Prediction of mutation trend in hemagglutinins and neuraminidases from influenza A viruses by means of cross-impact analysis. Biochem Biophys Res Commun 326:475–482. doi:10.1016/j.bbrc.2004.11.052
81. Gao N, Yan S, Wu G (2006) Pattern of positions sensitive to mutations in human haemoglobin α-chain. Protein Pept Lett 13:101–107. doi:10.2174/092986606774502090
82. Feller W (1968) An introduction to probability theory and its applications vol I, 3rd edn. Wiley, New York, pp 34–40
83. Butler P, McIntyre N, Mistry PK (2001) Molecular diagnosis of Wilson disease. Mol Genet Metab 72:223–230. doi:10.1006/mgme.2000.3143
84. Enzer S (1970) Delphi and cross-impact techniques: an effective combination for systematic futures analysis. Futures 3: 48–61
85. Enzer S (1972) Cross-impact techniques in technology assessment. Futures 4:30–51. doi:10.1016/0016-3287(72)90023-7
86. Gordon TG, Hayward H (1968) Initial experiments with the cross-impact matrix method of forecasting. Futures 1:100–116. doi:10.1016/S0016-3287(68)80003-5
87. Gordon TG (1969) Cross-impact matrices—an illustration of their use for policy analysis. Futures 2: 527–531
88. Sage AP (1977) Methodology for Large-scale Systems. McGraw-Hill, New York, pp 165–203
89. Wu G (2000) Application of cross-impact analysis to the relationship between aldehyde dehydrogenase 2 and flushing. Alcohol Alcohol 35:55–59. doi:10.1093/alcalc/35.1.55
90. Wu G, Yan S (2008) Building quantitative relationship between changed sequence and changed oxygen affinity in human hemoglobin β-chain. Protein Pept Lett 15:341–345. doi:10.2174/092986608784246498
91. Wikipedia—The free encyclopedia (2007) Bayes' theorem. en.wikipedia.org/wiki/Bayes'_theorem
92. Tanzi RE, Petrukhin K, Chernov I, Pellequer JL, Wasco W, Ross B et al (1993) The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. Nat Genet 5:344–350. doi:10.1038/ng1293-344
93. Forbes JR, Cox DW (2000) Copper-dependent trafficking of Wilson disease mutant ATP7B proteins. Hum Mol Genet 9:1927–1935. doi:10.1093/hmg/9.13.1927
94. Mercer JFB, Llanos RM (2003) Molecular and cellular aspects of copper transport in developing mammals. J Nutr 133: 1481S–1484S
95. La Fontaine S, Theophilos MB, Firth SD, Gould R, Parton RG, Mercer JFB (2001) Effect of the toxic milk mutation (tx) on the function and intracellular localization of Wnd, the murine homologue of the Wilson copper ATPase. Hum Mol Genet 10:361–370. doi:10.1093/hmg/10.4.361
96. Mercer JFB, Barnes N, Stevenson J, Strausak D, Llanos RM (2002) Copper-induced trafficking of the Cu-ATPases: a key mechanism for copper homeostasis. Biometals 16:175–184. doi:10.1023/A:1020719016675
97. Hahn S, Cho K, Ryu K, Kim J, Pai K, Kim M et al (2001) Identification of four novel mutations in classical Menkes disease and successful prenatal DNA diagnosis. Mol Genet Metab 73:86–90. doi:10.1006/mgme.2001.3169
98. Cox DW (1996) Molecular advances in Wilson disease. In: Boyer JL, Ockner RK (eds) Progress in liver diseases. Saunders, Philadelphia, pp 245–264