广西科学院学报
Journal of Guangxi Academy of Sciences

2009,25(3):183~186,191
Vol. 25,No. 3 August 2009

Quantitative Coupling of Human Antithrombin Ⅲ Mutations with Their Clinical Outcomes*
定量联结人抗凝血酶Ⅲ的变异及其临床表现

YAN Shao-min[1], WU Guang[2]
严少敏[1],吴 光[2]

(1. Guangxi Academy of Sciences, Nanning, Guangxi, 530007, China; 2. Computational Mutation Project, Dream Sci Tech Consulting, Shenzhen, Guangdong, 518054, China)
(1. 广西科学院,广西南宁 530007;2. 深圳市追梦科技咨询有限公司,广东深圳 518054)

Abstract: This study demonstrated how to convert a protein sequence into a numeric datum, and provided a descriptively probabilistic method to analyze the genotype-phenotype relationship of single gene disorder. The amino-acid distribution probability was used to quantify human antithrombin Ⅲ mutations, then the cross-impact analysis was used to couple the quantified mutations with their clinical outcomes, and finally the Bayesian equation was used to determine the probability that the antithrombin Ⅲ deficiency was determined under a mutation. The results showed that a person had a chance larger than 90% of being inherited antithrombin Ⅲ deficiency when a mutation was found in the antithrombin Ⅲ.

Key words: amino acid, antithrombin Ⅲ deficiency, single gene disorder, mutation, distribution probability, Bayes' law, cross-impact analysis

摘要:将蛋白质序列转换成数值数据后,建立一种描述性的概率方法分析单基因疾病的基因型与表型之间的关系。先用氨基酸分布概率定量变异的抗凝血酶Ⅲ,然后用交叉影响分析法联结人抗凝血酶Ⅲ的变异及其临床表现,再用贝叶斯公式计算变异条件下抗凝血酶Ⅲ缺乏的发病概率。结果显示当某个人被检测到抗凝血酶Ⅲ变异时将有大于90%的机会患遗传性抗凝血酶Ⅲ缺乏。

关键词:氨基酸 遗传性抗凝血酶Ⅲ缺乏 单基因疾病 变异 分布概率 贝叶斯法 交叉影响分析法
中图法分类号:Q75,R394.3 **文献标识码**:A **文章编号**:1002-7378(2009)03-0183-04

The antithrombin Ⅲ is a vitamin K-independent hepatocyte-synthesized protease inhibitor, which binds with and blocks the biologic activity of thrombin and other activated coagulation proteins involved in the clotting cascade[1].

The building of quantitative relationship between mutations in antithrombin Ⅲ and their clinical outcomes is intriguing, not only because 91 mutations have been documented in UniProtKB/Swiss-Prot entry[2], but also it would provide the basis for understanding and modeling of genotype-phenotype relationship. An important clinical outcome is the inherited antithrombin Ⅲ deficiency, which was first described in a Norwegian family[3]. Since then, many families with such deficiency have been described, where this disorder had caused severe venous thrombosis in successive generations, deep venous thrombus[4,5], cerebral venous thrombosis in neonatal[6,7], young[8] and adult patients[9], and so on.

However, it is understandable that a mutation in antithrombin Ⅲ could either cause a deficiency or have no clinical effects. In this context, a mutation has a yes-no relationship with its outcome, namely, an event with two possibilities, so the relationship between a single mutation with its outcome is quite simple. However, this would not be the case when there are many mutations involved, because each mutation occurred at different position with different substituted and substituting amino acids, which

收稿日期:2009-04-08
作者简介:严少敏(1958-),女,博士,研究员,主要从事定量诊断病理学和计算变异学研究。
*国际科技合作项目(2008DFA30710),广西自然科学基金项目(桂科攻0630003A2和桂科自0991080),广西科学院项目(09YJ17SW07)资助。

would result in a multiple yes-no relationship. Thus a method is needed in order to treat mutations as changed values rather than changed letters, which represent amino acids in different positions of antithrombin Ⅲ.

The simplest method would replace each amino acid in human antithrombin Ⅲ with a value that could be the physicochemical property of amino acid[10]. However, the physicochemical properties are not sensitive to mutations, thus our group has developed three methods based on random mechanism to transfer a protein sequence into a numeric sequence since 1999[11~14]. Hence, each mutation could be quantified with different value, and we could build a quantitative relationship between the mutations in the antithrombin Ⅲ and inherited antithrombin Ⅲ deficiency, which was designed as the aim of this study.

1 Materials and methods

1.1 Data

The human antithrombin Ⅲ sequence with its 91 mutations was obtained from UniProtKB/Swiss-Prot entry[2], of which 82 were missense mutations and 9 deletions.

1.2 Amino-acid distribution probability before and after mutation in antithrombin Ⅲ

The positions of amino acids in antithrombin Ⅲ could be associated with probability that was calculated according to $(r!)/(r_1! \times r_2! \times \cdots \times r_n!) \times (r!)/(q_0! \times q_1! \times \cdots \times q_n!) \times n^{-r}$[15], where r was the number of amino acids, n was the number of partitions, r_n was the number of amino acids in the n-th partition, q_n was the number of partitions with the same number of amino acids, and ! was the factorial function.

For example, there were five histidines in human antithrombin Ⅲ, locating at positions 33, 97, 152, 351 and 401, respectively. Accordingly, we could imagine dividing the antithrombin Ⅲ into five partitions because there were five histidines. As the antithrombin Ⅲ was composed of 464 amino acids, each partition would include about 93 amino acids (464/5 = 92.8). Then we counted how many histidines in each partition (column 3, Table 1), where one partition had two histidines, three partitions contained one histidine, and one partition had none, so we had $r_1 = 1, r_2 = 2, r_3 = 0, r_4 = 1, r_5 = 1, q_0 = 1, q_1 = 3, q_2 = 1, q_3 = 0, q_4 = 0, q_5 = 0$, and $(5!)/(1! \times 2! \times 0! \times 1! \times 1!) \times [(5!)/(1! \times 3! \times 1! \times 0! \times 0! \times 0!)] \times 5^{-5} = 120/(1 \times 2 \times 1 \times 1 \times 1) \times [120/(1 \times 6 \times 1 \times 1 \times 1 \times 1)] \times 1/3125 = 0.3840$. This was the distribution probability for the histidines in antithrombin Ⅲ before mutation.

If a mutation occurred related to histidine, its distribution probability would be changes. For example, a mutation in the antithrombin Ⅲ at position 425 substituted arginine for histidine leading to the type Ⅱ antithrombin Ⅲ deficiency[16]. In the mutant, there were six histidines with respect to their distributions in six partitions (column 5, Table 1), where $r_1 = 1, r_2 = 2, r_3 = 0, r_4 = 0, r_5 = 1, r_6 = 2, q_0 = 2, q_1 = 2, q_2 = 2, q_3 = 0, q_4 = 0, q_5 = 0, q_6 = 0$, i.e. $(6!)/(1! \times 2! \times 0! \times 0! \times 1! \times 2!) \times [(6!)/(\times 2! \times 2! \times 2! \times 0! \times 0! \times 0! \times 0!)] \times 6^{-6} = 0.3472$. Thus, this mutation decreased the distribution probability of histidines to 0.3472 from 0.384 (the last row in Table 1).

Table 1 Distributions and probabilities of arginines and histidines before and after R425H mutation in human antithrombin Ⅲ

Partition/ Probability	Before mutation		After mutation	
	Arginines	Histidines	Arginines	Histidines
Ⅰ	1	1	1	1
Ⅱ	0	2	0	2
Ⅲ	2	0	2	0
Ⅳ	2	1	2	0
Ⅴ	1	1	1	1
Ⅵ	0	-	0	2
Ⅶ	0	-	0	-
Ⅷ	2	-	2	-
Ⅸ	1	-	1	-
Ⅹ	0	-	1	-
Ⅺ	2	-	1	-
Ⅻ	0	-	0	-
ⅩⅢ	1	-	1	-
ⅩⅣ	3	-	3	-
ⅩⅤ	0	-	0	-
ⅩⅥ	0	-	0	-
ⅩⅦ	2	-	2	-
ⅩⅧ	0	-	1	-
ⅩⅨ	1	-	0	-
ⅩⅩ	0	-	2	-
ⅩⅪ	3	-	2	-
ⅩⅫ	2	-	0	-
ⅩⅩⅢ	0	-	-	-
Probability	0.0222	0.3840	0.0659	0.3472

Also, this mutation was related to arginines, whose distribution probabilities were 0.0222 and 0.0659 before and after mutation.

The overall effect of this mutation on the antithrombin Ⅲ was (0.0659−0.0222)+(0.3472−0.384)=0.0069, that was, the mutation increased the distribution probability for this mutant antithrombin Ⅲ. In this way, we had the amino-acid distribution probability for each mutation and its documented clinical outcome. Actually we have used this method in many our previous studies related to different subjects[11~14,17~22].

2 Results and discussion

After above computation, we had the amino-acid distribution probabilities of normal human antithrombin Ⅲ and its 91 mutants, of which 85 mutations were documented as inherited antithrombin Ⅲ deficiency. Thus, we could use the cross-impact analysis to couple the mutations with the occurrence/nonoccurrence of inherited antithrombin Ⅲ deficiency together, because the cross-impact analysis was particularly suited for coupling two relevant events[20,23~28].

Fig. 1 showed the scheme based on cross-impact analysis for events defined above. At the level of amino-acid distribution probability, $P(2)$ and $P(\bar{2})$ were the decreased and increased probabilities induced by mutations, and 36 and 55 mutations resulted in the distribution probability decreased and increased, respectively. At the level of clinical outcome: (i) $P(1|\bar{2})$ was the impact probability (conditional probability) that the inherited antithrombin Ⅲ deficiency occurred under the condition of increased distribution probability, and 52 mutations had such an effect. (ii) $P(\bar{1}|\bar{2})$ was the impact probability that no disease was documented under the condition of increased distribution probability, and 3 mutations worked in such a manner. (iii) $P(1|2)$ was the impact probability that the inherited antithrombin Ⅲ deficiency occurred under the condition of decreased distribution probability, and 33 mutations played such a role. (iv) $P(\bar{1}|2)$ was the impact probability that no disease was documented under the condition of decreased distribution probability, and 3 mutations fell into this category. At the level of combined events, the combined results with their frequency could be seen.

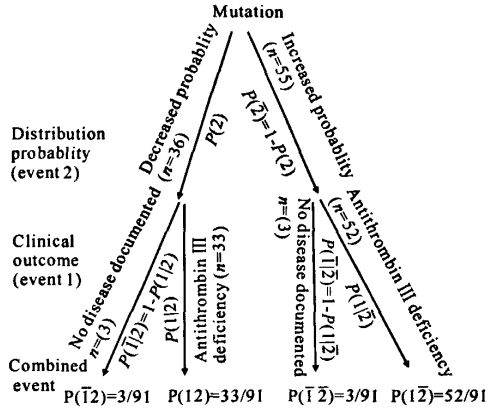

Fig. 1 Cross-impact relationship among mutations, clinical outcome, and their combination

Table 2 listed the calculated probabilities in the scheme in Fig. 1, from which several interesting points were found. (i) As $P(\bar{2})$ was larger than $P(2)$, a mutation had six tenths chance of increasing the distribution probability in mutant antithrombin Ⅲ. (ii) As $P(1|\bar{2})$ was 17 times of $P(\bar{1}|\bar{2})$, a mutation that increased the distribution probability had 0.95 chance of causing the inherited antithrombin Ⅲ deficiency. (iii) As $P(1|2)$ was remarkably larger than $P(\bar{1}|2)$, a mutation that decreased the distribution probability had more than nine tenths chance of causing the inherited antithrombin Ⅲ deficiency.

Table 2 Probability in cross-impact analysis in Fig. 1

Probability	Calculation from the data in Fig. 1	Result
$P(2)$	36/91	0.3956
$P(\bar{2})$	$1-P(2)=1-0.3956$ or 55/91	0.6044
$P(1\|\bar{2})$	52/55	0.9455
$P(\bar{1}\|\bar{2})$	$1-P(1\|\bar{2})=1-0.9455=3/55$	0.0545
$P(1\|2)$	33/36	0.9167
$P(\bar{1}\|2)$	$1-P(1\|2)=1-0.9167$ or 3/36	0.0833
$P(1\bar{2})$	$P(1\|\bar{2})\times P(\bar{2})=52/55\times55/91=52/91$	0.5714
$P(\bar{1}\bar{2})$	$P(\bar{1}\|\bar{2})\times P(\bar{2})=3/55\times55/91=3/91$	0.0330
$P(12)$	$P(1\|2)\times P(2)=33/36\times36/91=33/91$	0.3626
$P(\bar{1}2)$	$P(\bar{1}\|2)\times P(2)=3/36\times36/91=3/91$	0.0330

To this point, the Bayes' law[29], $P(1|2)=P(2|1)\dfrac{P(1)}{P(2)}$, could be applied to determine the probability, $P(1)$ that was the occurrence of inherited antithrombin Ⅲ deficiency under a mutation. For this equation, $P(2)$ and $P(1|2)$ could be found in cross-impact analysis, while $P(2|1)$ was the probability that the distribution probability

decreased under the condition of being inherited antithrombin Ⅲ deficiency. Thus, we had $P(1|2) = 33/36 = 0.9167$ (Table 2), and $P(2|1) = 33/(52+33) = 0.3882$, so $P(1) = \frac{P(1|2)}{P(2|1)}P(2) = \frac{0.9167 \times 0.3956}{0.3882} = 0.9342$, namely, a person had a chance larger than 90% of being inherited antithrombin Ⅲ deficiency when a mutation was found in human antithrombin Ⅲ. At present, the $P(1)$ calculated by our approach was very meaningful for clinical settings and benefit for diagnosis as well as counseling[30,31].

3 Conclusions

We demonstrated how to use the amino-acid distribution probability to convert a protein sequence into a numeric datum and provided a descriptively probabilistic method to couple the genotype-phenotype relationship of single gene disorder. The results showed that the probability of being inherited antithrombin Ⅲ deficiency was larger than 90% under a antithrombin Ⅲ mutation, which would benefit the clinical diagnosis.

References:

[1] Quinsey N S, Greedy A L, Bottomley S P, et al. Antithrombin: in control of coagulation [J]. Int J Biochem Cell Biol, 2004, 36: 386-389.

[2] Apweiler R, Bairoch A, Wu C H. Protein sequence databases[J]. Curr Opin Chem Biol, 2005, 8: 76-80.

[3] Egeberg O. Inherited antithrombin deficiency causing thrombophilia [J]. Thromb Diath Haemor, 1965, 13: 516.

[4] Bick R L. Prothrombin G20210A mutation, antithrombin, heparin cofactor Ⅰ, protein C, and protein S defects[J]. Hematol/Oncol Clin North Am, 2003, 17: 9-36.

[5] Kim V, Spandorfer J. Epidemiology of venous thromboembolic disease [J]. Emerg Med Clin North Am, 2001, 19: 839-859.

[6] Ahmad A. Genetics of cerebral venous thrombosis[J]. J Pakistan Med Assoc, 2006, 56: 488-490.

[7] Baud O, Picard V, Durand P, et al. Intracerebral hemorrhage associated with a novel antithrombin gene mutation in a neonate [J]. J Pediatr, 2001, 139: 741-743.

[8] Nagaraja D, Christopher R, Tripathi M. Plasma antithrombin Ⅲ deficiency in ischaemic stroke in the young [J]. Neurol India, 1999, 47: 155-156.

[9] Chen W H, Lan M Y, Chang Y Y, et al. The prevalence of protein C, protein S, and antithrombin Ⅲ deficiency in non-APS/SLE Chinese adults with noncardiac cerebral ischemia[J]. Clin Appl Thromb/Hemost, 2003, 9: 155-162.

[10] Chou K C. Structure bioinformatics and its impact to biomedical science [J]. Curr Med Chem, 2004, 11: 2105-2134.

[11] Wu G. The first and second order Markov chain analysis on amino acids sequence of human haemoglobin α-chain and its three variants with low O_2 affinity[J]. Comp Haematol Int, 1999, 9: 148-151.

[12] Wu G, Yan S. Randomness in the primary structure of protein: methods and implications[J]. Mol Biol Today, 2002, 3: 55-69.

[13] Wu G, Yan S. Mutation trend of hemagglutinin of influenza A virus: a review from computational mutation viewpoint[J]. Acta Pharmacol Sin, 2006, 27: 513-526.

[14] Wu G, Yan S. Lecture notes on computational mutation [M]. New York: Nova Science Publishers, 2008.

[15] Feller W. An introduction to probability theory and its applications Vol. 1 [M]. 3rd ed. New York: Wiley, 1968: 34-40.

[16] Erdjument H, Lane D A, Panico M, et al. Antithrombin Chicago, amino acid substitution of arginine 393 to histidine[J]. Thromb Res, 1989, 54: 613-619.

[17] Wu G, Yan S. 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 [J]. Comp Haematol Int, 2000, 10: 80-84.

[18] Wu G, Yan S. 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]. J Mol Model, 2001, 7: 318-323.

[19] Wu G, Yan S. Determination of sensitive positions to mutations in human p53 protein[J]. Biochem Biophys Res Commun, 2004, 321: 313-319.

[20] Wu G, Yan S. Prediction of mutation trend in hemagglutinins and neuraminidases from influenza A viruses by means of cross-impact analysis[J]. Biochem Biophys Res Commun, 2005, 326: 475-482.

[21] Gao N, Yan S, Wu G. Pattern of positions sensitive to mutations in human haemoglobin α-chain [J]. Protein Pept Lett, 2006, 13: 101-107.

[22] Wu G, Yan S. Prediction of mutations engineered by randomness in H5N1 neuraminidases from influenza A virus[J]. Amino Acids, 2007, 34: 81-90.

(下转第 191 页)

mation[J]. AmJ Pathol,2000,156(1):217-225.

[18] Nozoe T,Korenaga D,Kabashima A,et al. Significance of cyclinB1 expression as an independent prognostic indicator patients with squamous cell carcinoma of the esophagus[J]. Clin Cancer Res,2002,8(3):817-822.

[19] Dong Y,Sui L,Watanabe Y,et al. Clinical relevance of cyclinB1 overexpression in laryngeal squamous cell carcinoma[J]. Cancer Lett,2002,177(1):13-19.

[20] Hassan K A,Elnaggar A K,Soria J C,et al. Clinical significance of cyclinB1 protein expression in squamous cell carcinoma of the tongue[J]. Clin Cancer Res, 2001,7(8):2458-2462.

[21] 邵成伟，王培军. 肝癌细胞凋亡的研究[J]. 中国医学影像技术，2000,16(4):326-329.

[22] 秦三海，刘华钢，王博龙，等. 氯化两面针碱体外诱导肺癌 SPC-A-1、Tea8113 两种肿瘤细胞株凋亡的研究[J]. 中国药理学通报，2007,23(2):279-280.

[23] 刘华钢，秦三海，王博龙，等. 氯化两面针碱体外诱导两种鼻咽癌株的细胞凋亡[J]. 华西药学杂志，2007, 22(5):514-516.

[24] Nicholson D W,Alia,Thornberry N A,et al. Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis [J]. Nature,1995,376(6535):37-43.

[25] 邹夏慧. 中药抗肿瘤作用的分子机制研究进展[J]. 国外医学中医中药册，2004,26(6):331-332.

[26] 肖希斌，谢兆霞，秦群，等. MDR1 基因短发卡样 RNA 表达载体逆转 K562/A02 白血病细胞多药耐药的研究[J]. 中华肿瘤杂志，2006,28(6):425-426.

[27] Xiao X S,A Ntony S,Kohlhageng,et al. Design, synthesis and biological evaluation of cytotoxic 11-aminoalkenyl-inde-noisoquino line and 11-diam inoalkenylindeno isoquino line topoi-somerase I inhibitors[J]. Bioorg Med Chem,2004,12(19):5147-5160.

[28] 程轩轩，王冬梅，杨得坡，等. 异喹啉类生物碱的生物活性和构效关系研究进[J]. 中草药，2006,37(12): 1903-1904.

[29] 刘屏，陈凯先. 我国天然药物研究的现状与未来[J]. 中国药物应用与监测，2007,4(3):1-3.

（责任编辑：韦廷宗）

（上接第 186 页）

[23] Gordon T G,Hayward H. Initial experiments with the cross-impact matrix method of forecasting [J]. Futures,1968,1:100-116.

[24] Gordon T G. Cross-impact matrices an illustration of their use for policy analysis[J]. Futures, 1969, 2: 527-531.

[25] Enzer S. Delphi and cross-impact techniques:an effective combination for systematic futures analysis [J]. Futures,1970,3:48-61.

[26] Enzer S. Cross-impact techniques in technology assessment[J]. Futures,1970,4:30-51.

[27] Sage AP. Methodology for Large-Scale systems [M]. New York:McGraw-Hill,1977:165-203.

[28] Wu G. Application of cross-impact analysis to the relationship between aldehyde dehydrogenase 2 and flushing[J]. Alcohol Alcohol,2000,35:55-59.

[29] Bayes T. An essay towards solving a problem in the doctrine of chances. By the late Rev Mr Bayes, F R S communicated by Mr Price, in a letter to John Canton, A M F R S Giving some account of the present undertakings studies and labours of the ingenious in many considerable parts of the world[J]. Philos Trans Roy Soc London, 1763, 53: 370-418.

[30] Silver R M,Warren J E. Preconception counseling for women with thrombophilia[J]. Clin Obstet Gynecol, 2006,49:906-919.

[31] Varga E. Inherited thrombophilia:key points for genetic counseling[J]. J Genet Couns, 2007, 16: 261-277.

（责任编辑：尹 闯）