

人囊性纤维化跨膜电导调节子蛋白的变异模式研究*

Mutation Patterns in Human Cystic Fibrosis Transmembrane Conductance Regulator Protein

龙思宇,严少敏,吴光**

LONG Si-yu, YAN Shao-min, WU Guang

(广西科学院,非粮生物质酶解技术国家重点实验室,国家非粮生物质能源工程技术研究中心,广西生物质产业化工程院,广西生物炼制重点实验室,广西南宁 530007)

(State Key Laboratory of Non-food Biomass Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Biomass Industrialization Engineering Institute, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, Nanning, Guangxi, 530007, China)

摘要:【目的】编码囊性纤维化跨膜电导调节子(Cystic fibrosis transmembrane conductance regulator, CFTR)蛋白的基因突变可引起囊性纤维化,但该蛋白错义点突变的变异模式尚无报道。【方法】先用氨基酸对可预测性为指标将人CFTR蛋白及其178个错义点突变的氨基酸序列转换成标量序列,然后分析变异前后被替换掉的和替换出的氨基酸对的变化。【结果】97.19%的变异发生在不可预测的氨基酸对;87.08%的变异涉及1个或2个被替换掉的氨基酸对,其实际频率大于预测频率;15.17%的变异带来1个或2个替换出的氨基酸对,它们在正常的CFTR蛋白是不存在的;共有122个变异导致替换出的氨基酸对的实际频率小于预测频率。【结论】不可预测的氨基酸对对变异更敏感,变异的趋势是缩小氨基酸对实际频率和预测频率之间的差距,使氨基酸对的构成更加随机化,而人CFTR蛋白的这种退行性变导致了囊性纤维化。

关键词:氨基酸对 囊性纤维化 囊性纤维化跨膜电导调节子 变异模式

中图分类号:Q75, R394.3 **文献标识码:**A **文章编号:**1005-9164(2014)06-0671-06

Abstract:【Objective】Mutations in the gene that encodes the fibrosis transmembrane conductance regulator (CFTR) protein can cause cystic fibrosis but the patterns of missense mutations in CFTR protein haven't yet been reported.【Methods】The amino-acid pair predictability was used to convert the CFTR protein with its 178 missense mutations into scalar sequences, and then the substituted and substituting amino-acid pairs were analysed before and after mutation.【Results】97.19% of mutations occurred in unpredictable amino acid pairs. 87.08% of mutations targeted one or both substituted amino acid pairs whose actual frequency was larger than their predicted one. 15.17% mutations resulted in one or both substituting amino-acid pairs that

were absent before mutation. A total of 122 mutations brought about the substituting amino-acid pairs with their actual frequency smaller than predicted one.【Conclusion】The unpredictable amino-acid pairs are more sensitive to mutation, and the mutation trend is to narrow the difference between predicted and actual frequency of amino-acid pairs, thus the composition of amino acid pairs becomes more randomized, which

收稿日期:2014-08-10

修回日期:2014-09-09

作者简介:龙思宇(1984-),女,硕士,工程师,主要从事信息生物研究。

* 广西自然科学基金重点项目(2013GXNSFDA019007)和广西人才小高地建设专项基金项目资助。

** 通讯作者:吴光(1956-),男,博士,研究员,主要从事生物信息学、计算生物学和模型研究, E-mail: hongguanglishibahao@yahoo.com

leads the human CFTR protein degenerative and causes cystic fibrosis.

Key words: amino-acid pair, cystic fibrosis, cystic fibrosis transmembrane conductance regulator (CFTR), mutation pattern

【研究意义】囊性纤维化跨膜电导调节子(Cystic fibrosis transmembrane conductance regulator, CFTR)蛋白是一种氯离子和碳酸氢离子通道,它能够调节碳酸氢离子的分泌并使其分布在上皮细胞顶膜上,同时抑制氯离子通道的活性,在上皮细胞膜上离子的正常运输和水合作用起重要作用^[1]。作为蛋白信号传输网络中的一个节点,CFTR 不仅调控多种运输蛋白质,受控于多种激酶,同时还参与许多信号传输通路^[2]。编码囊性纤维化跨膜电导调节子蛋白的基因突变会导致囊性纤维化,这是一种在各种器官中常见的外分泌腺对分泌物进行轻微受损清除功能的失调现象^[3-6]。**【前人研究进展】**通过不同的方法在蛋白质水平上分析突变模式,最简单的方法是直接分析氨基酸存在的差异。例如,G551D 突变在囊性纤维化患者群体中相当常见,突变产生一个 CFTR 蛋白,该蛋白位于细胞质膜上,但未能开放以应对细胞信号^[7]。虽然该报道为 CFTR 蛋白变异模式分析提供了一定的参考,但却很难从中找到相关数据特征,而这些数据能通过常用的数学推导方法获得。标志性氨基酸序列既不是矢量数据也不是标量数据,然而数学分析方法使用的大多数数据模式都属于这两种数据范围。这就意味着我们需要将标志性蛋白质序列转变成标量性的蛋白质序列,进而系统地完整地分析蛋白质的变异模式。有多种方法可以将标志性蛋白质序列转变成标量数据,而最具影响力的一种是利用氨基酸的物理化学性质去取代蛋白质链上的每一个氨基酸,如分子量、熔点和旋光度^[8]。另一方面,我们的研究团队根据随机性原理建立了计算变异学,构建了 3 种将标志蛋白质序列转换为标量蛋白质序列的方法^[9-13]。**【本研究切入点】**CFTR 蛋白突变导致了不同的临床结果,也为我们提供了一个可以分析突变模式的模型,可以更好地了解随之发生的疾病的病理特征和临床表现机制。本研究采用计算变异学方法量化 CFTR 蛋白及其变异,系统地完整地分析 CFTR 蛋白的变异模式。**【拟解决的关键问题】**从随机性视角揭示蛋白质变异进化的机理。

1 材料与方法

1.1 数据

人囊性纤维化跨膜电导调节子(CFTR)蛋白的氨基酸序列(ID:P13569)和它们的 185 个自然变异

的相关数据来自 UniProtKB/Swiss-Prot 数据库(<http://www.uniprot.org/uniprot/P13569>),其中包括 5 个缺失、2 个插入和 178 个错义位点。本研究分析这 178 个错义点突变的变异模式。

1.2 转化标志性 CFTR 蛋白质序列为标量数据

组成蛋白质的天然氨基酸有 20 种。虽然可以利用氨基酸的理化性质来替换这 20 种氨基酸,但是被替换出来的 20 个数据不能表示蛋白质突变、序列长度、蛋白质组成、相邻氨基酸和氨基酸位点等信息的变化。因此,这一转化方法不适用于变异模式研究。我们对人 CFTR 蛋白氨基酸对进行数学排列,以此来确定一个氨基酸对的可预测性是否可以依据它在蛋白质上出现的频率来决定^[9-15]。人 CFTR 蛋白由 1480 个氨基酸组成。第 1 位和第 2 位氨基酸被认定为一个氨基酸对,第 2 位和第 3 位氨基酸为另一个氨基酸对,接下来是第 3 位和第 4 位,以此类推,直到第 1479 位和 1480 位,总共有 1479 个氨基酸对。举例说明,人 CFTR 蛋白中有 123 个丝氨酸“S”和 183 个亮氨酸“L”,那么氨基酸对 LS 的预测频率是 $15 (123/1480 \times 183/1479 \times 1479 = 15.209)$ 。而事实上 CFTR 蛋白质中恰巧有 15 个 LS 氨基酸对,所以利用数学排列的方法分类,氨基酸对 LS 是可预测的。与之相比,人 CFTR 蛋白中有 93 个谷氨酸“E”和 84 个甘氨酸“G”,氨基酸对 EG 的预测频率是 $5 (93/1480 \times 84/1479 \times 1479 = 5.278)$,也就是说 CFTR 蛋白质中应该有 5 个 EG 氨基酸对。但是氨基酸对 EG 实际出现了 12 次,因此利用数学排列的方法分类,氨基酸对 EG 是不可预测的。

1.3 可预测和不可预测氨基酸对的变异情况

一个点突变将两个氨基酸对替换为另外两个氨基酸对。例如,位点 336 上发生突变,异亮氨酸“I”突变成了赖氨酸“K”,两个氨基酸对 KI 和 IF 变成了 KK 和 KF,因为赖氨酸“K”位于位点 335 上而苯丙氨酸“F”位于 337 上。表 1 列出了这些氨基酸对的实际频率和预测频率,我们可以确定被替换掉的氨基酸对(KI 和 IF)和被替换出的氨基酸对(KK 和 KF)是可预测的还是不可预测的,并通过网络服务的方式进行运算(<http://www.nerc-nfb.ac.cn/calculation/fc.htm>),分析所有 CFTR 蛋白氨基酸对的突变情况。

1.4 预测频率和实际频率的差值

通过数据分析,我们算出受影响氨基酸对预测频

率(PF)和实际频率(AF)的差值,即 $\Sigma(PF- AF)$ 。如表1所示,突变前,被替换掉的氨基酸对的预测频率和实际频率的差值是 $(7-9)+(7-13)=-8$,而被替换出的氨基酸对是 $(6-5)+(5-4)=2$ 。突变后,则分别为 $(7-8)+(7-12)=-6$ 和 $(6-6)+(5-5)=0$ 。通过以上分析我们就可以比较突变对氨基酸对频率差值的影响。

表1 人CFTR蛋白I336K变异体变异前后对氨基酸对的影响

Table 1 I336K mutant and its effect on amino-acid pairs before and after mutation in human CFTR protein

分类 Classification	被替换掉的氨基酸对 Substituted pairs				被替换出的氨基酸对 Substituting pairs			
	KI		IF		KK		KF	
	PF	AF	PF	AF	PF	AF	PF	AF
变异前 Before mutation	7	9	7	13	6	5	5	4
变异后 After mutation	7	8	7	12	6	6	5	5

I: 异亮氨酸, K: 赖氨酸, F: 苯丙氨酸, PF: 预测频率, AF: 实际频率。
I: isoleucine, K: lysine, F: phenylalanine, PF: predicted frequency, AF: actual frequency.

1.5 统计分析

卡方检验(Chi-square test)用于比较可预测和不可预测的氨基酸对种类/氨基酸对的突变发生率, Kruskal-Wallis 单因素方差分析法(Kruskal-Wallis One Way Analysis of Variance on Ranks)用于比较氨基酸对不同组之间的差值, 同时邓恩检测法(Dunn's Method)用于氨基酸对的成对多重比较。当 $P < 0.05$ 时, 表示差异显著。

2 结果与分析

2.1 人CFTR蛋白中的氨基酸对

理论上, 20种自然氨基酸可以组成400种可能的氨基酸对(20^2)。通过上面的分析可知, 人CFTR蛋白有1479个氨基酸对, 这个数值远大于400, 说明其中的一些氨基酸对出现了多次。而且这400种理

表3 人CFTR蛋白变异发生在可预测和不可预测氨基酸对的比较

Table 3 Comparison of mutations at predictable/unpredictable kinds and pairs in human CFTR protein

分类 Classification	种类 Kinds		对数 Pairs		变异 Mutations		比率 Ratio	
	number	%	number	%	number	%	Mutations/Kinds	Mutations/Pairs
可预测的 Predictable	83	24.2	247	16.7	5	2.8	5/83=0.06	12/247=0.02
不可预测的 Unpredictable	260	75.8	1232	83.3	173	97.2	173/260=0.67	173/1232=0.14
总合 Total	343	100	1479	100	178	100	178/343=0.52	178/1479=0.12

注: 卡方检验表明可预测的和不可预测的氨基酸对变异的发生有显著的统计差异($P \leq 0.001$)。

Note: The Chi-square test indicates the highly statistical significance of occurrence of mutations between predictable and unpredictable kinds/pairs ($P \leq 0.001$).

论氨基酸对并不都存在于人CFTR蛋白质中。有57种理论氨基酸对不存在于人CFTR蛋白质中, 因此人CFTR蛋白中1479个氨基酸对只有343种氨基酸对种类($400-57=343$), 其中一些氨基酸对出现了多次(表2)。在实际运用上, 可预测/不可预测氨基酸对种类和氨基酸对的分类显得十分必要, 人CFTR蛋白的343种氨基酸对种类中, 83种是可预测的, 260种是不可预测的。人CFTR蛋白的1479个氨基酸对中, 247个是可预测的, 1232个是不可预测的。表3显示了相关的变异模式, 只有5种突变发生在可预测的氨基酸对中。

表2 氨基酸对种类及其在人CFTR蛋白的出现频率

Table 2 Kinds of amino-acid pairs and their appearance in human CFTR protein

出现频率 Appearance	氨基酸对种类 Kinds of amino-acid pairs	出现频率 Appearance	氨基酸对种类 Kinds of amino-acid pairs
1	33	10	9
2	41	11	7
3	45	12	4
4	40	13	5
5	19	14	1
6	16	15	1
7	16	16	1
8	13	18	1
9	8		

2.2 被变异替换掉的氨基酸对

如果一个氨基酸对因突变直接被替换掉, 并且在突变前只在蛋白质中出现一次, 那么这种氨基酸对将在突变后消失。但是如果一种氨基酸对在突变前出现多次, 那么这种氨基酸对在突变后仍然存在。一个位点突变一般只涉及到两个氨基酸对, 这就确保了其他的同一种氨基酸对在突变后仍然保留下来。表4列出了因突变而被替换掉的氨基酸对变异前后的分组情况, 前3列是根据可预测/不可预测和实际频率/预测频率将被替换掉的氨基酸对进行分组, 变异前后的各3列数值表示分组后的氨基酸对。通过比较变

异前后的出现频率得出结论,变异的目的是为了降低那些实际频率大于预测频率的氨基酸对的出现频率,例如,两个氨基酸对都属于此种类型的出现频率从74降低到31(表4第3行),同样的情况也出现在变异前的第4和第5行,87.07%的氨基酸对属于一个或两个被替换掉的氨基酸对,其实际频率大于预测频率,说明该突变的效果是通过降低实际频率来缩小实际频率和预测频率之间的差距来实现的。两个氨基酸对中仅有10种突变的实际频率小于预测频率,说明突变如果通过增加实际频率来缩小实际频率和预测频率之间的差距是比较难实现的。

2.3 被变异替换出的氨基酸对

表5中的结构与表4非常相似。例如,15.17%的变异带来了一个或两个氨基酸对,它们不存在于正

表4 人CFTR蛋白被替换掉的氨基酸对变异前后的比较

Table 4 Comparison of substituted amino-acid pairs before and after mutation in human CFTR protein

分类 Classification	频率 Frequency		变异前 Before mutation			变异后 After mutation		
	I	II	Appearance	%	total%	Appearance	%	Total%
可预测 Predictable	AF=PF	AF=PF	5	2.81	2.81	10	5.62	5.62
不可预测 Unpredictable	AF>PF	AF>PF	74	41.57	97.19	31	17.42	94.38
	AF>PF	AF=PF	34	19.10		35	19.66	
	AF>PF	AF<PF	47	26.40		51	28.65	
	AF<PF	AF=PF	8	4.49		30	16.85	
	AF<PF	AF<PF	10	5.62		21	11.80	

PF:预测频率,AF:实际频率。AF:Actual frequency,PF:Predicted frequency.

表5 人CFTR蛋白被替换出的氨基酸对变异前后的比较

Table 5 Comparison of substituting amino-acid pairs before and after mutation in human CFTR protein

频率 Frequency		变异前 Before mutation			变异后 After mutation		
I	II	Appearance	%	Total%	Appearance	%	Total%
AF=0,PF>0	AF=0,PF>0	2*	1.12	15.17	0	0	0.00
AF=0,PF>0	AF=PF=0	0*	0.00		0	0	
AF=0,PF>0	AF=PF>0	8*	4.49		0	0	
AF=0,PF>0	AF<PF,AF≠0	5*	2.81		0	0	
AF=0,PF>0	AF>PF	9*	5.06		0	0	
AF=PF=0	AF=PF=0	0	0.00		0	0	
AF=PF=0	AF=PF>0	2	1.12		0	0	
AF=PF=0	AF<PF,AF≠0	0*	0.00		0	0	
AF=PF=0	AF>PF	1	0.56		0	0	
AF<PF,AF≠0	AF<PF,AF≠0	34*	19.10	84.83	14	7.87	100.00
AF<PF,AF≠0	AF=PF>0	25*	14.04		15	8.43	
AF<PF,AF≠0	AF>PF	39*	21.91		39	21.91	
AF=PF>0	AF=PF>0	4	2.25		14	7.87	
AF=PF>0	AF>PF	23	12.92		51	28.65	
AF>PF	AF>PF	26	14.61		45	25.28	

* 共有122个变异导致了替换出的氨基酸对,它们的实际频率小于预测频率(68.54%)。

* Indicates the substituting amino-acid pairs with their actual frequency smaller than predicted one. The total of these mutations is 122(68.54%).

常的人CFTR蛋白中。另外,总共有122个变异(68.54%)替换出了实际频率小于预测频率的氨基酸对(用*表示)。

2.4 变异导致的氨基酸对的频率差异

图1展示了178个突变对氨基酸对预测频率与实际频率的影响情况,图2则展示了它们的统计学比较结果。变异前,被替换掉的氨基酸对的预测频率与实际频率差值的中值是-3,表明这些突变发生在实际频率大于预测频率的氨基酸对中。同时,被替换出的氨基酸对的相应值是0.5,表明突变导致了氨基酸对的构建更为随机。变异后,被替换掉和被替换出的氨基酸对的实际频率与预测频率差值的中值都是-1。

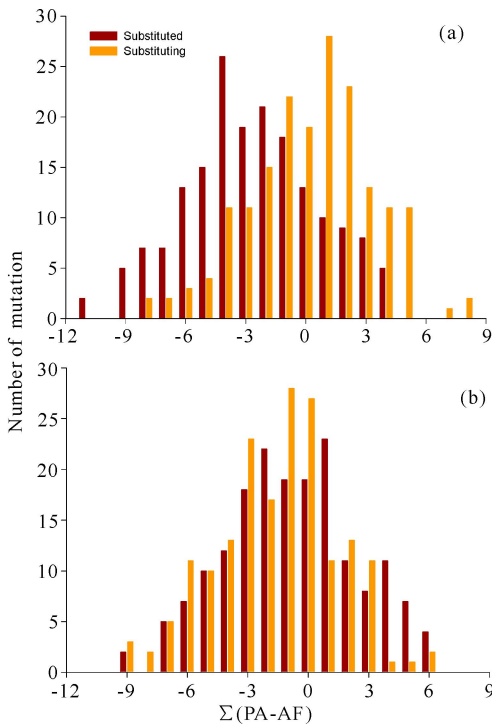


图1 人CFTR蛋白变异前后受影响氨基酸对的频率分布

Fig. 1 Frequency distribution of affected amino-acid pairs before and after mutation in human CFTR protein

(a)突变前;(b)突变后

(a)Before mutation;(b)After mutation

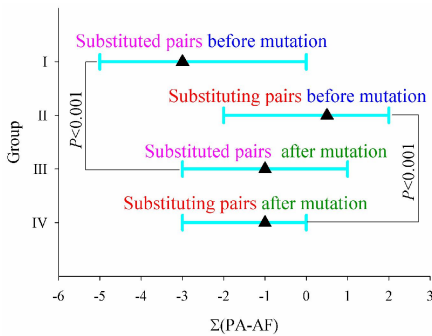


图2 人CFTR蛋白变异前后氨基酸对的频率比较

Fig. 2 Frequency comparison of affected amino-acid pairs before and after mutation in human CFTR protein

图中数据由中值和25%~75%区间值表示。相对应的各组间有显著的统计差异($P \leq 0.001$, Kruskal-Wallis 单因素方差分析法)。

The data are presented as median with interquartile interval. There is statistically significant difference between corresponding groups ($P \leq 0.001$, Kruskal-Wallis One Way Analysis of Variance on Ranks).

3 讨论

随着CFTR基因鉴定工作的发展,超过1500个不同的变异已被识别,依据它们的功能影响效果,临床分为6大类型。许多患有囊性纤维化的个体具有

CFTR蛋白质遗传异质性的特征^[16],然而,只有178个在蛋白质水平上的错义突变被记载。目前有两种观点解释了为什么有些氨基酸的突变比其他氨基酸更频繁。第一种是定向突变,由它定义的“热点”位置对内源及外源性突变剂都敏感^[17~19]。第二种是功能选择,认为蛋白质功能的损坏是由蛋白质突变的位点决定的^[20,21]。但是,这些观点始终不能全面地解释为什么某些氨基酸对突变更为敏感。通过随机理论分析,本研究恰好弥补了这一不足,且这一观点是可信的,不仅因为随机性在自然界中普遍存在^[22],而且随机可预测的氨基酸对体现了最大可能发生概率,同时也符合自然约束下构建氨基酸对所需要的时间和能量。毋庸置疑,蛋白质的功能位点更像是一个刻意进化的结果,因此它们的实际频率本应该不同于预测频率,而这也是我们可以用随机理论来分析氨基酸对的构成却不能用随机理论解释它们的功能的原因。

本研究结果说明,错义突变的趋势是氨基酸对的实际频率在一定程度上接近于它的预测频率。在自然状态下,实际频率不同于预测频率的氨基酸对很可能不适合存在,需要通过蛋白质突变来缩小预测频率与实际频率之间的差距,而这些突变造成的一些结果就是蛋白部分功能的缺失。但是,因突变而产生的氨基酸对可能会产生新的预测频率与实际频率的差距,此为突变提供了新的条件,从而进化可以继续下去。

如果我们可以找到使用标量蛋白序列的切入点,那么运用哪种方法来把氨基酸标志的蛋白序列转化成任意的标量蛋白序列并不重要,重要的是标量蛋白序列必须能够表示突变、蛋白质组成、蛋白质长度、相邻氨基酸、蛋白序列位点等信息,而我们的计算变异学方法则可以实现这些信息的定量化^[9~15],因此我们能够利用氨基酸对的可预测性来分析错义突变的趋势,从而为深入了解囊性纤维化疾病的基本机制提供了更好的研究模式。

4 结论

本研究用氨基酸对可预测性为指标将人CFTR蛋白及其178个错义点突变的氨基酸序列转换成标量序列,分析变异前后被替换掉的和替换出的氨基酸对的频率变化。结果显示不可预测的氨基酸对对变异更敏感,变异的趋势是缩小氨基酸对实际频率和预测频率之间的差距,使氨基酸对的构成更加随机化,而人CFTR蛋白的这种退行性变导致了囊性纤维化。

参考文献:

- [1] Ousingsawat J, Kongsuphol P, Schreiber R, et al. CFTR and TMEM16A are separate but functionally related Cl⁻ channels [J]. *Cellular Physiology and Biochemistry*, 2011, 28(4): 715-724.
- [2] Kunzelmann K, Mehta A. CFTR: A hub for kinases and crosstalk of cAMP and Ca²⁺ [J]. *FEBS Journal*, 2013, 280(18): 4417-4429.
- [3] Paccou J, Fardellone P, Cortet B. Cystic fibrosis-related bone disease [J]. *Current Opinion in Pulmonary Medicine*, 2013, 19(6): 681-686.
- [4] Chaaban M R, Kejner A, Rowe S M, et al. Cystic fibrosis chronic rhinosinusitis: A comprehensive review [J]. *Allergy*, 2013, 27(5): 387-395.
- [5] Galiotta L J. Managing the underlying cause of cystic fibrosis: A future role for potentiators and correctors[J]. *Paediatric Drugs*, 2013, 15(5): 393-402.
- [6] Wilschanski M, Novak I. The cystic fibrosis of exocrine pancreas[J]. *Cold Spring Harbor Perspectives in Medicine*, 2013, 3(5): a009746. doi: 10.1101/cshperspect.a009746.
- [7] Kotha K, Clancy J P. Ivacaftor treatment of cystic fibrosis patients with the G551D mutation: A review of the evidence[J]. *Therapeutic Advances in Respiratory Disease*, 2013, 7(5): 288-296.
- [8] Fasman G D. *Handbook of Biochemistry; Section D Physical Chemical Data* [M]. 3rd Edition. London: CRC Press, 1976.
- [9] Wu G, Yan S M. Randomness in the primary structure of protein: Methods and implications[J]. *Mol Biol Today*, 2002, 3: 55-69.
- [10] Wu G, Yan S. Fate of influenza A virus proteins[J]. *Protein and Peptide Letters*, 2006, 13(4): 377-384.
- [11] Wu G, Yan S. Mutation trend of hemagglutinin of influenza A virus: A review from computational mutation viewpoint[J]. *Acta Pharmacologica Sinica*, 2006, 27(5): 513-526.
- [12] Wu G, Yan S. *Lecture Notes on Computational Mutation*[M]. New York: Nova Science Publisher, 2008.
- [13] Yan S, Wu G. Creation and application of computational mutation[J]. *Journal of Guangxi Academy of Sciences*, 2010, 26(2): 130-139.
- [14] Yan S, Wu G. Determination of mutation pattern in human androgen receptor by means of amino-acid pair predictability [J]. *Protein and Peptide Letters*, 2009, 16(3): 289-296.
- [15] Yan S, Wu G. Mutation patterns in human galactosidase A [J]. *Molecular Diversity*, 2010, 14(1): 147-154.
- [16] Zaibak F, Williamson R. Genetics and genomics: Providing new opportunities for the translation of basic research to clinical care for cystic fibrosis[J]. *Expert Review of Respiratory Medicine*, 2008, 2(1): 1-5.
- [17] Schaefer E, Mehta A, Gal A. Genotype and phenotype in Fabry disease: Analysis of the Fabry Outcome Survey[J]. *Acta Paediatrica Supplement*, 2005, 94(447): 87-92.
- [18] Funnell A P, Wilson M D, Ballester B, et al. A CpG mutational hotspot in a ONECUT binding site accounts for the prevalent variant of hemophilia B Leyden[J]. *American Journal of Human Genetics*, 2013, 92(3): 460-467.
- [19] Damas J, Samuels D C, Carneiro J, et al. Mitochondrial DNA rearrangements in health and disease—a comprehensive study[J]. *Human Mutation*, 2014, 35(1): 1-14.
- [20] Dzikiewicz-Krawczyk A. The importance of making ends meet: Mutations in genes and altered expression of proteins of the MRN complex and cancer[J]. *Mutation Research*, 2008, 659(3): 262-273.
- [21] Cheng H, Rashid S, Yu Z, et al. Location of glycine mutations within a bacterial collagen protein affects degree of disruption of triple-helix folding and conformation [J]. *Journal of Biological Chemistry*, 2011, 286(3): 2041-2046.
- [22] Everitt B S. *Chance rules: An Informal Guide to Probability, Risk, and Statistics* [M]. New York: Springer, 1999.

(责任编辑:陆雁)