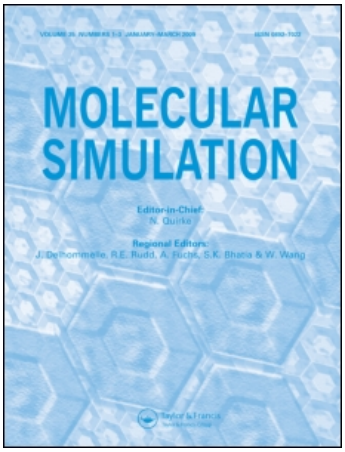

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Determination of Amino Acid Pairs Sensitive to Variants in Human Bruton's Tyrosine Kinase by means of a Random Approach

Guang Wu[a]; Shaomin Yan[b]
[a] Laboratoire de Toxicocinétique et Pharmacocinétique, Faculté de Pharmacie, Université de la Méditerranée Aix-Marseille II, Marseille, France [b] Cattedra di Anatomia Patologica, Dipartimento di Ricerche Mediche e Morfologiche, Facoltà di Medicina e Chirurgia, Università degli Studi di Udine, Udine, Italy

Online publication date: 26 October 2010

To cite this Article Wu, Guang and Yan, Shaomin(2003) 'Determination of Amino Acid Pairs Sensitive to Variants in Human Bruton's Tyrosine Kinase by means of a Random Approach', Molecular Simulation, 29: 4, 249 — 254
To link to this Article: DOI: 10.1080/0892702031000101321
URL: http://dx.doi.org/10.1080/0892702031000101321

Molecular Simulation, 2003 Vol. 29 (4), pp. 249–254

Determination of Amino Acid Pairs Sensitive to Variants in Human Bruton's Tyrosine Kinase by means of a Random Approach

GUANG WU[a,*] and SHAOMIN YAN[b]

[a]*Laboratoire de Toxicocinétique et Pharmacocinétique, Faculté de Pharmacie, Université de la Méditerranée Aix-Marseille II, Marseille, France;*
[b]*Cattedra di Anatomia Patologica, Dipartimento di Ricerche Mediche e Morfologiche, Facoltà di Medicina e Chirurgia, Università degli Studi di Udine, Udine, Italy*

(Received July 2002; In final form July 2002)

In this data-based theoretical analysis, we use a random approach to analyse amino acid pairs in human Bruton's tyrosine kinase (BTK) in order to determine which amino acid pairs are more sensitive to 112 variants with missense mutant in human BTK. The rationale of this study is based on our hypothesis and previous findings that variance is more likely to occur at randomly unpredictable amino acid pair position rather than at randomly predictable positions. This is reasonable to argue, as randomly predictable amino acid pairs are less likely to be deliberately evolved, whereas randomly unpredictable amino acid pairs are probably deliberately evolved in connection with protein function. A 91.96% of 112 variants occurred at randomly unpredictable amino acid pairs, which account for 72.49% of amino acid pairs in BTK, and the chance of a variant occurring is 3.4-fold higher in randomly unpredictable amino acid pairs than in predictable ones. Thus randomly unpredictable amino acid pairs are more sensitive to variance in human BTK. The results also suggest that the human BTK has a natural tendency towards variants.

Keywords: Agammaglobulinemia; Bruton's tyrosine kinase; Probability; Randomness

INTRODUCTION

The frequency of amino acid pairs has been introduced to predict protein secondary structure content [1,2]. In the past, we have used two probabilistic approaches to analyse the primary structure of proteins related to different diseases (for review see Ref. [3]). In general, our first approach can predict the presence and absence of amino acid sub-sequences in a protein primary structure. We argue that the randomly predictable present and absent sub-sequences were probably not deliberately evolved, whereas the randomly unpredictable present and absent sub-sequences were more likely to be deliberately evolved. Accordingly, our first approach can classify the present amino acid sub-sequences as randomly predictable and randomly unpredictable sub-sequences. We suggest that the randomly unpredictable amino acid sub-sequences are more related with protein function and the variants are more likely to occur at randomly unpredictable amino acid sub-sequences rather than at randomly predictable amino acid sub-sequences.

Recently, we used our approach to analyse the amino acid pairs in human haemoglobin α chain [4] and phenylalanine hydroxylase protein [5] to determine which amino acid sub-sequence is more sensitive to variance. The results show that randomly unpredictable amino acid pairs are more sensitive to variance. An intriguing question is brought about whether these phenomena are occasional or they represent some general sense. Thus, further studies are needed in order to obtain more information regarding this aspect.

Bruton's tyrosine kinase (BTK) plays a crucial role in signal transduction pathways [6,7]. Mutations in the gene encoding BTK are the cause of X-linked agammaglobulinemia (XLA), which is the prototype immunodeficiency disease and specifically affects the B-lineage [8–10].

*Corresponding author. E-mail: hongguanglishibahao@yahoo.com

ISSN 0892-7022 print/ISSN 1029-0435 online © 2003 Taylor & Francis Ltd
DOI: 10.1080/0892702031000101321

The majority of these patients have normal levels of pre-B cells in their bone marrow but virtually no circulating mature B-lymphocytes. This results in a lack of immunoglobulins of all classes and leads to recurrent bacterial infections [11–13]. So far, 115 variants in human BTK are documented in the Swiss-Protein data bank [14]. Of all variants, 112 belong to missense point mutations, two are small deletions of one amino acid and the last one is insertion of several amino acids.

In this study, we attempt to use a random approach to analyse amino acid pairs in human BTK with its 112 variants with missense mutants in order to determine which amino acid pairs are more sensitive to the variants.

MATERIALS AND METHODS

The amino acid sequence of the human BTK and its 112 variants with missense point mutants was obtained from the Swiss-Protein data bank [14] (access number Q06187, due to the limitation of space, we will not cite the numerous references related to human BTK). The detailed calculations and rationales have already been published in a number of our previous studies (for a review, see Ref. [3]). Briefly, the calculation procedure with its examples is as follows.

Amino Acid Pairs in Human BTK

The human BTK consists of 659 amino acids. The first and second amino acids are counted as an amino acid pair, the second and third as another amino acid pair, the third and fourth, until the 658th and 659th, thus, there are a total of 658 amino acid pairs. In general, there are 20 types of amino acids, any amino acid pair can be composed from any of 20 types of amino acids so, theoretically, there are 400 (20^2) possible amino acid pairs. Again there are 658 amino acid pairs in human BTK, which are more than 400 types of theoretical amino acid pairs, clearly some of 400 types of theoretical amino acid pairs should appear more than once. Meanwhile we may expect that some of 400 types of theoretical amino acid pairs are absent from human BTK.

Actual Frequency and Randomly Predicted Frequency in Human BTK

The randomly predicted frequency is calculated according to a simple permutation principle [15]. For example, there are 34 arginines (R) and 57 leucines (L) in human BTK, the random frequency of amino acid pair "RL" would be 3 ($34/659 \times 57/658 \times 658 = 2.941$). Actually we can find 3 "RL"s in human BTK, so the actual frequency of "RL" is 3. Hence we have 3 relationships between the actual and predicted frequencies, i.e. the actual frequency is smaller, equal to and larger than the predicted frequency, respectively.

Randomly Predictable Present Amino Acid Pairs

As described in the last section, the predicted frequency of randomly present amino acid pair "RL" would be 3 and "RL" does appear 3 times in human BTK, so the presence of "RL" is randomly predictable.

Randomly Unpredictable Present Amino Acid Pairs

There are 36 glutamines (G) in human BTK, the frequency of random presence of amino acid pair "RG" would be 2 ($34/659 \times 36/658 \times 658 = 1.857$), i.e. there would be 2 "RG"s in human BTK. But in fact the "RG" appears 4 times in human BTK, so the presence of "RG" is randomly unpredictable. In this case the actual frequency of "RG" is larger than the predicted frequency of "RG". In other case the actual frequency is smaller than the predicted frequency. For example, there are 58 glutamic acids (E) in human BTK and the randomly predicted frequency of "RE" is 3 ($34/659 \times 58/658 \times 658 = 2.992$), while the actual frequency is only 1.

Randomly Predictable Absent Amino Acid Pairs

There are 29 alanines (A) and 11 tryptophans (W) in human BTK, the frequency of random presence of "AW" would be 0 ($29/659 \times 11/658 \times 658 = 0.484$), i.e. the amino acid pair "AW" would not appear in human BTK, which is true in the real situation. Thus the absence of "AW" is randomly predictable.

Randomly Unpredictable Absent Amino Acid Pairs

There are 40 valines (V) in human BTK, the frequency of random presence of "RV" would be 2 ($34/659 \times 40/658 \times 658 = 2.064$), i.e. there would be 2 "RV"s in human BTK. However there is no "RV" in human BTK, therefore the absence of "RV" from human BTK is randomly unpredictable.

Variants in Randomly Predictable and Unpredictable Amino Acid Pairs

A variant with point mutation results in two amino acid pairs being substituted by another two pairs. After calculating the predicted frequency and comparing with the actual frequency, it can be determined that the substituted amino acid pairs

TABLE I Appearance of theoretical types of amino acid pairs in human BTK

Appearance	*Number of theoretical types of amino acid pairs*
0	107
1	111
2	87
3	46
4	25
5	14
6	6
7	3
8	1

belong to predictable/unpredictable amino acid pairs.

Difference Between Actual and Randomly Predicted Frequencies

For the numerical analysis, we calculate the difference between actual frequency (AF) and predicted frequency (PF) of affected amino acid pairs, i.e. $\Sigma(AF - PF)$. For instance, a variant at position 307 substitutes "R" to "T" which results in two amino acid pairs "VR" and "RD" changing to "VT" and "TD", because the amino acid is "V" at position 306 and "D" at position 308. The actual frequency and predicted frequency are 2 and 2 for "VR", 3 and 1 for "RD", 1 and 2 for "VT", and 0 and 1 for "TD", respectively. Thus, the difference between actual frequency and predicted frequency is 2 with regard to the substituted amino acid pairs, i.e. $(2 - 2) + (3 - 1) = 2$, and -2 for the substituting amino acid pairs, i.e. $(1 - 2) + (0 - 1) = -2$. In this way, we can compare the frequency difference in the amino acid pairs affected by variants.

RESULTS

General Information on Amino Acid Pairs and Variants in Human BTK

Of 400 types of theoretical amino acid pairs, 107 are absent from human BTK including 20 randomly predictable and 87 randomly unpredictable. Consequently, 658 amino acid pairs in human BTK include only 293 types of theoretical amino acid pairs ($400 - 107 = 293$), i.e. some amino acid pairs should appear more than once (Table I).

Of 293 types of theoretical amino acid pairs in human BTK, 107 types are randomly predictable and 186 are randomly unpredictable. As mentioned above, some types of amino acid pairs appear more than once, thus, of 658 amino acid pairs in human BTK, 181 pairs are randomly predictable and 477 pairs are randomly unpredictable. Therefore, the number of variants occurring with respect to these amino acid pairs in human BTK can be detected by probability (Table II).

Variants of Human BTK in Randomly Predictable and Unpredictable Present Amino Acid Pairs

As mentioned in "Materials and methods" section, a point mutant protein leads to two amino acid pairs being substituted by another two and their actual frequency can be smaller, equal to or larger than the predicable frequency. Tables III and IV detail the situations related to substituted and substituting amino acid pairs, respectively and the relationship between their actual and predicted frequencies.

Table III can be read as follows. The first column classifies the substituted amino acid pairs into randomly predictable and unpredictable. The second and third columns show in which type of amino acid pairs the variant occurs, for example, the first two cells in columns 2 and 3 indicate that the actual frequencies are equal to the predicated frequencies in both amino acid pairs I and II. The fourth and fifth columns indicate how many variants occur in amino acid pairs I and II, for instance, 9 variants (8.04%) occur at both amino acid pairs whose actual frequencies are equal to predicted frequencies. The sixth column indicates the percentage of 112 variants occurring at predictable and unpredictable amino acids.

Tables II and III show that 91.96% of variants occur at randomly unpredictable amino acid pairs and only 8.04% of variants occur in randomly predictable amino acid pairs. These results mean that 186 types of randomly unpredictable present amino acid pairs account for 91.96% variants in human BTK, whereas 107 types of randomly predictable present amino acid pairs account for only 8.04%. The phenomena

TABLE II Occurrence of variants with respect to randomly predictable and unpredictable amino acid pairs in human BTK

	Types		*Pairs*		*Variants*		*Ratio*	
Amino acids	*Number*	*(%)*	*Number*	*(%)*	*Number*	*(%)*	*Variants/Types*	*Variants/Pairs*
Predictable	107	36.52	181	27.51	9	8.04	$9/107 = 0.08$	$9/181 = 0.05$
Unpredictable	186	63.48	477	72.49	103	91.96	$103/186 = 0.55$	$103/477 = 0.22$
Total	293	100	658	100	112	100	$112/293 = 0.38$	$112/658 = 0.17$

TABLE III Classification of substituted amino acid pairs induced by variants in human BTK

	Amino acid pairs		*Variants*		
BTK	*I*	*II*	*Number*	*(%)*	*Total (%)*
Predictable	AF = PF	AF = PF	9	8.04	8.04
Unpredictable	AF > PF	AF > PF	24	21.43	91.96
	AF > PF	AF = PF	49	43.75	
	AF > PF	AF < PF	16	14.29	
	AF < PF	AF = PF	11	9.82	
	AF < PF	AF < PF	3	2.68	

AF, actual frequency; PF, predicted frequency.

support our rationale that the variants are more likely to occur at randomly unpredictable amino acid pair positions rather than at randomly predictable. Thus, the randomly unpredictable amino acid pair positions are more sensitive to the variants.

When looking at the unpredictable amino acid pairs in Table III, 79.46% of these pairs are characterised by one or both substituted pairs whose actual frequency is larger than their predicted frequency (the first three rows in unpredictable pairs). Comparing with the normal human BTK, the impact of variants is to narrow the difference between actual and predicted frequencies by means of reducing the actual frequency, which implies that the variants associated with the construction of amino acid pairs is randomly predictable. In other words, the variants result in the construction of amino acid pairs that are more likely to be naturally evolved. Also three variants occur in the amino acid pairs whose actual frequency is smaller than predicted frequency in both pairs. This interesting phenomenon suggests that it is difficult for variants to narrow the difference between actual and predicted frequencies by means of increasing the actual frequency. Commonly, reduction of actual frequency would lead to the construction of amino acid pairs against natural direction.

Table IV 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 occurring at amino acid pairs I and II and their percents, the fifth column shows total classifications.

Table IV shows that 41.96% of variants result in one or both substituting amino acid pairs are absent in normal human BTK (AF = 0). Furthermore 65.18% of variants target one or both substituting amino acid pairs with their actual frequency smaller than predicted frequency (see Table IV footnote). These phenomena indicate that the amino acid pairs in mutant BTK are more randomly constructed.

TABLE IV Classification of substituting amino acid pairs induced by variants in human BTK

Amino acid pairs		*Variants*		*Total*
I	*II*	*Number*	*(%)*	*(%)*
AF = 0, PF > 0	AF = 0, PF > 0	5*	4.46	41.96
AF = 0, PF > 0	AF = PF = 0	0*	0	
AF = 0, PF > 0	AF = PF > 0	8*	7.14	
AF = 0, PF > 0	AF < PF, AF ≠ 0	8*	7.14	
AF = 0, PF > 0	AF > PF	21*	18.75	
AF = PF = 0	AF = PF = 0	0	0	
AF = PF = 0	AF = PF > 0	2	1.79	
AF = PF = 0	AF < PF, AF ≠ 0	0*	0	
AF = PF = 0	AF > PF	3	2.68	
AF < PF, AF ≠ 0	AF < PF, AF ≠ 0	5*	4.46	58.04
AF < PF, AF ≠ 0	AF = PF > 0	17*	15.18	
AF < PF, AF ≠ 0	AF > PF	9*	8.04	
AF = PF > 0	AF = PF > 0	7	6.25	
AF > PF	AF > PF	14	12.50	
AF = PF > 0	AF > PF	13	11.61	

*Indicates the substituting amino acid pairs with their actual frequency smaller than predicted frequency. The total of these amino acid pairs is 73 (65.18%).

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 the positive difference, the more randomly unpredictable amino acid pairs are present; and (ii) the larger the negative difference, the more randomly unpredictable amino acid pairs are absent.

Considering all 112 variants, the difference between actual and predicted frequencies is 1.17 ± 0.15 (mean $\pm$ SE, ranging from -2 to 8) in substituted amino acid pairs. This means that the variants occur in the amino acid pairs which appear more than their predicted frequency. Meanwhile, the difference between actual and predicted frequencies is -0.02 ± 0.19 (mean $\pm$ SE, ranging from -4 to 6) in substituting amino acid pairs. This implies that the substituting amino acid pairs are more randomly constructed in the variants of BTK, as their actual and predicted frequencies are about the same. Striking statistical difference is found between the substituted and substituting amino acid pairs

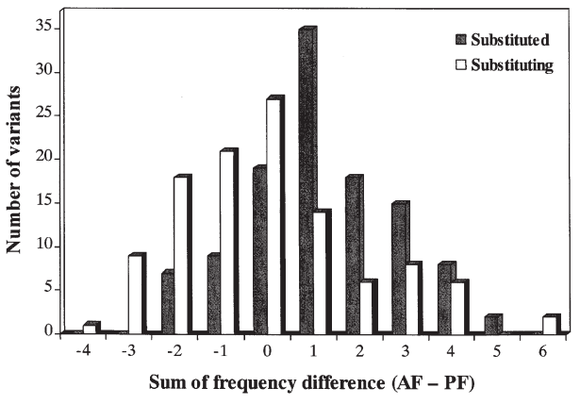

FIGURE 1 Frequency difference between substituted and substituting amino acid pairs induced by variants of human BTK.

($P < 0.0001$). Figure 1 shows the distribution of difference between actual and predicted frequencies.

DISCUSSION

Currently two explanations are commonly proposed to explain why some amino acids are mutated more frequently than the others. The first is targeted mutagenesis, which defined the "hotspot" sites sensitive to endogenous and exogenous mutagens [16–18]. The second is the function selection, which indicates the disruption of protein functions may depend upon the position of the mutation/variant in the protein [19–21]. However these explanations still do not answer why some amino acid sub-sequences are sensitive to variants.

This problem can be assessed from different approaches such as empirical (regression analysis), experimental (artificial and natural mutations) and computation (multiple sequence comparisons and alignments), etc. The probabilistic approach can contribute considerable understanding to this problem. By means of a random approach to estimate the variants from human haemoglobin α chain [4] and phenylalanine hydroxylase protein [5], we find that 94% of variants occur in randomly unpredictable amino acid pairs. In this study we, correspondingly, analyse the amino acid pairs in human BTK to determine which amino acid pairs are more sensitive to variants. The results are quite similar to the previous studies and confirm our hypothesis that the randomly unpredictable amino acid pairs are more sensitive to variants.

Based on our previous studies, our argument is that the functional amino acid pairs are more likely to be deliberately evolved and thus the actual frequency should be different from the randomly predicted frequency. As the randomly predicted frequency is the highest potential for construction of amino acid pairs, it is important to find whether the variant leads to the actual frequency to approach the randomly predicted frequency. If so, the protein has a natural trend to mutate; if not, the protein does not have a natural trend to mutate. The present study demonstrates that 79.46% of variants bring about one or both substituted amino acid pairs whose actual frequency is larger than predicted frequency, that 41.96% of variants result in one or both substituting amino acid pairs which are absent in normal human BTK (AF = 0) and that 65.18% of variants lead to one or both substituting amino acid pairs with their actual frequency smaller than predicted frequency. All of these results reveal that the human BTK has a natural trend to variance.

With respect to randomly unpredictable absent and present amino acid pairs, the difference between actual and predicted frequencies is interesting, because the randomly predicable absent and present frequency represents the more likely naturally occurring event, i.e. the construction of amino acid pairs should be the least energy- and time-consuming. Thus, the difference between actual and predicted frequencies should be engineered by the evolutionary process, i.e. the larger the difference, the larger the impact by the evolutionary process. Diminishing of difference between actual and predicted frequencies has been shown in this study (Fig. 1), thus the variants of human BTK in fact represent a degeneration process inducing XLA.

The current study highlights the changes in the frequency of amino acid pairs in mutant human BTK. In general, missense point mutations modify the configuration of amino acid pairs in a protein, which could target the changes in the secondary structure contents and consequently affect biological functions of the protein. Thus, our approach may provide useful insight into the molecular mechanisms of XLA.

We are trying to explore an interesting aspect regarding which amino acid sub-sequences are more or less sensitive to variants. If such a general rule could be drawn, then we could gain not only more insight into the relationship between protein variants and its related disorders but, more importantly we could offer more attention to these sensitive sub-sequences in order to prevent them from variants. Moreover, the possible sub-sequences sensitive to the, currently, unknown variants could be predicted.

References

[1] Chou, K.C. (1999) "Using pair-coupled amino acid composition to predict protein secondary structure content", *Protein Eng.* **12**, 1041–1050.
[2] Liu, W. and Chou, K.C. (1999) "Prediction of protein secondary structure content", *J. Protein Chem.* **18**, 473–480.

[3] Wu, G. and Yan, S.M. (2002a) "Randomness in the primary structure of protein: methods and implications", *Mol. Biol. Today* **3**, 55–69.
[4] Wu, G. and Yan, S.M. (2003) "Determination of amino acid pairs in human haemoglobin α chain sensitive to variants by means of a random approach", *Comp. Clin. Path.*, In Press.
[5] Wu, G. and Yan, S.M. (2002b) "Determination of amino acid pairs in human phenylalanine hydroxylase protein sensitive to variants by means of a random approach", *Peptides* **23**, 2085–2090.
[6] Rawlings, D.J. (1999) "Bruton's tyrosine kinase controls a sustained calcium signal essential for B lineage development and function", *Clin. Immunol.* **91**, 243–253.
[7] LeBien, T.W. (2000) "Fates of human B-cell precursors", *Blood* **96**, 9–23.
[8] Bruton, O. (1952) "Agammaglobulinemia", *Pediatrics* **9**, 722–728.
[9] Vetrie, D., Vorechovsky, I., Sideras, P., Holland, J., Davies, A., Flinter, F., Hammarstrom, L., Kinnon, C., Levinsky, R., Bobrow, M., Smith, C.I.E. and Bentley, D.R. (1993) "The gene involved in X-linked agammaglobulinaemia is a member of the *src* family of protein tyrosine kinase", *Nature* **361**, 226–233.
[10] Tsukada, S., Saffran, D.C., Rawlings, D.J., Parolini, O., Allen, R.C., Klisak, I., Sparkes, R.S., Kubagawa, H., Mohandas, T., Quan, S., Belmont, J.W., Cooper, M.D., Conley, M.E. and Witte, O.N. (1993) "Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia", *Cell* **72**, 279–290.
[11] Rosen, F.S., Wedgwood, R.J., Eibl, M.M., *et al.* (1997) "Primary immunodeficiency diseases: report of a WHO scientific group", *Clin. Exp. Immunol.* **109**(Suppl.), S1.
[12] Fischer, A. and Malissen, B. (1998) "Natural and engineered disorders of lymphocyte development", *Science* **280**, 237–243.
[13] Conley, M.E. and Cooper, M.D. (1998) "Genetic basis of abnormal B cell development", *Curr. Opin. Immunol.* **10**, 399–406.
[14] Bairoch, A. and Apweiler, R. (2000) "The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 2000", *Nucleic Acids Res.* **28**, 45–48.
[15] Feller, W. (1968) *An Introduction to Probability Theory and Its Applications*, 3rd Ed. (Wiley, New York) **Vol. I**.
[16] Rideout, W.M., Coetzee, G.A., Olumi, A.F. and Jones, P.A. (1990) "5-Methylcytosine as an endogenous mutagen in human LL receptor and p53 genes", *Science* **249**, 1288–1290.
[17] Montesano, R., Hainaut, P. and Wild, C.P. (1997) "Hepatocellular carcinoma: from gene to public health", *J. Natl Cancer Inst.* **89**, 1844–1851.
[18] Hainaut, P. and Pfeifer, G.P. (2001) "Patterns of p53 G → T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke", *Carcinogenesis* **22**, 367–374.
[19] Ory, K., Legros, Y., Auguin, C. and Soussi, T. (1994) "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**, 3496–3504.
[20] Forrester, K., Lupold, S.E., Ott, V.L., Chay, C.H., Band, V., Wang, X.W. and Harris, C.C. (1995) "Effects of p53 mutants on wild-type p53-mediated transactivation are cell type dependent", *Oncogene* **10**, 2103–2111.
[21] Aas, T., Borresen, A.L., Geisler, S., Smith-Sorensen, B., Johnsen, H., Varhaug, J.E., Akslen, L.A. and Lonning, P.E. (1996) "Specific p53 mutations are associated with *de novo* resistance to doxorubicin in breast cancer patients", *Nat. Med.* **2**, 811–814.