FULL-LENGTH PAPER

Mutation patterns in human α-galactosidase A

Shaomin Yan · Guang Wu

Abstract A way to study the mutation pattern is to convert a 20-letter protein sequence into a scalar protein sequence, because the 20-letter protein sequence is neither vector nor scalar while a promising way to study patterns is in numerical domain. In this study, we use the amino-acid pair predictability to convert α-galactosidase A with its 137 mutations into scalar sequences, and analyse which amino-acid pairs are more sensitive to mutation. Our results show that 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.

Keywords Amino-acid pair · α-Galactosidase A · Fabry disease · Mutation · Pattern

Introduction

α-Galactosidase A is a lysosomal enzyme that catalyzes the hydrolysis of melibiose to form galactose whose deficiency leads to progressive accumulation of globotriaosylceramide, digalactosyl ceramide, group B, B1, and P1 glycolipids in many tissues [1] classified clinically as Fabry disease [2,3]. Most affected patients suffer with severe peripheral pain in childhood or early adult life [4–6]. Later manifestations of the disease include skin lesions [7], cardiomyopathy [8–10], distressing gastrointestinal symptoms [11], end-stage renal

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

failure [12], and central neurological defects as a consequence of cerebrovascular disease [13–16].

Clearly, mutations in α-galactosidase A not only lead to various clinical outcomes, but also provide a model to analyze mutation patterns to understand their consequent diseases for better clinical managements.

Actually, we can analyze mutation patterns at protein level in several different ways, and the most straightforward way is to directly analyze mutation patterns in terms of difference in amino acids. For example, we record a mutation at position 231 of α-galactosidase A, which changes aspartic acid "D" to asparagine "N" [17]. Although this record could provide some pattern as the documentation increases, it is hard to find numeric features that are generally obtained through mathematical deduction.

This is so because the symbolized amino acids are neither vector data nor scalar data, while most patterns found with mathematical tools are in the data domain. This means that we need to perform some conversion to change symbolized protein sequences into scalar protein sequences, then we would have a full ability to analyse the mutation patterns.

There are several ways to transform the symbolized protein sequences into scalar data, of which the most profound one is to use the physicochemical property of amino acids to replace each amino acid in a protein sequence, for example, molecular weight, melting point, optical rotation [18].

On the other hand, our group has developed three approaches to convert a symbolized protein sequence into a scalar protein sequence based on random mechanism (for review, see [19–22]).

Moreover, many studies have indicated that mathematical and computational approaches such as diffusion-controlled reaction simulation [23], graph/diagram approach [24–31], bio-macromolecular internal collective motion simulation [32–34], structural bioinformatics [18,35], molecular

docking [36], molecular packing [37,38], pharmacophore modelling [39,40], Monte Carlo simulated annealing approach [41–44], QSAR [45,46], protein subcellular location prediction [47–52], protein structural class prediction [53–56], identification of membrane proteins and their types [57,58], identification of enzymes and their functional classes [59], identification of GPCR and their types [60–62], identification of proteases and their types [63,64], protein cleavage site prediction [65–67], and signal peptide prediction [68,69] can timely provide very useful information and insights for both basic research and drug design and hence are widely welcome by science community.

In this study, we apply our approach to study mutation patterns in hopes that it can throw some light on mutation patterns.

Materials and methods

The amino acid sequences of the human α-galactosidase A and its 137 missense point mutants are obtained from the UniProtKB/Swiss-Prot entry [70].

Conversion of symbolized human α-galactosidase A into scalar data

There are 20 types of naturally occurring amino acids in proteins. Although we can, for example, use physicochemical properties to replace 20 types of amino acids, the replaced 20 numbers might not be subject to mutation, length of protein sequence, composition of protein, neighboring amino acids, and amino-acid position in protein. Thus, this type of conversion might not be suited to study mutation patterns.

The approach we use is to apply the permutation of amino-acid pairs in human α-galactosidase A to determine if an amino-acid pair is predictable or unpredictable in terms of its appearance in human α-galactosidase A [19–22,71–75]. Human α-galactosidase A consists of 429 amino acids. The first and second amino acids can be counted as an amino-acid pair, the second and third as another amino-acid pair, the third and fourth and so forth until the 428th and 429th, thus there is a total of 428 amino-acid pairs.

Thereafter, for example, there are 30 aspartic acids "D" and 48 leucines "L" in human α-galactosidase A, the appearance amino-acid pair DL would be 3 ($30/429 \times 48/428 \times 428 = 3.357$). Actually we do find three DLs in α-galactosidase A, so DL is predictable by permutation. By contrast, there are 22 arginines "R" and 23 glutamines "Q" in human α-galactosidase A, the appearance of RQ would be 1 ($22/429 \times 23/428 \times 428 = 1.179$), i.e., there would be one RQ in α-galactosidase A. However, the RQ pair appears four times indicating that its appearance is unpredictable by permutation.

Table 1 D231N mutation and its effect on amino-acid pairs before and after mutation in human α-galactosidase A

	Substituted pairs				Substituting pairs			
	AD		DI		AN		NI	
	PF	AF	PF	AF	PF	AF	PF	AF
Before mutation	2	5	2	3	2	1	1	0
After mutation	2	4	1	2	2	2	1	1

A Alanine, *D* aspartic acid, *N* asparagine, *I* isoleucine, *PF* predicted frequency, *AF* actual frequency

Mutations at predictable and unpredictable amino-acid pairs

A point mutation results in two amino-acid pairs being replaced by another two pairs. For example, there is a mutation at position 231 changing aspartic acid "D" to asparagine "N" [17]. This mutation results in two amino-acid pairs AD and DI changing to AN and NI, because the amino acid is alanine "A" at position 230 and isoleucine "I" at position 232. The actual and predicted frequencies of these amino-acid pairs are shown in Table 1, and we can determine whether the substituted amino-acid pairs (AD and DI) and substituting amino-acid pairs (AN and NI) belong to predictable or unpredictable amino-acid pairs. In this way, we can analyse all of the amino-acid pairs housing other mutations [76].

Difference between predicted and actual frequency

For the numerical analysis, we calculate the difference between predicted frequency (PF) and actual frequency (AF) of affected amino-acid pairs, i.e., $\sum(PF - AF)$. As seen in Table 1, before mutation the difference between predicted and actual frequency is $(2 - 5) + (2 - 3) = -4$ for substituted amino-acid pairs, and $(2 - 1) + (1 - 0) = 2$ for substituting amino-acid pairs. After mutation, they are $(2 - 4) + (1 - 2) = -3$ and $(2 - 2) + (1 - 1) = 0$. Thus, we can compare mutation effects on the frequency difference.

Statistics

The Chi-square test was used to compare the occurrence of mutation in predictable and unpredictable kind/pair, and the Mann–Whitney U test for two groups. $p < 0.05$ is considered significant.

Results

Amino-acid pairs in human α-galactosidase A

Theoretically, 20 types of amino acids can construct 400 kinds of possible amino-acid pairs. As the human

Table 2 Mutations at predictable/unpredictable kinds and pairs in human α-galactosidase A

Amino acids	Kinds		Pairs		Mutations		Ratio	
	Number	%	Number	%	Number	%	Mutations/kinds	Mutations/pairs
Predictable	111	46.44	148	34.58	12	8.76	12/111 = 0.11	12/148 = 0.08
Unpredictable	128	53.56	280	65.42	125	91.24	125/128 = 0.98	125/280 = 0.45
Total	239	100	428	100	137	100	137/239 = 0.57	137/428 = 0.32

The Chi-square test indicates the highly statistical significance of occurrence of mutations between predictable and unpredictable kinds/pairs

Table 3 Substituted amino-acid pairs before and after mutation in human α-galactosidase A

Amino-acid pairs	I	II	Before mutation			After mutation		
			Appearance	%	Total %	Appearance	%	Total %
Predictable	AF = PF	AF = PF	12	8.76	8.76	16	11.68	11.68
Unpredictable	AF > PF	AF > PF	54	39.42	91.24	11	8.03	88.33
	AF > PF	AF = PF	56	40.88		33	24.09	
	AF > PF	AF < PF	9	6.57		22	16.06	
	AF < PF	AF = PF	6	4.38		39	28.47	
	AF < PF	AF < PF	0	0.00		16	11.68	

AF Actual frequency, *PF* predicted frequency

α-galactosidase A has 428 amino-acid pairs, which are more than 400 kinds of theoretical amino-acid pairs, some of 400 types of theoretical amino-acid pairs should appear more than once. Meanwhile, we may expect that some of 400 kinds of theoretical amino-acid pairs are absent from human α-galactosidase A.

Out of the 400 kinds of theoretical amino-acid pairs, 161 are absent in human α-Galactosidase A, so 428 amino-acid pairs in human α-galactosidase A include only 239 kinds of theoretical amino-acid pairs ($400 - 161 = 239$), which furthermore means that some amino-acid pairs should appear more than once. Actually, out of the 428 amino-acid pairs in human α-galactosidase A, 119 kinds appear once, 77 kinds twice, 28 kinds three times, 8 kinds four times, 5 kinds five times, and 2 kinds seven times.

Naturally, a further classification appears necessary, say, predictable/unpredictable kind and predictable/unpredictable pair. Out of the 239 kinds of theoretical amino-acid pairs in human α-galactosidase A, 111 kinds are predictable and 128 are unpredictable. Out of the 428 amino-acid pairs in human α-galactosidase A, 148 pairs are predictable and 280 pairs are unpredictable. Hence, the mutation pattern can be found in this regard in Table 2.

Amino-acid pair targeted by mutation

If an amino-acid pair, which is directly targeted by mutation, appears once before mutation, this kind of amino-acid will disappear after mutation. However, if a kind of amino-acid pair appears more than once before mutation, this kind of amino-acid pair will still appear after mutation. Moreover, a point mutation is generally related to two pairs, which warrant the remaining of a kind of amino-acid pair after mutation.
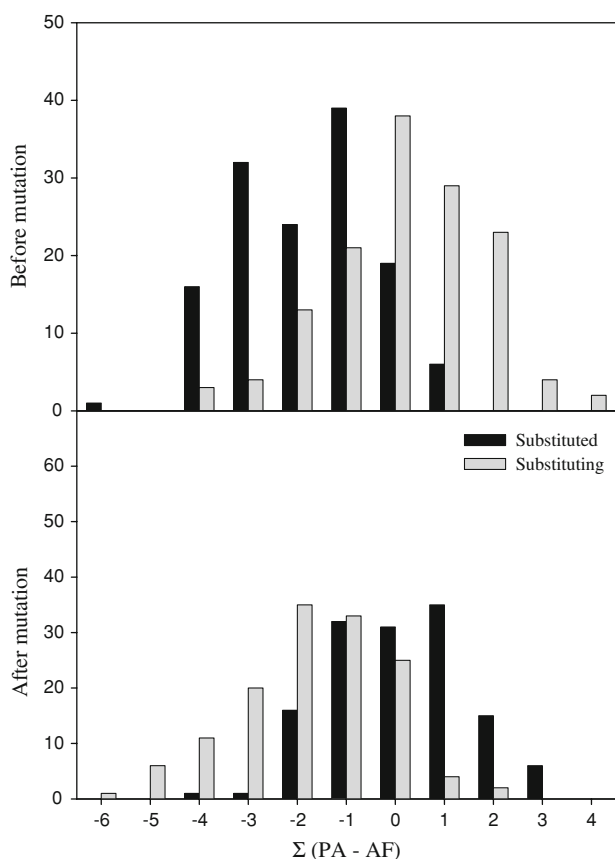
Table 3 lists the grouped amino-acid pairs, which are targeted by mutations, before and after mutation. This table can be read as follows. The first three columns group the substituted amino-acid pairs according to predictable/unpredictable as well as actual and predicted frequency. The three columns under before mutation are the grouped amino-acid pairs, and the last three columns under after mutation are also the grouped amino-acid pairs.

By comparing the appearance before and after mutation, we can see the aim of mutation in this regard, for example, 137 mutations dramatically reduced the appearance of amino-acid pairs, whose actual frequency is larger predicted frequency in both pairs, from 54 to 11 (the third line in Table 3), also from row 4 to row 6 under before mutation, 86.86% of these pairs are characterised by one or both substituted pairs whose actual frequency is larger than their predicted one. These results suggest that the impact of mutations is to narrow the difference between actual and predicted frequency by means of reducing the actual frequency. No mutation occurs 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 mutations

Table 4 Substituting amino-acid pairs before and after mutation in human α-galactosidase A

Amino-acid pairs		Before mutation			After mutation		
I	II	Appearance	%	Total %	Appearance	%	Total%
AF = 0, PF > 0	AF = 0, PF > 0	19[a]	13.87	59.85	0	0	0.00
AF = 0, PF > 0	AF = PF = 0	3[a]	2.19		0	0	
AF = 0, PF > 0	AF = PF > 0	21[a]	15.33		0	0	
AF = 0, PF > 0	AF < PF, AF ≠ 0	8[a]	5.84		0	0	
AF = 0, PF > 0	AF > PF	23[a]	16.79		0	0	
AF = PF = 0	AF = PF = 0	2	1.46		0	0	
AF = PF = 0	AF = PF > 0	3	2.19		0	0	
AF = PF = 0	AF < PF, AF ≠ 0	0[a]	0.00		0	0	
AF = PF = 0	AF > PF	3	2.19		0	0	
AF < PF, AF ≠ 0	AF < PF, AF ≠ 0	0[a]	0.00	40.15	2	1.46	100.00
AF < PF, AF ≠ 0	AF = PF > 0	4[a]	2.92		4	2.92	
AF < PF, AF ≠ 0	AF > PF	7[a]	5.11		7	5.11	
AF = PF > 0	AF = PF > 0	13	9.49		22	16.06	
AF = PF > 0	AF > PF	10	7.30		48	35.04	
AF > PF	AF > PF	21	15.33		54	39.42	

[a] Indicates the substituting amino-acid pairs with their actual frequency smaller than predicted one. The total of these amino-acid pairs is 85 (62.04%)

Fig. 1 Frequency difference between substituted and substituting amino-acid pairs before and after mutation in human α-galactosidase A

to narrow the difference between actual and predicted frequency by means of increasing the actual frequency.
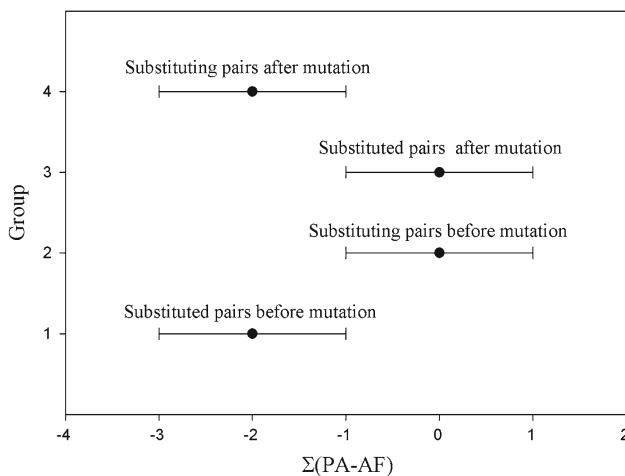
Amino-acid pairs appeared through mutations

Table 4 lists the grouped amino-acid pairs, which appeared through mutation, before and after mutation. Actually, the format of results and underlined implication in Table 4 are very similar to Table 3, for example, 59.85% mutations result in one or both substituting amino-acid pairs are absent before mutation.

Frequency difference of amino-acid pairs affected by mutations

Figure 1 illustrates the difference between predicted and actual frequency in the amino-acid pairs that are influenced by 137 mutations, besides Fig. 2 shows their statistical comparison. Before mutation, the median of difference between predicted and actual frequency is −2 in substituted amino-acid pairs. This means that the mutations occur in the amino-acid pairs that appear more than their predicted frequency. Meanwhile, the corresponding value is 0 in substituting amino-acid pairs indicating that the mutations lead to the construction of amino-acid pairs randomly.

After mutation, the median of difference between actual and predicted frequency is 0 in substituted amino-acid pairs, and their corresponding value is −2 in substituting amino-acid

Fig. 2 Sum of difference between predicted and actual frequency [$\Sigma(\text{PF} - \text{AF})$] of substituted and substituting amino-acid pairs before and after mutation in human α-galactosidase A. The data are presented as median with interquartile interval. There is statistically significant difference between corresponding groups ($p < 0.001$, Mann–Whitney U test)

pairs. This implies that these amino-acid pairs are more randomly constructed in the mutants, as their predicted and actual frequencies are about the same.

Discussion

The gene encoding α-galactosidase A has been sequenced and more than 300 different mutations were identified in affected individuals [77,78], and the genetic heterogeneity of α-galactosidase A contributes to the different phenotypes of Fabry disease [79,80]. However, only 137 mutations have been documented at protein level, otherwise we would have a more comprehensive view.

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 [81–83]. The second is the function selection, which indicates the disruption of protein functions may depend upon the position of the mutation in the protein [84–86]. However, these explanations still do not fully answer why some amino acids are more sensitive to mutation.

This study explains why some amino acids are more sensitive to mutation from random viewpoint. This is very plausible, not only because pure chance is now considered to lie at the very heart of nature [87] but also because the randomly predictable amino-acid pair suggests the maximal probability of occurrence, which requires the least time and energy for construction of amino-acid pair being consistent with nature parsimony.

Needless to say, the functional sites in protein are more likely to be deliberately evolved, thus their actual frequency should be different from the predicted frequency because the amino-acid pair, which can be explained by randomness, may not be explained by its function.

Our results suggest that the trend is that the mutation leads the actual frequency to approach to the predicted frequency to some degree. Likely, nature feels uncomfortable to have pairs, whose actual frequency is different from the predicted frequency, and requires the protein to mutate to narrow the difference between predicted and actual frequency at the expense of losing a certain function. However, the amino-acid pairs, which appear through mutation, might lead to the new difference between predicted and actual frequency, which offers the new opportunity of mutation, thus the evolution continues.

It really does not matter which method to use to convert the symbolized protein sequence into any scalar protein sequence if we can find something interesting using the scalar protein sequence. However, it is very important that the scalar protein sequence is subject to mutation, composition of protein, length of protein, neighboring amino acid, position in protein sequence, etc., which can be met by our approaches [19–22], hence we use the amino-acid pair predictability in this study.

Conclusions

In this study, we methodologically demonstrate how to study mutation patterns in proteins using an approach that converts a protein sequence into a numeric sequence. Then we find out the mutation pattern through the analysis of numeric sequence, by which we theoretically find that the mutation pattern in human α-galactosidase A is to narrow the difference between predicted and actual frequency of amino-acid pairs.

References

1. Eng CM, Germain DP, Banikazemi M, Warnock DG, Wanner C, Hopkin RJ, Bultas J, Lee P, Sims K, Brodie SE, Pastores GM, Strotmann JM, Wilcox WR (2006) Fabry disease: guidelines for the evaluation and management of multi-organ system involvement. Genet Med 8:539–548. doi:10.1097/01.gim.0000237866.70357.c6
2. Levin M (2006) Fabry disease. Drugs Today (Barc) 42:65–70. doi:10.1358/dot.2006.42.1.957357
3. Clarke JT (2007) Narrative review: Fabry disease. Ann Intern Med 146:425–433
4. Schiffmann R (2006) Neuropathy and Fabry disease: pathogenesis and enzyme replacement therapy. Acta Neurol Belg 106:61–65

5. Horowitz SH (2007) The diagnostic workup of patients with neuropathic pain. Med Clin North Am 91:21–30. doi:10.1016/j.mcna.2006.10.002

6. Ramaswami U (2008) Fabry disease during childhood: clinical manifestations and treatment with agalsidase alfa. Acta Paediatr Suppl 97:38–40. doi:10.1111/j.1651-2227.2008.00658.x

7. Happle R (2006) X-chromosome inactivation: role in skin disease expression. Acta Paediatr Suppl 95:16–23. doi:10.1080/08035320600618775

8. Shah JS, Elliott PM (2005) Fabry disease and the heart: an overview of the natural history and the effect of enzyme replacement therapy. Acta Paediatr Suppl 94:11–14. doi:10.1080/08035320510028058

9. Bodary PF, Shayman JA, Eitzman DT (2007) Alpha-galactosidase A in vascular disease. Trends Cardiovasc Med 17:129–133. doi:10.1016/j.tcm.2007.02.006

10. Chimenti C, Hamdani N, Boontje NM, DeCobelli F, Esposito A, Bronzwaer JG, Stienen GJ, Russo MA, Paulus WJ, Frustaci A, van der Velden J (2008) Myofilament degradation and dysfunction of human cardiomyocytes in Fabry disease. Am J Pathol 172:1482–1490. doi:10.2353/ajpath.2008.070576

11. Hoffmann B, Keshav S (2007) Gastrointestinal symptoms in Fabry disease: everything is possible, including treatment. Acta Paediatr Suppl 96:84–86

12. Breunig F, Wanner C (2008) Update on Fabry disease: kidney involvement, renal progression and enzyme replacement therapy. J Nephrol 21:32–37

13. Ballabio E, Bersano A, Bresolin N, Candelise L (2007) Monogenic vessel diseases related to ischemic stroke: a clinical approach. J Cereb Blood Flow Metab 27:1649–1662. doi:10.1038/sj.jcbfm.9600520

14. Francis J, Raghunathan S, Khanna P (2007) The role of genetics in stroke. Postgrad Med J 83:590–595. doi:10.1136/pgmj.2007.060319

15. Møller AT, Jensen TS (2007) Neurological manifestations in Fabry's disease. Nat Clin Pract Neurol 3:95–106. doi:10.1038/ncpneuro0407

16. Moore DF, Kaneski CR, Askari H, Schiffmann R (2007) The cerebral vasculopathy of Fabry disease. J Neurol Sci 257:258–263. doi:10.1016/j.jns.2007.01.053

17. Redonnet-Vernhet I, Ploos van Amstel JK, Jansen RPM, Wevers RA, Salvayre R, Levade T (1996) Uneven X inactivation in a female monozygotic twin pair with Fabry disease and discordant expression of a novel mutation in the alpha-galactosidase A gene. J Med Genet 33:682–688. doi:10.1136/jmg.33.8.682

18. Fasman GD (1976) Handbook of biochemistry: section D physical chemical data. 3rd edn. CRC Press, London

19. Wu G, Yan SM (2002) Randomness in the primary structure of protein: methods and implications. Mol Biol Today 3:55–69

20. Wu G, Yan S (2006) Fate of influenza A virus proteins. Protein Pept Lett 13:377–384. doi:10.2174/092986606775974474

21. 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

22. Wu G, Yan S (2008) Lecture notes on computational mutation. Nova Science Publisher, New York

23. Chou KC, Zhou GP (1982) Role of the protein outside active site on the diffusion-controlled reaction of enzyme. J Am Chem Soc 104:1409–1413. doi:10.1021/ja00369a043

24. Chou KC, Forsen S (1980) Graphical rules for enzyme-catalyzed rate laws. Biochem J 187:829–835

25. Zhou GP, Deng MH (1984) An extension of Chou's graphical rules for deriving enzyme kinetic equations to system involving parallel reaction pathways. Biochem J 222:169–176

26. Myers D, Palmer G (1985) Microcomputer tools for steady-state enzyme kinetics. Bioinformatics 1:105–110. doi:10.1093/bioinformatics/1.2.105

27. Chou KC (1989) Graphical rules in steady and non-steady enzyme kinetics. J Biol Chem 264:12074–12079

28. Chou KC (1990) Review: applications of graph theory to enzyme kinetics and protein folding kinetics. Steady and non-steady state systems. Biophys Chem 35:1–24. doi:10.1016/0301-4622(90)80056-D

29. Althaus IW, Chou JJ, Gonzales AJ, Diebel MR, Chou KC, Kezdy FJ, Romero DL, Aristoff PA, Tarpley WG, Reusser F (1993) Steady-state kinetic studies with the non-nucleoside HIV-1 reverse transcriptase inhibitor U-87201E. J Biol Chem 268:6119–6124

30. Chou KC, Kezdy FJ, Reusser F (1994) Review: steady-state inhibition kinetics of processive nucleic acid polymerases and nucleases. Anal Biochem 221:217–230. doi:10.1006/abio.1994.1405

31. Andraos J (2008) Kinetic plasticity and the determination of product ratios for kinetic schemes leading to multiple products without rate laws: new methods based on directed graphs. Can J Chem 86:342–357. doi:10.1139/V08-020

32. Chou KC (1988) Review: low-frequency collective motion in biomacromolecules and its biological functions. Biophys Chem 30:3–48. doi:10.1016/0301-4622(88)85002-6

33. Chou KC (1989) Low-frequency resonance and cooperativity of hemoglobin. Trends Biochem Sci 14:212. doi:10.1016/0968-0004(89)90026-1

34. Zhou GP (1989) Biological functions of soliton and extra electron motion in DNA structure. Phys Scr 40:698–701. doi:10.1088/0031-8949/40/5/021

35. Chou KC (2004) Modelling extracellular domains of GABA-A receptors: subtypes 1, 2, 3, and 5. Biochem Biophys Res Commun 316:636–642. doi:10.1016/j.bbrc.2004.02.098

36. Chou KC, Wei DQ, Zhong WZ (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. Biochem Biophys Res Commun 308:148–151. doi:10.1016/S0006-291X(03)01342-1 (Erratum: ibid, 2003, vol. 310, 675)

37. Chou KC, Nemethy G, Scheraga HA (1984) Energetic approach to packing of a-helices: 2. General treatment of nonequivalent and nonregular helices. J Am Chem Soc 106:3161–3170. doi:10.1021/ja00323a017

38. Chou KC, Maggiora GM, Nemethy G, Scheraga HA (1988) Energetics of the structure of the four-alpha-helix bundle in proteins. PNAS USA 85:4295–4299. doi:10.1073/pnas.85.12.4295

39. Sirois S, Wei DQ, Du QS, Chou KC (2004) Virtual Screening for SARS-CoV Protease Based on KZ7088 Pharmacophore Points. J Chem Inf Comput Sci 44:1111–1122. doi:10.1021/ci034270n

40. Chou KC, Wei DQ, Du QS, Sirois S, Zhong WZ (2006) Review: progress in computational approach to drug development against SARS. Curr Med Chem 13:3263–3270. doi:10.2174/092986706778773077

41. Chou KC (1992) Energy-optimized structure of antifreeze protein and its binding mechanism. J Mol Biol 223:509–517. doi:10.1016/0022-2836(92)90666-8

42. Wu G (1998) Application of the queueing theory with Monte Carlo simulation in inhalation toxicology. Arch Toxicol 72:330–335. doi:10.1007/s002040050510

43. Wu G (1998) Application of queueing theory with Monte Carlo simulation to the study of the intake and adverse effects of ethanol. Alcohol 33:519–527. doi:10.1093/alcalc/33.5.519

44. Wu G, Yan SM (2001) Estimation of the rate of arrivals of ions at a single-channel. J Biochem Mol Biol Biophys 5:505–511

45. Du QS, Huang RB, Chou KC (2008) Review: recent advances in QSAR and their applications in predicting the activities of

chemical molecules, peptides and proteins for drug design. Curr Protein Pept Sci 9:248–259. doi:10.2174/138920308784534005

46. Prado-Prado FJ, Gonzalez-Diaz H, de la Vega OM, Ubeira FM, Chou KC (2008) Unified QSAR approach to antimicrobials. Part 3: First multi-tasking QSAR model for input-coded prediction, structural back-projection, and complex networks clustering of anti-protozoal compounds. Bioorg Med Chem 16:5871–5880. doi:10.1016/j.bmc.2008.04.068

47. Chou KC, Shen HB (2008) Cell-PLoc: a package of web-servers for predicting subcellular localization of proteins in various organisms. Nat Prot 3:153–162. doi:10.1038/nprot.2007.494

48. Chou KC, Shen HB (2007) Review: recent progresses in protein subcellular location prediction. Anal Biochem 370:1–16. doi:10.1016/j.ab.2007.07.006

49. Chou KC, Shen HB (2006) Hum-PLoc: a novel ensemble classifier for predicting human protein subcellular localization. Biochem Biophys Res Commun 347:150–157. doi:10.1016/j.bbrc.2006.06.059

50. Chou KC, Shen HB (2007) Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. J Proteome Res 6:1728–1734. doi:10.1021/pr060635i

51. Chou KC, Shen HB (2006) Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic K-nearest neighbor classifiers. J Proteome Res 5:1888–1897. doi:10.1021/pr060167c

52. Zhou GP, Doctor K (2003) Subcellular location prediction of apoptosis proteins. Proteins: Struct Func Genet 50:44–48

53. Chou KC (1995) A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space. Proteins: Struct Func Genet 21:319–344

54. Chou KC, Zhang CT (1995) Review: prediction of protein structural classes. Crit Rev Biochem Mol Biol 30:275–349. doi:10.3109/10409239509083488

55. Zhou GP (1998) An intriguing controversy over protein structural class prediction. J Protein Chem 17:729–738. doi:10.1023/A:1020713915365

56. Zhou GP, Assa-Munt N (2001) Some insights into protein structural class prediction. Proteins: Struct Func Genet 44:57–59

57. Chou KC (2001) Prediction of protein cellular attributes using pseudo amino acid composition. Proteins: Struct Func Genet 43:246–255 (Erratum: ibid, 2001, vol. 44, 60)

58. Chou KC, Shen HB (2007) MemType-2L: a Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem Biophys Res Commun 360:339–345. doi:10.1016/j.bbrc.2007.06.027

59. Shen HB, Chou KC (2007) EzyPred: a top–down approach for predicting enzyme functional classes and subclasses. Biochem Biophys Res Commun 364:53–59. doi:10.1016/j.bbrc.2007.09.098

60. Chou KC, Elrod DW (2002) Bioinformatical analysis of G-protein-coupled receptors. J Proteome Res 1:429–433. doi:10.1021/pr025527k

61. Chou KC (2005) Prediction of G-protein-coupled receptor classes. J Proteome Res 4:1413–1418. doi:10.1021/pr050087t

62. Xiao X, Wang P, Chou KC (2008) GPCR-CA: a cellular automaton image approach for predicting G-protein-coupled receptor functional classes. J Comput Chem 30:1414–1423. doi:10.1002/jcc.21163

63. Chou KC, Shen HB (2008) ProtIdent: a web server for identifying proteases and their types by fusing functional domain and sequential evolution information. Biochem Biophys Res Commun 376:321–325. doi:10.1016/j.bbrc.2008.08.125

64. Shen HB, Chou KC (2009) Identification of proteases and their types. Anal Biochem 385:153–160. doi:10.1016/j.ab.2008.10.020

65. Chou KC (1993) A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. J Biol Chem 268:16938–16948

66. Chou KC (1996) Review: prediction of HIV protease cleavage sites in proteins. Anal Biochem 233:1–14. doi:10.1006/abio.1996.0001

67. Shen HB, Chou KC (2008) HIVcleave: a web-server for predicting HIV protease cleavage sites in proteins. Anal Biochem 375:388–390. doi:10.1016/j.ab.2008.01.012

68. Chou KC, Shen HB (2007) Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. Biochem Biophys Res Commun 357:633–640. doi:10.1016/j.bbrc.2007.03.162

69. Shen HB, Chou KC (2007) Signal-3L: a 3-layer approach for predicting signal peptide. Biochem Biophys Res Commun 363:297–303. doi:10.1016/j.bbrc.2007.08.140

70. UniProtKB/Swiss-Prot (2008) http://www.expasy.org/uniprot/P06280, accession number P06280, annotations were last modified on September 23, 2008, entry version 108

71. Wu G, Yan S (2004) Fate of 130 hemagglutinins from different influenza A viruses. Biochem Biophys Res Commun 317:917–924. doi:10.1016/j.bbrc.2004.03.134

72. 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

73. 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

74. Wu G, Yan S (2007) 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

75. Wu G, Yan S (2008) Prediction of mutations engineered by randomness in H5N1 hemagglutinins of influenza A virus. Amino Acids 35:365–373. doi:10.1007/s00726-007-0602-4

76. Amino-acid pair predictability (2008) http://www.dreamscitech.com/Service/rationale.htm

77. Ashley GA, Shabbeer J, Yasuda M, Eng CM, Desnick RJ (2001) Fabry disease: twenty novel alpha-galactosidase A mutations causing the classical phenotype. J Hum Genet 46:192–196. doi:10.1007/s100380170088

78. Germain DP (2007) Genetics of Fabry disease: diagnostic and therapeutic implications. Presse Med 36(1):1S14–S19

79. Branton M, Schiffmann R, Sabnis S, Murray GJ, Quirk JM, Altarescu G, Goldfarb L, Brady RO, Balow JE, Austin HA III, Kopp JB (2002) Natural history of Fabry renal disease: influence of α-galactosidase A activity and genetic mutations on clinical course. Medicine 81:122–138. doi:10.1097/00005792-200203000-00003

80. Schaefer E, Mehta A, Gal A (2005) Genotype and phenotype in Fabry disease: analysis of the Fabry outcome survey. Acta Paediatr Suppl 94:87–92. doi:10.1080/08035320510031045

81. Rogozin I, Kondrashov F, Glazko G (2001) Use of mutation spectra analysis software. Hum Mutat 17:83–102. doi:10.1002/1098-1004(200102)17:2<83::AID-HUMU1>3.0.CO;2-E

82. Rogozin IB, Pavlov YI (2003) Theoretical analysis of mutation hotspots and their DNA sequence context specificity. Mutat Res 544:65–85. doi:10.1016/S1383-5742(03)00032-2

83. Acharya N, Abu-Nasr NF, Kawaguchi G, Imai M, Yamamoto K (2007) Frameshift mutations produced by 9-aminoacridine in wild-type, uvrA and recA strains of Escherichia coli; specificity within a hotspot. J Radiat Res (Tokyo) 48:361–368. doi:10.1269/jrr.07036

84. Dzikiewicz-Krawczyk A (2008) The importance of making ends meet: mutations in genes and altered expression of proteins of the MRN complex and cancer. Mutat Res 659:262–273. doi:10.1016/j.mrrev.2008.05.005

85. Levy LS (2008) Advances in understanding molecular determinants in FeLV pathology. Vet Immunol Immunopathol 123:14–22. doi:10.1016/j.vetimm.2008.01.008

86. Martinez-Picado J, Martínez MA (2008) HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. Virus Res 134:104–123. doi:10.1016/j.virusres.2007.12.021

87. Everitt BS (1999) Chance rules: an informal guide to probability, risk, and statistics. Springer, New York