

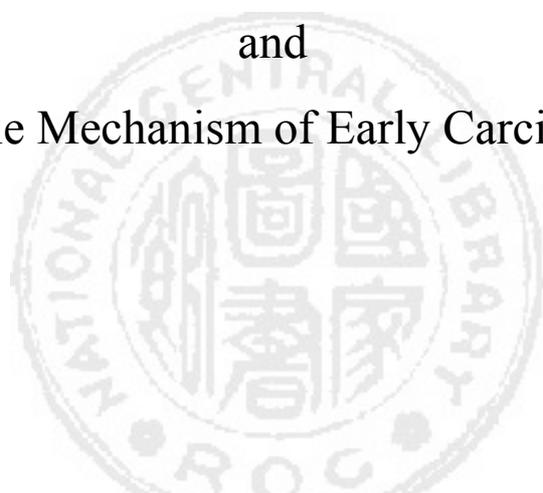
東海大學物理學系碩士論文

癌症基因的電性變化與早期癌變的可能機制

Charge Transport in Cancer-Related Genes

and

A Possible Mechanism of Early Carcinogenesis



研究生：鄭詠尹

指導教授：施奇廷博士

中華民國九十八年六月

誌謝

還記得在辛苦除錯後，寫出了可以分析的程式後，有點恐懼去執行程式，這是因為害怕分析結果不如預期，但若得到了預期的結果，心中會泛起莫名的興奮。我想我有體會到做研究的快樂吧！

感謝施奇廷教授給予了我足夠去分析的知識，並持續不斷的鼓勵我思考新分析方法，我才得以寫出這篇論文；感謝曾郁欽先生指點我程式，減少我摸索程式的時間；感謝家人一直讓我自由發展，讓我做自己想做的事。

2008年台中東海

癌症基因的電性變化與早期癌變的可能機制

研究生：鄭詠尹

指導教授：施奇廷 教授

東海大學物理研究所

摘要

利用 tight-binding model 以及 transfer matrix method，我們計算了 20 種腫瘤抑制基因和 15 種與癌症相關基因的所有可能的點突變所引起的電性變化，統計結果顯示，會引發癌症病變的基因突變，其電性變化平均而言小於其他的突變。實驗顯示，BER(base excision repair) 酵素能透過長距離氧化還原反應檢測 DNA 片段電性，進而快速找出到基因損傷的位置。我們推論，致病突變因其電性變化小，可能因此而躲過 DNA 的損傷修補機制，而得以留存，進而引發癌症。

Charge Transport in Cancer-Related Genes
and
A Possible Mechanism of Early Carcinogenesis

Student: Yun-yin Cheng

Advisor: Dr.Chi-Tin Shin

Department of Physics, Tunghai University, Taichung, Taiwan ROC

Abstract

In this thesis, DNA sequences are approached by tight-binding model. Changes of charge transport(CT) of DNA mutations are studied by transfer matrix method. We focus our investigations on all possible point mutations of 20 tumor suppressor genes and other 15 cancer-associated genes. Our statistical results exhibit that average change charge transport of carcinogenic mutations is less than non-carcinogenic mutations. Furthermore, experiments showed that the BER(base excision repair) enzymes can detect damage in DNA sequence by probing electric change. We conjecture that mutations of smaller changes in CT properties, may circumvent the DNA repair mechanisms and to be retained, and finally cause cancers.

目錄

誌謝	i
中文摘要	ii
英文摘要	iii
目錄	iv
表目	v
圖目	vi
第一章 緒論	1
1.1 研究動機與源起	1
1.2 DNA 簡介	1
1.3 BER 酵素與 DNA 的電性	3
1.4 HGMD 資料庫	3
1.5 OMIM 資料庫	4
1.6 癌症基因簡介	5
第二章 DNA 的穿透係數計算	6
2.1 DNA 模型	6
2.2 穿隧矩陣	8
2.3 傳輸係數變化定義	12
第三章 統計結果分析	14
3.1 單一位置突變前後的 charge transport 變化比較	14
3.2 致癌突變的傳輸係數在 codon 中的排名	16
3.3 致癌突變的傳輸係數在序列中的排名	20
第四章 致癌 DNA 序列與序列的相依性	27
第五章 結論	32
參考文獻	33

附錄 1：癌症相關基因的致癌突變在 codon 上的電性排名分佈-----	35
附錄 2：p53 序列全部可能突變與致癌突變引起之電性變化排名分佈-----	52
附錄 3：基因序列與含致癌突變之序列的相依性-----	72

表目錄

表 1.1 癌症基因簡介-----	5
表 3.1 p53 序列 14585 _{th} 位置所有可能突變所引起的傳輸係數改變-----	15
表 3.2 胺基酸與 codon 之對應-----	16
表 3.3 在 p53 序列中的所有 12 種可能突變與致癌突變的各別次數-----	21
表 3.4 p53 序列上所有長度 80，鹼基由 T 突變成 A，與該序列致癌突變為 T 變 A 的傾向值計算-----	23
表 3.5 p53 序列上的 12 種突變的傾向值-----	24
表 4.1 各種子序列出現之數-----	27
表 4.2 各種計次方式所得出含致癌突變的子序列出現次數-----	29
表 4.3 癌症序列與其含致癌突變點之序列的相依性-----	31

圖目錄

圖 1.1 DNA 結構-----	2
圖 1.2 BER 酵素修復 DNA 損傷的機制-----	3
圖 2.1 DNA 單股模型-----	6
圖 2.2 DNA 雙股模型-----	6
圖 2.3 DNA 魚骨模型-----	7
圖 2.4 DNA 梯子模型-----	8
圖 2.5 DNA 外接金屬-----	8
圖 2.6 電子穿透 DNA 的波函數-----	9
圖 2.7 DNA 序列突變前後的傳輸係數比較-----	11
圖 2.8 DNA 子序列之定義-----	12
圖 2.9 DNA 子序列突變前後之義-----	12
圖 2.10 固定長度下，包含突變點的所有子序列，其突變所引起的傳輸係數 變化-----	13
圖 3.1 p53 序列突變位置與癌症例之統計-----	14
圖 3.2 p53 序列 14585 _{th} 位置上突變前傳輸係數與該位置所有可能突變的傳 輸係數-----	15
圖 3.3 codon 上所有點突變所得到的電性改變-----	17
圖 3.4 BRCA1 序列的致癌突變所引起的電性變化，在該突變點所屬之 codon 上的各種點突變所引起之電性變化的排名分佈-----	17
圖 3.5 NF1 序列的致癌突變所引起的電性變化，在該突變點所屬之 codon 上的各種點突變所引起之電性變化的排名分佈-----	18
圖 3.6 35 筆序列的致癌突變所引起的電性變化，在該突變點所屬之 codon 上的各種點突變所引起之電性變化的排名分佈-----	18

圖 3.7 WT1 序列的致癌突變所引起的電性變化，在該突變點所屬之 codon 上的各種點突變所引起之電性變化的三種排名分佈-----	19
圖 3.8 CDH1 序列的致癌突變所引起的電性變化，在該突變點所屬之 codon 上的各種點突變所引起之電性變化的三種排名分佈-----	19
圖 3.9 35 筆序列的致癌突變所引起的電性變化，在該突變點所屬之 codon 上的各種點突變所引起之電性變化的三種排名分佈-----	20
圖 3.10 在長度 80 時，p53 序列的所有可能突變與致癌突變所引起的電性變化排名分佈-----	21
圖 3.11 在長度 80 時，p53 序列的所有可能突變與致癌突變中由 T 突變成 A 所引起的電性變化排名分佈-----	22
圖 3.12 傾向值計算示意圖-----	23
圖 3.13 在長度 20 時，p53 序列的所有可能突變與致癌突變中由 C 突變成 T 所引起的電性變化排名分佈-----	24
圖 3.14 在長度 40 時，p53 序列的所有可能突變與致癌突變中由 C 突變成 T 所引起的電性變化排名分佈-----	25
圖 3.15 在長度 60 時，p53 序列的所有可能突變與致癌突變中由 C 突變成 T 所引起的電性變化排名分佈-----	25
圖 3.16 在長度 80 時，p53 序列的所有可能突變與致癌突變中由 C 突變成 T 所引起的電性變化排名分佈-----	26
圖 4.1 以長度 3 的視窗掃過序列-----	27
圖 4.2 含致癌突變的子序列之計算方式說明-----	28
圖 4.3 TSHR 序列與其含致癌突變點的子序列之相依性-----	30

第一章 序論

1.1 研究動機與緣起

DNA 由於具有自我組成電性與序列相關的特性，且尺度小，有發展成奈米電子元件的潛力，吸引了許多科學家研究它的電性[1]，然而目前為止 DNA 的電性仍無明確定論，絕緣體[2-5]、半導體[6,7]、導體[7-13]甚至超導體[14]的行為的說法都有。DNA 分子的電荷傳輸性質與長程氧化機制扮演了重要的角色 [1,15]，實驗顯示 DNA 的電性可能是配置錯誤的鹼基對會不會被修復的關鍵，所以施奇廷教授與他的合作者在 p53 基因的研究中[16]，計算了正常 p53 序列的電荷穿透係數與所有可能的點突變的穿透係數，比較突變前後的電性改變量，統計的結果顯示致癌突變的電性改變量，平均而言小於其他不會致癌的點突變電性改變量。他們據此提出一個癌症形成機制：致癌點突變所引起的電性改變量相對於其他不會致癌的點突變來得小，所以躲過了 DNA 修復酵素的檢測機制。為瞭解這現象是否有普遍性，我們決定大量的研究其他癌症基因序列的電性。

1.2 DNA 簡介

DNA (DNA – Deoxyribonucleic Acid) 具自我複製 (replication) 的能力，在生物繁衍或分裂的過程中扮演了攜帶遺傳訊息的角色，使得生物子代擁有相同之特徵。DNA 在生物體內的作用，係藉由轉錄 (transcription) 合成核糖核酸 (RNA – Ribonucleic Acid)，RNA 轉譯 (translation) 合成蛋白質 (protein)，蛋白質再執行生物體內的所需的功能。

DNA 呈雙股螺旋轉狀，每股的基本單元為核苷酸 (nucleotide)，核苷酸由五個碳糖 (S) 與一磷酸根 (P) 和含氮鹼基所組成。鹼基分嘌呤 (purine)和

嘧啶（pyrimidine），嘌呤再分為腺嘌呤（adenine—A）和鳥嘌呤（guanine—G）；嘧啶則分為胞嘧啶（cytosine—C）與胸腺嘧啶（thymine—T）。在正常的 DNA 分子互補的雙股中，A 只與 T，C 只與 G 以氫鍵結合，形成鹼基對（base pair）。

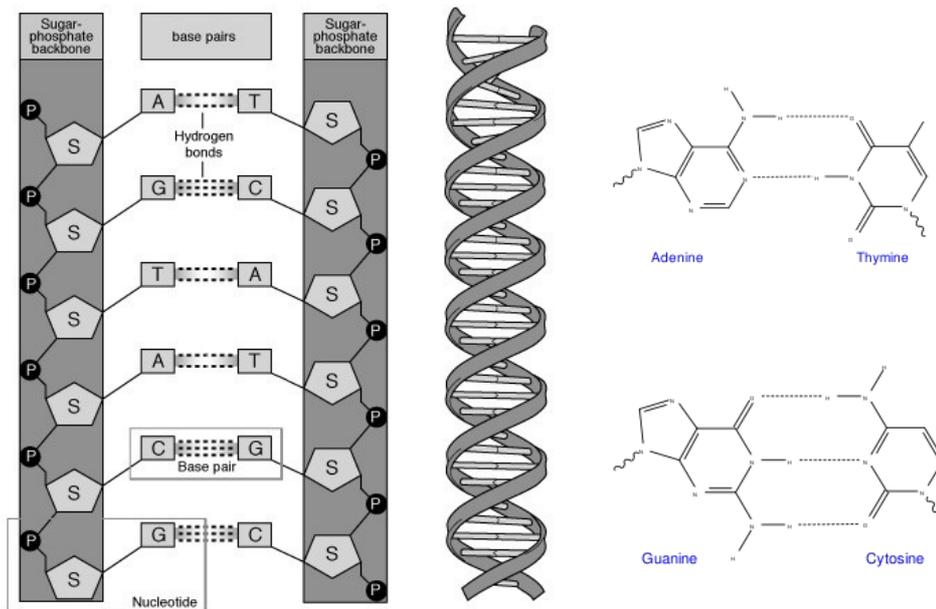


圖 1.1 DNA 為雙股螺旋結構，攤開呈現如上圖左，兩側 Sugar-phosphate，稱之為主幹，兩側主幹以中間的 base pairs 連接，鹼基所能形成的 base pairs 只有 A-T 與 G-C 如圖右。

在基因序列上，分外顯子（exon）與內含子（intron），其中外顯子與編碼胺基酸（Amino Acid）有關。

在 DNA 序列上，每三個鹼基負責編碼一種胺基酸。胺基酸是構成蛋白質的基本成份，因此 DNA 序列上的異常突變，可能無法正確編碼所需要的胺基酸，進而影響蛋白質的合成，使得細胞無法正常執行功能，成為生物體內的危害。

1.3 BER 酵素與 DNA 的電性

BER (base excision repair) 酵素的實驗結果顯示，BER 酵素藉由電子在 DNA 序列上的傳導特性，找出基因序列上配置錯誤的鹼基對，並加以修正[17]。在 BER 酵素中有 FeS^{2+} 的原子團，且在接觸到 DNA 後，放出一個電子成為 FeS^{3+} ，並附著在序列上移動，在透過與另一 BER 酵素交換電子後，即恢復成 FeS^{2+} 後離開 DNA 分子。若基因序列上有配置錯誤 (mismatch, 例如 A 接 C) 的部份，將形成電子阱，使得附著於基因序列上的 BER 酵素持續保持 FeS^{3+} ，無法離開該基因序列，直至 BER 酵素移動到配置錯誤的鹼基對，將其錯誤修復後，方始脫離。

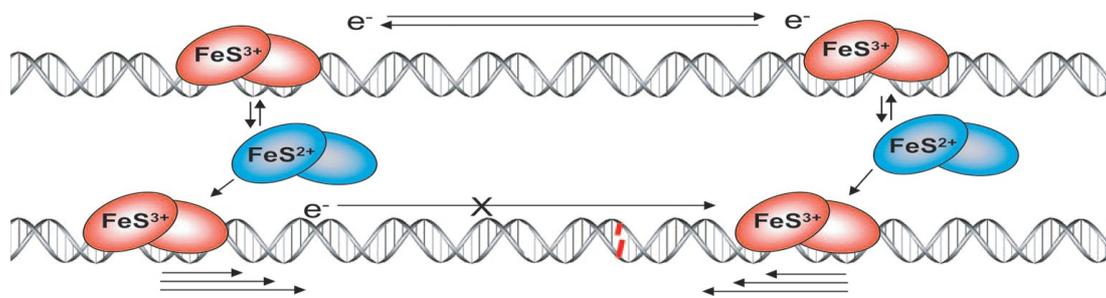


圖 1.2 實驗顯示 BER (base excision repair) 酵素可以透過交換電子的方式快速找到 DNA 損傷的位置。

1.4 HGMD 資料庫

HGMD (The Human Gene Mutation Database) [18]此網站是由 Cardiff 大學所維護，其資料包含正常人體基因序列所擁有的外顯子序列、致癌突變位置與該位置突變前後，所造成編碼胺基酸的改變，由 D.N.Cooper, E.V.Ball, P.D.Stenson, A.D.Phillips, K.Howells 以及 N.S.T.Thomas 共同維護。可在此找到會致癌突變的資料。

HGMD 網址：<http://www.hgmd.cf.ac.uk/ac/index.php>

1.5 OMIM 資料庫

OMIM (Online Mendelian Inheritance in Man) 這是一個全面性，具權威性且即時性彙整人類基因特性的網站，公佈所有已知的孟德爾式 (Mendelian) 遺傳疾病和超過 12000 個已知的基因。OMIM 重點在於表現型和基因型之間的關係[19]。這是每天更新，而且條目包含許多通往其他基因庫的鏈結。這資料庫最初始於 60 年代初，由 Dr. Victor A. McKusick 依孟德爾性狀與疾病的目錄，命名為 MIM (Mendelian Inheritance in Man)。第 12 版的 MIM 圖書發表在 1966 年和 1998 年之間。在線上的版本即 OMIM，是由美國國立醫學圖書館和 Welch Medical 圖書館合作，於 1985 年創立於 Johns Hopkins 大學，在 NCBI (The National Center for Biotechnology Information) 國家生物技術信息中心的協助下，MIM 普及於 1987 年，並在 1995 年改版為全球資訊網頁版本 OMIM。OMIM 是在 Ada Hamosh 博士的指導下，由 Johns Hopkins 大學的 McKusick-Nathans 遺傳醫學研究所撰寫與編輯的。

OMIM 網址：<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>。

1.6 癌症基因簡介

表 1.1 所列的 35 筆癌症相關基因是我們此次主要分析的對象。

Tumor Suppressor				Cancer-associated			
Name	長度	致癌突變數	Tumor Type	Name	長度	致癌突變數	Tumor Type
p53	20303	122	與半數以上癌症相關	BCAM	12341	14	上皮細胞癌
RB	178144	133	視網膜母細胞瘤等疾病	CDH1	98250	30	胃癌和乳癌
p16	26740	85	黑色素瘤等疾病	DPYD	843317	25	5-flourouracil toxicity
APC	108353	222	結腸癌	FUT1	7380	22	結腸癌和胃癌
PTCH	73984	59	基底細胞癌	FUT3	8587	11	結腸癌，胃癌，乳癌
SMAD4	49535	20	胰腺癌與結腸癌	LDHB	22501	11	乳癌
PTEN	105338	80	膠質母細胞瘤，前列腺癌	LHCGR	68951	31	睪丸間質細胞腺瘤
TSC1	53285	44	腎細胞癌	MC1R	2360	24	黑色素瘤鎮痛
TSC2	40724	165	腎細胞癌	MSH6	23872	54	結腸癌
NF1	282701	200	神經纖維肉瘤，膠質瘤	SCN4A	34365	43	重症肌無力綜合症狀
NF2	95023	72	神經纖維肉瘤，聽神經瘤	TGFBR2	87641	14	結腸癌
WT1	47763	56	Wilm's tumor	TSHR	190778	46	甲狀腺癌
MSH2	80098	170	胃癌，卵巢癌，膀胱癌	WRN	22637	62	沃納綜合徵
MLH1	57359	73	胃癌，卵巢癌，膀胱癌	STK11	140199	22	黑斑息肉症，腸道癌
ATM	146268	169	淋巴網狀細胞瘤	BRIP1	180771	13	乳癌
CHK2	54092	20	Li-Fraumeni 綜合症狀				
BRCA1	81155	301	乳癌與卵巢癌				
BRCA2	84193	163	乳癌與卵巢癌				
VHL	10444	172	腎細胞癌				
MEN1	7779	175	內分泌腺瘤病				

表 1.1

第二章 DNA 的穿透係數計算

2.1 DNA 模型

爲了能夠計算長距離的 DNA 傳輸係數，將 DNA 粗粒化成以下四種模型：圖 2.1 單股模型（1L，one-leg）、圖 2.2 雙股模型（2L，two-leg）、圖 2.3 魚骨模型（FB，fish bone）、圖 2.4 梯子模型（LM，ladder model）[20]，分別詳述如下：

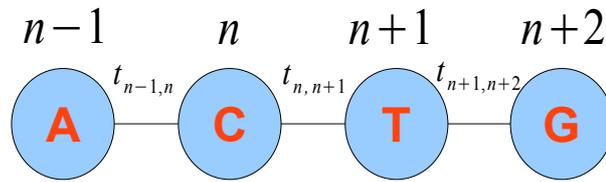


圖 2.1 單股模型（1L）單獨只看 DNA 一股上的鹼基，只考慮該股鹼基上的電子躍遷行爲。

$$H = \sum_n \varepsilon_n c_n^\dagger c_n - \sum_n t_{n,n+1} (c_n c_{n+1}^\dagger + h.c) \quad (2.1)$$

1L 模型是將 DNA 簡化成以鹼基對爲基本單元的鹼基對長鏈，在（2.1）式中， n 是位置， t 是鹼基對之間的躍遷積分，這裡取 0.4eV ， c 與 c^\dagger 是湮滅與創生算符， ε 是位能與鹼基對的種類有關，其中 $\varepsilon_A = 8.24\text{eV}$ 、 $\varepsilon_T = 9.14\text{eV}$ 、 $\varepsilon_C = 8.87\text{eV}$ 、 $\varepsilon_G = 7.75\text{eV}$ [21]。

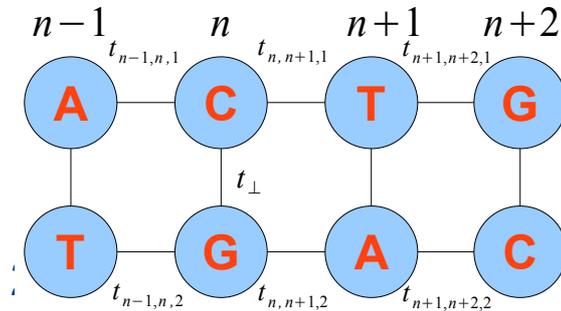


圖 2.2 雙股模型（2L）考慮 DNA 上的雙股各自的鹼基，並考慮鹼基之間電子的躍遷行爲。

$$H = \sum_{n,l=1,2} \varepsilon_{n,l} c_{n,l}^\dagger c_{n,l} - \sum_{n,l=1,2} t_{n,n+1,l} (c_{n,l} c_{n+1,l}^\dagger + h.c.) - \sum_n t_{n,\perp} (c_{n,l} c_{n,2}^\dagger + h.c.) \quad (2.2)$$

2L 模型是將 DAN 簡化成以鹼基為單位。(2.2) 式中， $l=1,2$ 表示 DNA 的兩股， $t_{n,n+1,l}$ 是同股鹼基之間的躍遷積分，相似的鹼基對 (ex:AT/AT,GC/CG) 取 0.35eV ，不相似的鹼基對 (ex:AT/CG,AT/GC) 取 0.17eV ， $t_{n,\perp}$ 表示鹼基對之間的躍遷積分，取 0.1eV 。

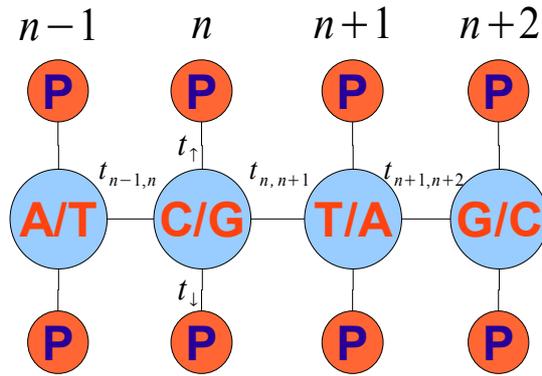


圖 2.3 魚骨模型 (FB) 除原本的單股模型外，再多考慮電子在 base pair 與主鏈上磷酸的躍遷行為。

$$H = \sum_n \varepsilon_n c_n^\dagger c_n + \sum_{n,u=\uparrow,\downarrow} \varepsilon_{n,u} c_{n,u} c_{n,u}^\dagger - \sum_n t_{n,n+1} (c_n c_{n+1}^\dagger + h.c.) - \sum_{n,u=\uparrow,\downarrow} t_{n,u} (c_n c_{n,u}^\dagger + h.c.) \quad (2.3)$$

FB 模型是除原本 1L 模型外，同時考慮主鏈上的位能和電子從鹼基對躍遷到主鏈上的影響。(2.3) 式中， $u=\uparrow$ 或 \downarrow 分別代表上排或下排的主幹， $t_{n,u}$ 代表鹼基對到主幹之間的躍遷積分， $\varepsilon_{n,u}$ 代表主幹上磷酸的位能，取 8.5eV 。

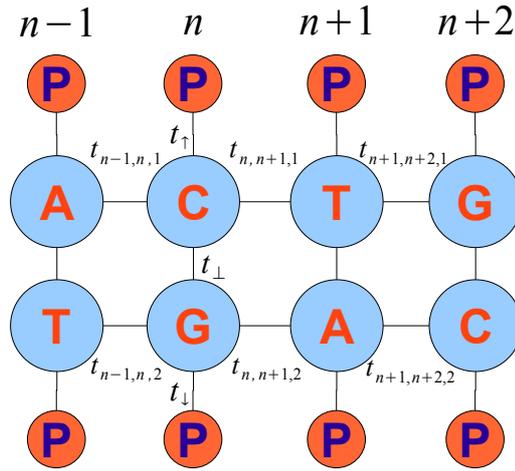


圖 2.4 梯子模型 (LM) 考慮雙股模型外，也考慮電子在鹼基與主幹上的躍遷。

$$H = \sum_{n,l=1,2} \varepsilon_{n,l} c_{n,l}^\dagger c_{n,l} - \sum_{n,l=1,2} t_{n,n+1,l} (c_{n,l} c_{n+1,l}^\dagger + h.c.) - \sum_n t_{n,\uparrow} (c_{n,1} c_{n,2}^\dagger + h.c.) - \sum_n t_{n,\downarrow} (c_{n,2} c_{n,1}^\dagger + h.c.) \quad (2.4)$$

LM 模型是原本的 2L 模型再多考慮了主幹的影響。(2.4) 式中， $t_{1,\uparrow}$ 與 $t_{2,\downarrow}$ 代表鹼基與主幹間的躍遷積分。

2.2 穿隧矩陣 (Transfer Matrix Method)

單股模型 (1L) 導電係數公式計算

$$H = \sum_n \varepsilon_n c_n^\dagger c_n - \sum_n t_{n,n+1} (c_n c_{n+1}^\dagger + h.c.)$$

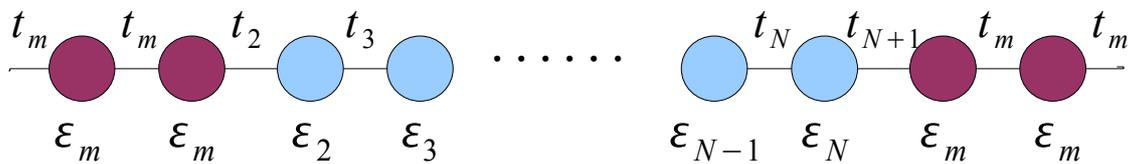


圖 2.5 DNA 外接金屬電極，對應於在生物體內放出與吸收電子的 BER 酵素

圖 2.5 是 1L 模型兩側接金屬電極，第 2 個到第 N 個位置是 DNA，依鹼基對不同取不同的位能， A 取 8.24eV 、 T 取 9.14eV 、 C 取 8.87eV 、 G

取 7.75eV ; $(-\infty, 1]$ 與 $[N+1, +\infty)$ 是電極，其位能取 7.75eV 。 $t_{n,n+1}$ 是 n 與 $n+1$ 個之間的遷躍項，這裡取 0.4eV 。

Wave function $|\Psi\rangle = \sum a_n |n\rangle$ ， $|n\rangle$ 表示電子在第 n 個位置之態， a_n 爲 Wave function amplitude 。

$$\begin{aligned}
 H|\psi\rangle &= E|\psi\rangle \\
 \Rightarrow \sum_n \varepsilon_n a_n |n\rangle + \sum_n a_{n+1} t_{n,n+1} |n\rangle + \sum_n a_n t_{n,n+1} |n+1\rangle &= E \sum_n a_n |\psi\rangle \\
 \Rightarrow \varepsilon_n a_n + a_{n+1} t_{n,n+1} + a_{n-1} t_{n-1,n} &= E a_n \\
 \Rightarrow a_{n+1} &= \frac{E - \varepsilon_n}{t_{n,n+1}} - \frac{t_{n-1,n}}{t_{n,n+1}} a_{n-1} \\
 \Rightarrow \begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} &= \begin{pmatrix} \frac{E - \varepsilon_n}{t_{n,n+1}} & -\frac{t_{n-1,n}}{t_{n,n+1}} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = M_n \begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = M_n M_{n-1} \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix} \\
 \Rightarrow \begin{pmatrix} a_{N+2} \\ a_{N+1} \end{pmatrix} &= M_{N+1} M_N \dots M_1 \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} = P(N) \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} \quad (2.5)
 \end{aligned}$$

如圖 2.6 在電極的區段中爲自由電子，故其波函數爲平面波。

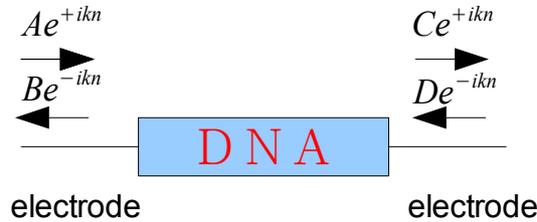


圖 2.6

$$a_n = Ae^{ikn} + Be^{-ikn} \quad \text{當 } n \leq 1 \quad (2.6)$$

$$a_n = Ce^{ikn} + De^{-ikn} \quad \text{當 } n \geq N \quad (2.7)$$

當 $E = \varepsilon_m + 2t_m \cos k$

$$\Rightarrow \begin{pmatrix} a_{N+2} \\ a_{N+1} \end{pmatrix} = \begin{pmatrix} e^{-ik} & e^{ik} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} e^{-ik(N+1)} & D \\ e^{ik(N+1)} & C \end{pmatrix} = \begin{pmatrix} e^{-ik} & e^{ik} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} e^{-ik(N+1)} & 0 \\ 0 & e^{ik(N+1)} \end{pmatrix} \begin{pmatrix} D \\ C \end{pmatrix} \equiv S\theta^{-1} \begin{pmatrix} D \\ C \end{pmatrix} \quad (2.8)$$

$$\begin{pmatrix} a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} e^{-ik} & e^{ik} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} B \\ A \end{pmatrix} = S \begin{pmatrix} B \\ A \end{pmatrix} \quad (2.9)$$

將 (2.8) (2.9) 式代入 (2.5) 式中

$$\begin{aligned} \begin{pmatrix} a^{N+2} \\ a^{N+1} \end{pmatrix} &= P(N) \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} \\ \Rightarrow S \Theta^{-1} \begin{pmatrix} D \\ C \end{pmatrix} &= P(N) S \begin{pmatrix} B \\ A \end{pmatrix} \\ \Rightarrow \begin{pmatrix} D \\ C \end{pmatrix} &= \Theta S^{-1} P(N) S \begin{pmatrix} B \\ A \end{pmatrix} \end{aligned}$$

$P(N) = M_{N+1} \cdot M_N \cdot M_{N-1} \cdots M_2 \cdot M_1$ 爲一個 2×2 矩陣 $\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$ 。

$$\begin{aligned} S &= \begin{pmatrix} e^{-ik} & e^{ik} \\ 1 & 1 \end{pmatrix} \Rightarrow S^{-1} = \frac{1}{e^{-ik} - e^{ik}} \begin{pmatrix} 1 & -e^{ik} \\ -1 & e^{-ik} \end{pmatrix} = \frac{1}{2i \sin k} \begin{pmatrix} -1 & e^{ik} \\ 1 & -e^{-ik} \end{pmatrix} \\ \Rightarrow \Theta S^{-1} P(N) S &= \begin{pmatrix} e^{ik(N+1)} & 0 \\ 0 & e^{-ik(N+1)} \end{pmatrix} \frac{1}{2i \sin k} \begin{pmatrix} -1 & e^{ik} \\ 1 & -e^{-ik} \end{pmatrix} \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} e^{-ik} & e^{ik} \\ 1 & 1 \end{pmatrix} = \tau(N) \end{aligned}$$

其中

$$\begin{pmatrix} -1 & e^{ik} \\ 1 & -e^{-ik} \end{pmatrix} \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} e^{ik} & e^{-ik} \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -P_{11}e^{-ik} - P_{12} + P_{21} + P_{22}e^{ik} & -P_{11}e^{ik} - P_{12} + P_{21}e^{2ik} + P_{22}e^{ik} \\ P_{11}e^{-ik} + P_{12} - P_{21}e^{-2ik} - P_{22}e^{-ik} & P_{11}e^{ik} + P_{12} - P_{21} - P_{22}e^{-ik} \end{pmatrix}$$

(2.10)

考慮一個平面波由左向右入射 $\Rightarrow A=1, B=\gamma, C=t, D=0$

$$\begin{pmatrix} 0 \\ t \end{pmatrix} = \tau(N) \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 0 \\ t \end{pmatrix} = \begin{pmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{pmatrix} \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \Rightarrow \begin{aligned} \gamma \tau_{11} + \tau_{12} &= 0 \\ \gamma \tau_{21} + \tau_{22} &= t \end{aligned}$$

$$\Rightarrow \gamma = -\frac{\tau_{12}}{\tau_{11}}, t = \tau_{22} - \frac{\tau_{12}\tau_{21}}{\tau_{11}} = \frac{\tau_{22}\tau_{11} - \tau_{12}\tau_{21}}{\tau_{11}} = \frac{\det(\tau)}{\tau_{11}}$$

$$\det(\tau) = \det(\Theta S^{-1} P(N) S) = \det \Theta \prod_{i=1}^{N+1} \det M_i \det(S^{-1}) \det(S)$$

$$\Rightarrow \det(\tau) = \prod_{i=1}^{N+1} \begin{pmatrix} E - \varepsilon & -t_i \\ t_{i+1} & t_{i+1} \\ 1 & 0 \end{pmatrix} = \prod_{i=1}^{N+1} \frac{t_i}{t_{i+1}} = \frac{t_1}{t_{N+2}} = \frac{t_m}{t_m} = 1$$

$$\Rightarrow t = \frac{1}{\tau_{11}} \Rightarrow \text{transmission coefficient} \circ$$

$$T_N(E) = |t|^2 = \frac{1}{|\tau_{11}(N)|^2}$$

由 (2.10) 式可得 $\tau_{11}(N) = \frac{e^{ik(N+1)}}{2i \sin k} (-p_{11}e^{-ik} - P_{12} + P_{21} + P_{22}e^{ik})$

$$\Rightarrow T_N(E) = \frac{1}{|\tau_{11}(N)|^2} = \frac{4 \sin^2 k}{(P_{11}e^{-ik} + P_{12} - P_{21} - P_{22}e^{ik})(P_{11}e^{ik} + P_{12} - P_{21} - P_{22}e^{-ik})}$$

$$\Rightarrow T_N(E) = \frac{4 - \left(\frac{E - \epsilon_m}{t_m}\right)^2}{\sum_{i,j=1,2} p_{i,j}^2 + 2 - \left(\frac{E - \epsilon_m}{t_m}\right)^2 P_{11}P_{21} + \left(\frac{E - \epsilon_m}{t_m}\right)(P_{11} - P_{22})(P_{12} - P_{21})} \quad (2.11)$$

在導出傳輸係數公式後，我們可以計算出 DNA 點突變前後的變化。例如下圖紅線是 $(GC)_{30}$ 對應不同的入射能量所得的 $T(E)$ 綠線則是前述序列中的第三十個鹼基，由 C 替換成 G 所得到的 $T(E)$ 。

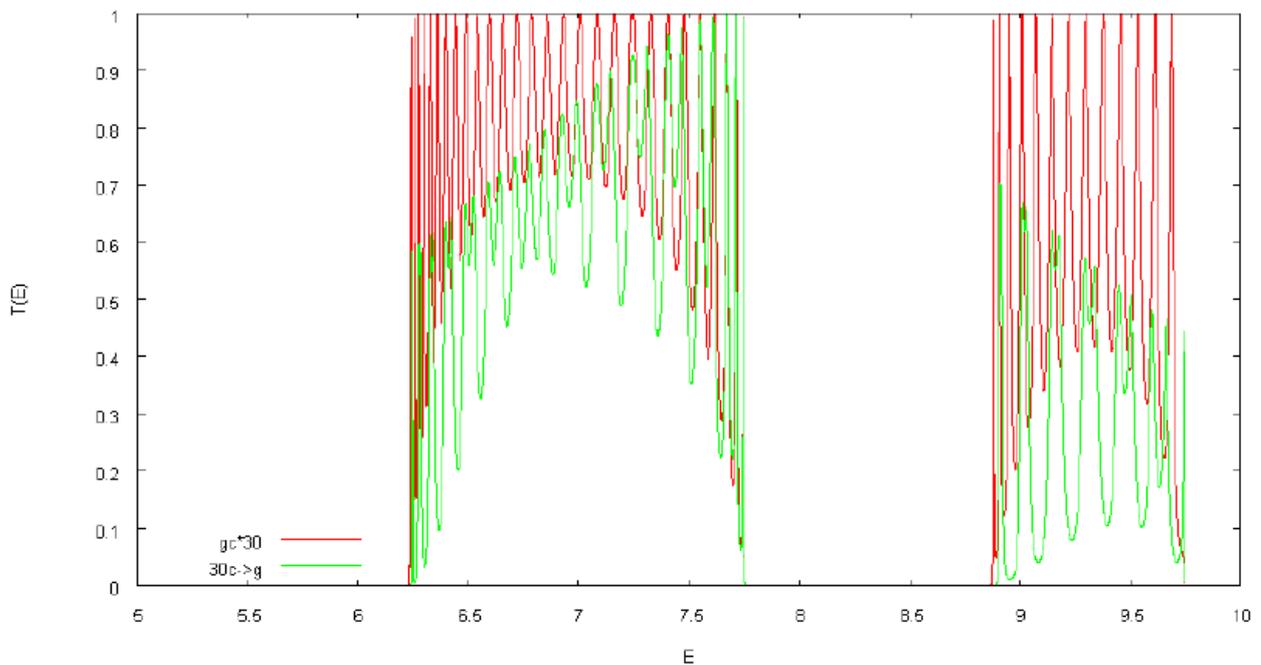


圖 2.7 計算連續 30 個 GC 串連總長度 60 個鹼基單位的 DNA 序列，在對應不同的入射能量時的導電係數，得到紅線的圖形；將前述序列的第 30 個鹼基由 C 換成 G，則得到綠線的圖形，由此可以看出，即使只改變一個鹼基，也可能造成傳輸係數上明顯的變化。

2.3 傳輸係數變化定義

我們將一個總長度為 N 的完整 DNA 序列定義為 $S=(s_1,s_2,\dots,s_N)$ ， S 是該 DNA 序列所有鹼基的集合， s_n 為該 DNA 第 n 個位置的鹼基，即為 A 、 T 、 C 、 G 。我們將起始位置為 n ，長度 L 的 DNA 子序列定義為 $S_{n,L}$ ， $S_{n,L}(m)=S(n-1+m)$ ， $S_{n,L}(m)=S(n-1+m)$ 。

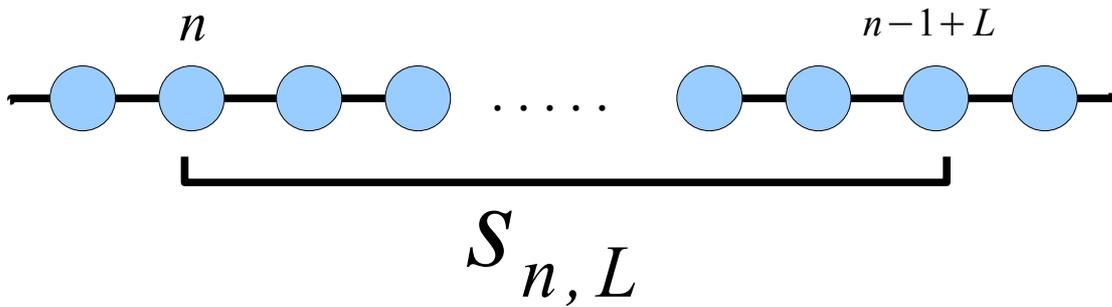


圖 2.8 $S_{j,L}$ 定義：以第 j 個位置為起始位置，長度 L 的 DNA 片段序列。

$S_{n,L}$ 所對應的傳輸係數我們定義為 $T_{n,L}(E)$ 。

假設在序列中的 k_{th} 位置為突變點，包含此位置的子序列為 $S_{j,L}$ ，當 k_{th} 的鹼基由 s_k 突變為 s 時，此序列我們稱之為 $S_{j,L}^{k,s}$ ，突變前後的傳輸係數各別為 $T_{j,L}(E)$ 和 $T_{j,L}^{k,s}(E)$ 。

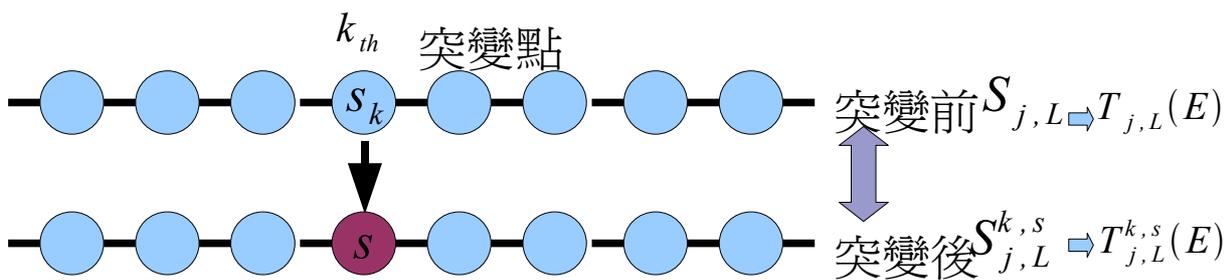


圖 2.9 突變前與突變後的，對應不同的入射能量，各自計算出傳輸係數後再相減，可以得到突變前與突變後的傳輸係數改變量。

藉由所有可能入射能量，計算出 $s_{j,L}$ 和 $s_{j,L}^{k,s}$ 所對應的 $T_{j,L}(E)$ 和 $T_{j,L}^{k,s}(E)$ 我們得以計算出 k_{th} 位置突變前後的傳輸係數改變量 $\bar{\Delta}_{j,L}^{k,s}$ 。

$$\bar{\Delta}_{j,L}^{k,s} = \frac{1}{E_1 - E_0} \int_{E_0}^{E_1} |T_{j,L}(E) - T_{j,L}^{k,s}(E)|^2 dE \quad (2.11)$$

再以 (2.11) 式，計算出包含突變 k_{th} 位置，長度 L 的所有子序列，得到突變 (k,s) 所引起的 CT 變化，我們將之定義為 $\Gamma(k,s,L)$

$$\Gamma(k,s,L) = \frac{1}{L} \sum_{j=k-L+1}^k \bar{\Delta}_{j,L}^{k,s} \quad (2.12)$$

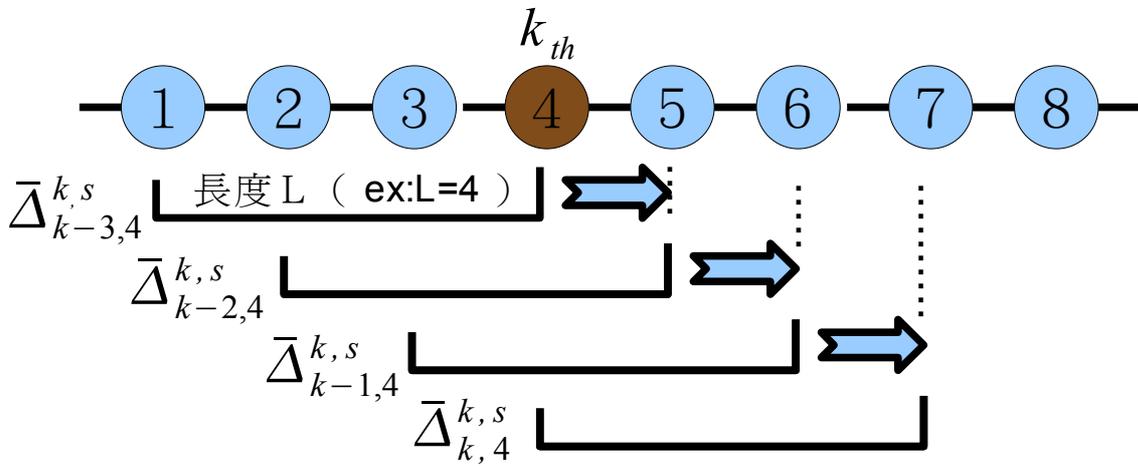


圖 2.10 例如對所有包含第 4 個鹼基，且長度 4 的子序列各別計算出 $\bar{\Delta}_{j,4}^{k,s}$ 。

第三章 統計結果分析

3.1 單一位置突變前後的 charge transport 變化比較

在正常的 DNA 序列的一個位置上，可以有三種不同的突變。為容易比較致癌突變與非致癌突變的差異，我們選 p53 序列中的 14585th 個鹼基來當做例子，該位置其正常的鹼基為 C，且唯有突變成 T 時才會致癌，在 p53 所有恰有一種突變會致癌的鹼基中，該位置的病例數最多。

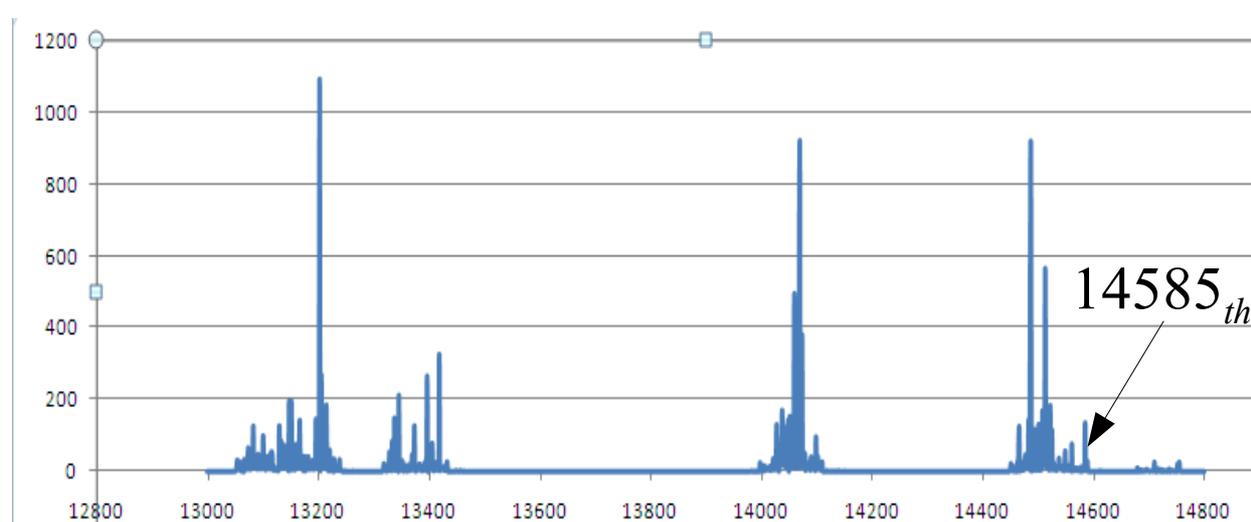


圖 3.1 統計癌症病患們的 p53 序列上突變的位置，得到其分佈的圖形。

縱軸是病例數，橫軸是 p53 序列發生突變的位置。

我們計算包含該位置長度為 20 的所有可能突變鹼基所得到的序列，在對應不同的入射能量下的傳輸係數，與原先的序列的傳輸係數做比較，得到圖 3.2 的結果。

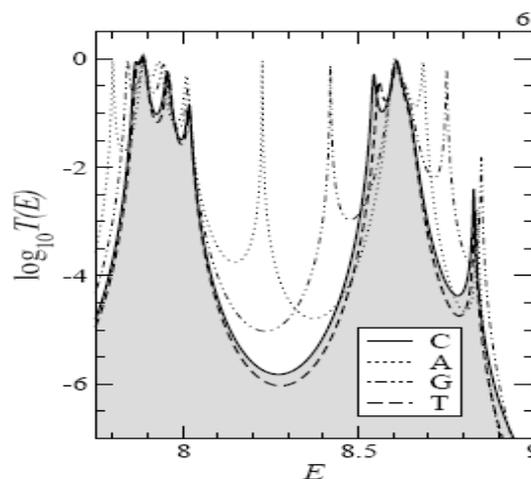


圖 3.2 在可能的入射能量下,所計算出來 14585_{th} 位置上鹼基為 C 與所有可能突變的傳輸係數。其中致癌突變 T 與原始鹼基 C 的傳輸係數差異最小。

由圖 3.2 看出致癌突變所引起的傳輸係數改變似乎比非致癌突變所引起的改變來的少。在不同長度下，我們也得到同樣的結果如表 3.1。

s	L	$1L$
$C \rightarrow A$	20	23.1
$C \rightarrow G$	20	37.6
$C \rightarrow T$	20	5.63
$C \rightarrow A$	30	15.7
$C \rightarrow G$	30	21.4
$C \rightarrow T$	30	9.14
$C \rightarrow A$	40	1.16
$C \rightarrow G$	40	2.21
$C \rightarrow T$	40	0.40

表 3.1 在 14585_{th} 的位置上，我們取不同的長度，算出該位置各種可能突變所引起的傳輸數改變量，結果顯示出該位置上的致癌突變比非致癌突變所引起的傳輸係數改變量來的小。

由此結果我們提出一個假設，有些突變可以避免被探測CT(charge transport)變化的DNA損傷修補機制所修復，並且得以生存，而繼續發展成癌症腫瘤。對不同的傳輸長度 $L=10, \dots, 160$ 時，我們發現都具有此傾向。

3.2 致癌突變的傳輸係數在 codon 中的排名

我們從只比較致癌突變點上的所有 Γ 推廣到比較該突變點所屬 codon 的所有 Γ 。DNA 上的三個鹼基為一個 codon，一個 codon 編碼一個胺基酸，胺基酸共有二十種，表 3.2 為胺基酸與編碼該胺基酸的 codon 種類對照。

Lys	AAA	Ser	AGT	Cys	TGT	Pro	CCA
Lys	AAG	Ser	AGC	Cys	TGC	Pro	CCT
Asn	AAT	Ser	TCA	Glu	CAA	Pro	CCC
Asn	AAC	Ser	TCT	Glu	CAG	Pro	CCG
Ile	ATA	Ser	TCC	Glu	GAA		
Ile	ATT	Ser	TCG	Glu	GAG		
Ile	ATC	Term	TAA	Val	GTA		
Met	ATG	Term	TAG	Val	GTT		
Thr	ACA	Term	TGA	Val	GTC		
Thr	ACT	Phe	TTT	Val	GTG		
Thr	ACC	Phe	TTC	Asp	GAT		
Thr	ACG	Len	TTA	Asp	GAC		
Arg	AGA	Len	TTG	Ala	GCA		
Arg	AGG	Len	CTA	Ala	GCT		
Arg	CGA	Len	CTT	Ala	GCC		
Arg	CGT	Len	CTC	Ala	GCG		
Arg	CGC	Len	CTG	Gly	GGA		
Arg	CGG	Trp	TGG	Gly	GGT		
Tyr	TAT	His	CAT	Gly	GGC		
Tyr	TAC	His	CAC	Gly	GGG		

表 3.2 codon 的長度為 3 共有 64 種組合，共編碼 20 種胺基酸，有些 codon 會編碼相同的胺基酸。

由於一個 codon 編碼一個胺基酸，因此當 codon 上的一個鹼基發生突變時，有可能影響胺基酸的正常編碼。一個 codon 上有三個鹼基，每個鹼基有三個不同的可能突變，因此總共有九種突變，九種突變的 Γ 我們由小到大給予排名。

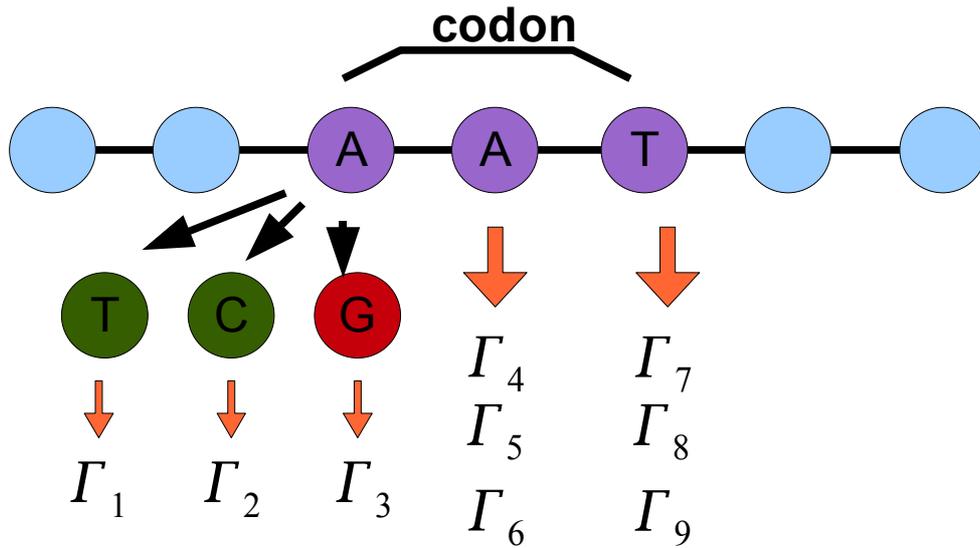


圖 3.3 假設在此 AAT 的 codon 上唯有第一個 A 突變成 G 時才會致癌，我們將依此比較突變成 Γ_3 與在該 codon 上的其他非致癌突變所算出來的 Γ ，由小到大做排名。

我們比較致癌突變在不同長度時，其 Γ 在 codon 上所有可能突變的排名，以 BRCA1、NF1 為例，得到圖 3.4、圖 3.5。

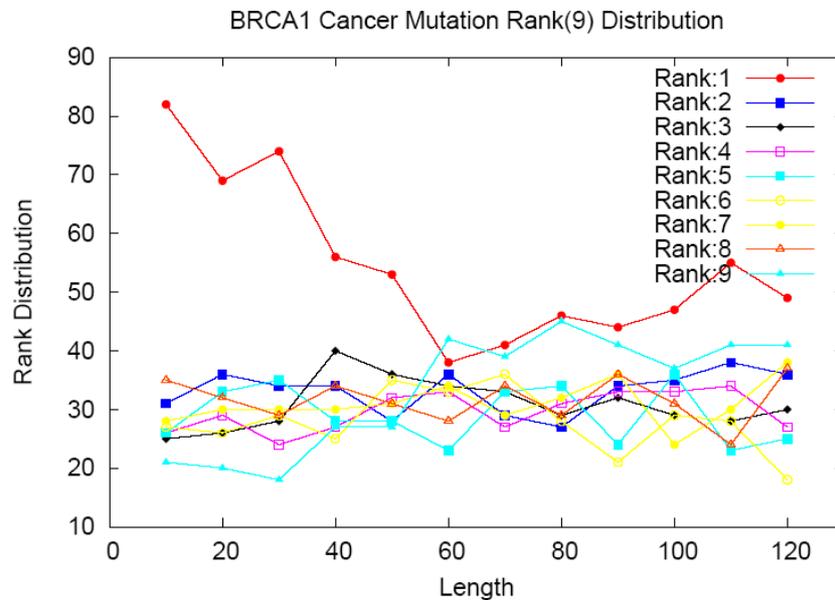


圖 3.4

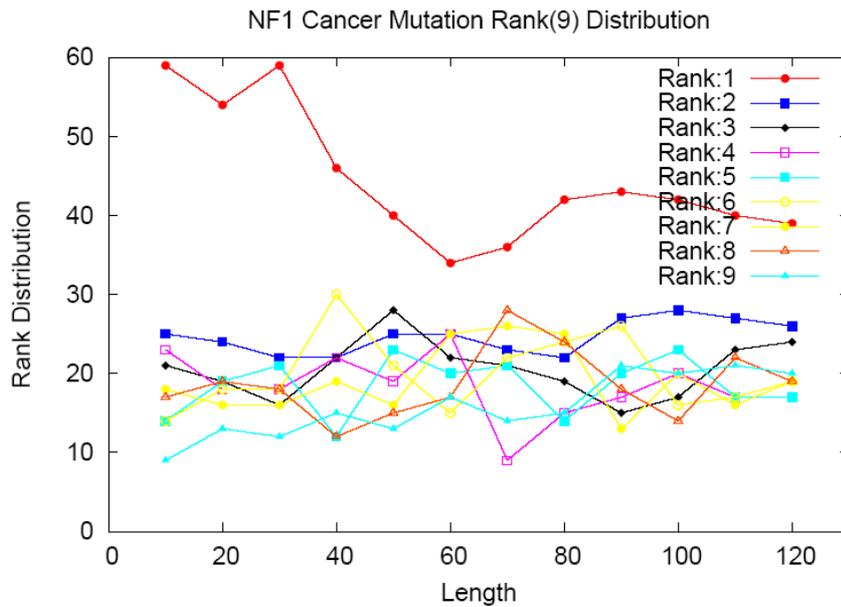


圖 3.5

以相同方式分析 35 個癌症相關基因序列上的所有致癌突變，得到 圖 3.6。

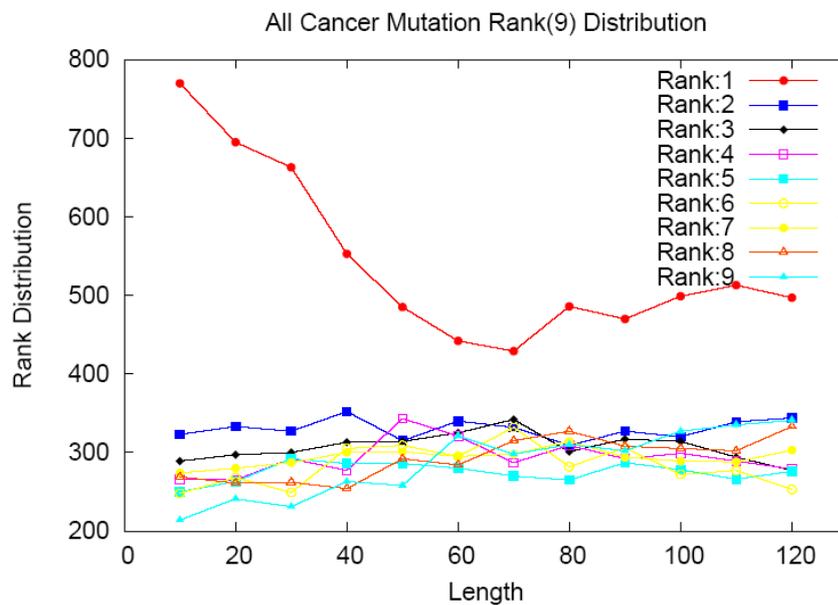


圖 3.6 全部 2903 個致癌突變的 I ，在 codon 上的排名分佈。

對於致癌突變數小於 100 筆的 DNA 序列，我們僅將排名分成三類， I 名次從 1~3 名的為一類，4~6 名為一類，7~9 名的算一類。以下為 WT1、CDH1 的圖形，圖 3.7、圖 3.8。

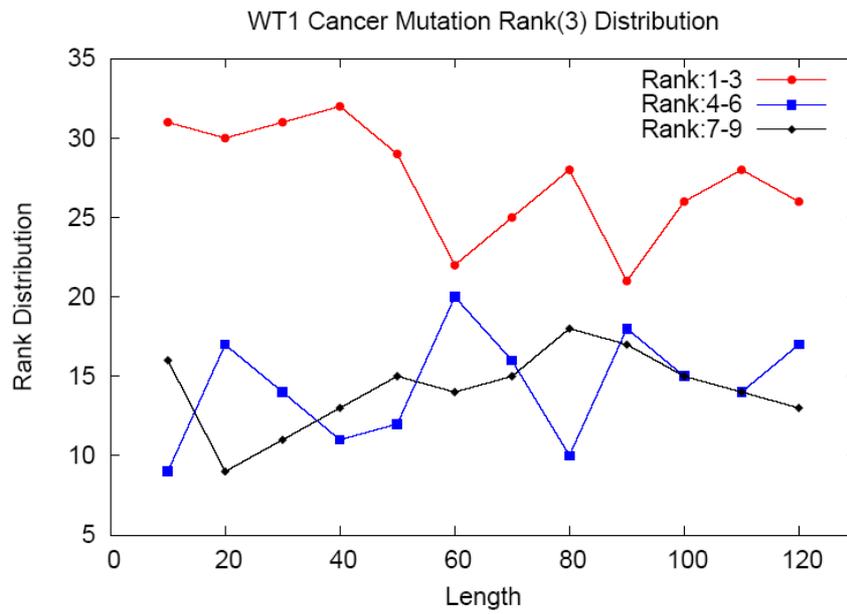


圖 3.7

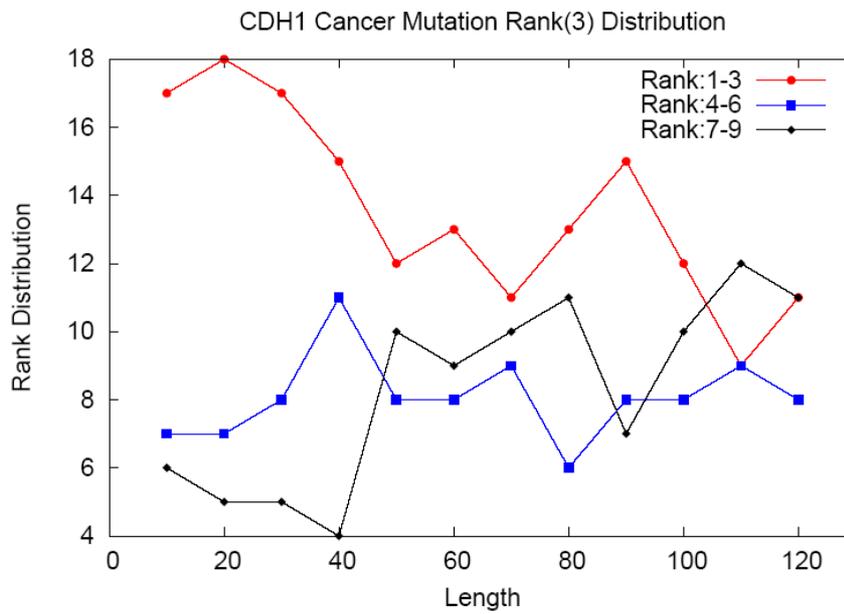


圖 3.8

以相同的分類方式，再次分析全部 35 個癌症相關基因，得到 圖 3.9。

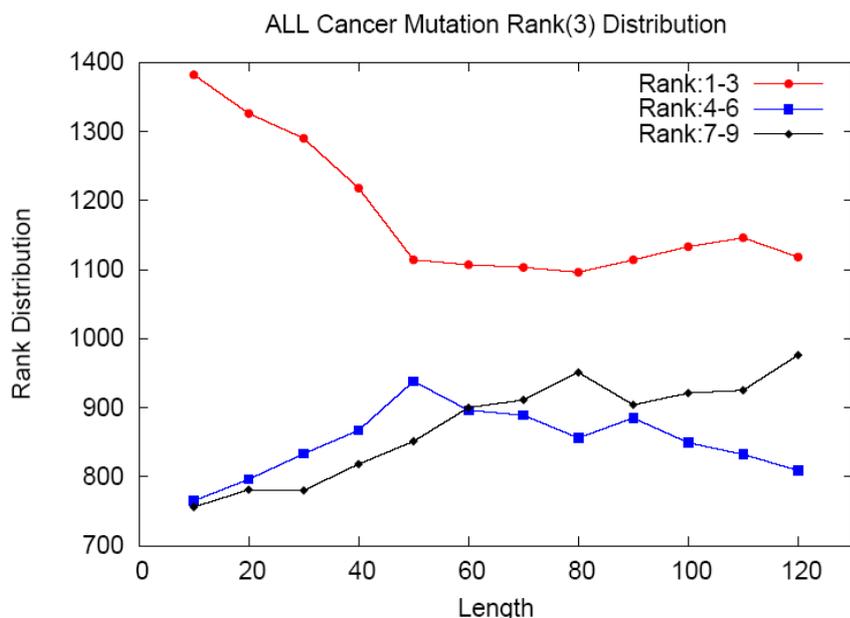


圖 3.9

更多的分析結果詳見附錄一，在不同的長度下計算所得到的 Γ ，其致癌突變的 Γ 在 codon 上的名次都是分佈在第一名的最多。我們猜測正因為是致癌突變的 Γ 小，所以未被人體的 DNA 電性變化檢測機制所發現，而未將其修復，最終發展成癌症。

3.3 致癌突變的傳輸係數在序列中的排名

以 p53 序列為例，全長 20303 的 p53 序列，共有 60909 種可能的點突變，依其突變所得到的 Γ ，由小到大做排名。將排名以 0~10%，.....，90~100% 做區隔，所有排名平均分佈在各個區間中，得到圖 3.10 的紅色長條圖。

在此使用 IARC(International Agency for Resesrch on Cancer)網站上 p53 序列的 2003 個致癌突變點為分析資料[22]。p53 序列中會致癌的 2003 個點突變 Γ 值，其排名分佈則為圖 3.10 的紅色長條圖。由圖可以看出 30% 致癌突變，其 Γ 排名是分佈在前 10% 的，這結果支持我們的假說。

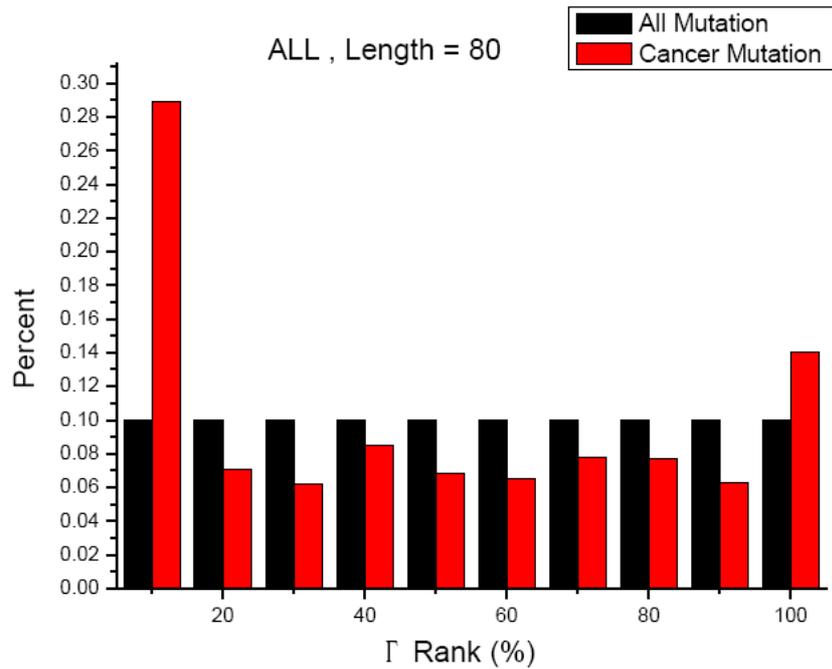


圖 3.10 計算出 p53 序列 60909 種可能突變的 Γ ，並由小至大給予名次，並找出致癌突變的 Γ 排名，以此畫出排名的分佈，似乎 30% 的致癌突變的 Γ 都排在前 10%。

依突變的前後的鹼基之不同，分成由 A 突變到 T，由 A 突變到 C，.....，由 G 突變到 C 等 12 種突變。在此序列，12 種突變的整體可能突變和致癌突變的整理於表 3.3。

原鹼基	突變成其他鹼基	整體可能突變的次數	會致癌突變的次數
A	T	5246	137
A	C	5246	113
A	G	5246	158
T	A	4980	121
T	C	4980	142
T	G	4980	109
C	A	4970	165
C	T	4970	284
C	G	4970	156
G	A	5107	252
G	T	5107	202
G	C	5107	164

表 3.3 在 p53 序列中所有 12 種可能突變與致癌突變的各別次數。

將 p53 序列的所有可能突變分類為上述的 12 種，並依其在全部突變中的排名做分佈，以由 T 突變為 A，長度 80 為例，得到圖 3.11 的結果。

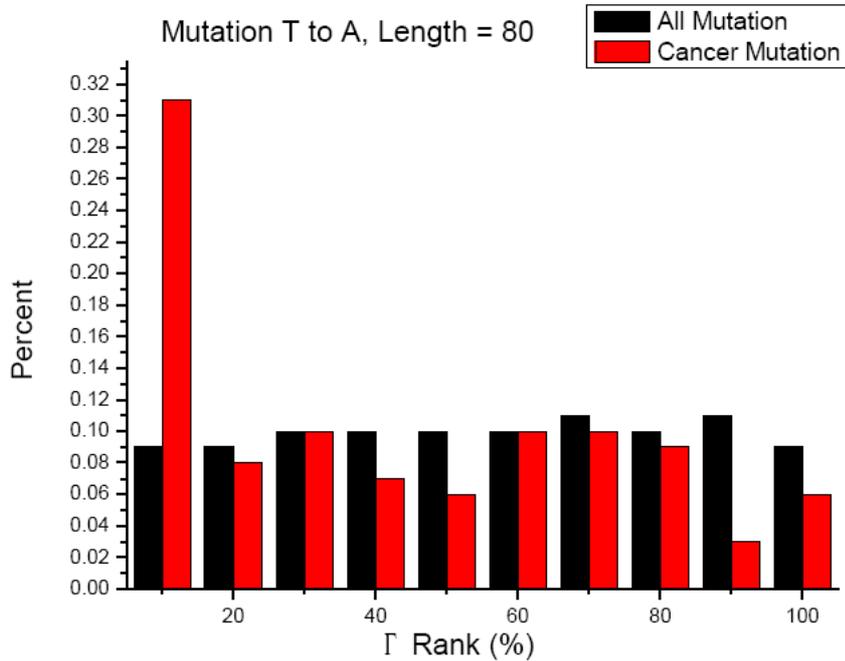


圖 3.11 分析的方式同圖 3.2，但在這圖中，我們只看致癌突變與全體突變中由 T 突變到 A 的排名分佈。

更多分析結果，請見附錄二，由圖 3.11 觀察出，致癌突變分佈傾向於排名前 50%，為量化此一傾向我們依圖 3.12，以 55% 的位置做為支點，向左延伸到 10%，向右延伸到 100%，成為力臂。以每個區間中的致癌突變分佈百分比 p_{cancer} 和全部突變分佈百分比 p_{all} 相減，再乘以距離支點的長度做為加權，算出每個區間中對致癌突變傾向於 Γ 較小的數值。得到表 3.4，最後再將每個區間所貢獻的數值相加起來，以此為例，得到 14.23 的數字，我們將此數值定義為傾向值 Λ ，正值越大越支持我們的假說。

$$\Lambda = \sum_{n=1}^{10} (p_{cancer} - p_{all}) \times (55 - 10n), n=1, \dots, 10 \quad (3.1)$$

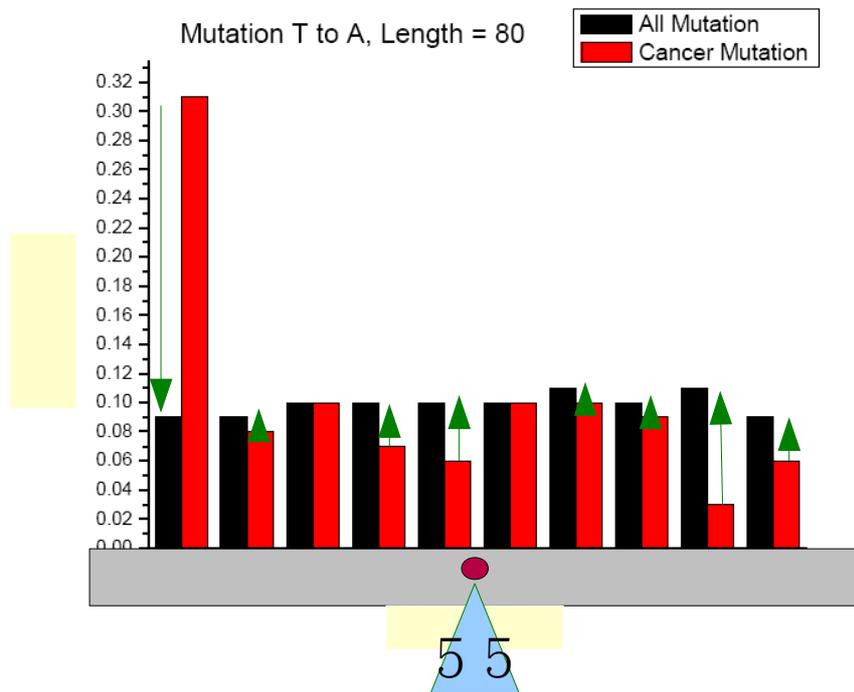


圖 3.12 以各位置上的致癌突變百分比減掉全部突變的分比，再乘以該位置離支點的距離作加權。

排名百分比	致癌突變百分比分佈	全部突變百分比分佈	計算	結果
0~10%	3.14E-001	0.09	$() \times (55 - 10)$	12.59
10~20%	8.26E-002	0.09	$() \times (55 - 20)$	-0.46
20~30%	9.92E-002	0.1	$() \times (55 - 30)$	0.02
30~40%	6.61E-002	0.1	$() \times (55 - 40)$	-1.68
40~50%	5.79E-002	0.1	$() \times (55 - 50)$	-2.57
50~60%	9.92E-002	0.1	$() \times (55 - 60)$	-0.26
60~70%	9.92E-002	0.11	$() \times (55 - 70)$	-0.58
70~80%	9.09E-002	0.1	$() \times (55 - 80)$	-0.74
80~90%	3.31E-002	0.11	$() \times (55 - 90)$	-4.4
90~100%	5.79E-002	0.09	$() \times (55 - 100)$	-1.93
				Sum=14.23

表 3.4

以上述方式計算在 TP53 中 12 種突變在不同長度時的數值，結果如表 3.5 所列，僅有兩個 Δ 為負，以反白標示，這與我們的假說十分一致。

A → T		A → C		A → G	
Length	Δ	Length	數字	Length	Δ
20	1.41	20	3.36	20	4.05
40	4.04	40	4.94	40	5.56
60	3.48	60	4.38	60	4.45
80	5.54	80	7.45	80	6.07
T → A		T → C		T → G	
Length	Δ	Length	數字	Length	Δ
20	10.91	20	12.66	20	13.6
40	13.89	40	13.39	40	14.26
60	15.45	60	14.28	60	14.61
80	14.23	80	13.21	80	12.15
C → A		C → T		C → G	
Length	Δ	Length	Δ	Length	Δ
20	5.49	20	0.17	20	3.05
40	9.37	40	-0.95	40	6.62
60	9.96	60	-1.39	60	6.18
80	12.54	80	2.96	80	12.69
G → A		G → T		G → C	
Length	Δ	Length	Δ	Length	Δ
20	4.58	20	3.66	20	7.94
40	3.07	40	2.39	40	6.55
60	2.58	60	1.11	60	6.2
80	1.31	80	2.95	80	5.83

表 3.5

對於唯二的負值，我們觀察 TP53 序列由 C 突變 T 的排名分佈圖，圖 3.13、圖 3.14、圖 3.15、圖 3.16。

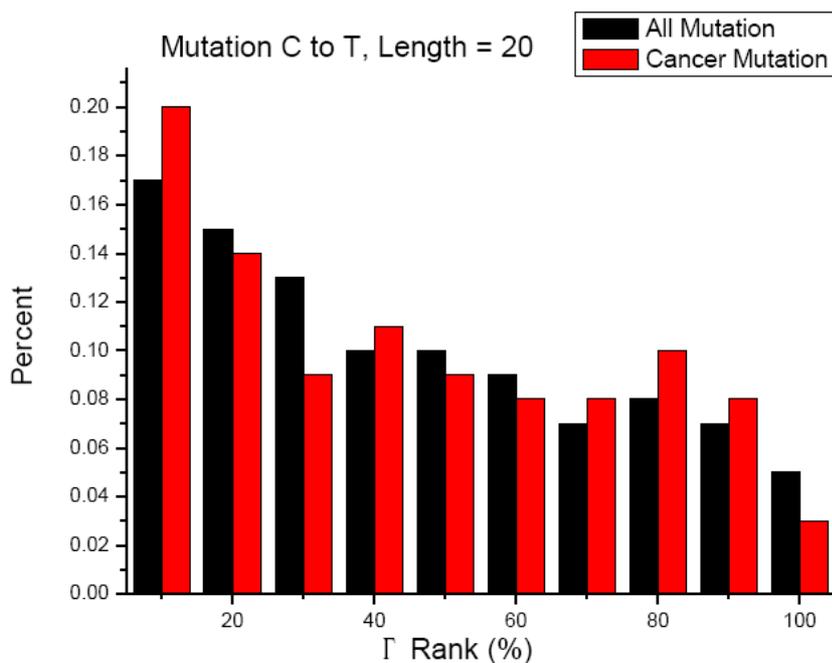


圖 3.13

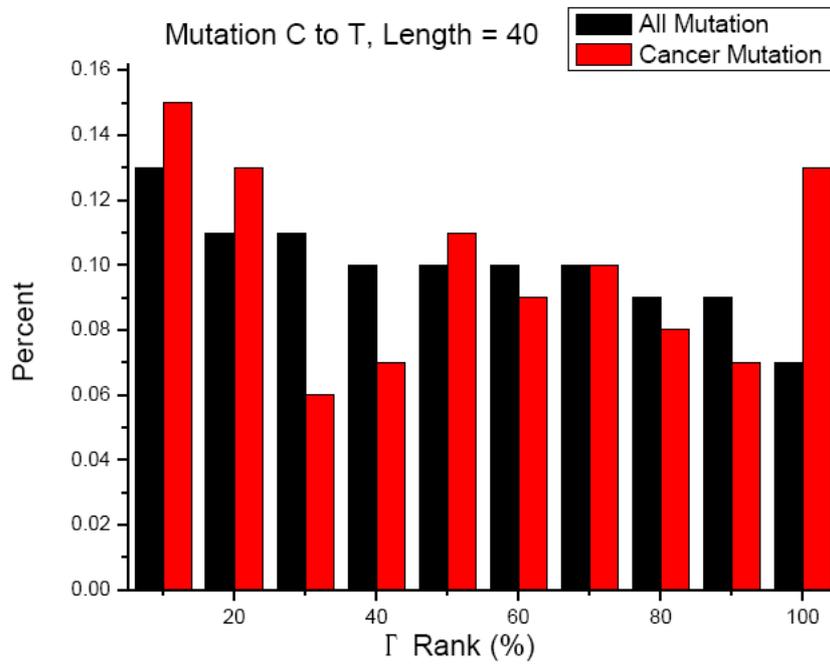


圖 3.14

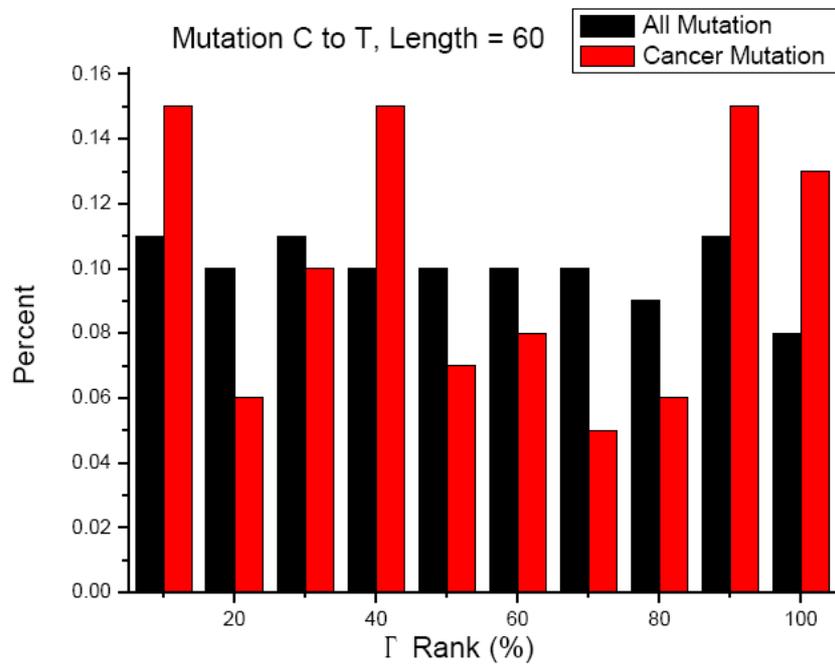


圖 3.15

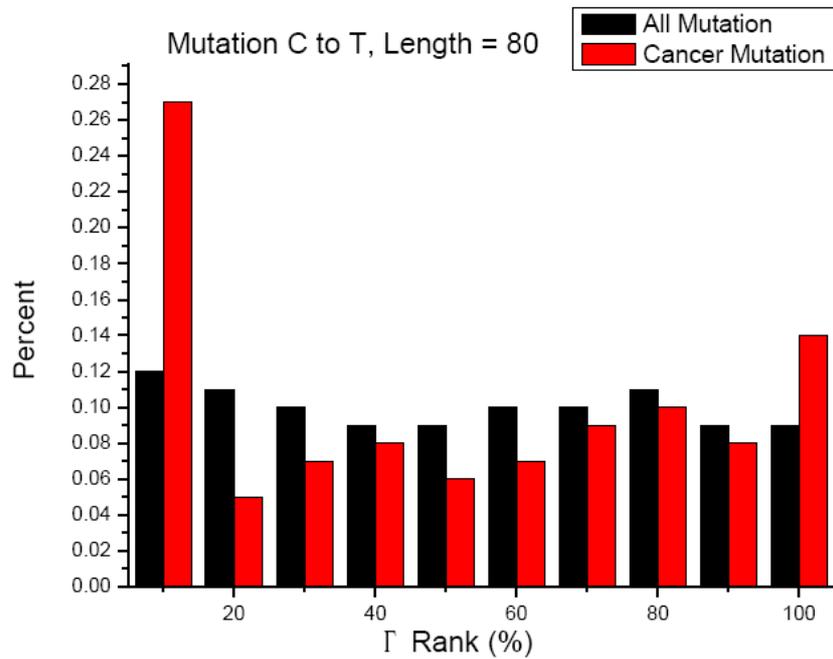


圖 3.16

由圖形可以看出在 TP53 序列上所有由 C 突變成 T 所計算出來的 Γ ，其分佈名次較前的數量為多，因此在此方式的分析中，沒有其他突變有明顯的正傾向值。

為何由 C 突變 T 的 Γ 排名傾向在前？我們認為是因為 C 與 T 的位能很接近，所以當由 C 突變成 T 時，造成的傳輸係數變化平均而言不大，因而使得由 C 突變 T 的 Γ 排名傾向在前。

第四章 致癌 DNA 序列與序列的相依性

分析方法：在相同長度下，計算每一種序列在該基因中出現的次數，同時也計算每一致病突變出現在每一種序列中的次數，然後計算出兩者的 correlation。

以長度 3 為例，則可能出現的序列組合共有 $4^3=64$ 種，如 AAA、AAT、AAG...GGC、GGG 等。

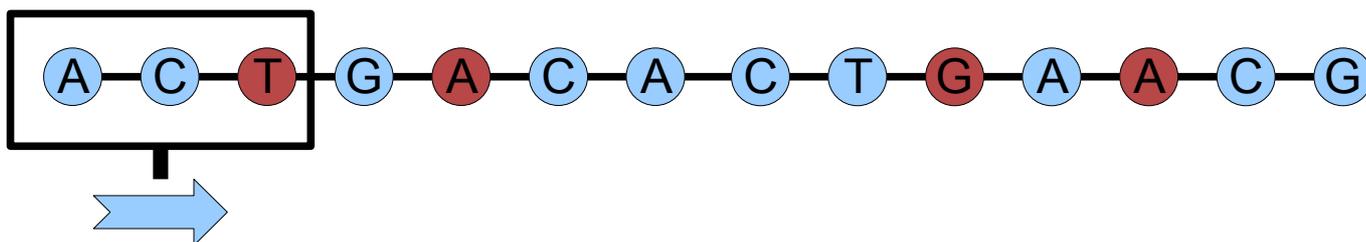


圖 4.1 以一個長度 3 的視窗掃過整條序列，如以此圖為例，則依序讀到的序列是 ACT、CTG、TGA、GAC、ACA、CAC、ACT、CTG、TGA、GAA、AAC、ACG，則記為 ACT 為 2 次，CTG 為 2 次，TCA 為 2 次，GAC 為 1 次，ACA 為 1 次，CAC 為 1 次，GAA 為 1 次，AAC 為 1 次，ACG 為 1 次。

計數每一種組合的序列在整條 DNA 序列中出現的次數 x_i ， i 為所有子序列，序列的 x_i 值如表 4.1 所示。

各種序列	出現次數
AAC	1
ACA	1
ACT	2
ACG	1
TGA	2
CAC	1
CTG	2
GAA	1
GAC	1

表 4.1

同時我們也計算包含致病突變點的各種序列出現的次數 y ，由於在一段長度上的致病突變點可能不止一個，又同一位置可能不只一種致病突變，爲了可以全部討論，我們設了三種不同的 y 值計算方式。

只問該片段序列是否有包括致病突變點的次數，稱之爲 y_s ；問該片段序列包括多少致病突變點，其計次法爲該片段中包含的致病突變位置全部納入計次，稱次爲 y_c ；問該片段序列包括了多少致病突變位置外，每個致病位置若含有一種以上會致病的突變，亦全部納入計次，稱次爲 y_m 。

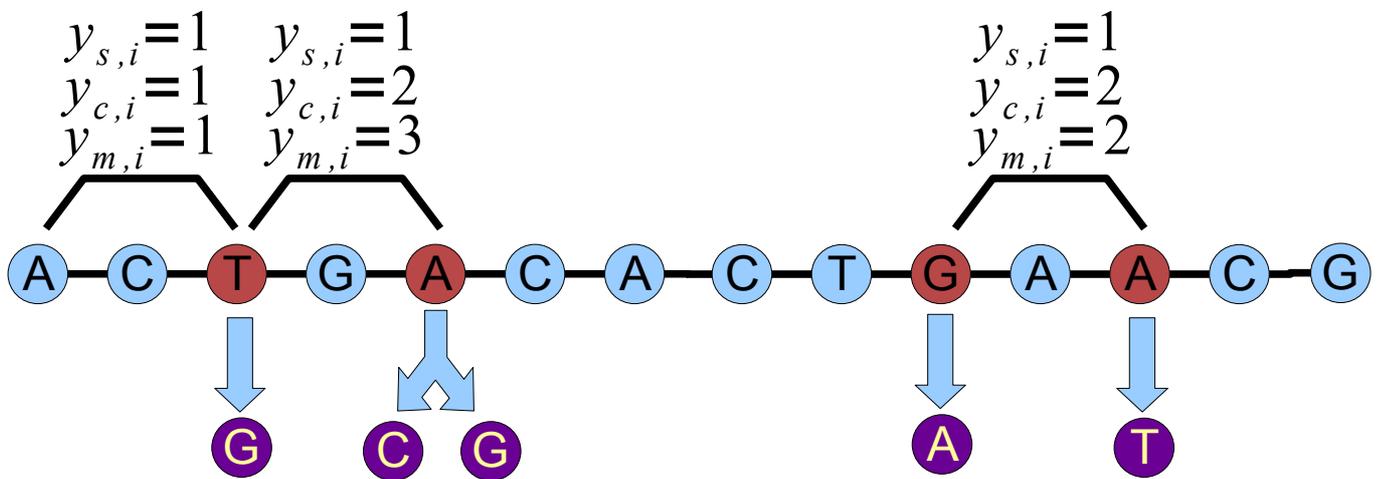


圖 4.2 致癌突變點以紅色標示之，紫色爲該點會致癌的突變鹼基。在長度取三的視窗掃描下，第一個視窗讀到的是 ACT 的序列，因爲含有致癌突變點，所以該序列的 y_s 總數加 1，因爲僅包含一個致癌突變點，所以該序列的 y_c 總數加 1，因爲包含的致癌突變點，只有一種突變會致癌，所以該序列的 y_m 總數加 1；第二個視窗讀到的 CTC 序列亦同；在第三個視窗所讀到的 TGA 序列，因爲含有致癌突變點，所以該序列的 y_s 總數加 1，因爲包含兩個致癌突變點，所以該序列的 y_c 總數加 2，因爲包含的致癌突變點，有三種突變會致癌，所以該序列的 y_m 總數加 3。

計數每一種包含致癌突變的序列在整條 DNA 序列中出現的 y_s 、 y_c 、 y_m 次數，如表 4.2。

各種序列	是否有致病突變 y_s	有多少個致病突變點 y_c	每個致病突變點有多少種致病突變 y_m
AAC	1	1	1
ACA	1	1	2
ACT	1	1	1
ACG	1	1	1
TGA	2	3	4
CTG	2	2	2
GAA	1	2	2
GAC	1	1	2

表 4.2

依公式 (4.1)，我們可以計算出在長度 3 時，以 y_s 、 y_c 、 y_m 各別代入式中的 y 算出 correlation 值，各別稱之為 c_s 、 c_c 、 c_m 。

$$correl(x, y) = \frac{\sum_1^{4^L} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^{4^L} (x_i - \bar{x})^2 \sum_1^{4^L} (y_i - \bar{y})^2}} \quad (4.1)$$

在 (4.1) 式中的 L 為我們所取的長度， $\bar{x} = \frac{\sum x_i}{L}$ ， $\bar{y} = \frac{\sum y_i}{L}$ ，這個公式是在計算基因所包含的各種序列數與致癌突變的序列的相關性，其值的大小範圍為：

$$-1 \leq correl(x, y) \leq 1$$

若呈現負相關，代表整串序列中與包含致癌突變點的序列相同的序列比例為低，意即包含致癌突變點的序列是屬於特別的序列，這樣的序列容易發生突變；而呈現正相關則代表整串序列中與包含致癌突變點的序列相同的序列比例為高，表示包含致癌突變點的序列只是在整串序列中出現的次數多，所以成為致癌突變的機會也高而已。

以此方式計算長度 1~7 的 correlation，以 TSHR 為例，可畫成圖 4.2，更多結果請參見附錄三。

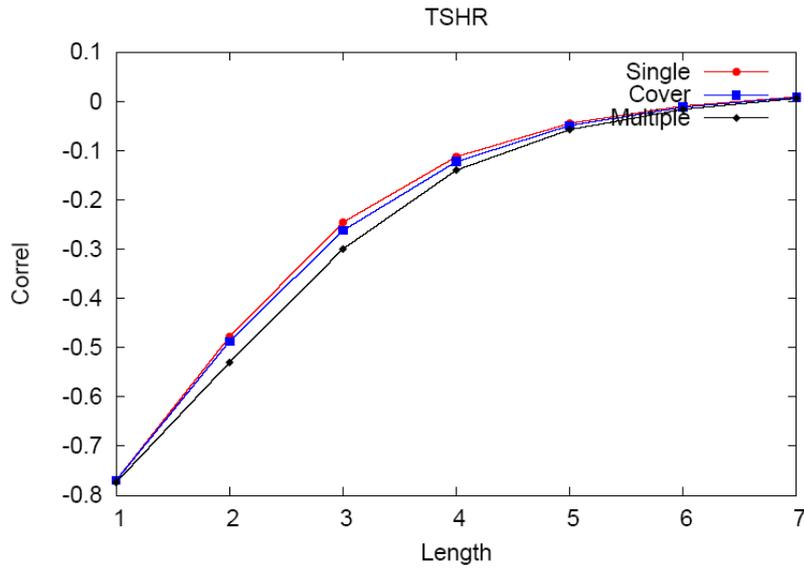


圖 4.3 線段-Single、線段-Cover、線段-Multiple 各別為 c_s 、 c_c 、 c_m 在取不同長度時的函數圖形。

更多分析結果，請看附錄三，從圖形可看出在 TSHR 中，我們取長度 1~7 時，致癌突變序列與整段序列中出現過的序列次數相較，都是呈現負相關。我們觀察三十五個癌症相關基因的 correlation 圖形，在取 c_s 、 c_c 或 c_m 作圖的情況下，差異皆不大。因此在表 4.3 中，我們僅例出三十五個癌症基因的 c_s 與長度之關係。

Length\Name	APC	ATM	BCAM	BRCA1	BRCA2	BRIP1	CHK2
1	<u>-0.76</u>	<u>-0.63</u>	1	<u>-0.93</u>	<u>-0.66</u>	<u>-0.91</u>	<u>-9.13E-001</u>
2	<u>-0.14</u>	<u>-0.03</u>	<u>-0.06</u>	0.25	0.19	<u>-0.59</u>	<u>-7.87E-001</u>
3	0.12	0.34	0.06	0.43	0.45	-0.19	<u>-3.76E-001</u>
4	0.2	0.4	0.07	0.42	0.47	-0.02	<u>-1.51E-001</u>
5	0.22	0.29	0.06	0.3	0.42	0.03	<u>-1.99E-003</u>
6	0.19	0.21	0.07	0.2	0.32	0.03	2.58E-002
7	0.14	0.14	0.08	0.14	0.21	0.01	4.22E-002
Length\Name	CDH1	DPYD	FUT1	FUT3	LDHB	LHCGR	MC1R
1	<u>-0.73</u>	<u>-0.28</u>	9.99E-001	9.96E-001	9.79E-001	8.30E-002	0.89
2	<u>-0.6</u>	<u>-0.28</u>	2.28E-001	<u>-3.10E-001</u>	<u>-7.88E-002</u>	1.29E-001	0.09
3	<u>-0.24</u>	<u>-0.1</u>	2.92E-001	<u>-1.28E-001</u>	1.60E-001	2.53E-001	0.24
4	<u>-0.02</u>	0.04	1.98E-001	1.19E-002	9.30E-002	1.84E-001	0.31
5	0.06	0.06	1.76E-001	4.18E-002	3.24E-002	1.33E-001	0.24
6	0.07	0.04	1.38E-001	4.65E-002	2.62E-002	8.15E-002	0.25
7	0.06	0.02	1.07E-001	6.51E-002	3.71E-002	5.83E-002	0.27
Length\Name	MEN1	MLH1	MSH2	MSH6	NF1	NF2	p16
1	9.81E-001	<u>-0.97</u>	<u>-0.81</u>	<u>-0.93</u>	<u>-0.5</u>	<u>-0.92</u>	<u>-0.95</u>
2	7.20E-001	<u>-0.23</u>	<u>-0.21</u>	<u>-0.74</u>	0.02	<u>-0.07</u>	<u>-0.77</u>
3	7.69E-001	0.11	0.05	<u>-0.21</u>	0.16	0.15	<u>-0.52</u>
4	6.93E-001	0.15	0.13	<u>-0.07</u>	0.18	0.14	<u>-0.31</u>
5	5.28E-001	0.14	0.12	0.01	0.17	0.12	<u>-0.14</u>
6	4.09E-001	0.11	0.09	0.05	0.12	0.1	<u>-0.03</u>
7	3.42E-001	0.07	0.06	0.07	0.07	0.07	0.04
Length\Name	p53	PTCH	PTEN	RB	SCN4A	SMAD4	STK11
1	<u>-0.27</u>	<u>-0.69</u>	<u>-0.09</u>	<u>-8.48E-001</u>	8.35E-001	<u>-0.94</u>	0.9
2	0.58	0.01	0.17	<u>-1.59E-001</u>	<u>-1.34E-001</u>	<u>-0.54</u>	0.5
3	0.63	0.1	0.31	1.36E-001	9.42E-002	<u>-0.14</u>	0.4
4	0.54	0.13	0.23	2.17E-001	8.65E-002	<u>-0.06</u>	0.3
5	0.41	0.12	0.16	2.01E-001	9.54E-002	0	0.23
6	0.28	0.09	0.11	1.56E-001	1.04E-001	0.02	0.15
7	0.19	0.08	0.07	1.12E-001	9.26E-002	0.03	0.11
Length\Name	TGFBR2	TSC1	TSC2	TSHR	VHL	WRN	WT1
1	<u>-9.15E-001</u>	<u>-0.82</u>	0.67	<u>-0.77</u>	0.26	<u>-0.57</u>	<u>-0.71</u>
2	<u>-8.77E-001</u>	<u>-0.4</u>	0.41	<u>-0.48</u>	<u>-0.12</u>	<u>-0.38</u>	<u>-0.43</u>
3	<u>-5.45E-001</u>	<u>-0.06</u>	0.5	<u>-0.25</u>	0.06	<u>-0.01</u>	<u>-0.12</u>
4	<u>-2.62E-001</u>	0.05	0.48	<u>-0.11</u>	0.12	0.08	0.03
5	<u>-1.15E-001</u>	0.07	0.41	<u>-0.04</u>	0.14	0.08	0.06
6	<u>-4.23E-002</u>	0.06	0.31	<u>-0.01</u>	0.11	0.06	0.06
7	<u>-7.32E-003</u>	0.05	0.22	0.01	0.1	0.04	0.05

表 4.3 三十五個癌症基因的 c_s 與長度之關係，

其中若 c_s 呈負值，我們予以標示底線

在 35 筆分析中，在長度 1 有時有 24 筆的 correlation 是呈現負相關，這也許意味著有特定的鹼基容易發生突變，而這些鹼基在序列中是屬於少數的。又或者，負責編碼胺基酸的外顯子序列所含有的鹼基與自身序列所含有的鹼基呈負相關；而在長度 2 時，有 23 筆的 correlation 是呈現負相關，這或許意味序列有特別的編列偏好，而非依其偏好的序列，似乎有較大的可能性突變，或有較大的可能性包含致癌突變位置。

第五章 結論

我們假設致癌突變所造成的 Γ 值最小，所以未被人體的 DNA 電性檢測機制所偵測出來，但在第三章第二節中，關於致癌突變在 codon 中的排名分析，致癌突變的 Γ 未必都是 codon 上九種突變計算出來的最小值，我們猜測那些 Γ 值比致癌突變更小的非致癌突變，其絕大多數所編碼的胺基酸與未突變前的 codon 相同。不過，即使我們將那些編碼相同胺基酸的突變納入排名的計算中，致癌突變的 Γ 在 codon 上是屬於最小的，所佔有的比例仍是最多。在第三章第三節中，TP53 序列上，致癌突變的 Γ 排名在該序列所有可能突變中，是傾向排名較前的分佈，我們依突變前後鹼基的不同，將之分為 12 類，以傾向值的定義去分析，也得到定性上的結果。這些結果十分支持我們的假設，對於人體內部有檢測 DNA 電性變化的機制的猜想，應該不是無稽之談。

第四章的分析看出，長度 1 時有 68.6% 的 DNA 序列是呈現負相關，長度 2 時有 65.7% 的 DNA 序列是呈現負相關，也許值得更進一步的去研究外顯子序列與其基因序列之間的相依性，目前數據不足，我無法提出肯切的推論。

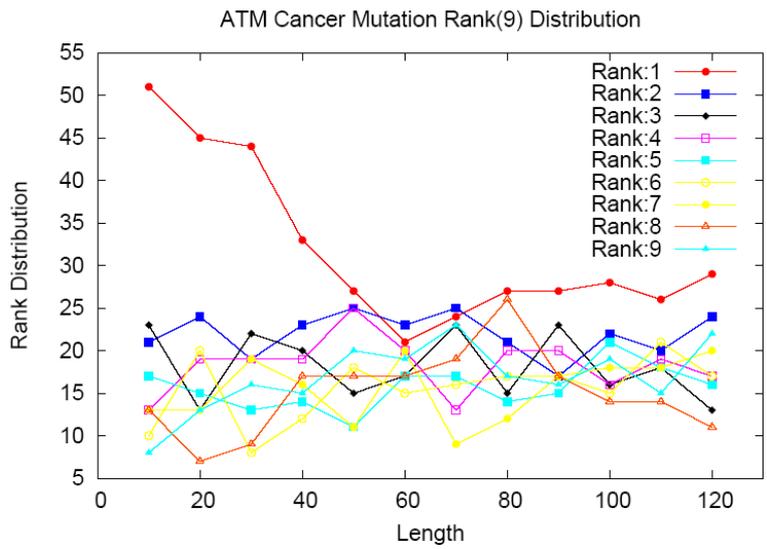
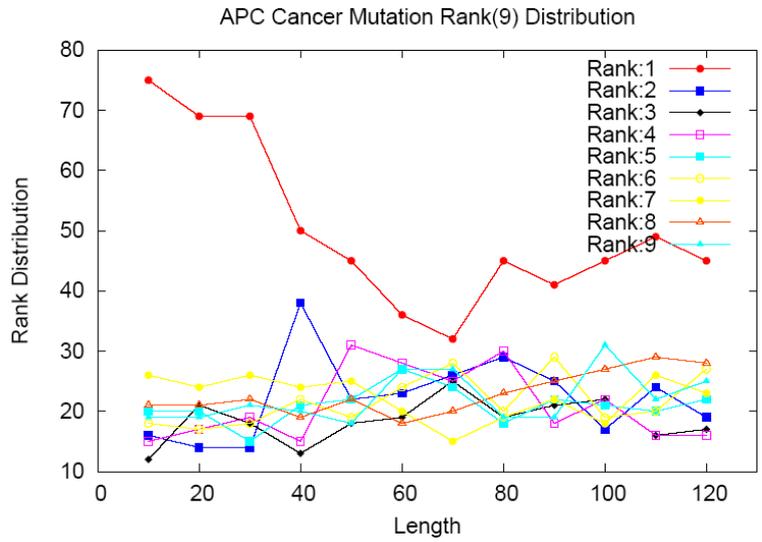
參考文獻：

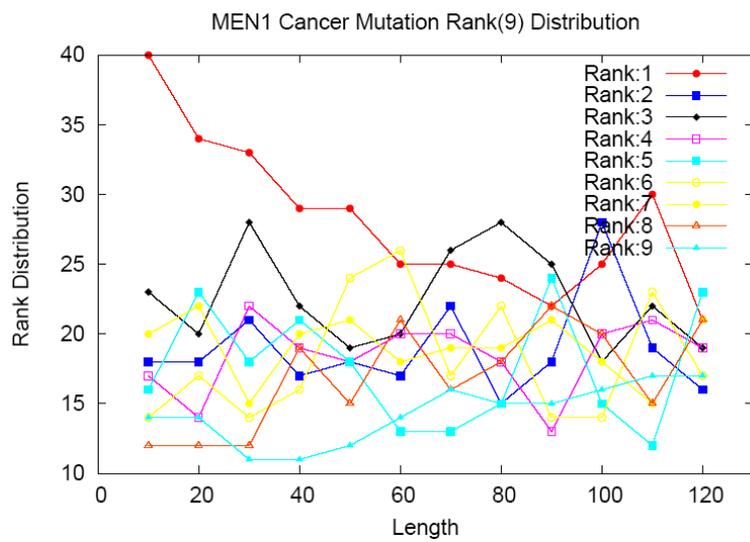
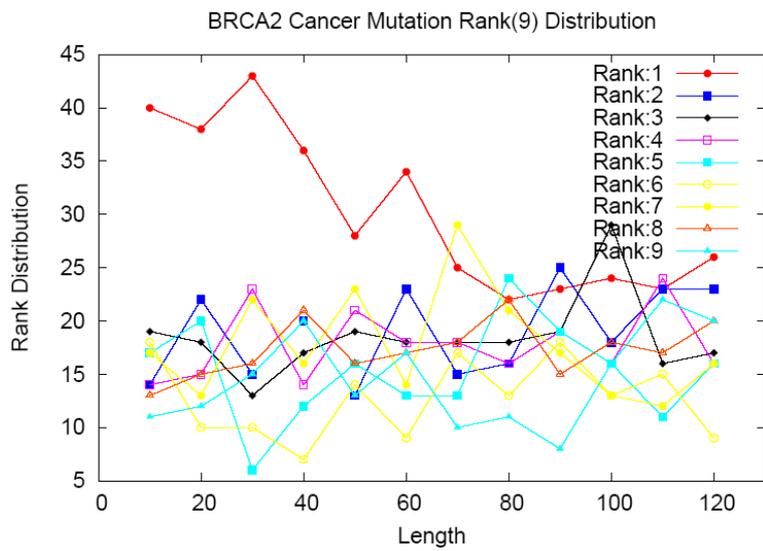
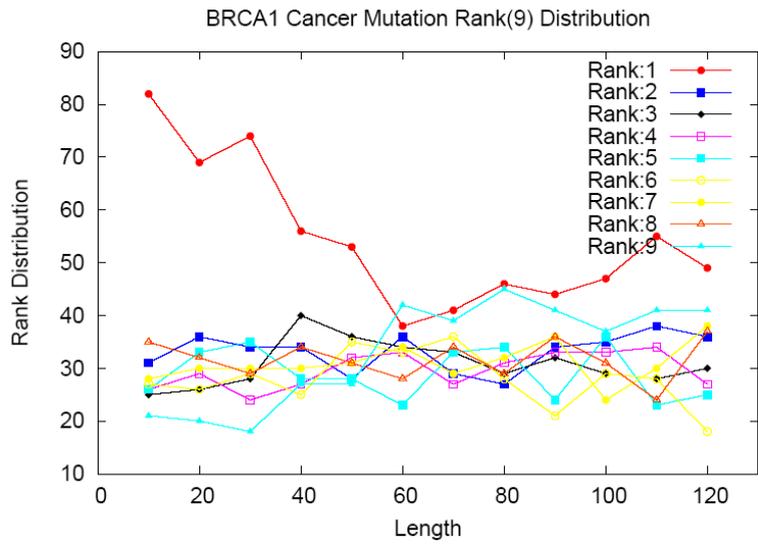
- [1] R. G. Endres, D. L. Cox, and R. P. Singh, *Rev. Mod. Phys.* **76**, 195 (2004).
- [2] Braun, E., Y. Eichen, U. Sivan, and G. Ben-Yoseph, 1998, *Nature (London)* **391**,775.
- [3] de Pablo, P. j. F. Moreno-Herrero, J. Colchero, J. Gómez Herrero, P. Herrero, A. M. Bar, P. Ordejon, J. M. Soler, and E. Artacho, 2000, *Phys. Rev. Lett.* **85**, 4992.
- [4] Storm, A. J., J. van Noort, S. de Vries, and C. Dekker, 2001, *Appl. Phys. Lett.* **79**, 3881.
- [5] Zhang. Y. R. H. Austin, J. Kraeft, E. C. Cox, and N. P. Ong, 2002, *phys. Rev. Lett.* **89**, 198102.
- [6] Proath, D., A. Bezryadin, S. De Vries, and C. Decker, 2000, *Nature (London)* **403**, 635.
- [7] Rakitin, A., P. Aich, C. Papadopoulos, Y. Kobzar, A.S. Vedeneev, J. S. Lee, and J. M. Xu, 2001, *Phys. Rev. Lett.* **86**, 3670.
- [8] Fink, H. W. , and C. Schönenberger, 1999, *Nature (London)* **398**, 407.
- [9] Hartzell, B., B. McCord, D.Asare, H. Chen, J. J. Heremans, and V. Soghomonian, 2003a, *Appl. Phys Lett.* **82**, 4800.
- [10] Hartzell, B., B. McCord, D.Asare, H. Chen, J. J. Heremans, and V. Soghomonian, 2003b, *Appl. Phys Lett.* **94**, 2764.
- [11] Tran, P., B. Alavi, and G. Grüner, 2000, *Phys. Rev. Lett.* **85**, 1564.
- [12] Cai, L., H. Tabata, and T. Kawai, 2000, *Appl. Phys. Lett.* **77**, 3105.
- [13] Yoo, K.-H., D. H. Ha, J.-o. Lee, J. W. Park, J. Kim, J. J. Kim, H.-Y.Lee, T. Hawaii, and H. Y. Choi, 2001, *Phys.Rev. Lett.* **87**, 198102.
- [14] Kasumov, A. Y., M. K. Kociak, S. Gueron, B. Reulet, and V. T. Volkov, 2001, *Science* **291**, 280.
- [15] *Charge Migration in DNA: Perspectives from Physics, Chemistry and Biology*, edited by T. Chakraborty (Springer-Verlage, Berlin, 2007).

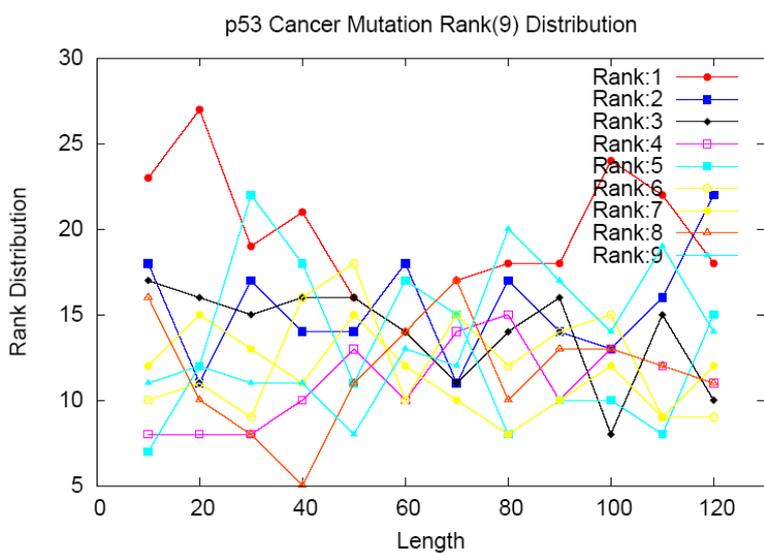
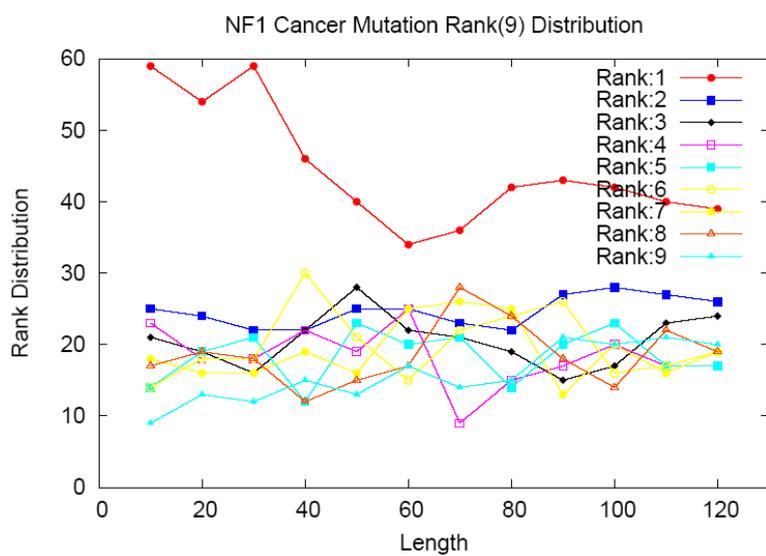
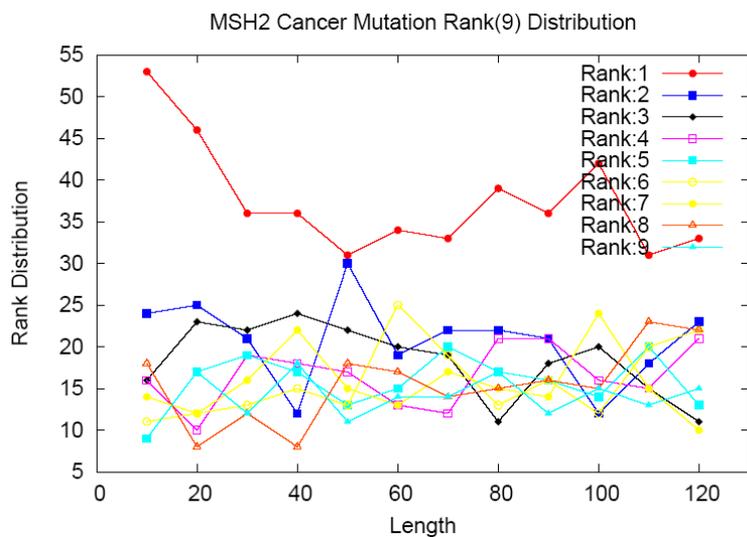
- [16] C. T. Shih, S. Roche, and R. A. Römer, Phys. Rev. Lett. **100**, 018105 (2008)
- [17] E. Yavin et al.. Proc. Natl .Acad. Sci. **102**, 3456
- [18] HGMD (The Human Gene Mutation Database) 資料庫
<http://www.hgmd.cf.ac.uk/ac/index.php>
- [19] OMIM (Online Mendelian Inheritance in Man) 資料庫
<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>
- [20] Roche, S., 2003, Phys. Rev. Lett. **91**, 108101.
- [21] A.A. Voityuk, J. Jortner, M. Boxin, and N. Rösch, J. Chem. Phzs. **114**, 5614 (2001); H. Sugiyama and I. Saito, J. Am. Chem. Soc. **118**, 7063 (1996)
- [22] IARC (International Agency for Resesrch on Cancer) 資料庫
<http://www-p53.iarc.fr/index.html>

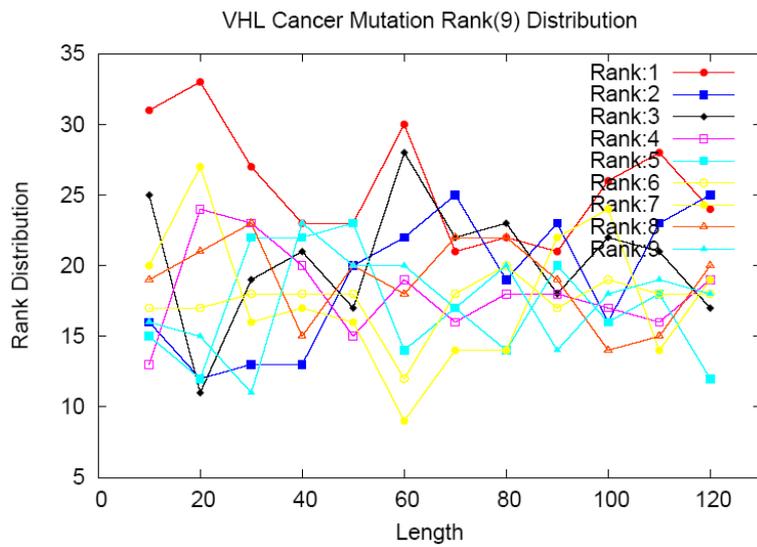
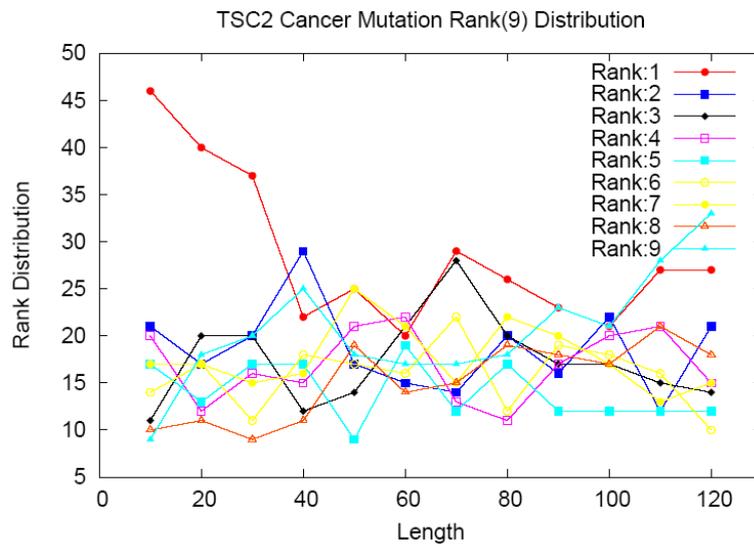
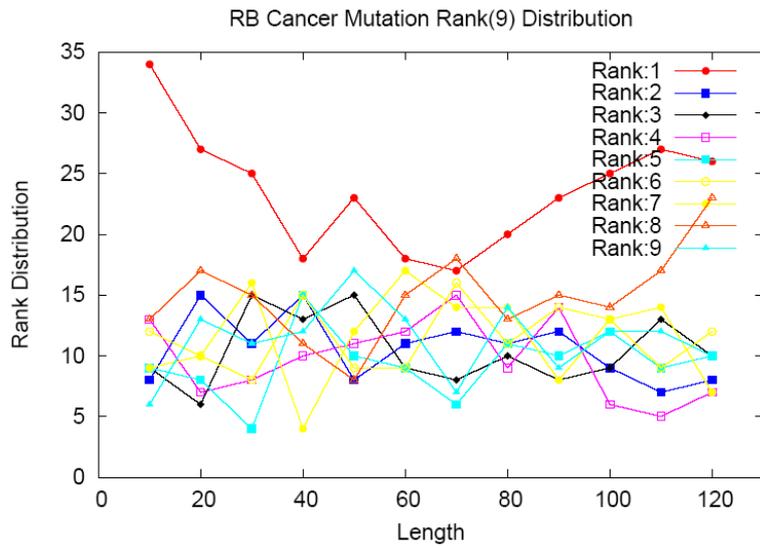
附錄 1 :

癌症相關基因的致癌突變在 codon 上的九種排名分布（僅繪製致癌突變數大於 100 筆的）：

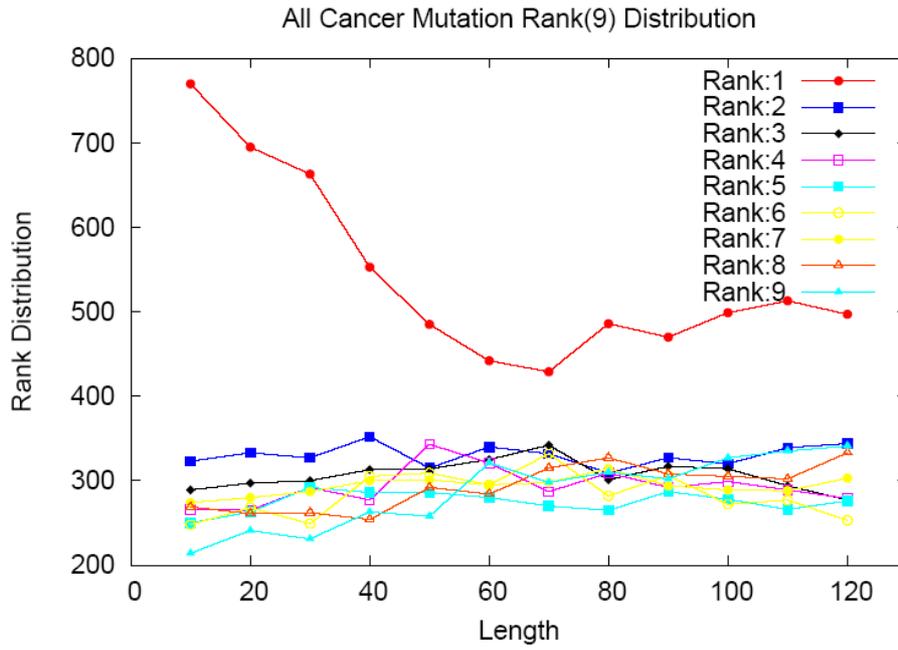




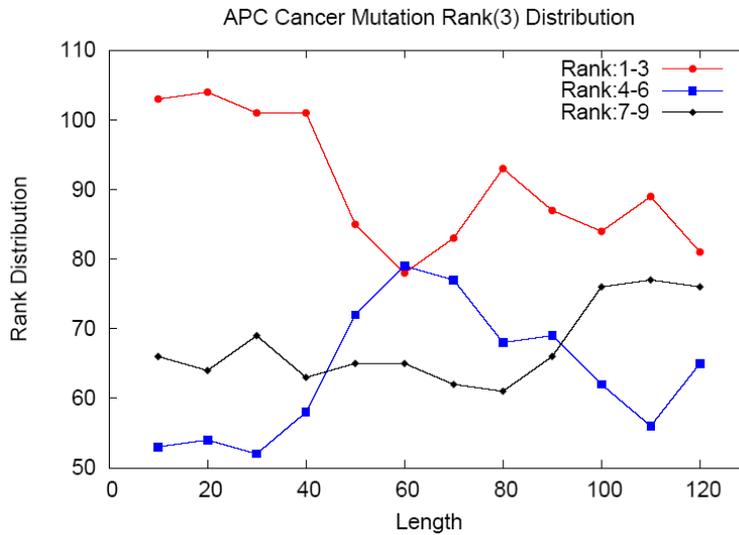


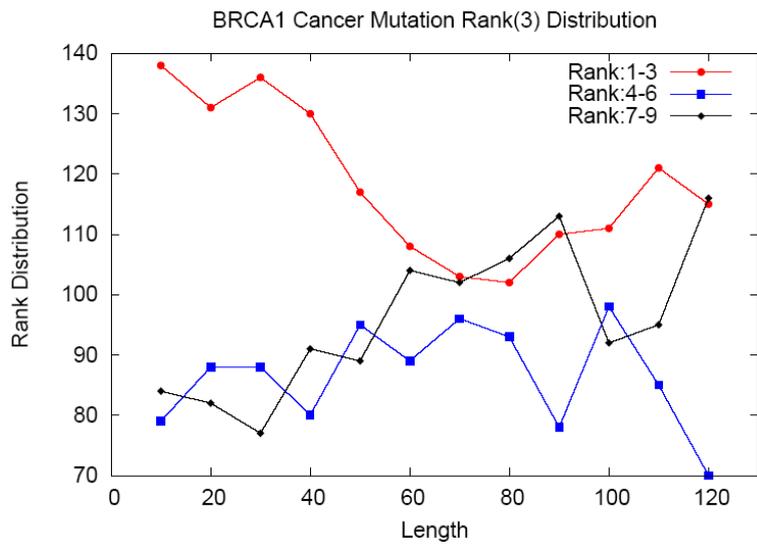
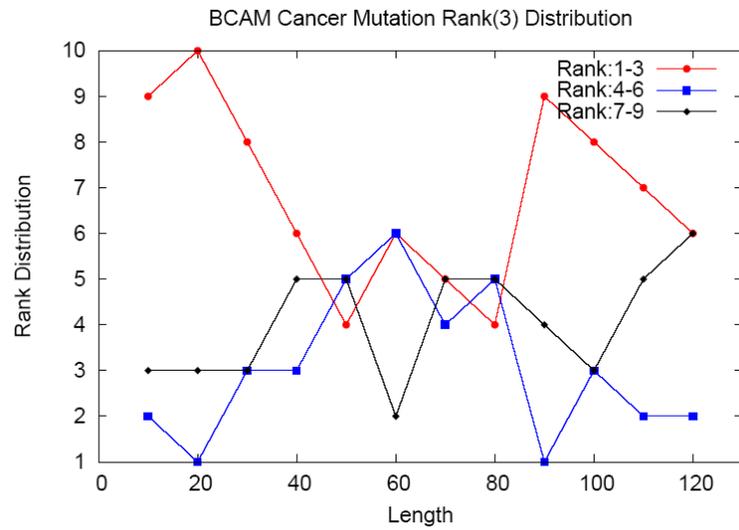
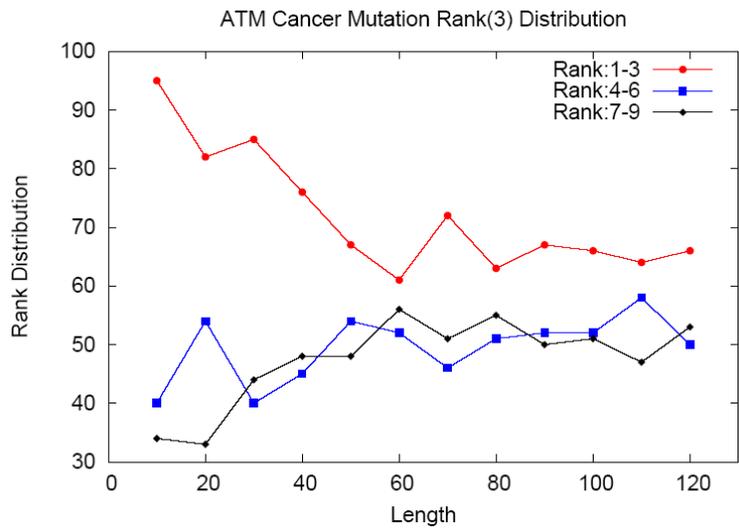


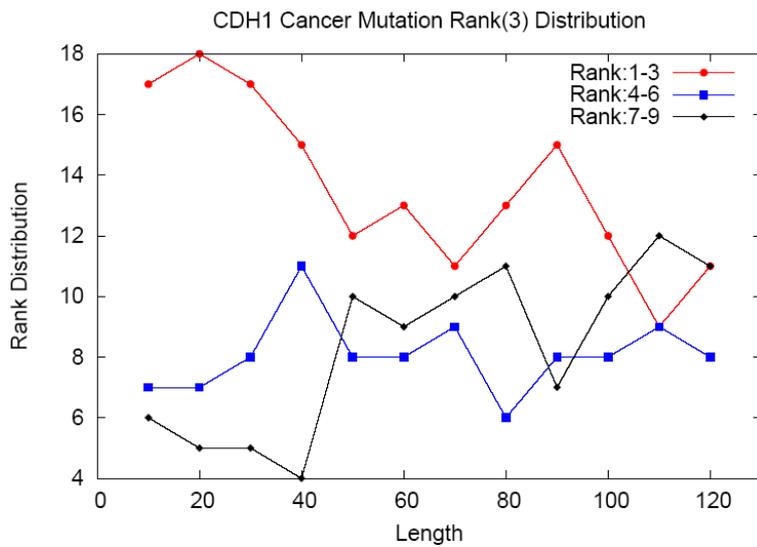
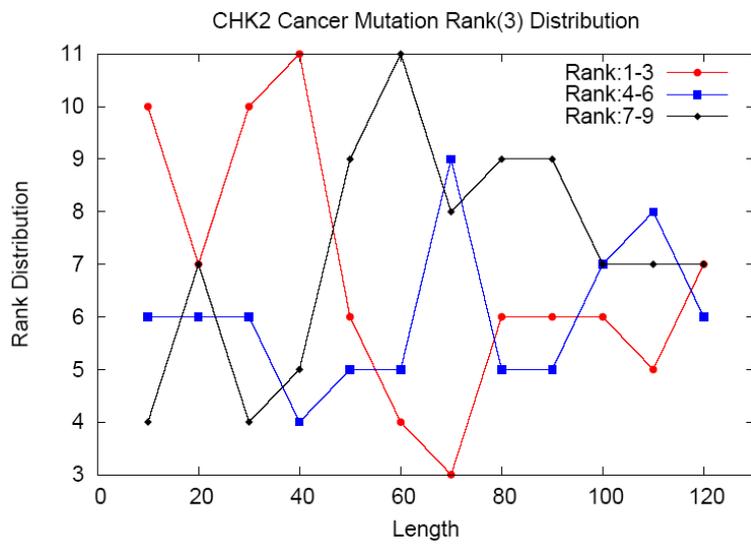
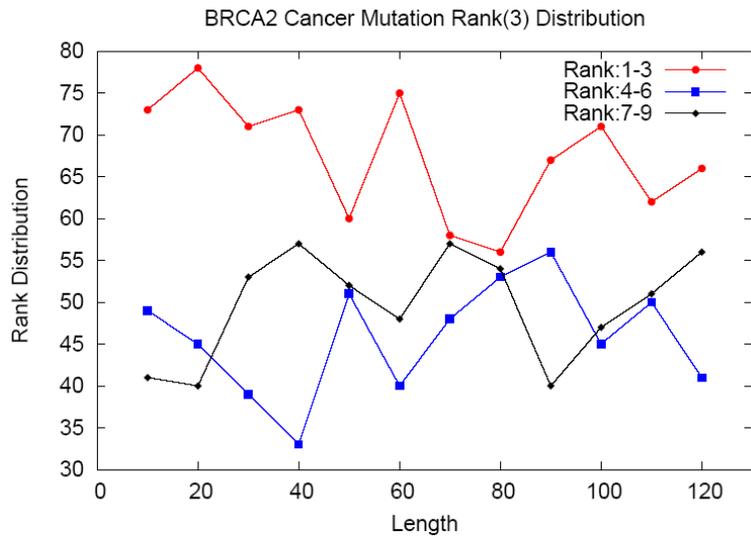
35 筆癌症相關基因的致癌突變在 codon 上的九種排名分布：

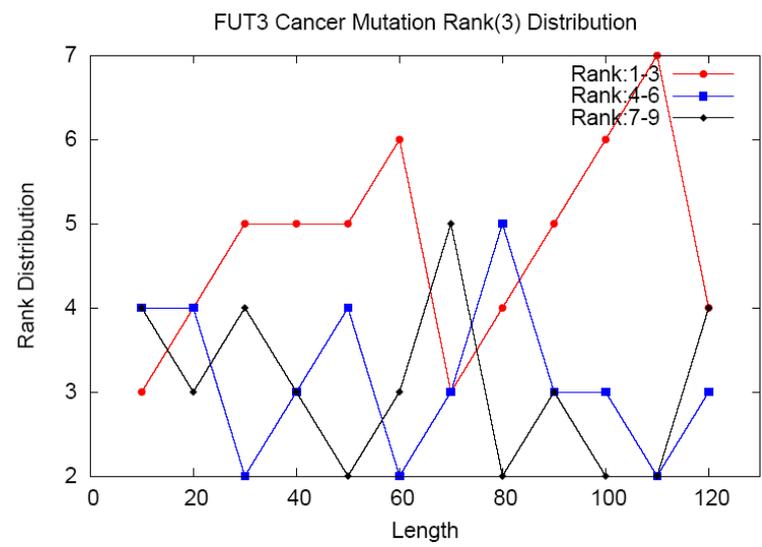
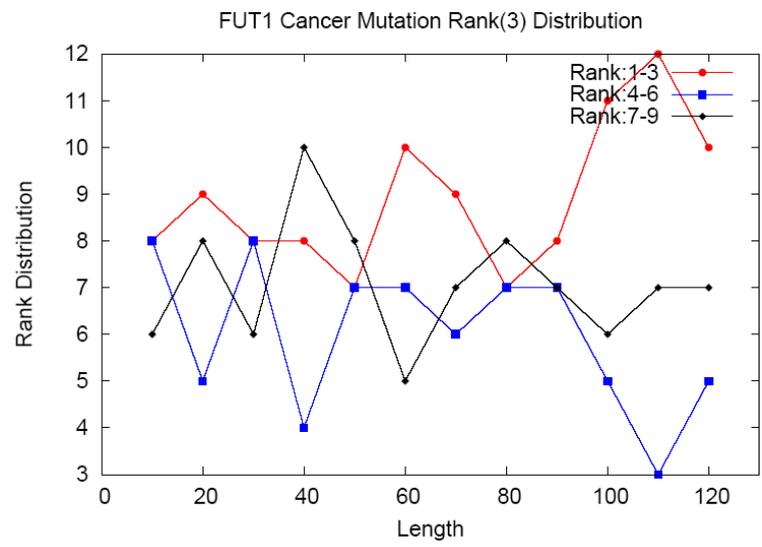
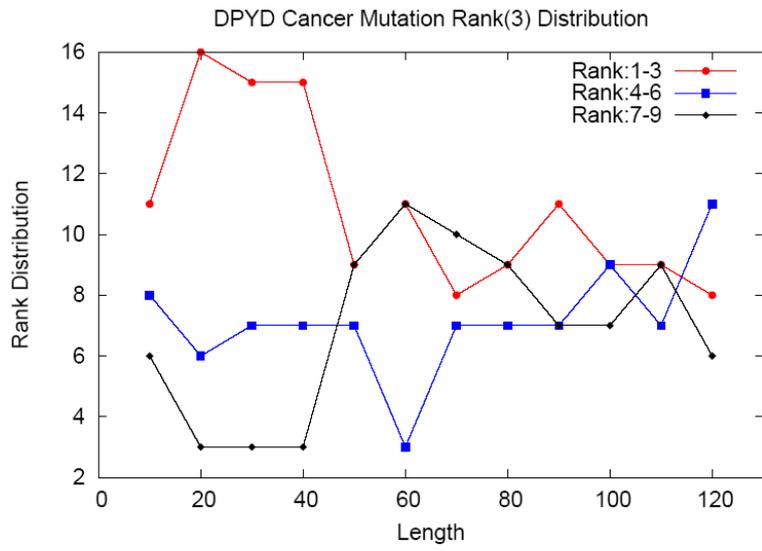


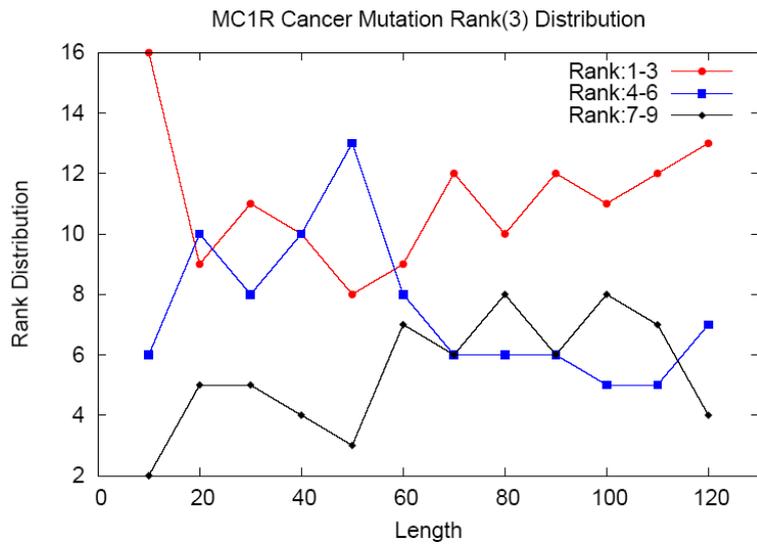
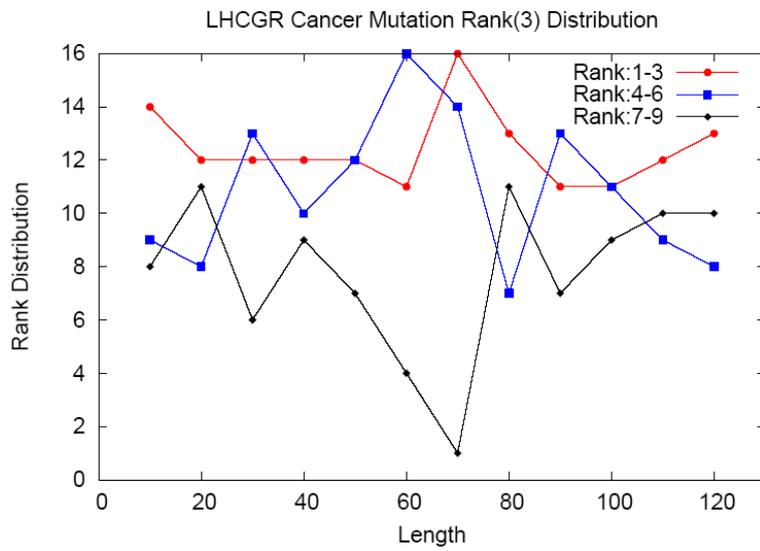
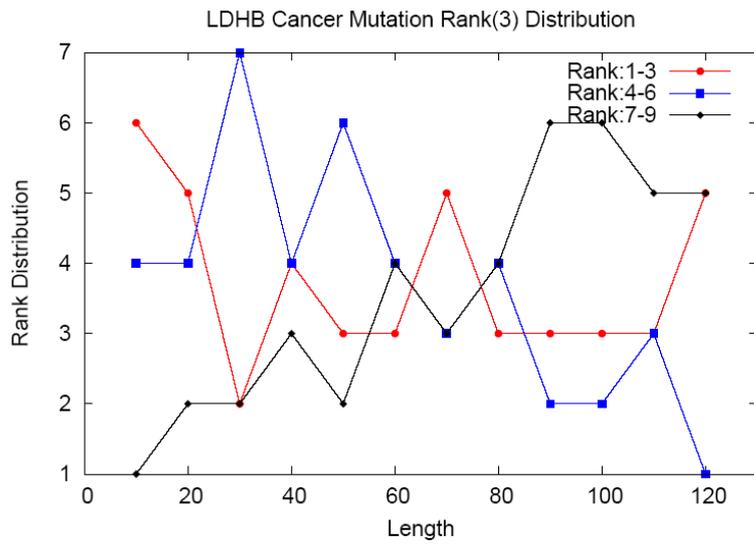
癌症相關基的致癌突變在 codon 上的 3 種排名分布：

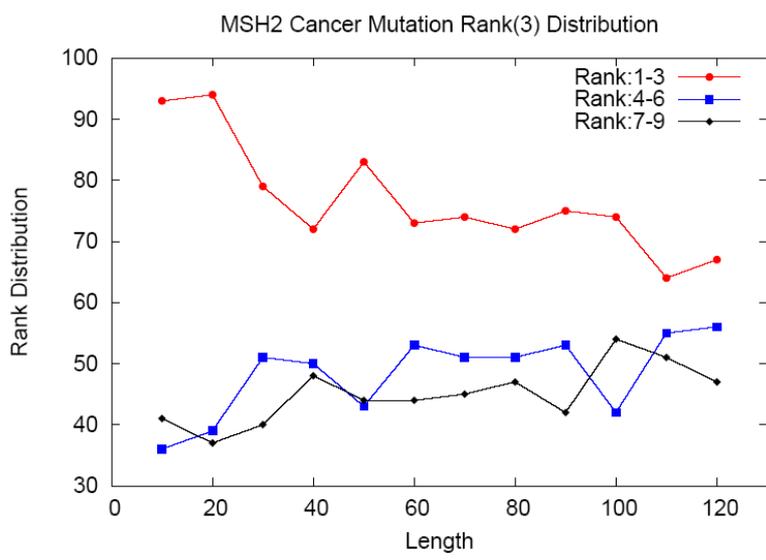
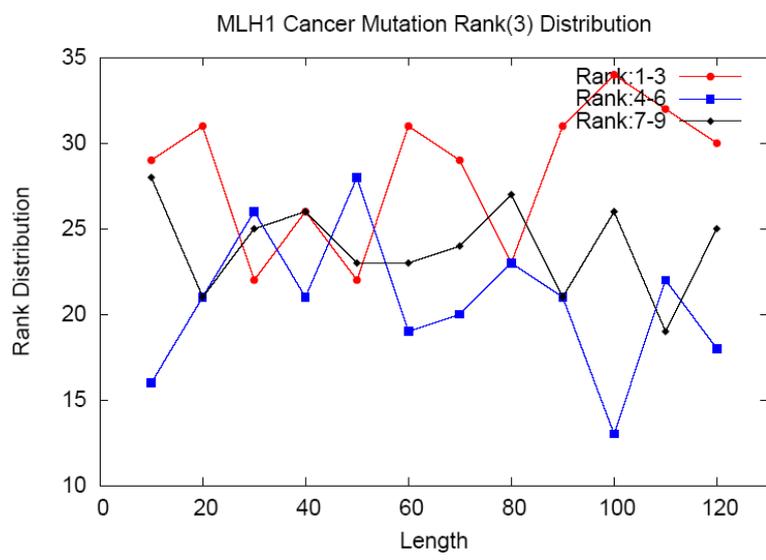
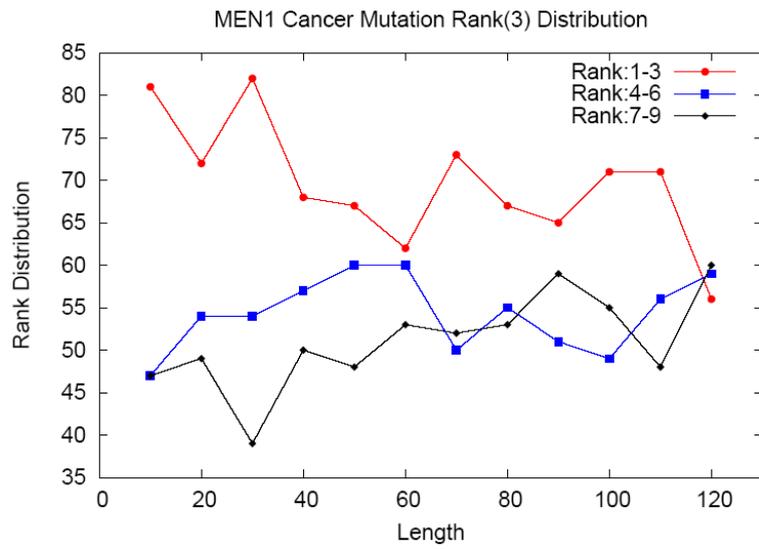


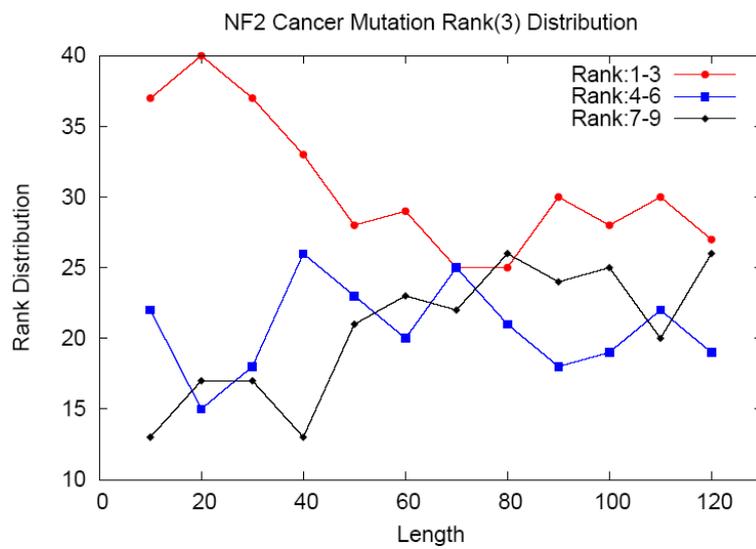
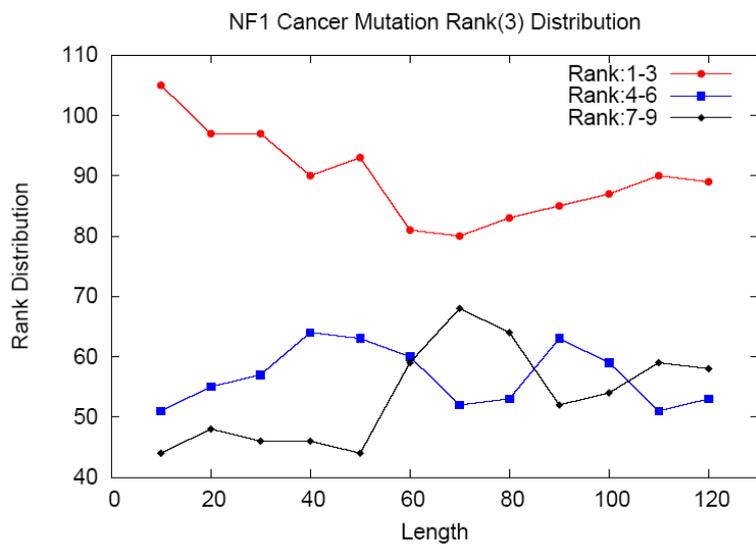
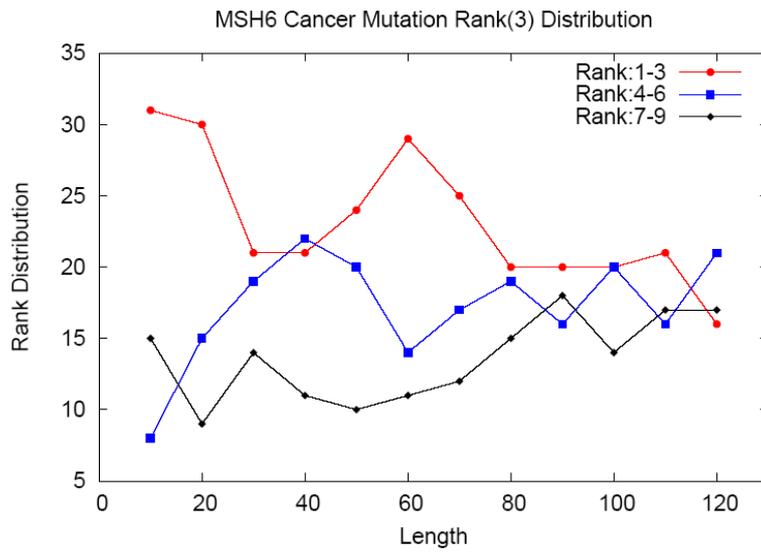


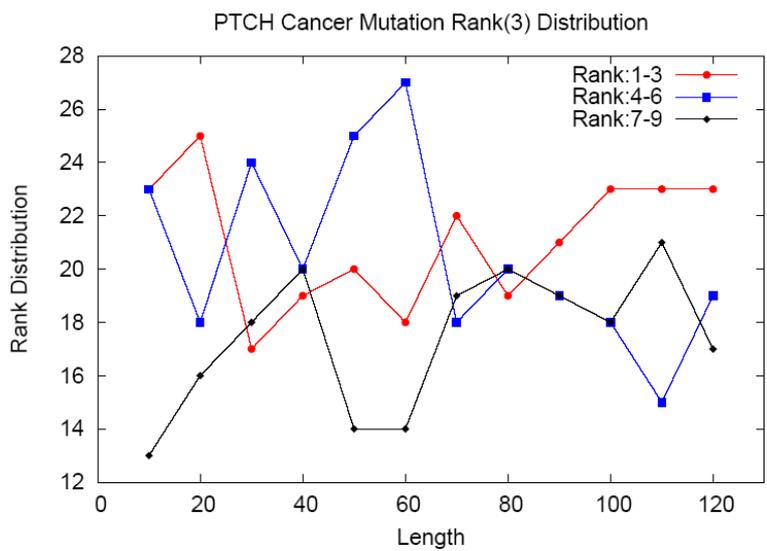
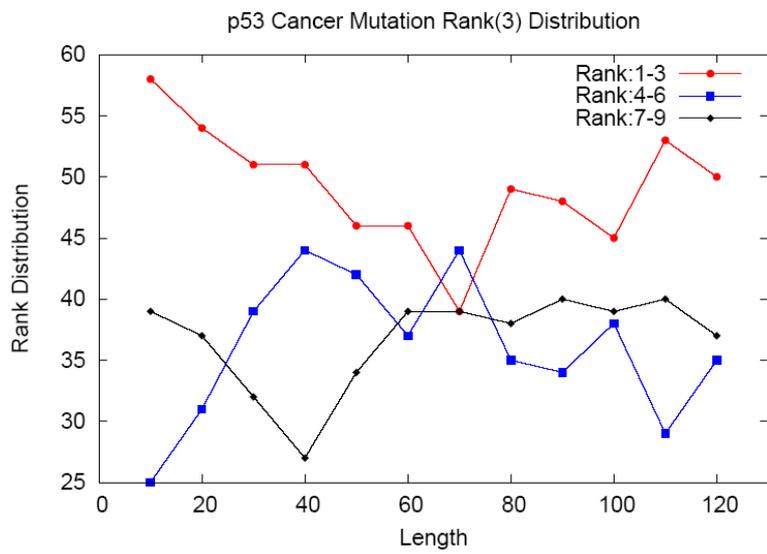
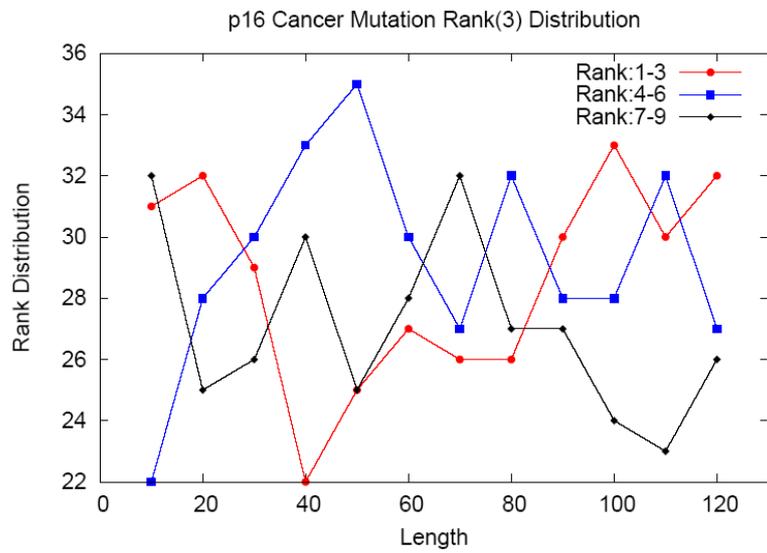


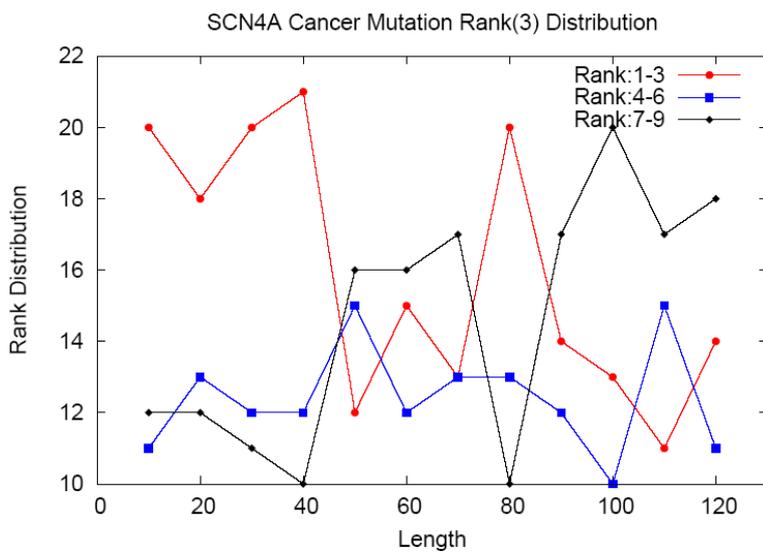
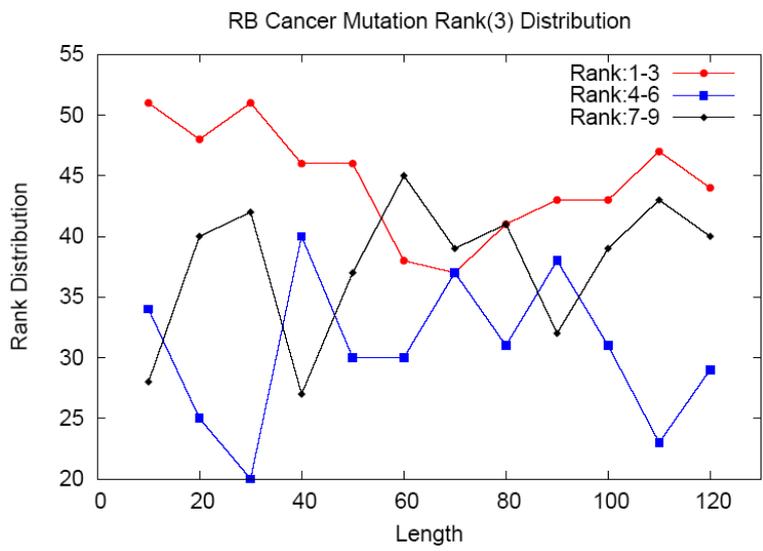
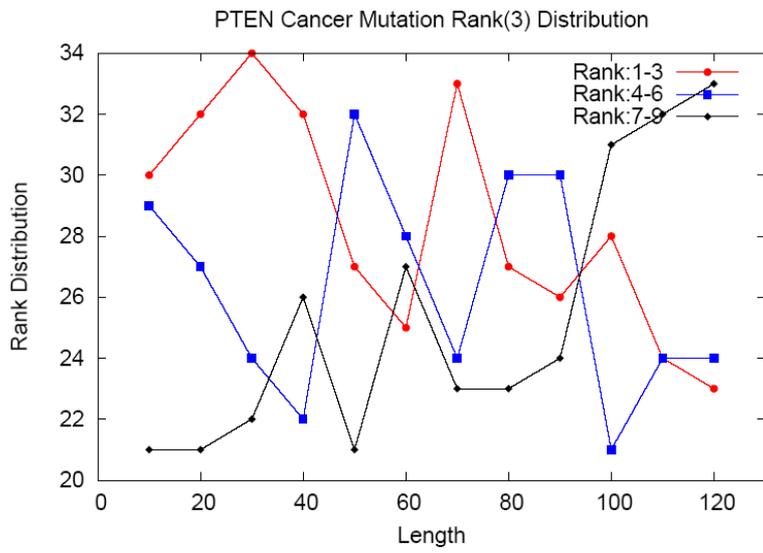


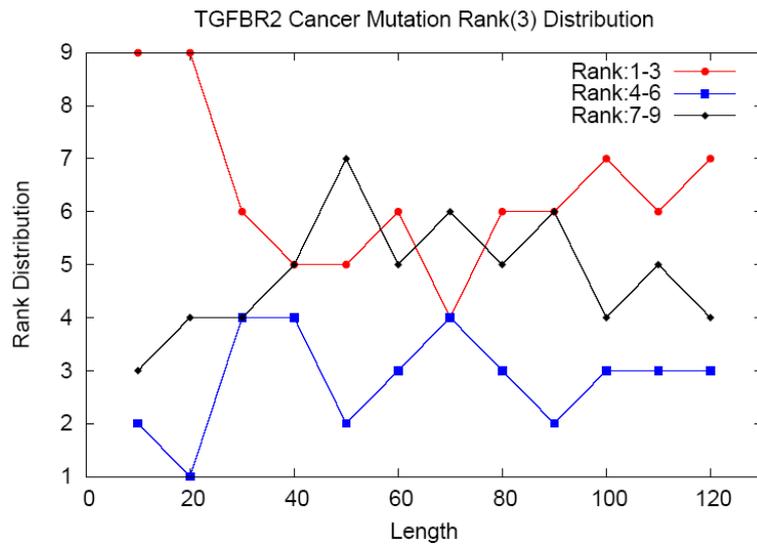
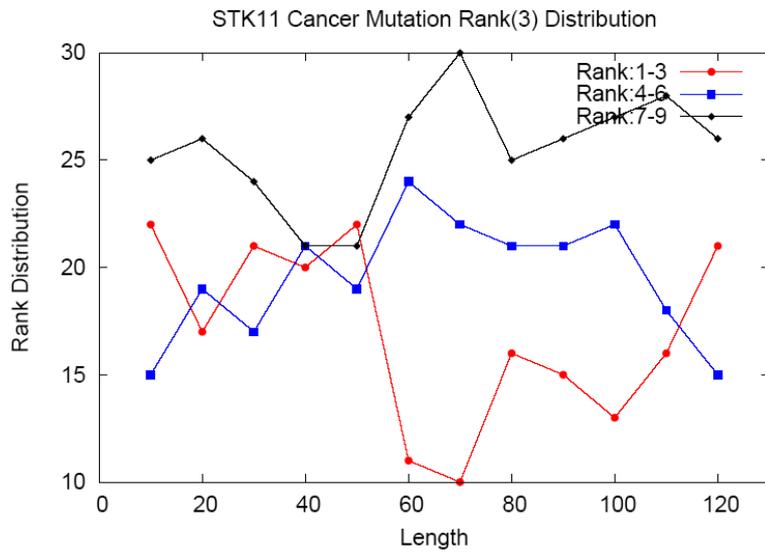
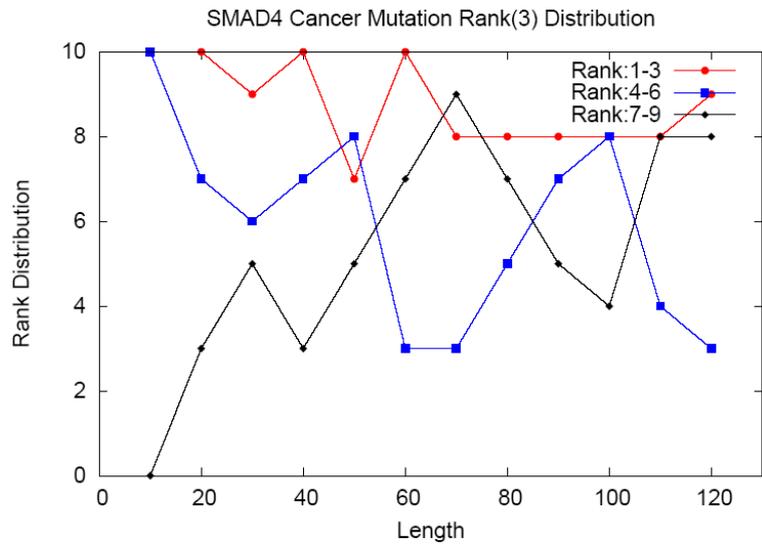


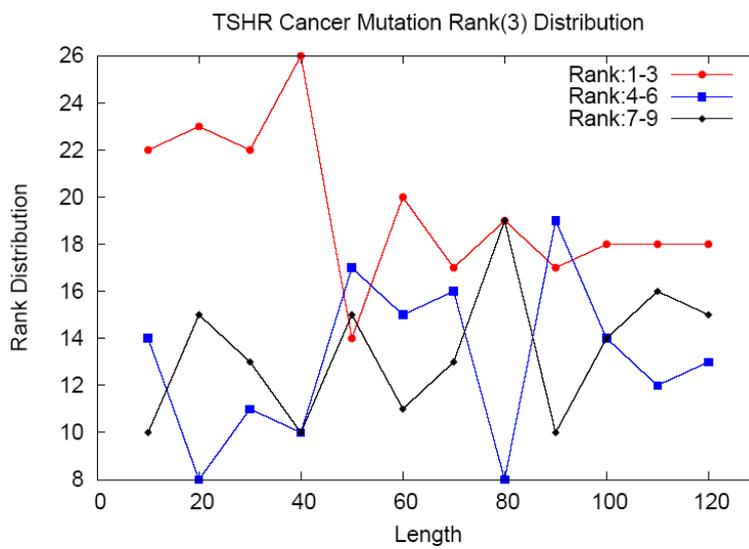
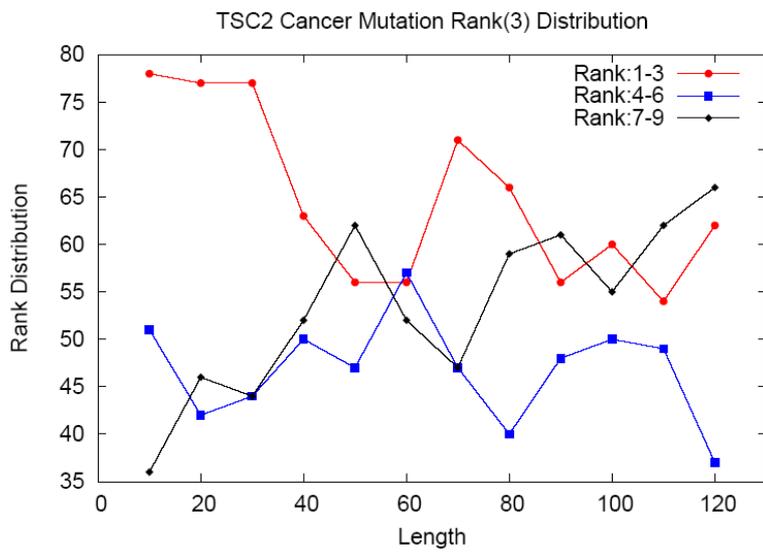
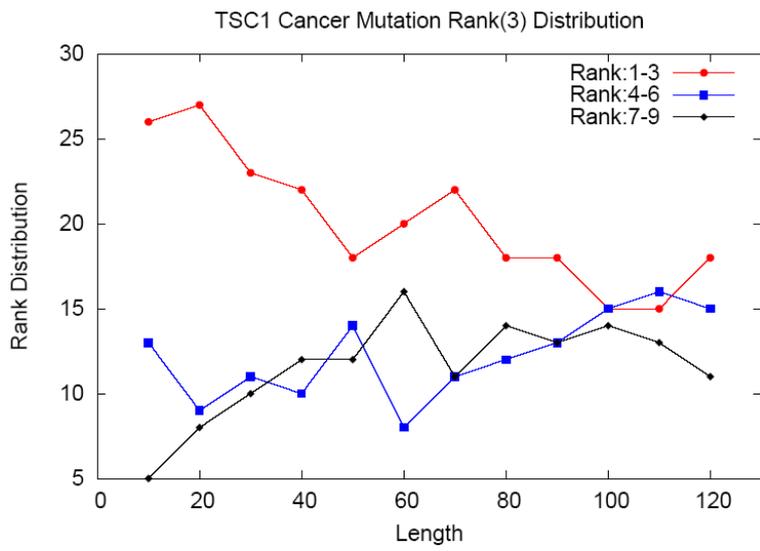


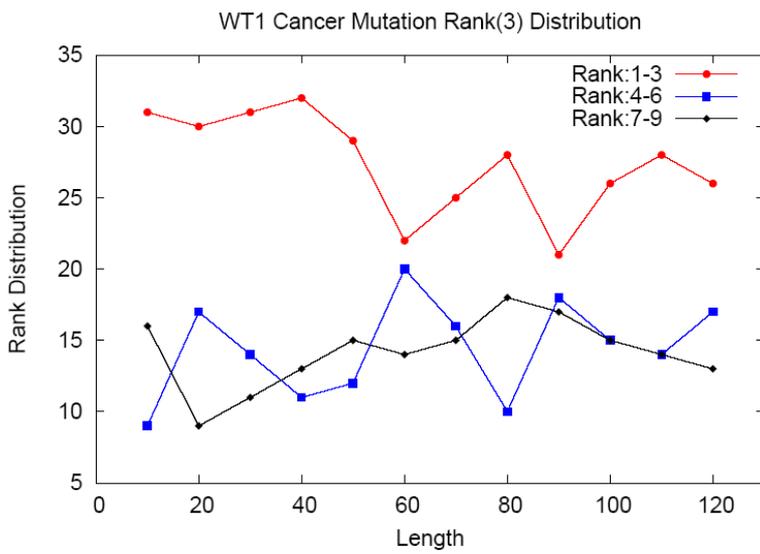
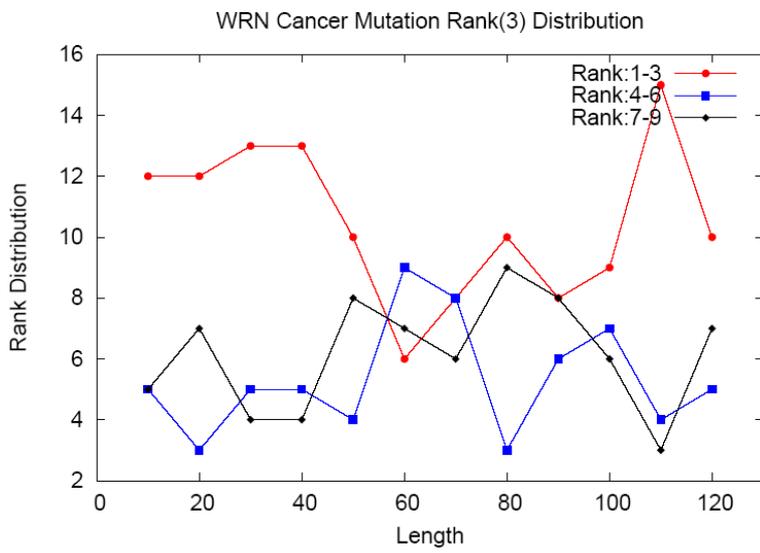
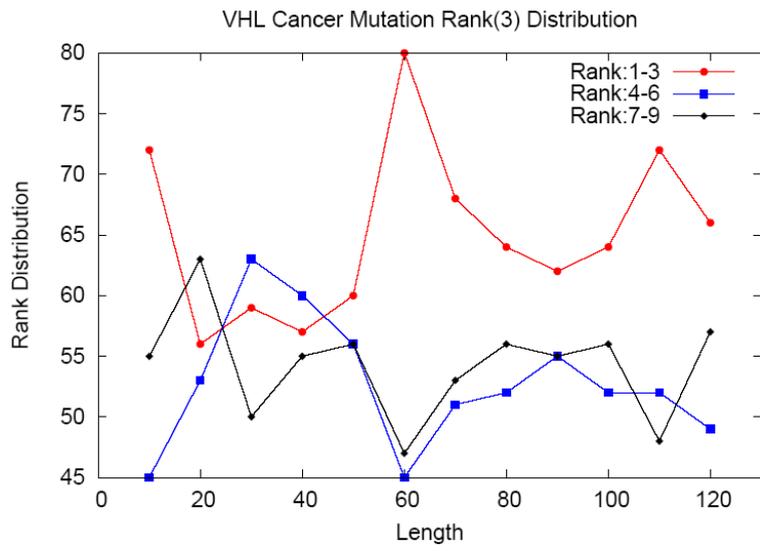


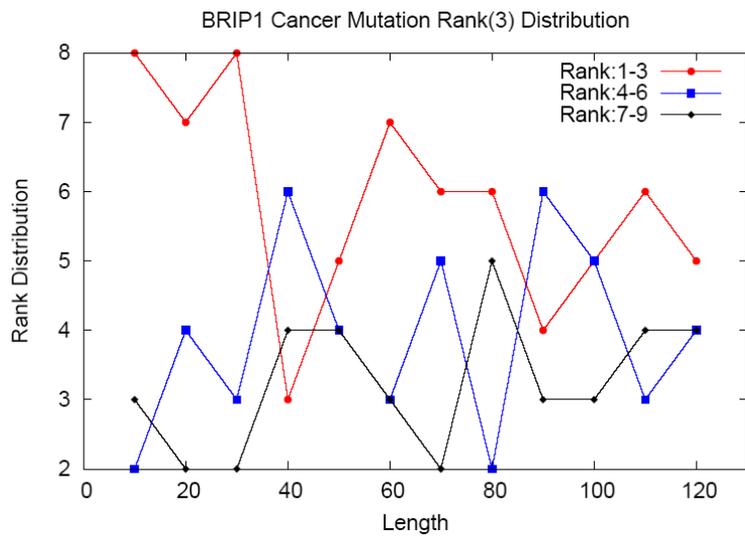




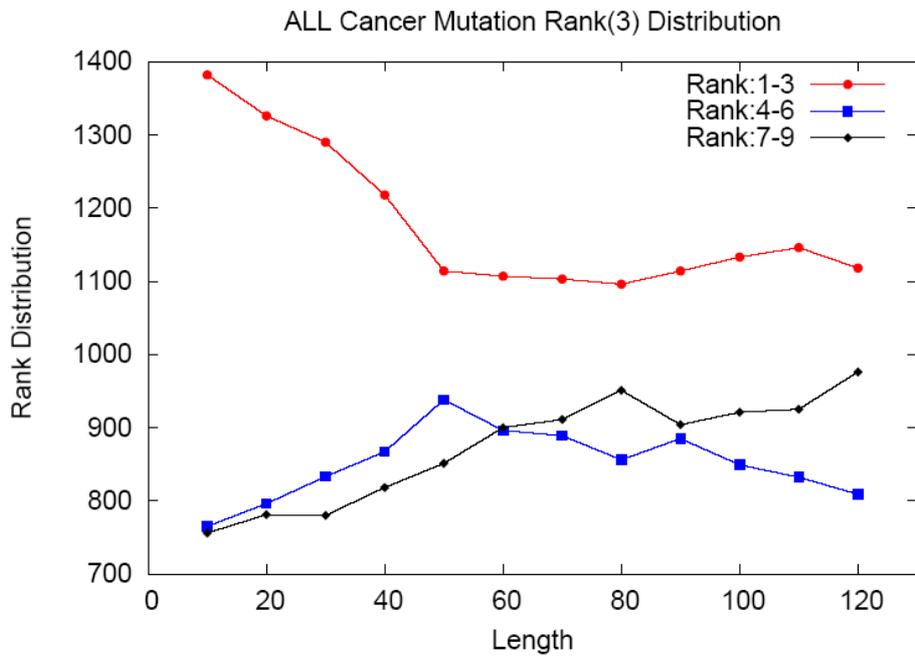






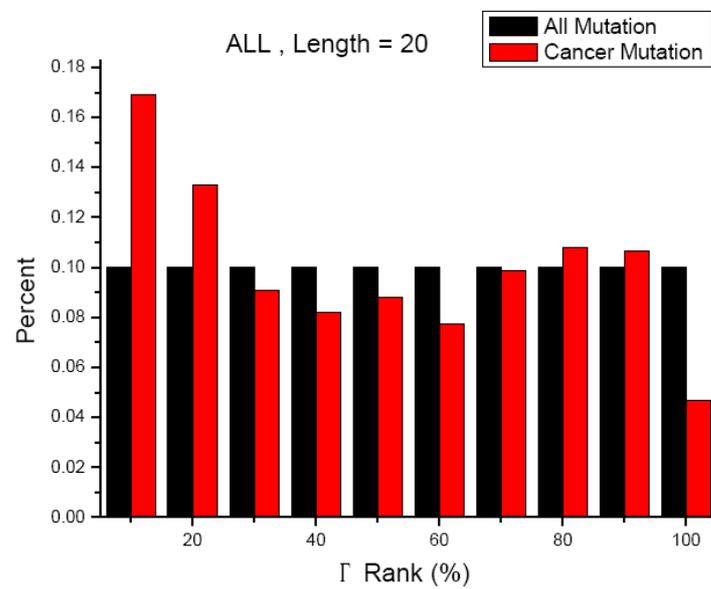
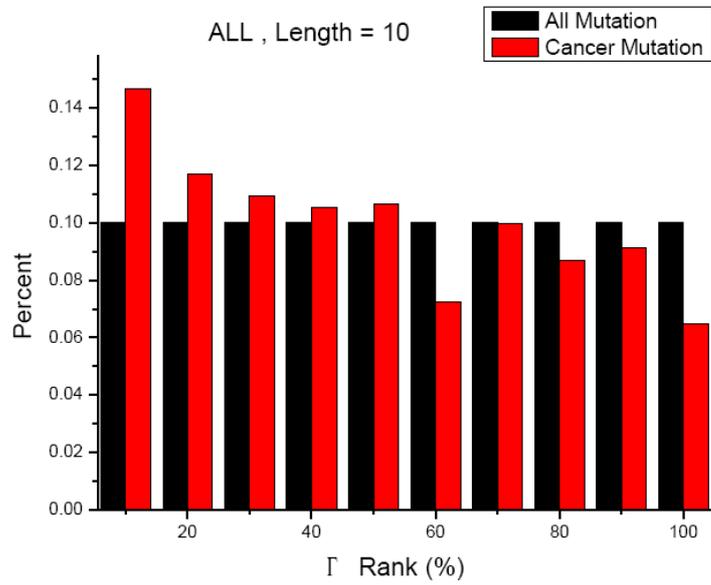


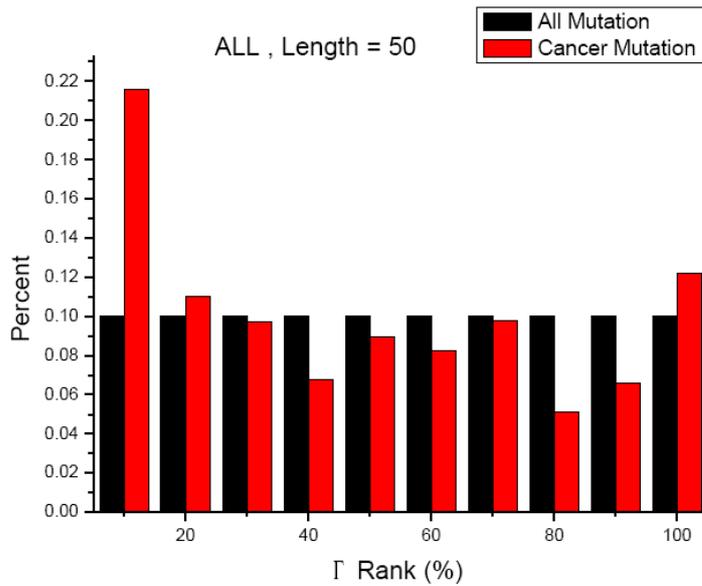
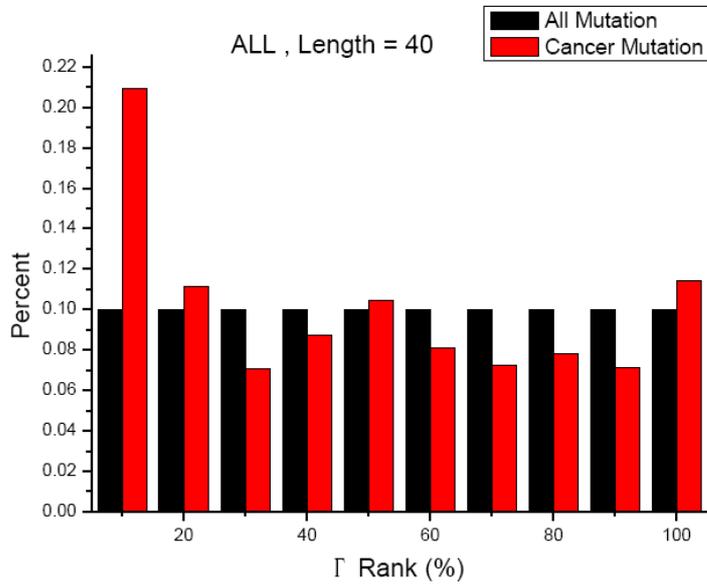
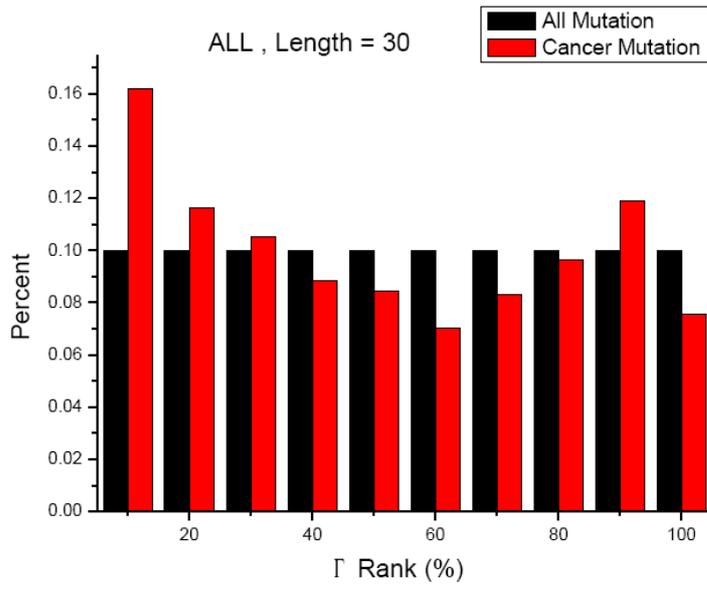
35 筆癌症相關基因的致癌突變在 codon 上的三種排名分布：

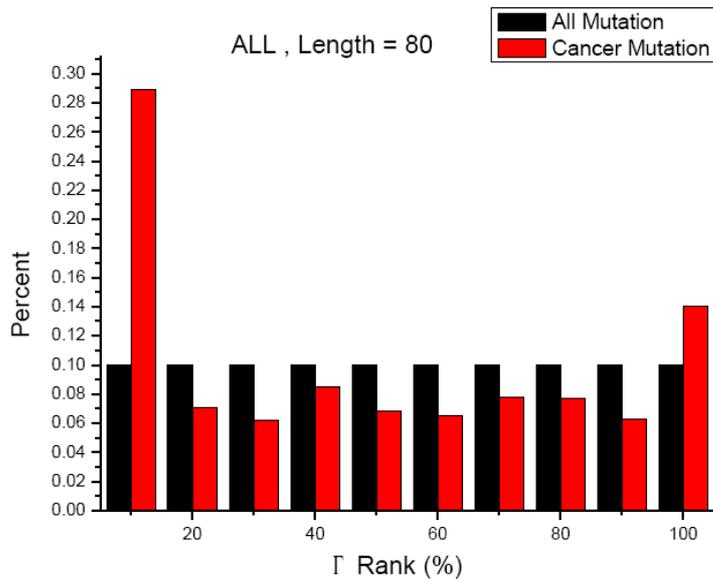
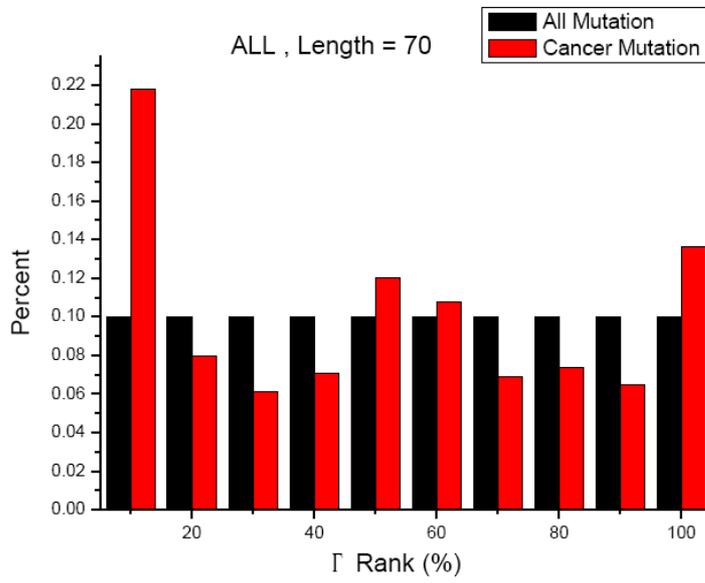
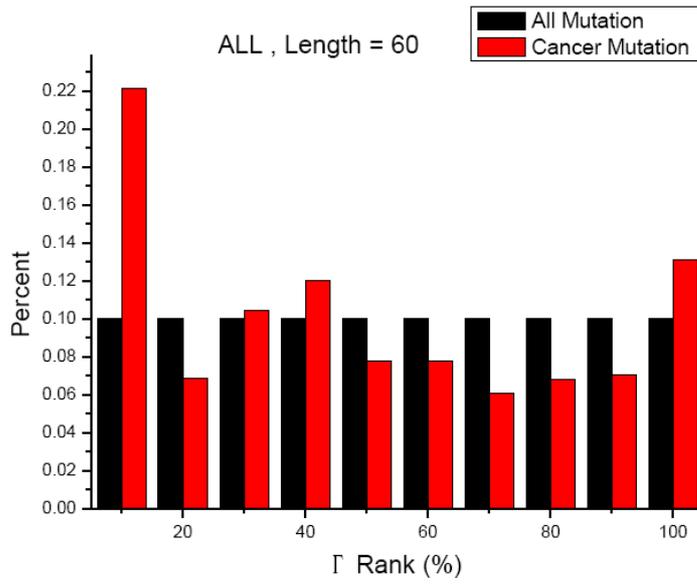


附錄 2 :

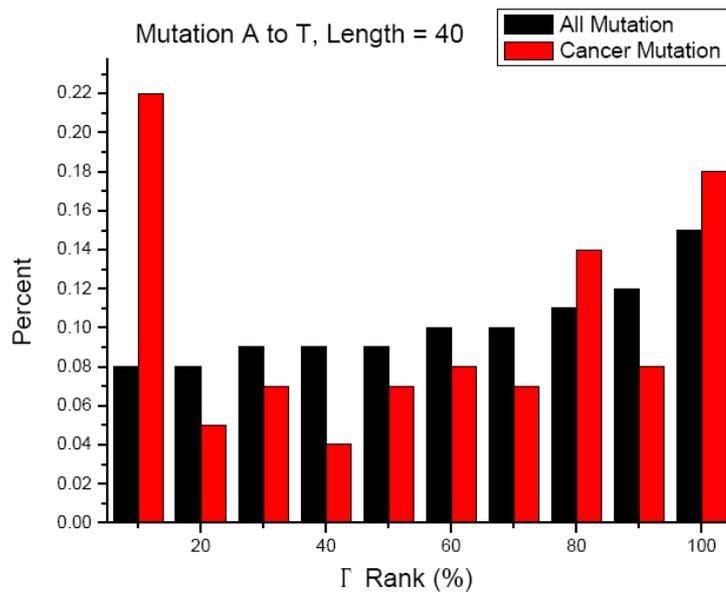
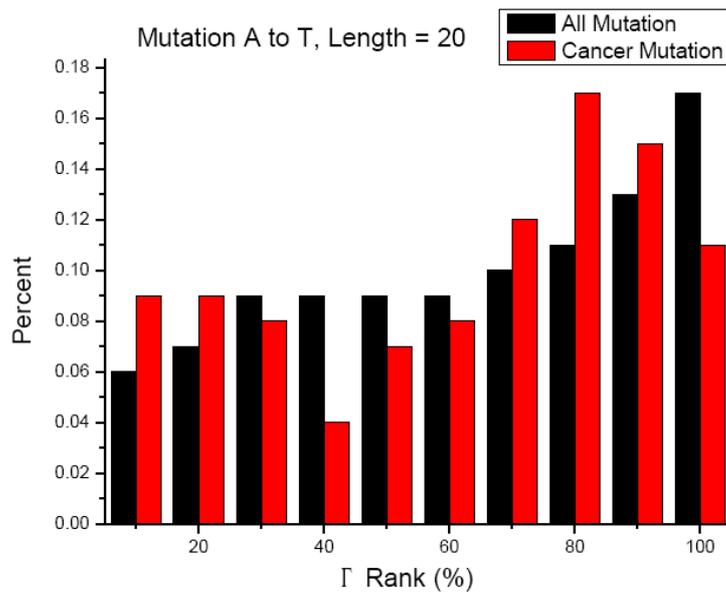
TP53 致癌突變與全部可能突變，在不同長度時的 Γ 名次分佈：

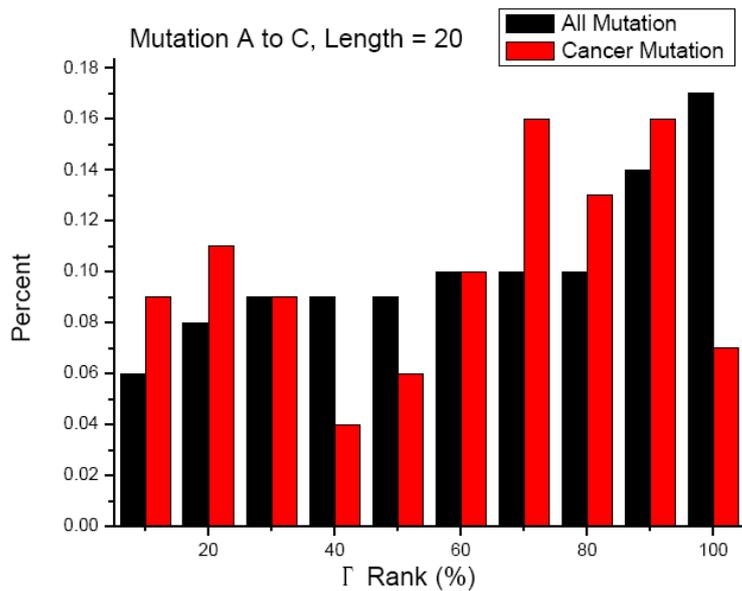
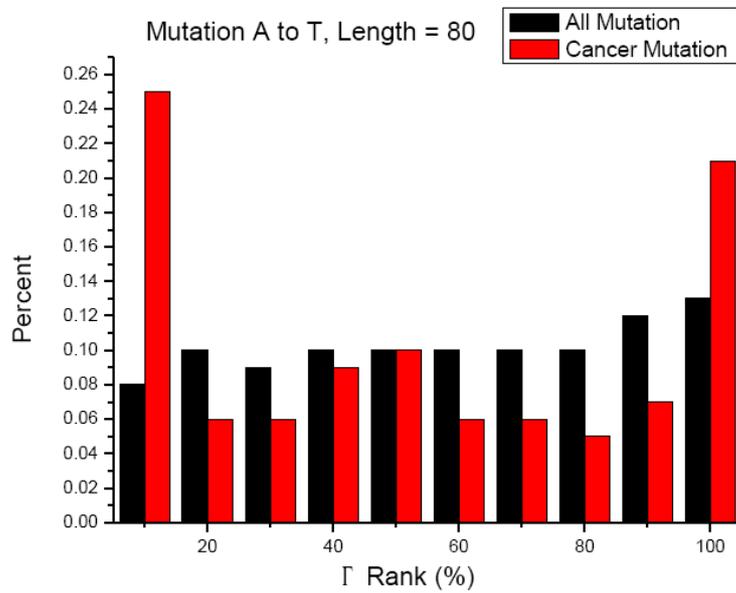
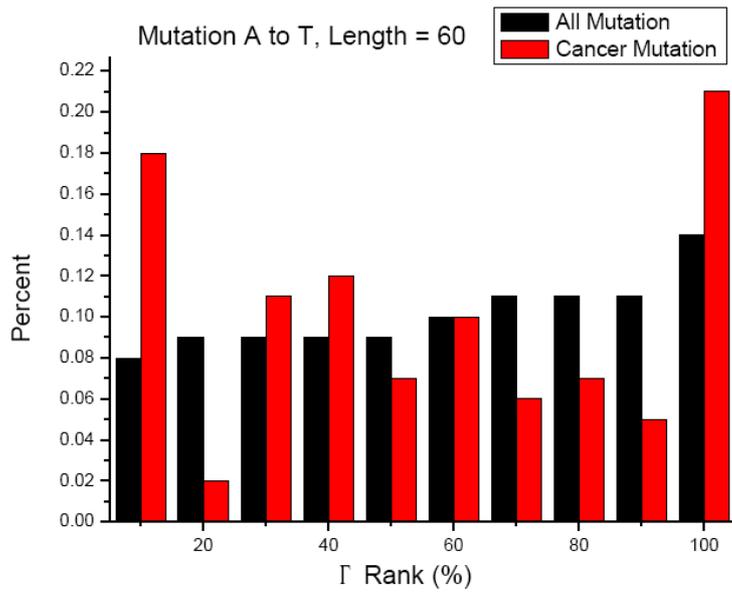


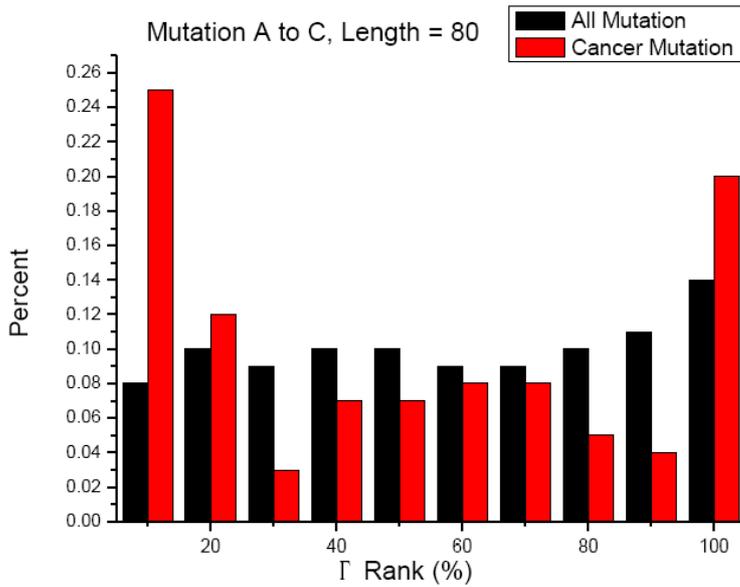
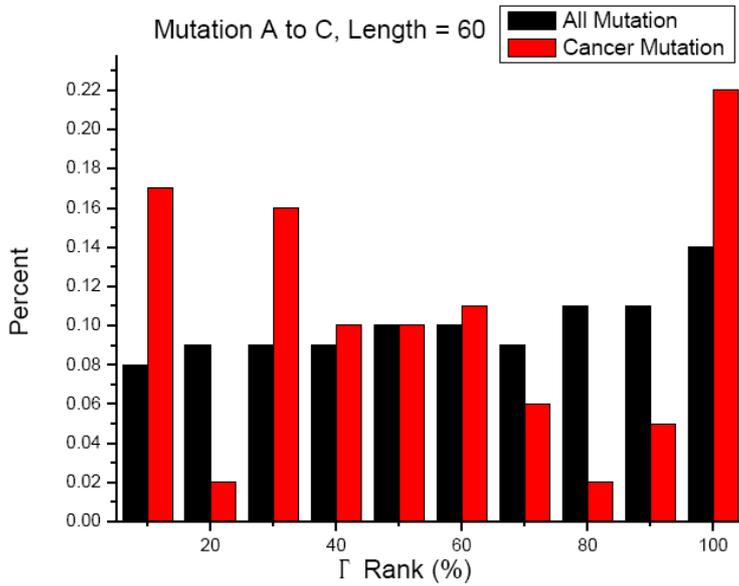
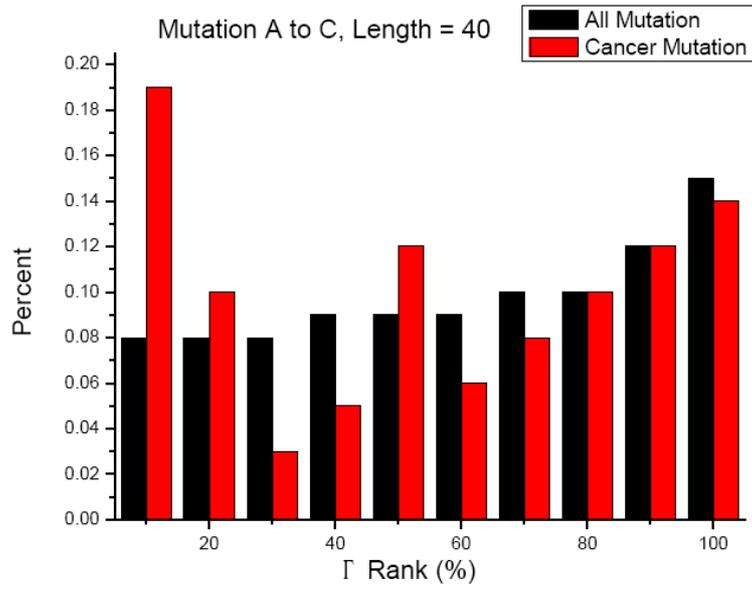


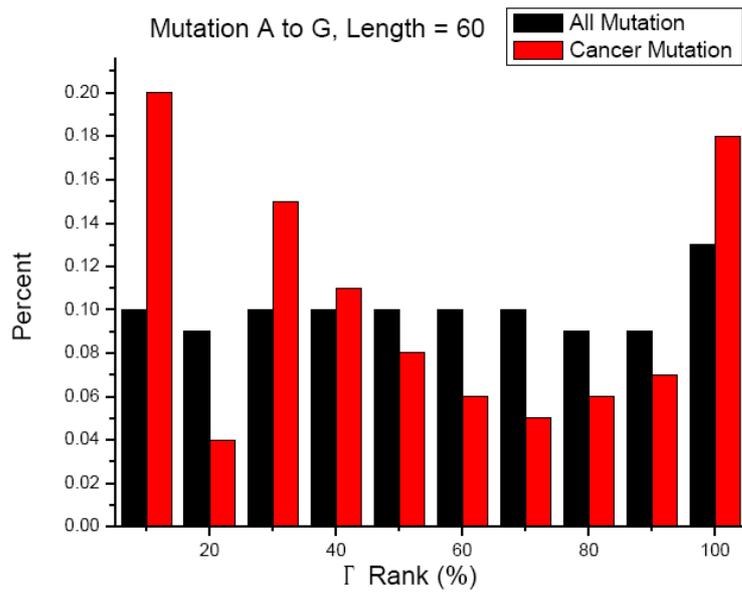
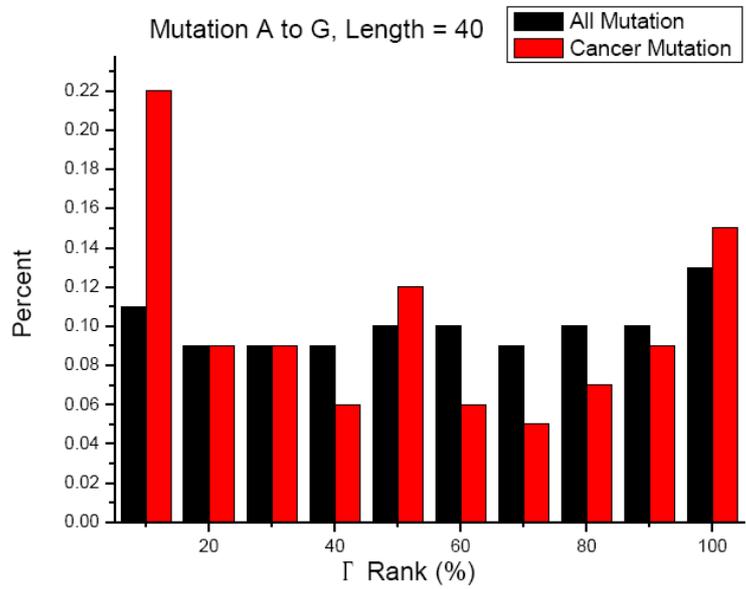
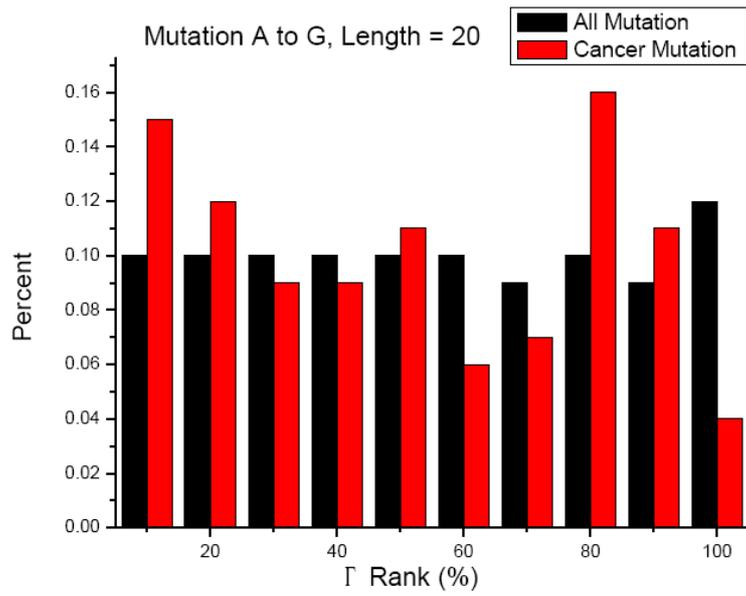


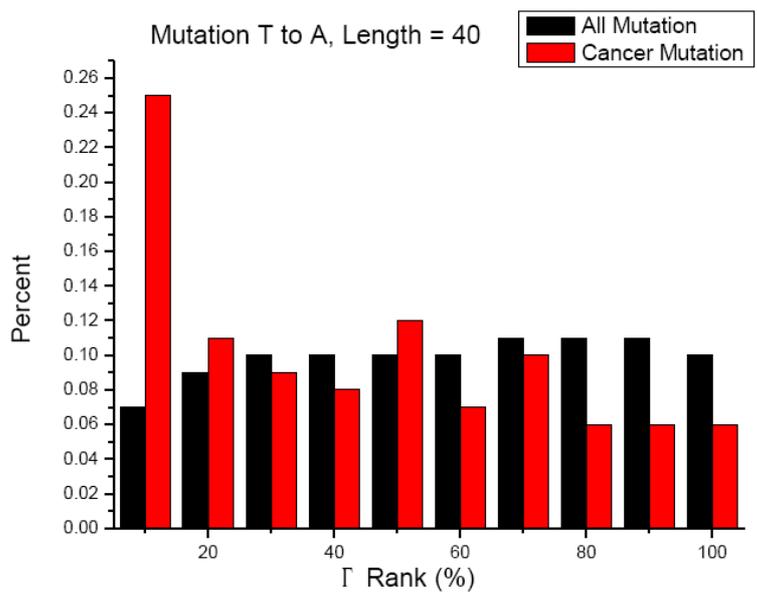
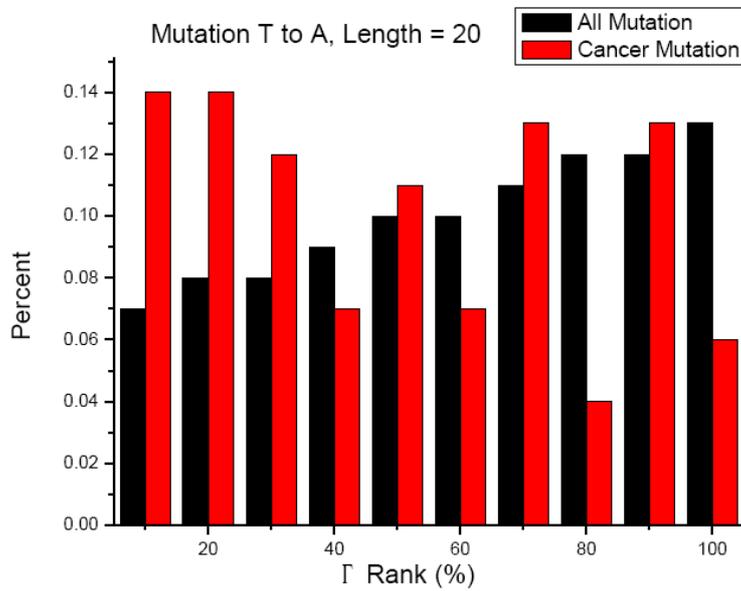
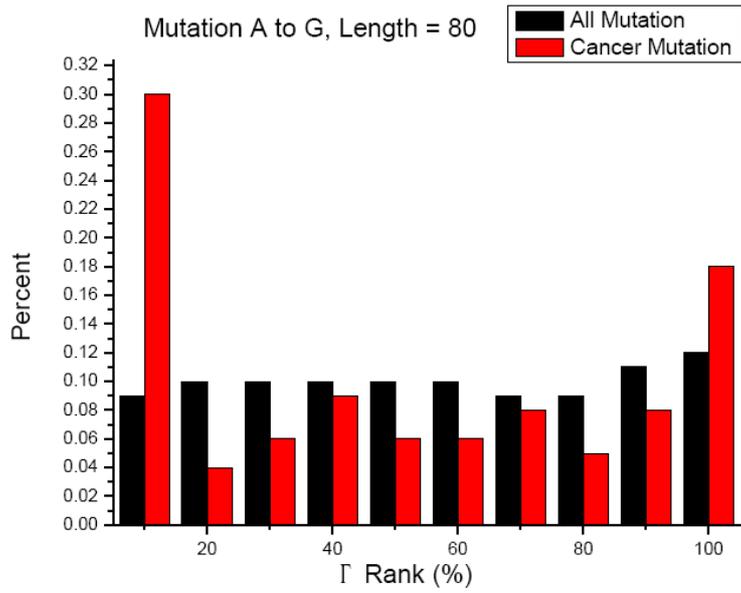
TP53 的 12 種致癌突變與全部 12 種可能突變，在不同長度時的 Γ 名次分佈：

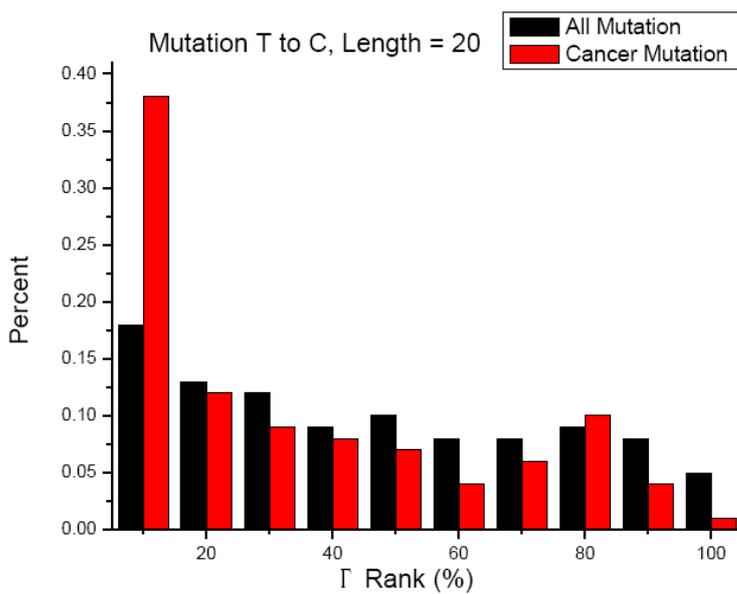
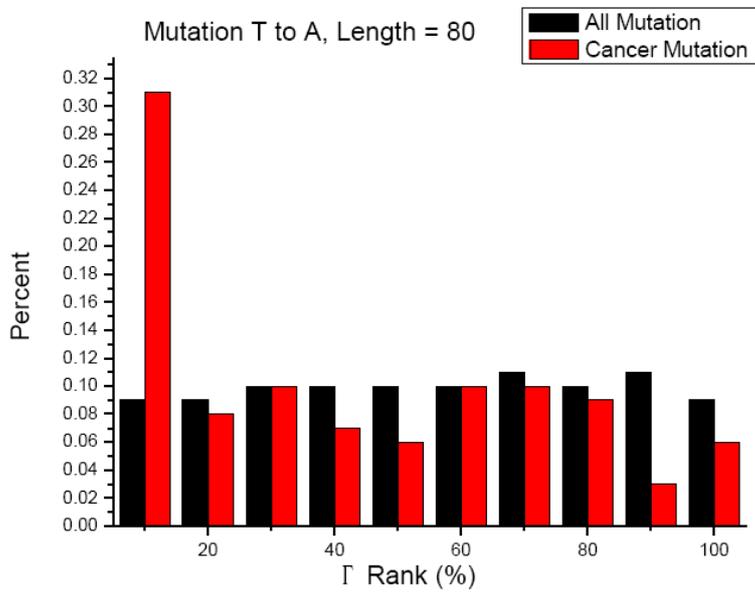
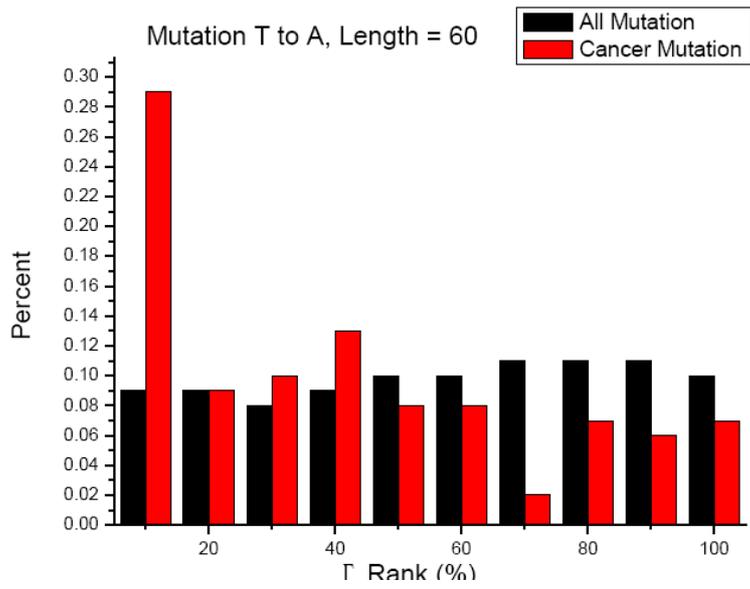


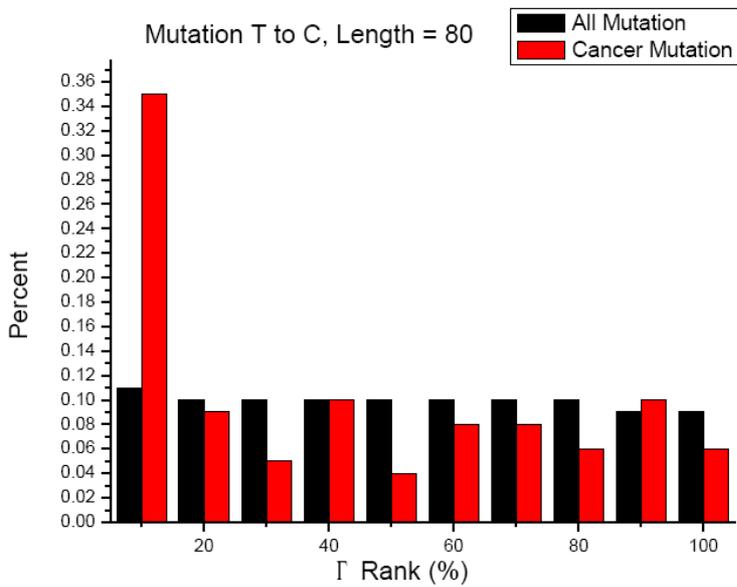
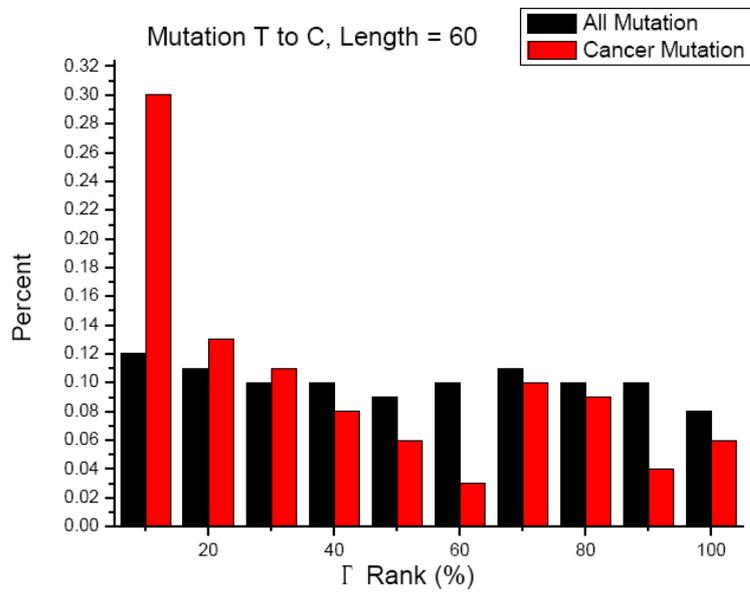
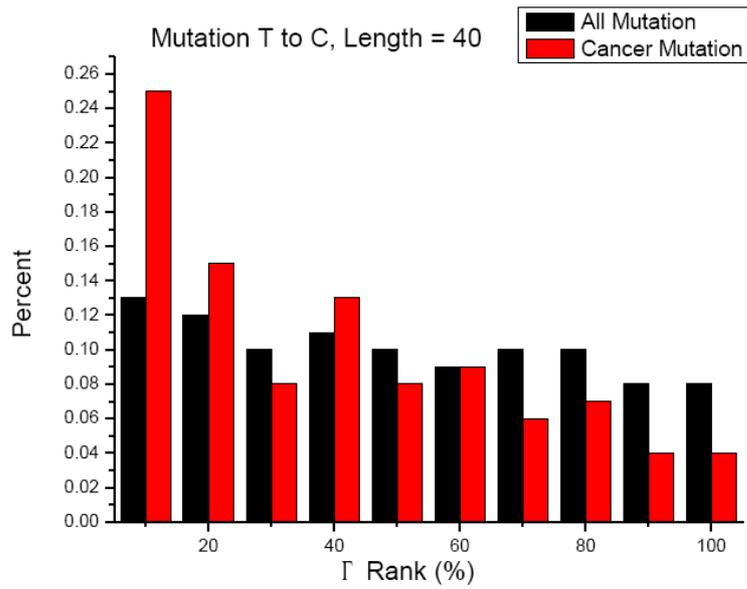


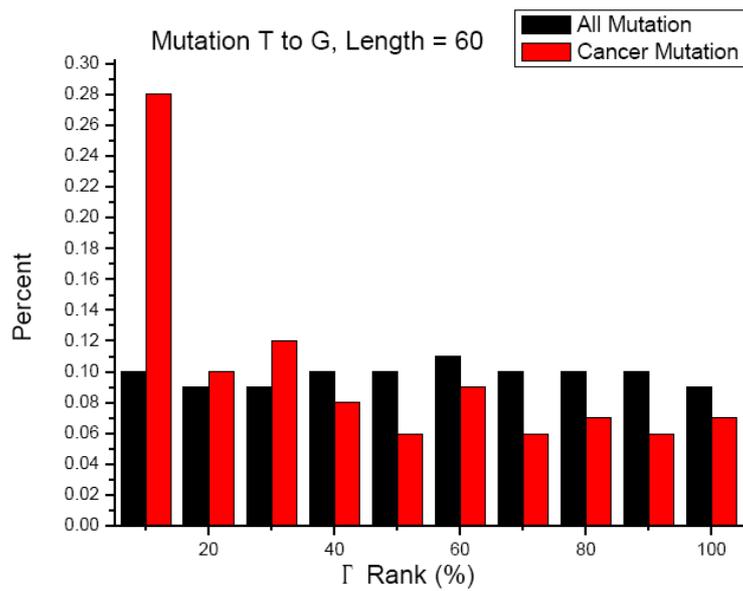
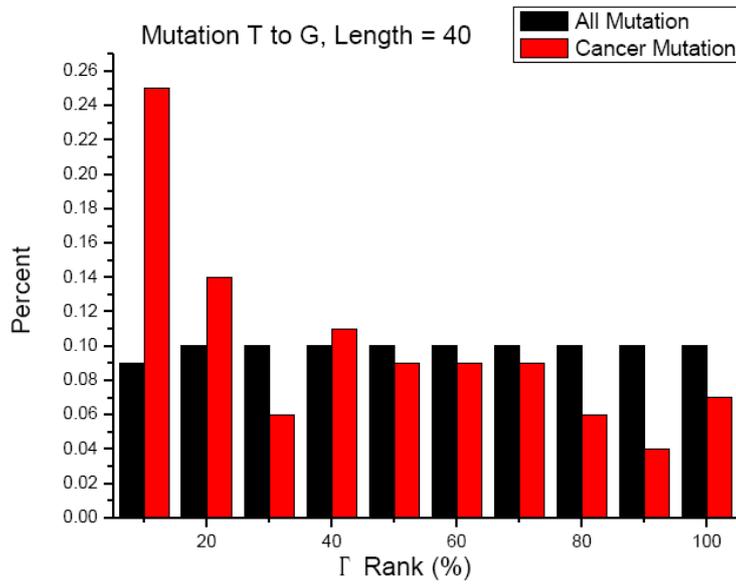
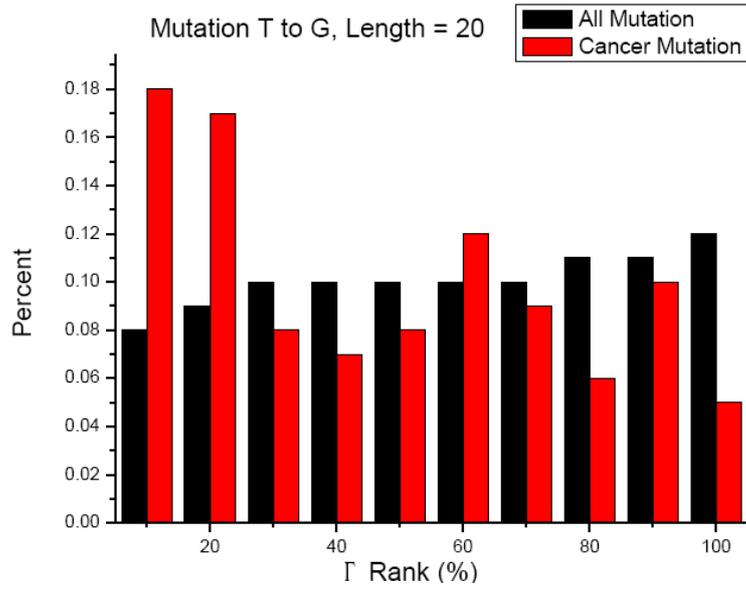


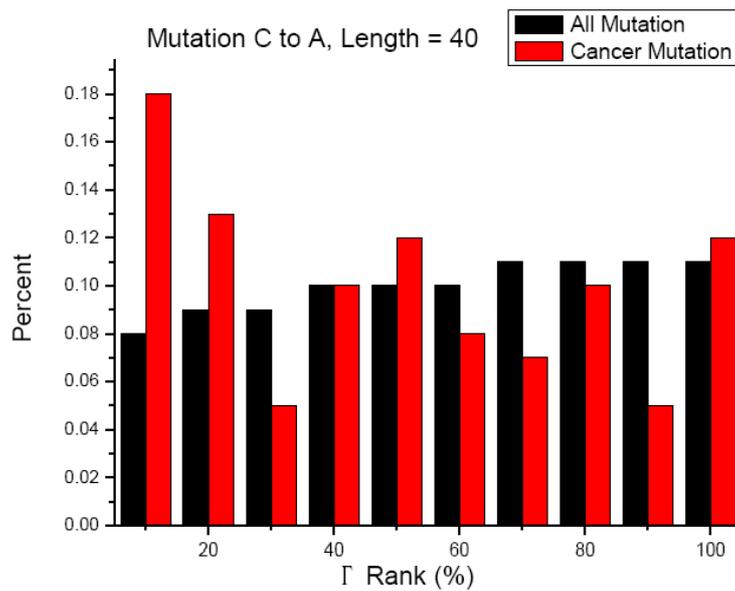
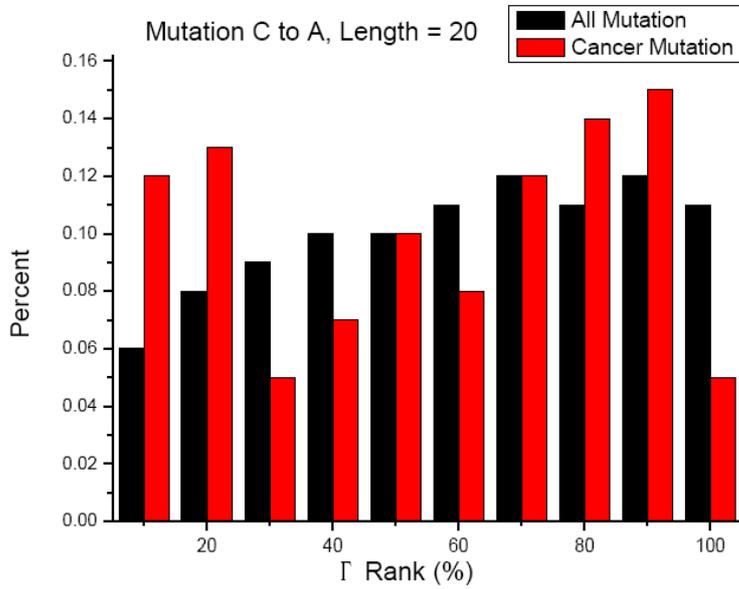
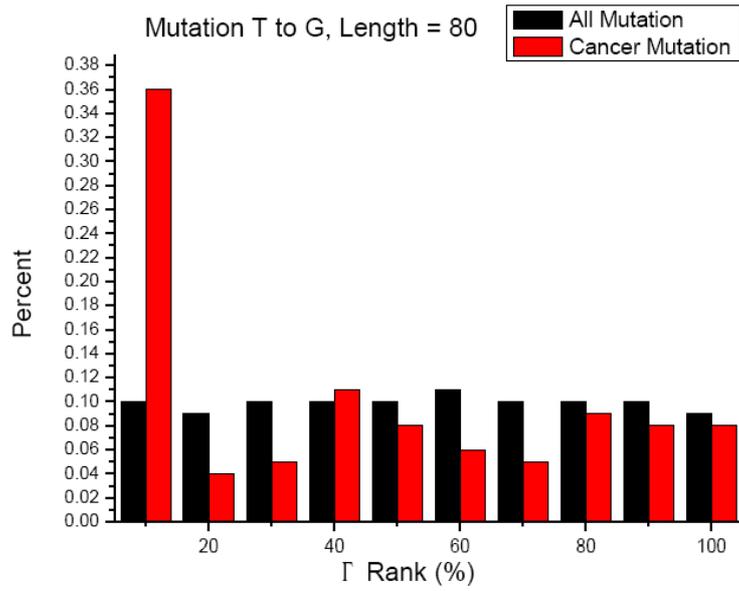


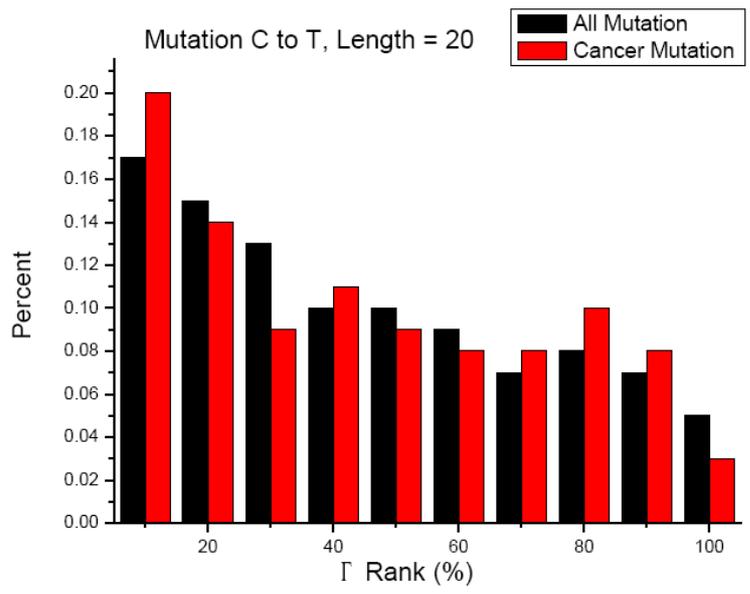
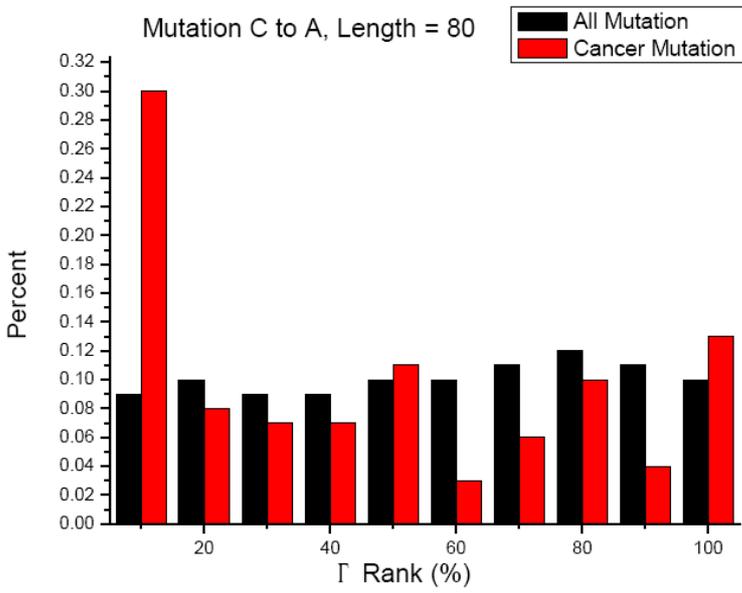
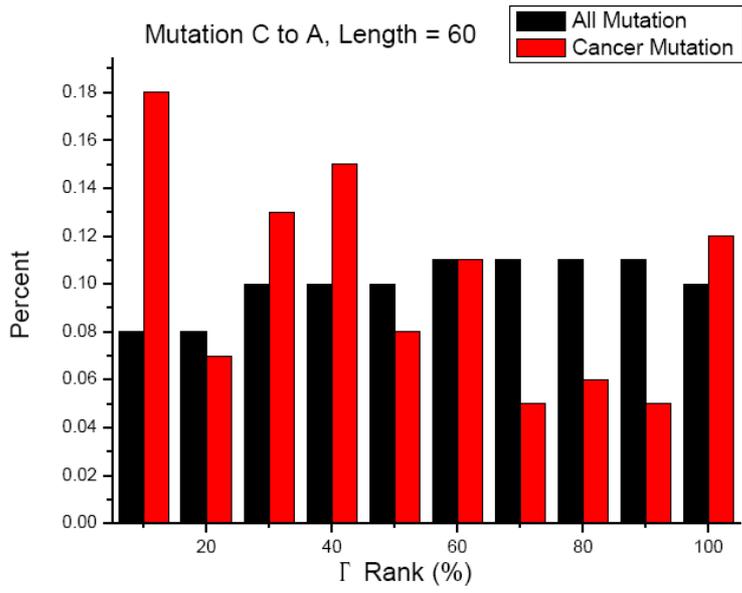


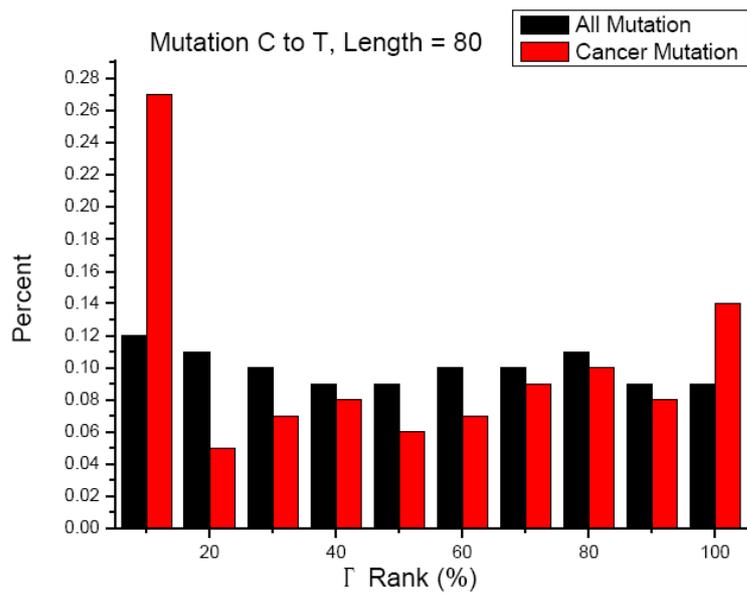
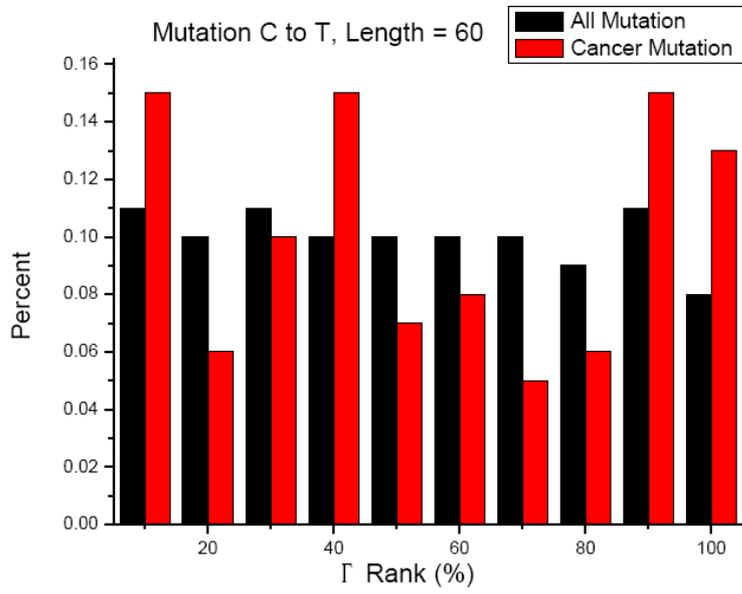
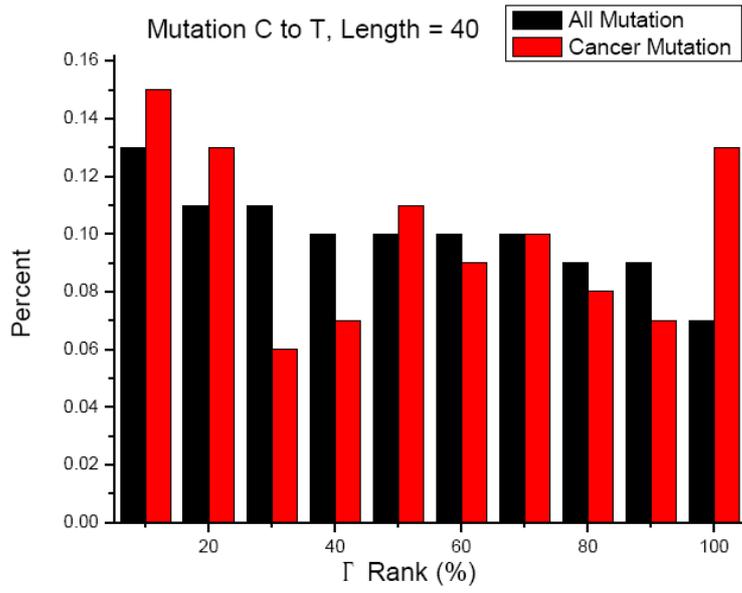


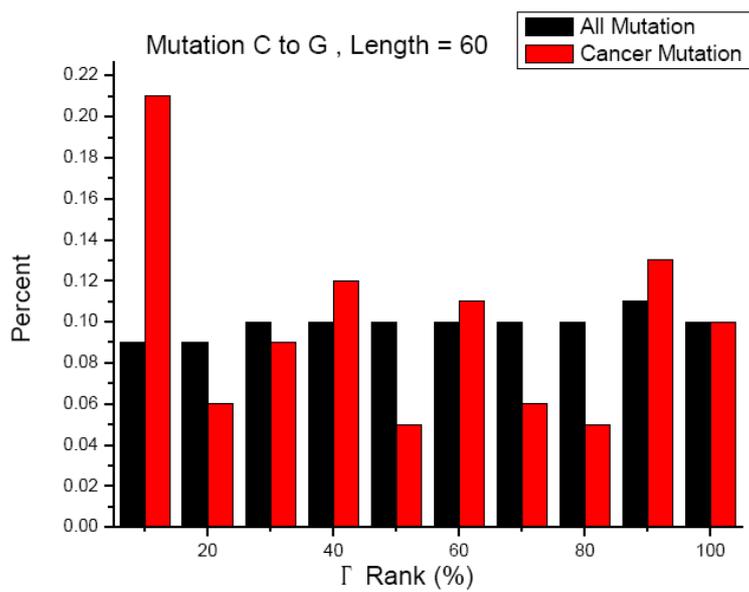
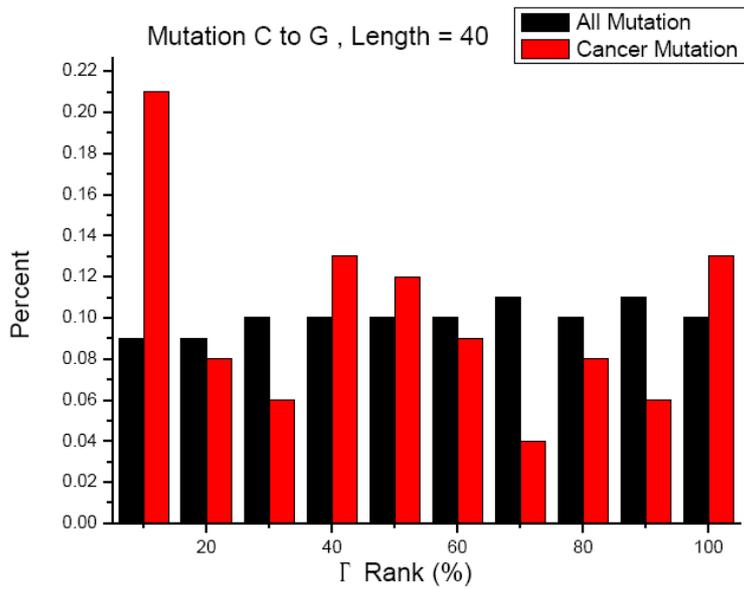
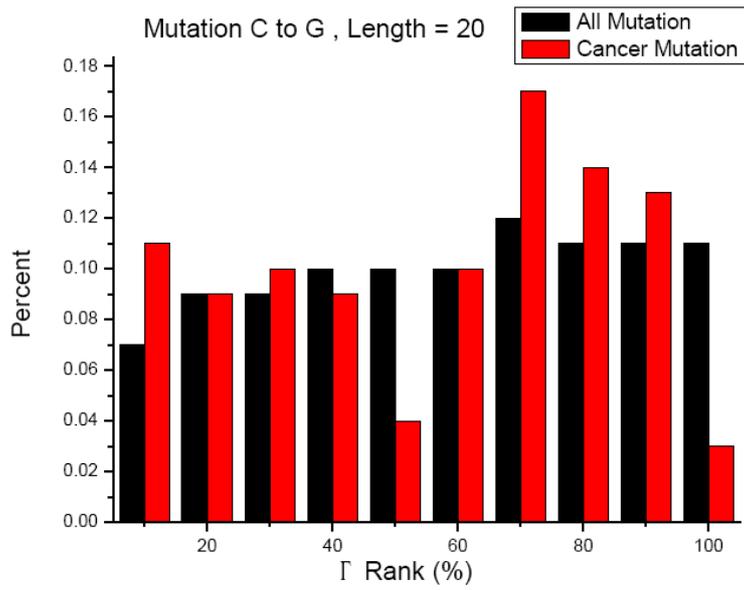


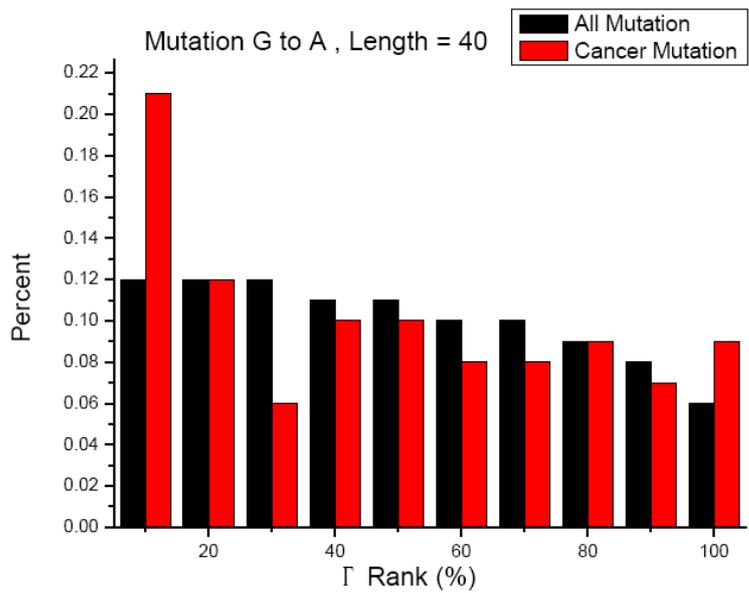
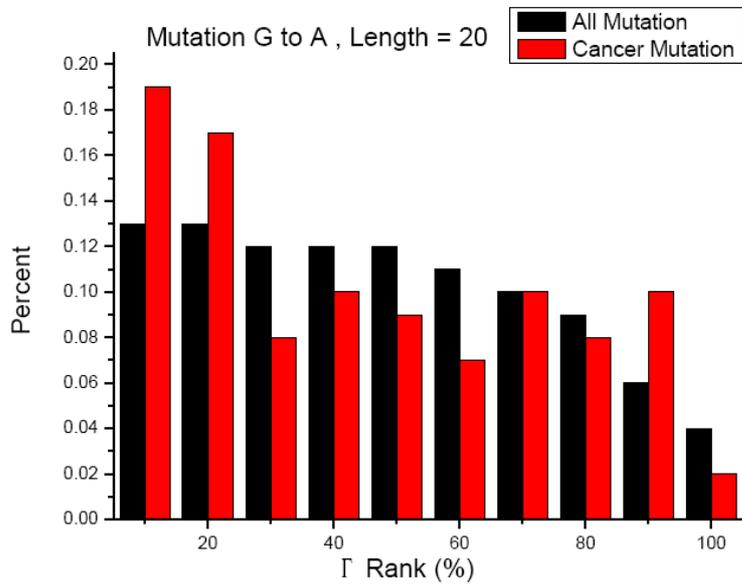
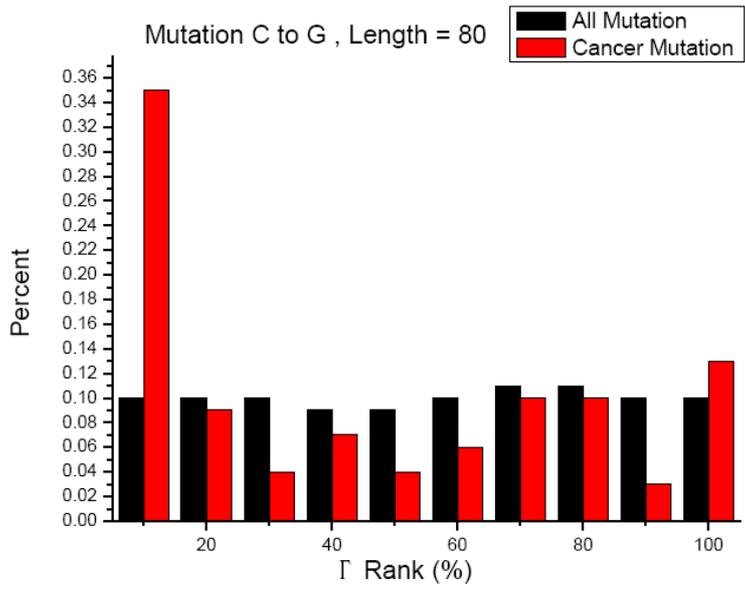


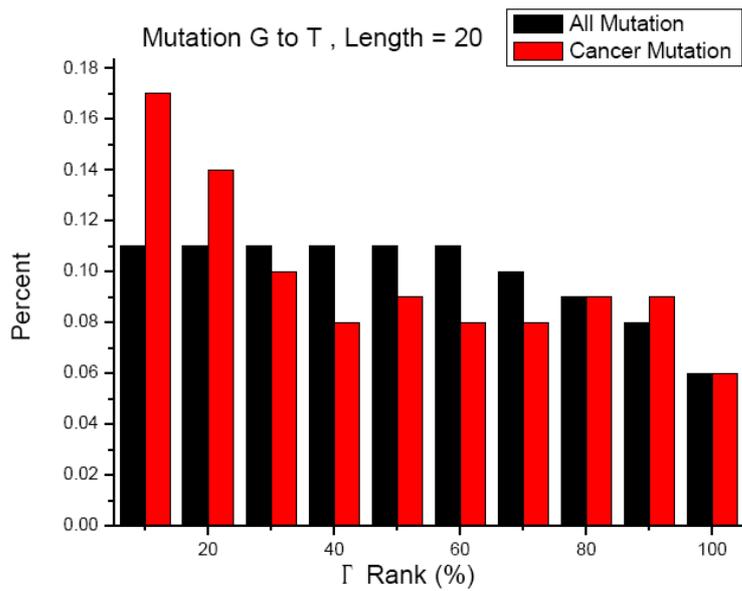
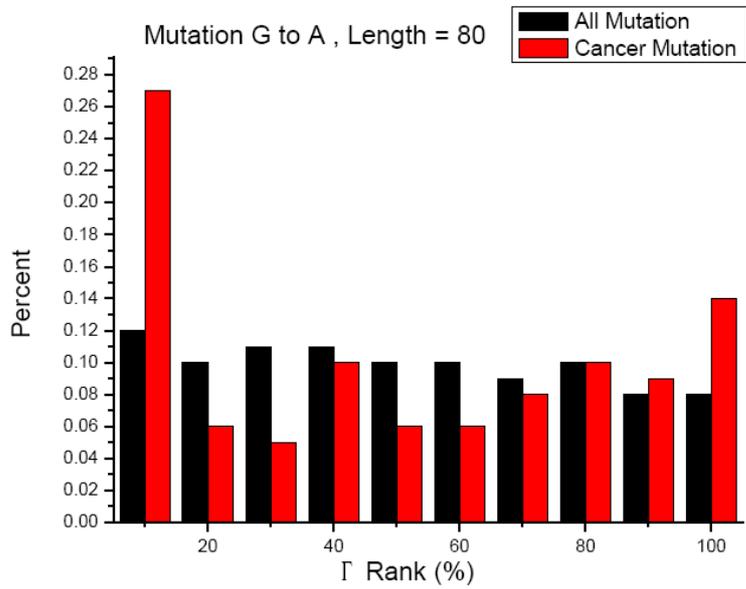
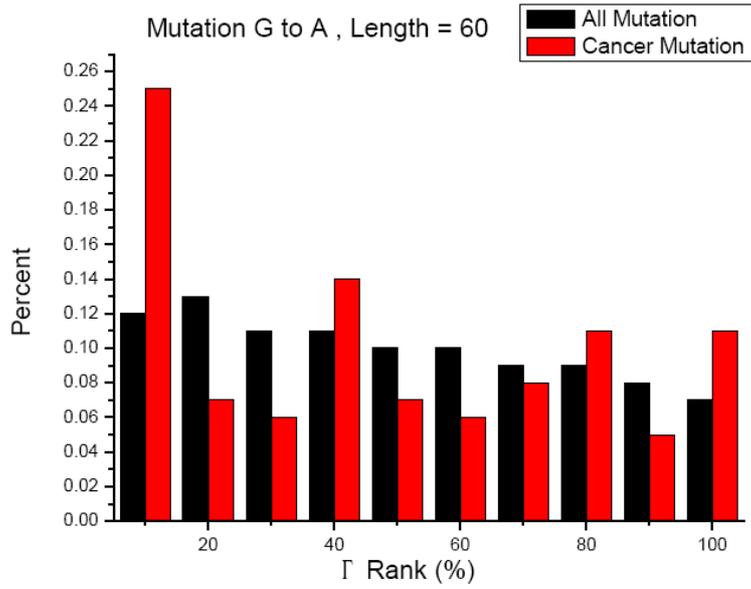


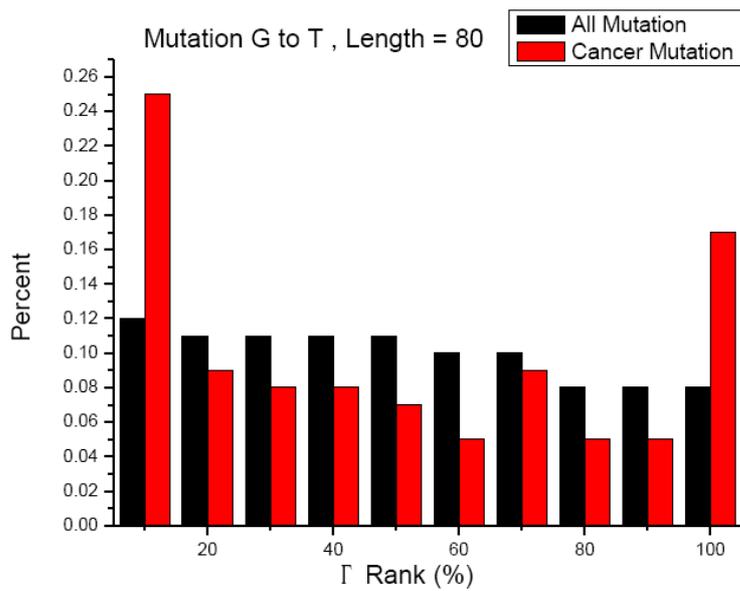
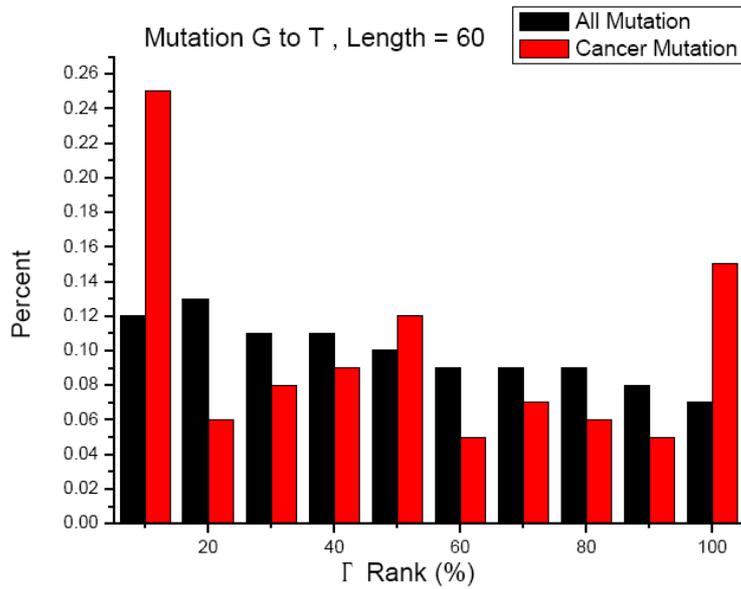
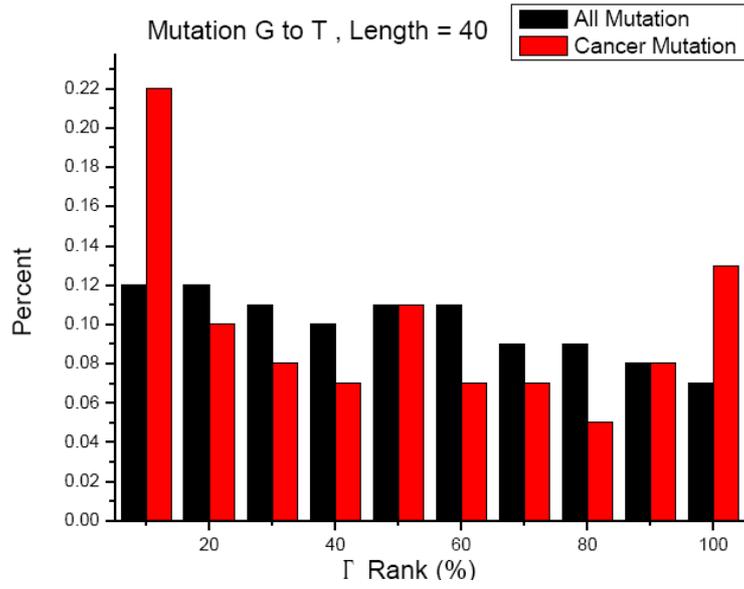


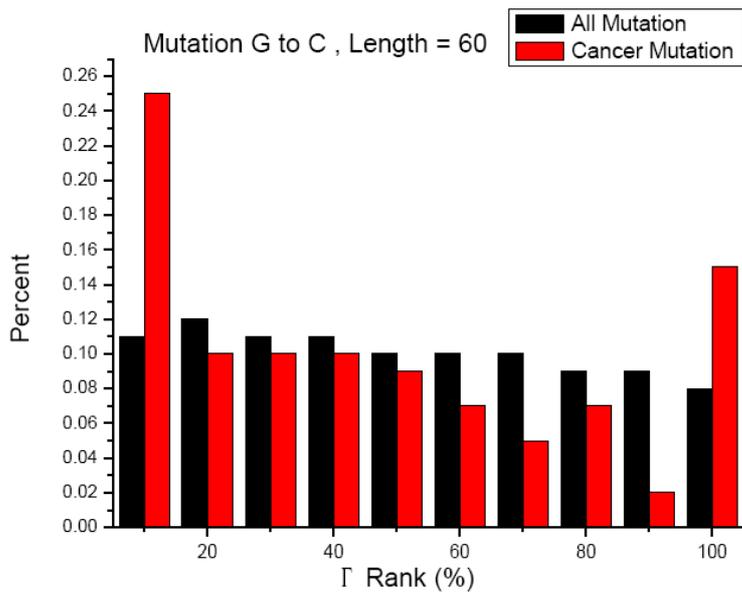
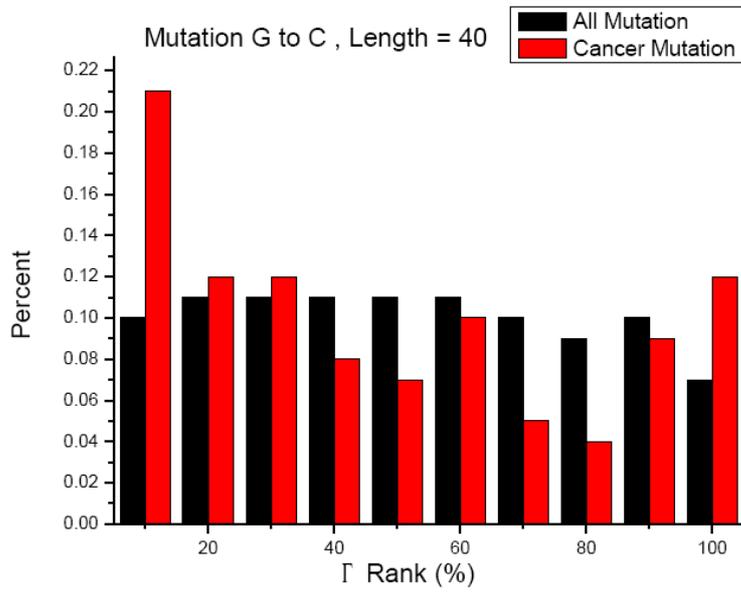
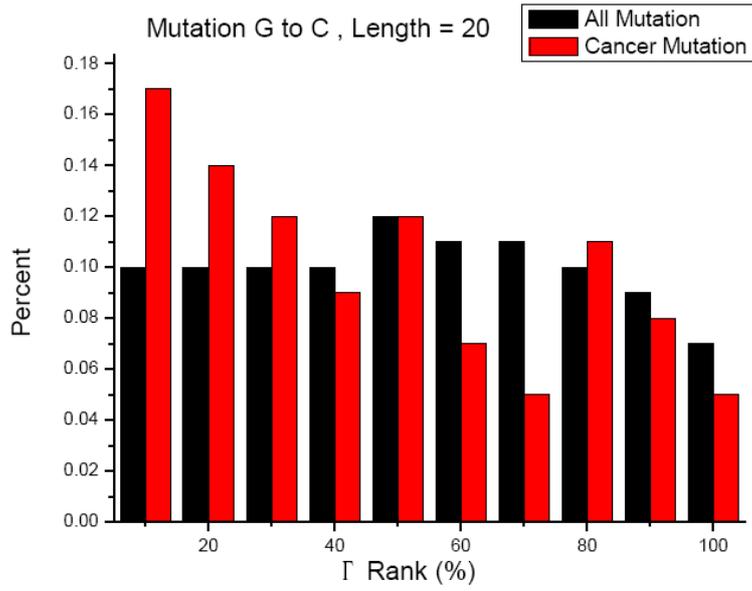


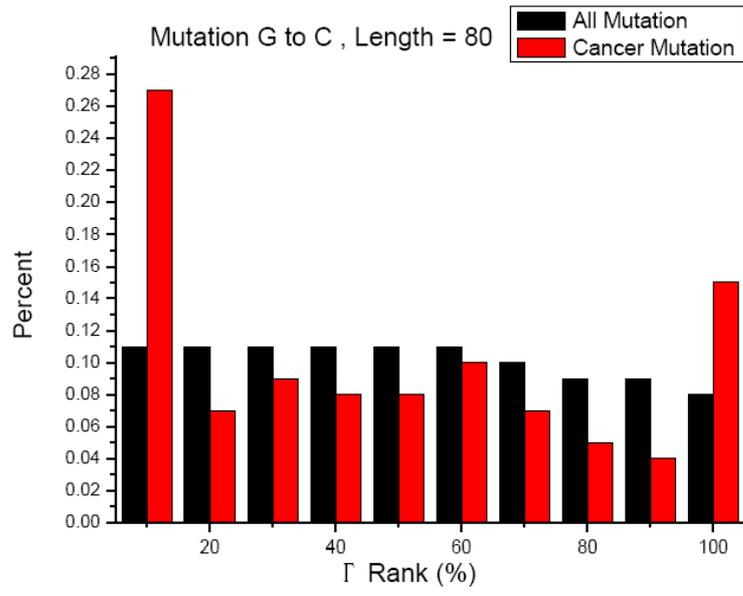












附錄 3

致病 DNA 序列與序列的相依性：

