

# 生物資訊學第五次作業

## 1. 方法一：透過關鍵字搜尋選取序列

### 我選擇的基因與物種

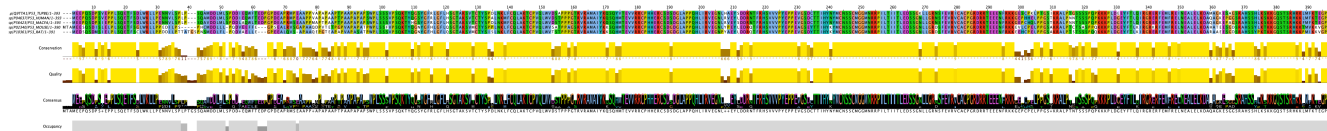
使用 UniProt 的進階搜尋功能，輸入基因名稱「TP53」並限定資料為 Reviewed (Swiss-Prot)，選出以下五種物種的 TP53 序列：

- 人 (*Homo sapiens*)
- 小鼠 (*Mus musculus*)
- 大鼠 (*Rattus norvegicus*)
- 食蟹獼猴 (*Macaca fascicularis*)
- 樹鼩 (*Tupaia belangeri*)

```
1 >sp|P02340|P53_MOUSE Cellular tumor antigen p53 OS=Mus musculus OX=10090 GN=TP53
PE=1 SV=4 MTAMEESQSDISLELPLSQETFSGLWKLLPPEDILPSPHCMDDLLLPQDVVEEFFEGPSEA
LRVSGAPAAQDPVTETPGPVAPAPATPWPLSSSFVPSQKTYQGNYGFHLGFLQSGTAKSVM
CTYSPPLNKLFCQLAKTCPVQLWVSATPPAGSRVRAMAIYKKSQHMTEVVRRCPHHERCS
DGDGLAPPQHLIRVEGNLYPEYLEDQRQTFRHSVVVPYEPPEAGSEYTTIHYKYMCSNSSCM
GGMNRRPILTIITLEDSSGNLLGRDSFEVRVCACPRDRRTEENFRKKEVLCPELPPGS
AKRALPTCTSASPPQKKKPLDGEYFTLKIRGRKRFEMFRELNEALELKDAHATEESGDSR
AHSSYLKTKKGQSTSRHKKTMVKKVGPDS >sp|P04637|P53_HUMAN Cellular tumor antigen p53
OS=Homo sapiens OX=9606 GN=TP53 PE=1 SV=4
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRGLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGRVVRAMAIYKKSQHMTEVVRRCPHHE
RCSDS DGLAPPQHLIRVEGNLRVEYLDNRNTRFRHSVVVPYEPPEVGS DCTTIHYNMCSN
SCMGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPRDRRTEENLRKKGEPPHEL
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD >sp|P10361|P53_RAT Cellular tumor antigen p53
OS=Rattus norvegicus OX=10116 GN=TP53 PE=1 SV=1
MEDSQSDMSIELPLSQETFSCLWKLLPPDDILPTTATGSPNSMEDLFLPQDVAELLEGPE
EALQVSAPAAQEPGTEAPAPVAPASATPWPLSSSVPSQKTYQGNYGFHLGFLQSGTAKSV
MCTYSISLNKLFCQLAKTCPVQLWVTSTPPPGRVVRAMAIYKKSQHMTEVVRRCPHHERC
SDGDGLAPPQHLIRVEGNPYAEYLDNRQTFRHSVVVPYEPPEVGS DYTIIHYKYMCSN
MGMNRRPILTIITLEDSSGNLLGRDSFEVRVCACPRDRRTEENFRKKEEHCPCLPPG
SAKRALPTSTSSSPQPKKKPLDGEYFTLKIRGRERFEMFRELNEALELKDARAAEESGDS
RAHSSYPKTKKGQSTSRHKKPMIKKVGPDSD >sp|P56423|P53_MACFA Cellular tumor antigen p53
OS=Macaca fascicularis OX=9541 GN=TP53 PE=2 SV=2
MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPLPSQAVDDLMLSPDDLAQWLTEDPGP
DEAPRMSEAAPPMAPTPAAPTAPAPAPAPSWPLSSSVPSQKTYHGSYGFRGLGFLHSGTAK
SVTCTYSPDLNKMFCQLAKTCPVQLWVDSTPPPGSRVVRAMAIYKKSQHMTEVVRRCPHHE
RCSDS DGLAPPQHLIRVEGNLRVEYSDNRNTRFRHSVVVPYEPPEVGS DCTTIHYNMCSN
SCMGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPRDRRTEENFRKKGEPC HQLP
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPA
GSRAHSSHLKSKKGQSTSRHKKFMFKTEGPDSD >sp|Q9TTA1|P53_TUPBE Cellular tumor antigen p53
OS=Tupaia belangeri OX=37347 GN=TP53 PE=2 SV=1
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRGLGFLHSGTAK
```

SVTCTYSPDLNKLFCQLAKTCPVQLWVDSAPPPGTRVRAMAIYKQSQYVTEVRRCPHHE  
RCSDSDDLAPPQHLIRVEGNLHAEYSDRNTFRHSVVVPYEPPEVGSDCTTIHNYMCNS  
SCMGMNRRPILTIITLEDSSGKLLGRNSFEVRICACPRDRRTEENFRKKGESCPKLP  
TGSIKRALPTGSSSSPQPKKKPLDEEYFTLQIRGRERFEMLEINEALELKDAMAGKESA  
GSRAHSSHLKSKKGQSTSRHRKLMFKTEGPDSD

觀察MSA結果。



```
1  CLUSTAL O(1.2.4) multiple sequence alignment
2
3
4  sp|Q9TTA1|P53_TUPBE      ---MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLP---
   SQAMDDLMLSPDDIEQWF
5  sp|P04637|P53_HUMAN      ---MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLP---
   SQAMDDLMLSPDDIEQWF
6  sp|P56423|P53_MACFA      ---MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPLP---
   SQAVDLMLSPDDLAQWL
7  sp|P02340|P53_MOUSE      MTAMEESQSDISLELPLSQETFSGLWKLLPPEDILPS-----PHCMDDL---
   PQDVEEFF
8  sp|P10361|P53_RAT        ---MEDSQSDMSIELPLSQETFSCLWKLLPPDDILPTTATGSPNSMEDLFL-
   PQDVAELL
9
   ** : *** *: * ***** ***** :: *          :. :. **: * **: :
   :
10
11 sp|Q9TTA1|P53_TUPBE
   TEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRGLGFL
12 sp|P04637|P53_HUMAN
   TEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRGLGFL
13 sp|P56423|P53_MACFA
   TEDPGPDEAPRMSEAAPPMAPTPAAPTPAAPAPAPSWPLSSSVPSQKTYHGSYGFRGLGFL
14 sp|P02340|P53_MOUSE      E---
   GPSEALRVSGAPAAQDPVTETPGPVAPAPATPWPLSSFVPSQKTYQGNYGFHLGFL
15 sp|P10361|P53_RAT        E---GPEEALQVS-
   APAAQEPGTEAPAPVAPASATPWPLSSSVPSQKTYQGNYGFHLGFL
16
   **, ** :: *      *      : * *, *** *      *****
   ***** : *, ** : ****
17
18 sp|Q9TTA1|P53_TUPBE
   HSGTAKSVTCTYSPDLNKLFCQLAKTCPVQLWVDSAPPPGTRVRAMAIYKQSQYVTEVVR
19 sp|P04637|P53_HUMAN
   HSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGRVRAMAIYKQSQHMTEVVR
20 sp|P56423|P53_MACFA
   HSGTAKSVTCTYSPDLNKMFCQLAKTCPVQLWVDSTPPPGSRVRAMAIYKQSQHMTEVVR
21 sp|P02340|P53_MOUSE
   QSGTAKSVMCTYSPPLNKLFCQLAKTCPVQLWVSATPPAGSRVRAMAIYKKSQHMTEVVR
22 sp|P10361|P53_RAT
   QSGTAKSVMCTYSISLNKLFCQLAKTCPVQLWVTSTPPPGRVRAMAIYKKSQHMTEVVR
23
   : ***** ***** *** : ***** ***** :: **
   * : ***** : ** : *****
24
```

```

25 sp|Q9TTA1|P53_TUPBE
RCPHHERCSDSDGLAPPQHLIRVEGNLHAEYSDDRNTFRHSVVVPYEPPEVGSDCTTIHY
26 sp|P04637|P53_HUMAN
RCPHHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHY
27 sp|P56423|P53_MACFA
RCPHHERCSDSDGLAPPQHLIRVEGNLRVEYSDDRNTFRHSVVVPYEPPEVGSDCTTIHY
28 sp|P02340|P53_MOUSE
RCPHHERCSDGDGLAPPQHLIRVEGNLYPEYLEDQRQTFRHSVVVPYEPPEAGSEYTTIHY
29 sp|P10361|P53_RAT
RCPHHERCSDGDGLAPPQHLIRVEGNPYAEYLDDRQTFRHSVVVPYEPPEVGSDYTTIHY
30 *****.***** ** **:*****.**:
****
31
32 sp|Q9TTA1|P53_TUPBE
NYMCNSSCMGMNRRPILTIITLEDSSGKLLGRNSFEVRICACGRDRRTEENFRKKGE
33 sp|P04637|P53_HUMAN
NYMCNSSCMGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACGRDRRTEENLRKKGE
34 sp|P56423|P53_MACFA
NYMCNSSCMGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACGRDRRTEENFRKKGE
35 sp|P02340|P53_MOUSE
KYMCNSSCMGMNRRPILTIITLEDSSGNLLGRDSFEVRVCACGRDRRTEENFRKKEV
36 sp|P10361|P53_RAT
KYMCNSSCMGMNRRPILTIITLEDSSGNLLGRDSFEVRVCACGRDRRTEENFRKKEE
37 :*****:****:*****:*****:***
38
39 sp|Q9TTA1|P53_TUPBE
SCPPLPTGSIKRALPTGSSSPQPKKKPLDEEYFTLQIRGRERFEMFREINEALELKDAM
40 sp|P04637|P53_HUMAN
PHHELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELEALELKDAQ
41 sp|P56423|P53_MACFA
PCHQLPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELEALELKDAQ
42 sp|P02340|P53_MOUSE
LCPELPPGSAKRALPTCTASPPQKKKPLDGEYFTLKIRGRKRFEMFRELEALELKDAH
43 sp|P10361|P53_RAT
HCPELPPGSAKRALPTSTSSSPQKKKPLDGEYFTLKIRGRERFEMFRELEALELKDAR
44 **: ** *****. **: ** *****
*****:****:****:**:*****
45
46 sp|Q9TTA1|P53_TUPBE AGKESAGSRAHSSHLKSKKGQSTSRHRKLMFKTEGPDS
47 sp|P04637|P53_HUMAN AGKEPGGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDS
48 sp|P56423|P53_MACFA AGKEPAGSRAHSSHLKSKKGQSTSRHKKFMFKTEGPDS
49 sp|P02340|P53_MOUSE ATEESGDSRAHSSYLKTKKGQSTSRHKKTMVKKVGPDS
50 sp|P10361|P53_RAT AAEEGDSRAHSSYPKTKKGQSTSRHKKPMIKKVPDS
51 * : * ..*****: *:*****:* *. *. *****

```

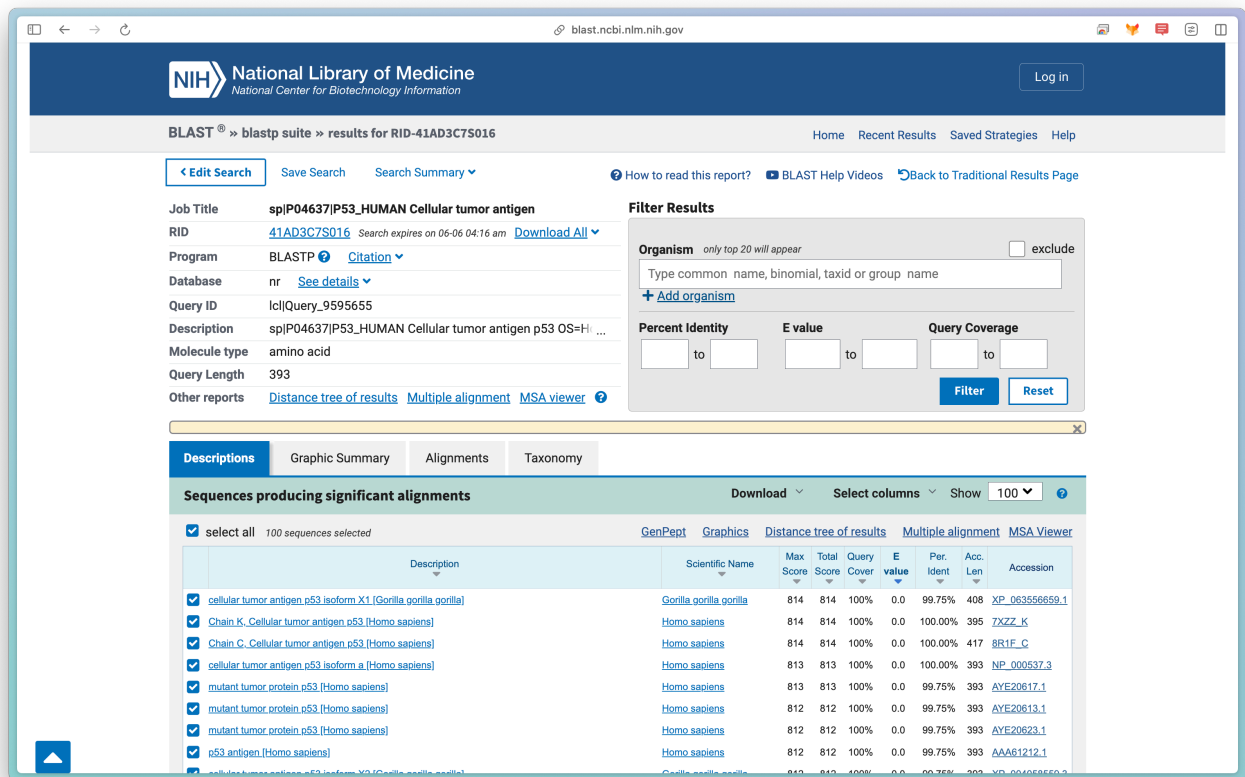
- **保守熱點**：整個 TP53 序列中，DNA-binding domain（第 50–300 位）和 Tetramerization domain（第 300–350 位）幾乎完全保守，演化壓力在多數哺乳動物中都集中於維持這兩段與基因調控及蛋白四聚體化直接相關的功能。
- **演化可塑區**：N 端（第 1–50 位）與 C 端（第 350–末端）在五條序列裡可見較多的插入/刪除、氨基酸替換或序列展延，表示這些調控區在不同物種中對蛋白質交互作用、後轉譯修飾需求各異，演化上允許更大幅度的改變。
- **整體保守度**：若以柱狀保守度圖量化，DNA-binding 及 Tetramerization 兩段的平均保守度均在 0.9–1.0（即 9–10 分制），而 NTD/CTD 的保守度約介於 0.4–0.8，差異顯著；這樣的保守分佈正對應 TP53 在腫瘤抑制功能與生物體差異需求之間維持平衡的演化策略。

總之，這次多序列比對結果強烈顯示 TP53 的「DNA-binding domain」和「Tetramerization domain」是五種哺乳動物所共通的功能核心，幾乎未見重大變異，突顯其作為腫瘤抑制蛋白的關鍵地位；而 N 端與 C 端的調控片段則因應不同物種特異的調控網絡而演化出多樣性。這樣的序列保守/可塑分佈，反映了 TP53 在保持基本功能的同時，也必須具備一定的彈性來對接不同生理需求。

## 2. 方法二：透過 BLASTp 篩選同源序列

## BLAST 搜尋與篩選

將人類 TP53 FASTA 上傳至 NCBI BLASTp, 篩選非人類的前幾名同源序列, 選取以下物種:



取前5個進行multiple sequence alignment

```

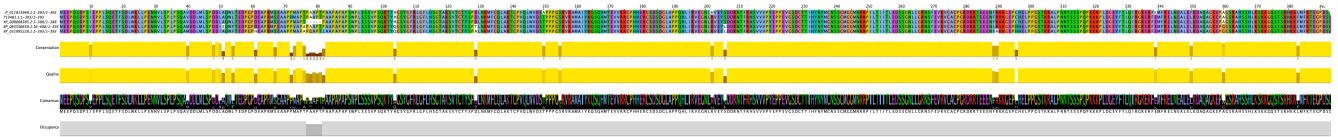
1 CLUSTAL O(1.2.4) multiple sequence alignment
2
3
4 XP_011813948.1:1-393
MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPLPSQAVDDLMLSPDDLAQWLTEDPGP
5 P13481.1:1-393
MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPLPSQAVDDLMLSPDDLAQWLTEDPGP
6 XP_008008385.2:1-388
MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPLPSQAVDDLMLSPDDLAQWLTEDPGP
7 XP_063556659.1:16-408
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
8 XP_031995228.1:1-393
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPEDIAQWFTEDPGP
9
*****:*****:*****:*
**:*****

```

```
11 XP_011813948.1:1-393
DEAPRMSEAAPPMAPTAAPTPAAPAPAPSWPLSSSVPSQKTYHGSYGFRLGFLHSGTAK
12 P13481.1:1-393
DEAPRMSEAAPPMAPTAAPTPAAPAPAPSWPLSSSVPSQKTYHGSYGFRLGFLHSGTAK
13 XP_008008385.2:1-388 DEAPRMSEAAPPMAPT——
PAAPAPAPSWPLSSSVPSQKTYHGSYGFRLGFLHSGTAK
14 XP_063556659.1:16-408
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
15 XP_031995228.1:1-393
HEAPRMSEAAPPMAAPAGAPTLAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
16 .***** ***** :** :
***** :*****
17
18 XP_011813948.1:1-393
SVTCTYSPDLNKMFCQLAKTCPVQLWVDSAPPPGSRVRAMAIYKQSQHMTEVVRRCPHHE
19 P13481.1:1-393
SVTCTYSPDLNKMFCQLAKTCPVQLWVDSTPPPGSRVRAMAIYKQSQHMTEVVRRCPHHE
20 XP_008008385.2:1-388
SVTCTYSPDLNKMFCQLAKTCPVQLWVDSTPPPGSRVRAMAIYKQSQHMTEVVRRCPHHE
21 XP_063556659.1:16-408
SVTCTYSPTLNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
22 XP_031995228.1:1-393
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
23 *****
***** :***** :*****
24
25 XP_011813948.1:1-393
RCSDSDDLAPPQHLIRVEGNLHVEYSDDRNTFRHSVVVPYEPPEVGSDCTTIHYNMNCNS
26 P13481.1:1-393
RCSDSDDLAPPQHLIRVEGNLHVEYSDDRNTFRHSVVVPYEPPEVGSDCTTIHYNMNCNS
27 XP_008008385.2:1-388
RCSDSDDLAPPQHLIRVEGNLHVEYSDDRNTFRHSVVVPYEPPEVGSDCTTIHYNMNCNS
28 XP_063556659.1:16-408
RCSDSDDLAPPQHLIRVEGNLHVEYLLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNMNCNS
29 XP_031995228.1:1-393
RCSDSDDLAPPQHLIRVEGNLHVEYLLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNMNCNS
30 ***** :***
*****
31
32 XP_011813948.1:1-393
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGDRRTEENFRKKGEPCHHELP
33 P13481.1:1-393
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGDRRTEENFRKKGEPCHHELP
34 XP_008008385.2:1-388
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGDRRTEENFRKKGEPCHHELP
35 XP_063556659.1:16-408
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGDRRTEENLKKKGEPHHELP
36 XP_031995228.1:1-393
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGDRRTEENFHKKKGEPHHELP
37 ***** :*****
****
38
39 XP_011813948.1:1-393
PGSTKRALPNNTSSSPQPKKPLDGEYFTLQIRGRERFKMFRELNEALEFKDAQAGKEPA
```



40	P13481.1:1-393	PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPA
41	XP_008008385.2:1-388	PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPA
42	XP_063556659.1:16-408	PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
43	XP_031995228.1:1-393	PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
44		*****:*****:*****.
45		
46	XP_011813948.1:1-393	GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
47	P13481.1:1-393	GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
48	XP_008008385.2:1-388	GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
49	XP_063556659.1:16-408	GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
50	XP_031995228.1:1-393	GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD



這次對齊的五條 TP53 序列從 N 端調控區至 C 端調控區共約 393 個殘基，整體可分為四大區域：

首先，第 1–50 殘基的 **N-terminal Transactivation Domain (NTD)** 顯示較多插入/缺失與氨基酸替換，尤其是靠近第 10–20 位的 Proline-rich region (Pro/Ser/Lys 聚集) 差異最多，例如小鼠和大鼠常出現額外的 Ser-Pro 插入或 P → L 置換，顯示此區在不同物種中對細胞應激、蛋白質互動和磷酸化修飾的需求不盡相同，保守度落在 0.4–0.7 左右。

接著，第 50–300 殘基對應於 **DNA-binding Domain**，此區經過強烈演化壓力保護：對齊中近乎每個位置均標示 \* (完全相同) 或 : (相似替換)，尤其第 90–135 位包含 L1、L2 迴圈和 zinc-finger Motif (Cys/His 四配位 Residue，如 “Cys 176, Cys 238, His 179, His 264”) 完全保留，平均保守度 > 0.9。這段區域中人類、靈長類和齧齒類之間僅在第 104、135、204 等少數位點出現同義替換 (例如 Gly ↔ Ala、Ser ↔ Thr)，但所有能與 DNA 結合的 α 螺旋與 β 折板結構都被完好保留，因而突顯 TP53 作為腫瘤抑制因子必須維持對 DNA 的精準鍵合。

第 300–350 殘基構成 **Tetramerization Domain**，該段包含 “ILTIITLE” 與 “LRVCACPG” 等四聚體界面關鍵序列，五條序列中依然維持高度同質 (標示多數 \* )，僅在少數 loop 或鏈環位置 (例如第 315 位左右) 可見 Pro → Ser、Arg → Lys 等保守性替換 (保守度約 0.8–0.9)。由於 p53 必須以四聚體形式結合 DNA，該域的完整性與精密互補性對維持功能至關重要，故演化極不容許重大變異。

最後，第 350–393 殘基的 **C-terminal Regulatory Domain (CTD)** 保守度再次下降，多見 Lys/Arg-rich 序列 (例如 “KRKKPLDGEYFTLQI”) 雖有部分保留，但在位於第 370–380 的泛素化或甲基化修飾位點常出現細微差異 (如 Thr → Ser、Glu → Asp)，顯示此區在不同物種對下游蛋白質互動與後轉譯調控的需求具備彈性。

綜合而言，TP53 的 DNA-binding 及 Tetramerization 兩大核心功能域在樹鼩、人類、食蟹獼猴、小鼠與大鼠之間幾乎零差異 (保守度 > 90%)，僅在輔助迴圈或結構邊緣位置有少許相似性置換；相比之下，N-terminal 和 C-terminal 調控區則展現了多樣性插入、缺失和非同義替換，反映出這兩段較不直接介入 DNA 鍵合與四聚體形成，演化上得以維持高度可塑性以適應不同細胞環境及修飾機制。