Ứng dụng CNTT trong Công nghệ Sinh học

PGS.TS. Trần Văn Lăng
Viện Cơ học và Tin học ứng dụng
Viện Hàn lâm Khoa học và Công nghệ Việt Nam
langtv@vast.vn

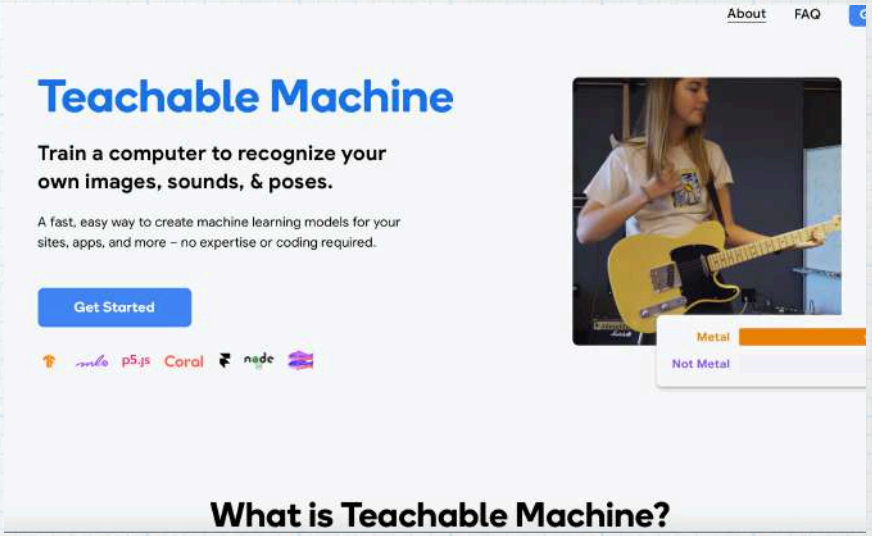
Nội dung

* Góc nhìn của Tin học về Sinh học phân tử
  - Khoa học sự sống
  - Protein và vai trò của protein trong sự sống
  - Cấu trúc của protein
  - Dự đoán cấu trúc bậc II protein (hay cấu trúc thứ cấp)

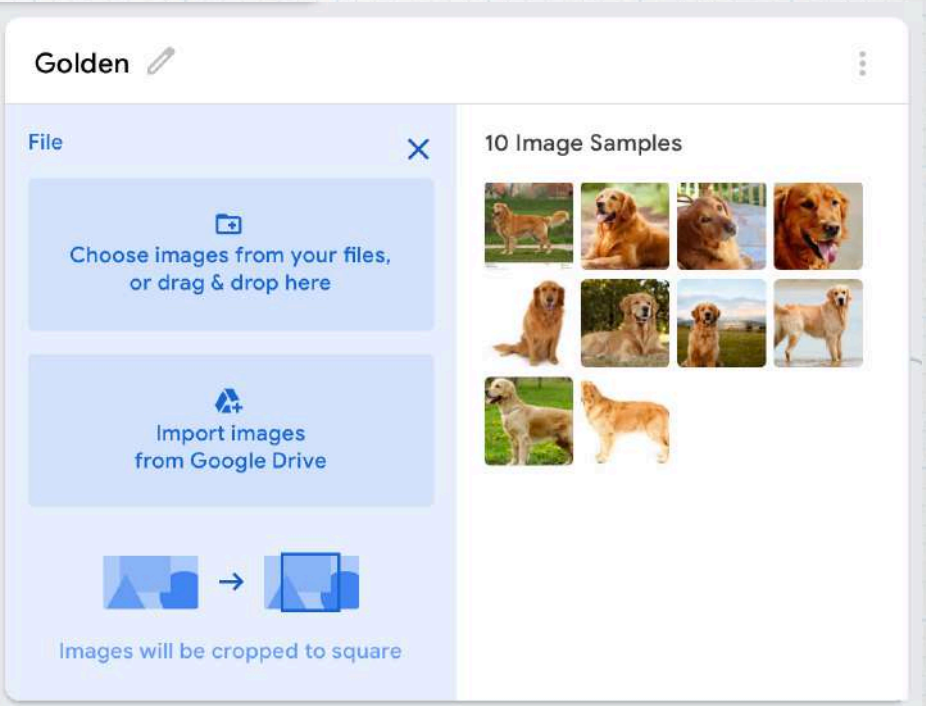
Trước hết

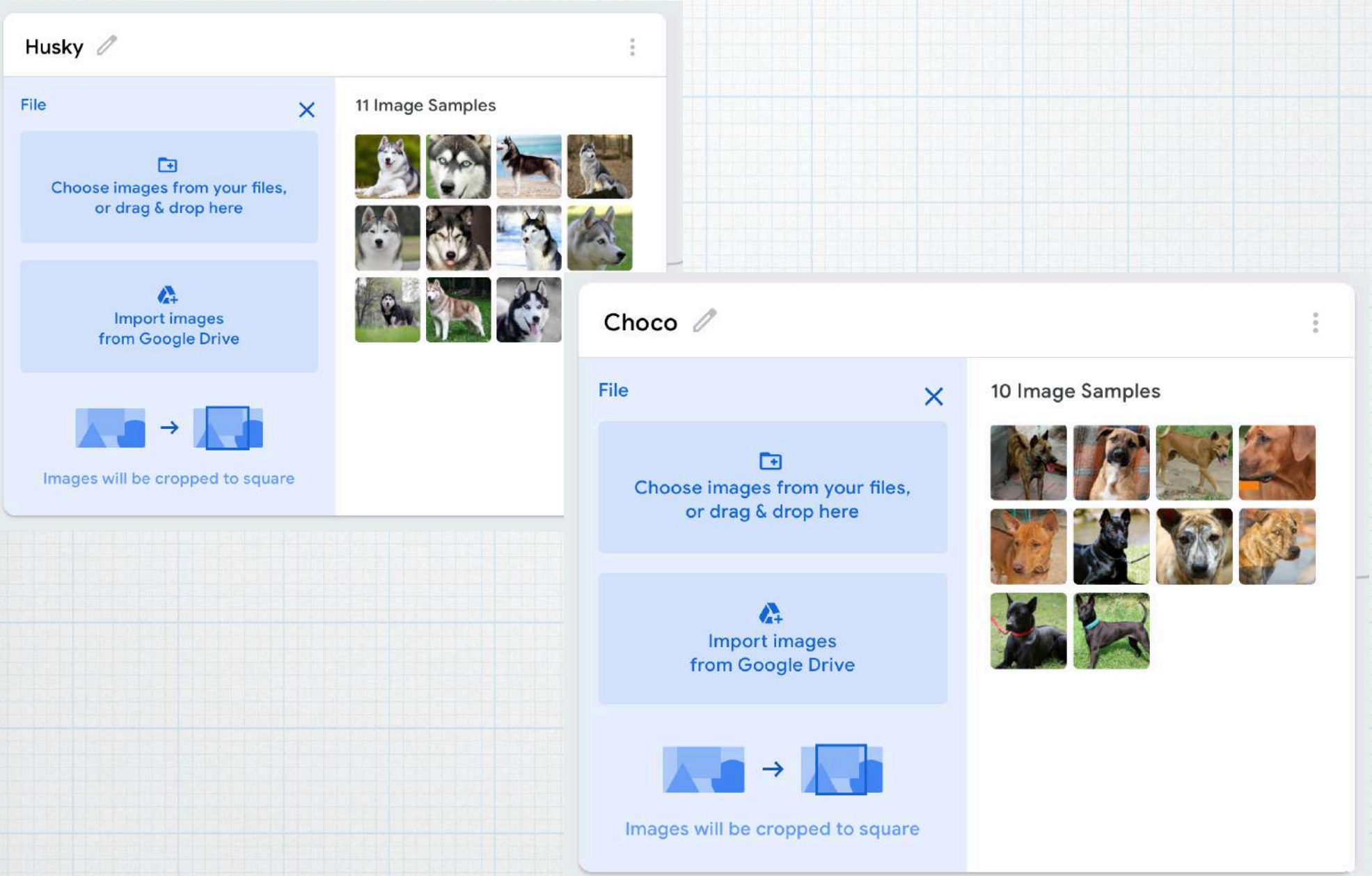
* Công nghệ hiện nay đã và đang phát triển khá vượt bật.

* Chẳng hạn, trong lĩnh vực AI (Thông minh nhân tạo), Google cũng đã có những đóng góp rất cụ thể về công nghệ; đó là Teachable Machine (https://teachablemachine.withgoogle.com)

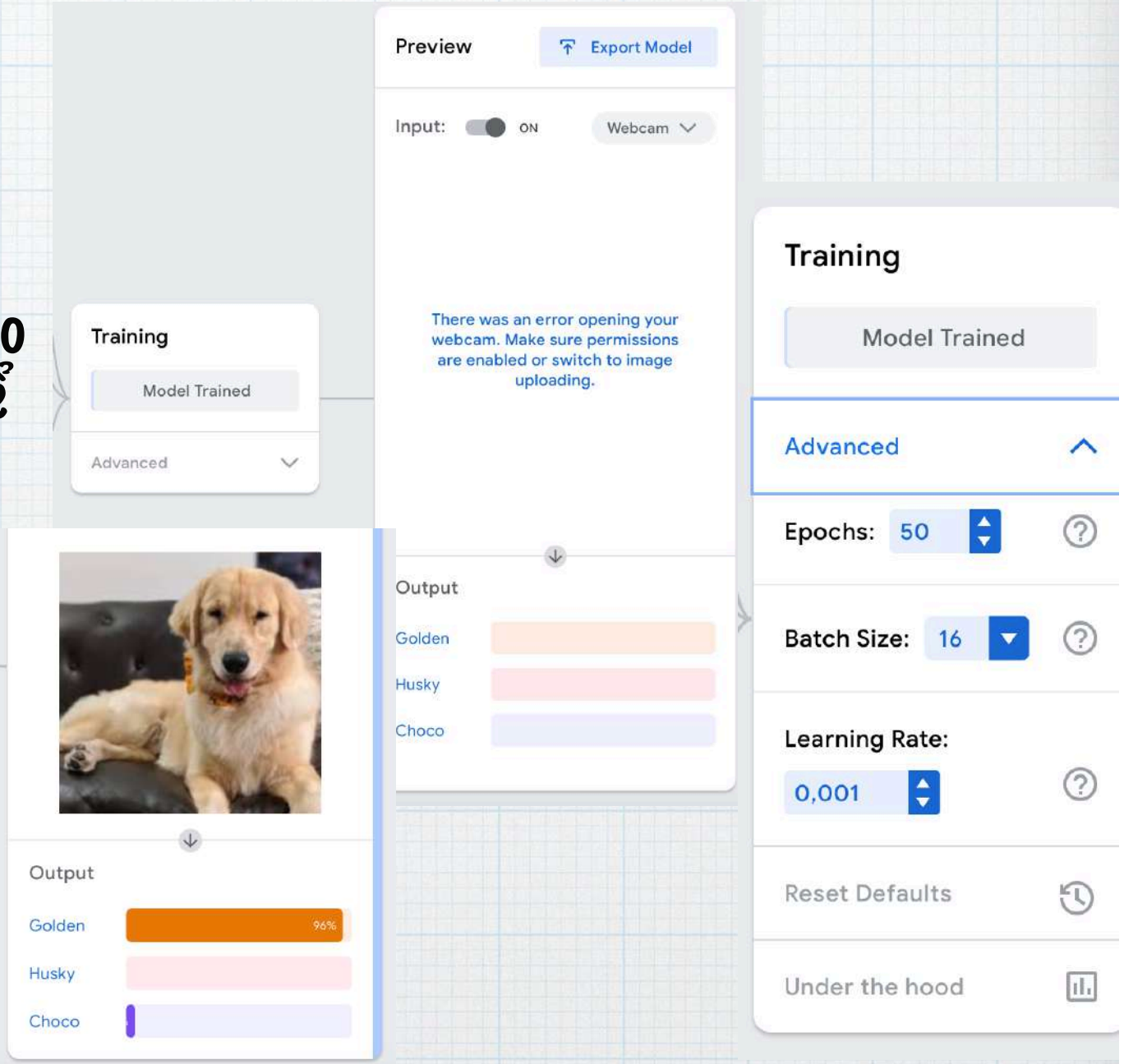


* Chúng ta có thể nhanh chóng tạo ra sản phẩm AI mà không cần lập trình bằng cách đưa vào nhiều bước đơn giản.
* Đưa các dữ liệu để huấn luyện




Husky
File
Choose images from your files, or drag & drop here
Import images from Google Drive
Images will be cropped to square
11 Image Samples
Choco
File
Choose images from your files, or drag & drop here
Import images from Google Drive
Images will be cropped to square
10 Image Samples

* Sau đó bước tiếp theo nhập các thông số để huấn luyện
* Sau đó kiểm thử kết quả.



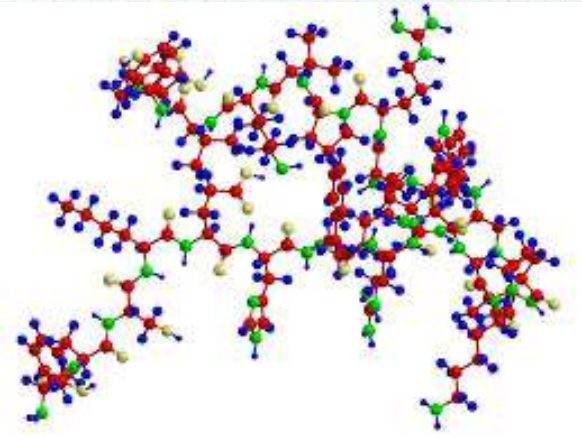
Nét chung

* Protein như một bài văn thậm chí là một đoạn văn
* Trong đó câu văn là một trình tự peptide
* Các từ trong câu văn là các amino acid (aa)
* Những từ này được tạo thành từ 4 ký tự là A, C, G và T
  - Lưu ý, các từ có chiều dài bằng nhau là 3 chữ (3 ký tự)
  - Vấn đề là làm sao hiểu được quy tắc ngữ pháp tạo nên câu và bài văn

* Trong sự sống, có 4 đại phân tử sinh học không thể thiếu:
  - Protein
  - Nucleic acid
  - Polysaccharide
  - Pipid

* Về mặt tổ chức đây là những hợp chất bao gồm các đơn phân tử cùng loại.
* Những đơn phân tử này liên kết với nhau theo liên kết cộng hoá trị.

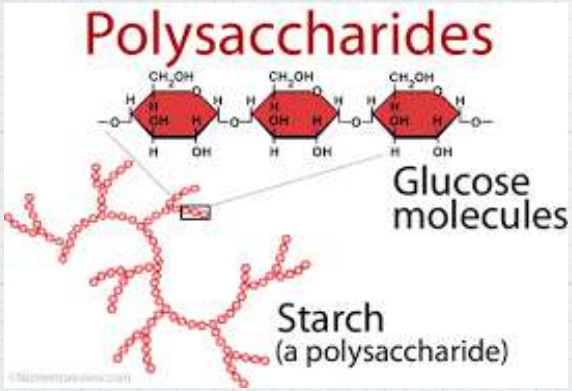
* Quan trọng hơn cả là:
  - Nucleic acid: lưu trữ thông tin di truyền
  - Protein: biểu hiện của vật chất sống

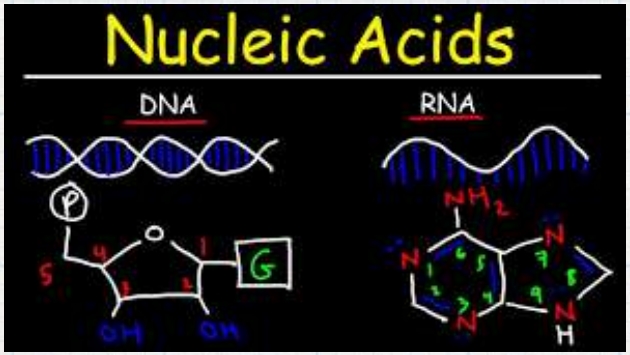


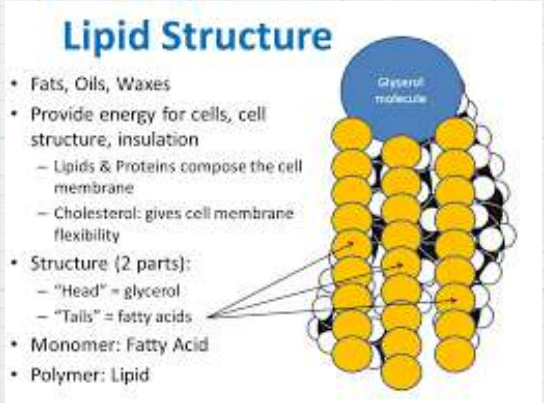
* Còn:
  - Polysaccharide: tham gia cấu tạo tế bào, là nguồn dự trữ năng lượng chính.
  - Lipid: thành phần của màng tế bào, được cấu tạo từ các acid béo; là nhân tố chính để hình thành các màng sinh học.

* Một đặc điểm quan trọng là cấu trúc và tính chất hoá lý của các Nucleic acid, Lipid, Polysaccharide tương đối đồng nhất

* Nhưng Protein lại đa dạng về cấu trúc và chức năng



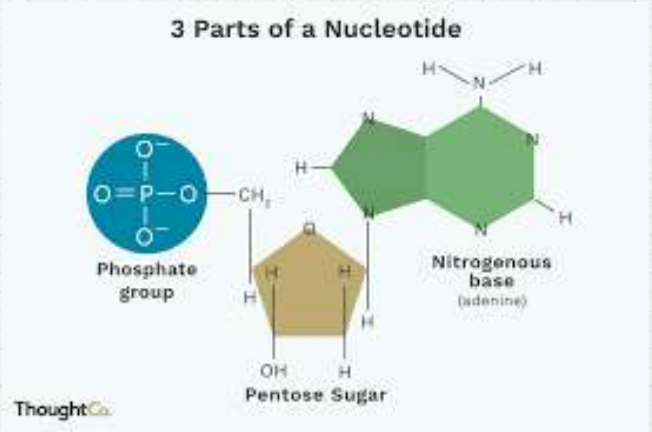




* Nucleic acid là vật chất mang thông tin di truyền của các cơ thể sống, được hình thành từ các phân tử nucleotide.

* Mỗi nucleotide gồm 3 thành phần:

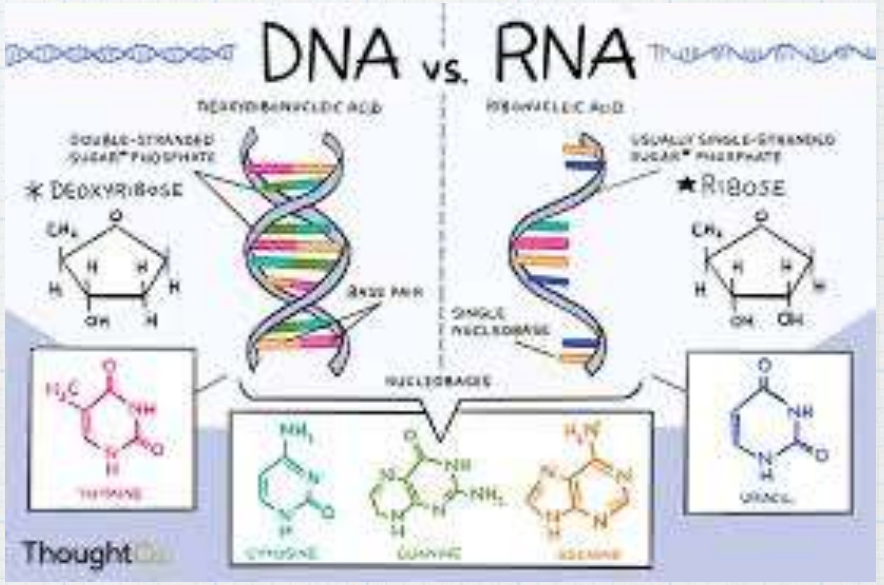
  - Phosphate
  - Đường
  - Và một trong 4 base hữu cơ là Adenine (A), Cytosine (C), Guanine (G) và Thymine (T)



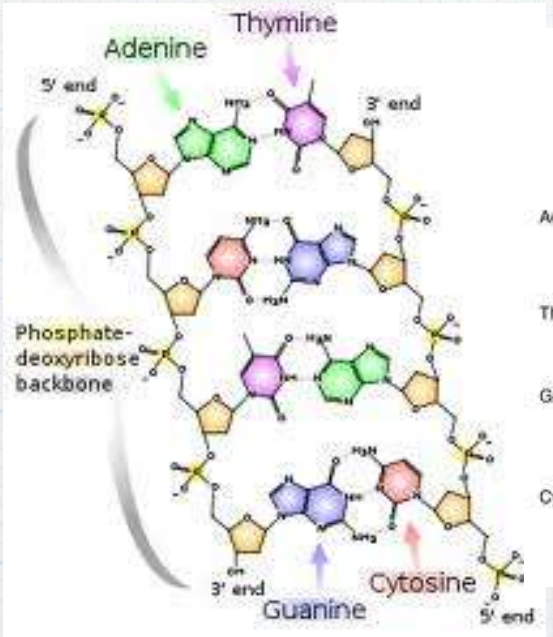
* Do các Nucleotide chỉ khác nhau ở thành phần base hữu cơ, nên thỉnh thoảng người ta thường dùng thuật ngữ Base thay cho Nucleotide.

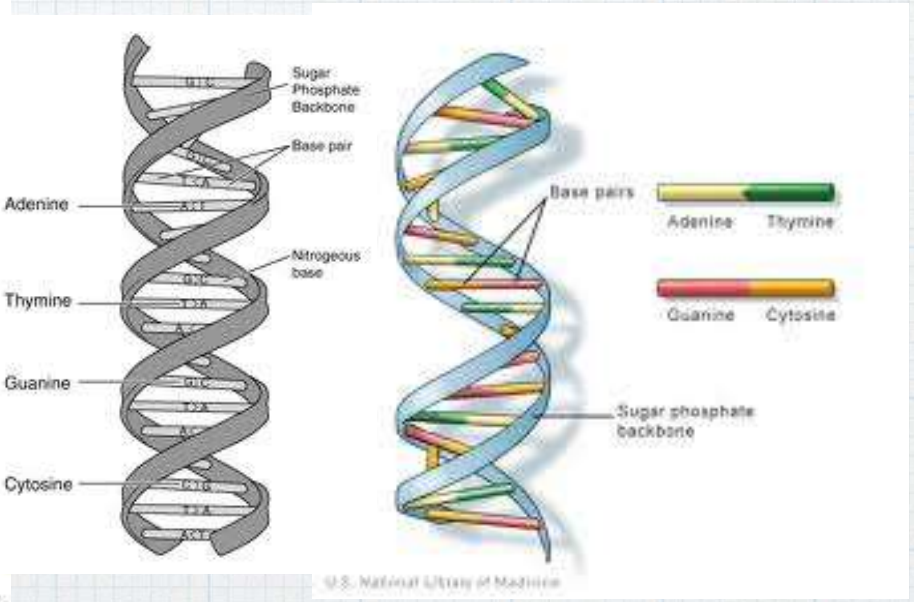
* Đại phân tử Nucleic acid gồm 2 loại:

  - DNA: Deoxyribonucleic Acid
  - RNA: Rebonucleic Acid



* Đại phân tử DNA là chuỗi xoắn kép gồm 2 mạch đơn, mỗi mạch đơn là một chuỗi nucleotide

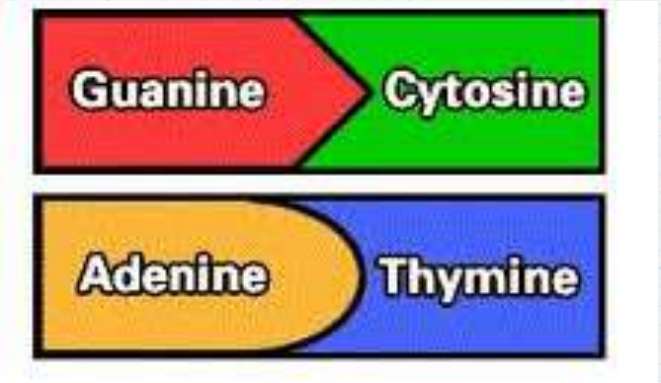




* Các nucleotide trong một mạch đơn của DNA liên kết với nhau bằng liên kết cộng hóa trị; là liên kết được hình thành giữa đường của nucleotide này với phosphate của nucleotide kế tiếp.

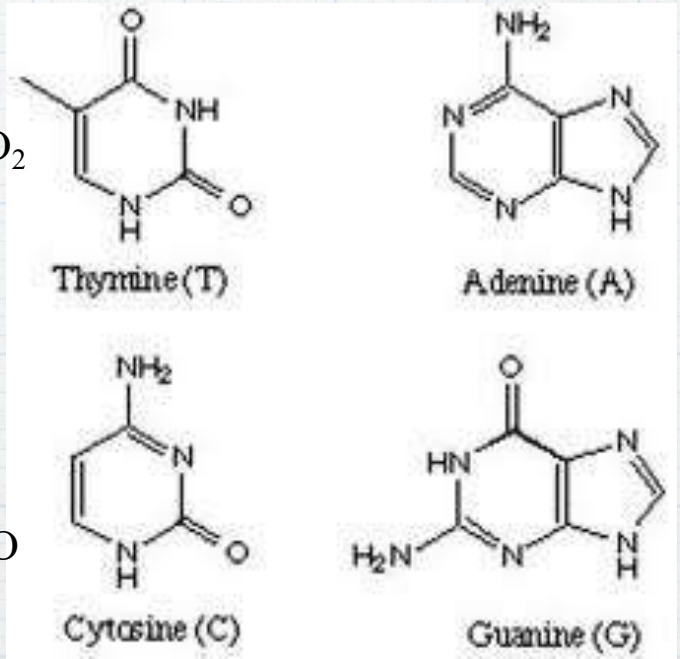
* Hai mạch đơn liên kết với nhau bằng liên kết hydro hình thành giữa các base; là tương tác tĩnh điện yếu giữa phần tử hydro mang điện tích dương với phần tử mang điện tích âm

* Trong hai mạch đơn liên kết với nhau thì:
  - G của mạch này liên kết với C của mạch kia
  - A của mạch này liên kết với T của mạch kia



* Do các Nucleotide chỉ khác nhau thành phần base hữu cơ, nên đại phân tử DNA như là một trình tự sinh học (Biology sequence) gồm các base là:

- A (Adenine),
- C (Cytosine),
- G (Guanine),
- T (Thymine).



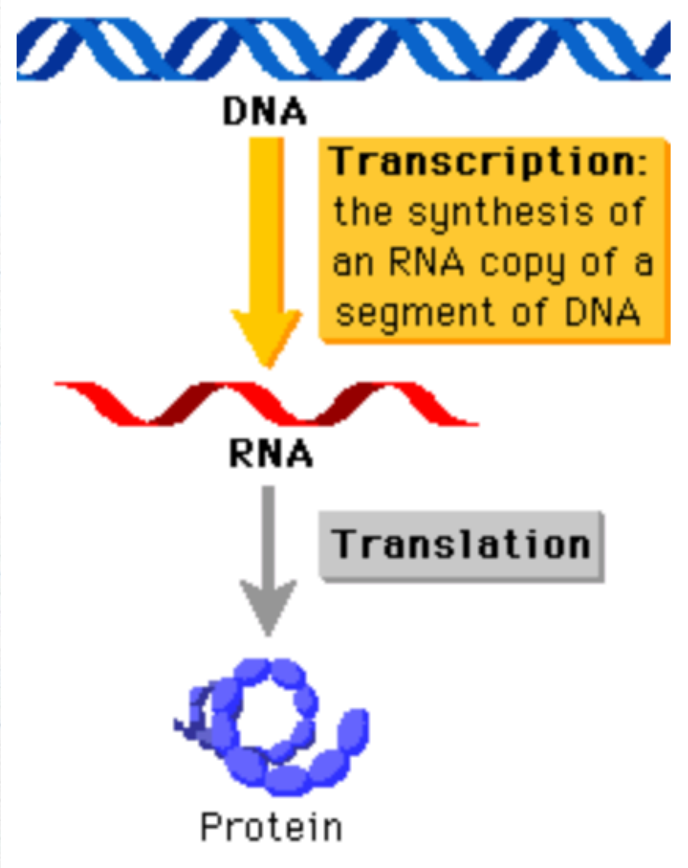
* Điều này rất thuận lợi khi biểu diễn các đại phân tử DNA trên máy tính bằng chuỗi ký tự chứa ký tự chữ A, C, G, T.

* Như vậy, với một chuỗi nucleotide được người nghiên cứu về tin học coi đó như là một chuỗi dài chứa 4 ký tự chữ như trên.

* Chẳng hạn một chuỗi có 10 nucleotide, thì số loại DNA khác nhau là $4^{10} = 2^{20} = 1.048.576$ loại.

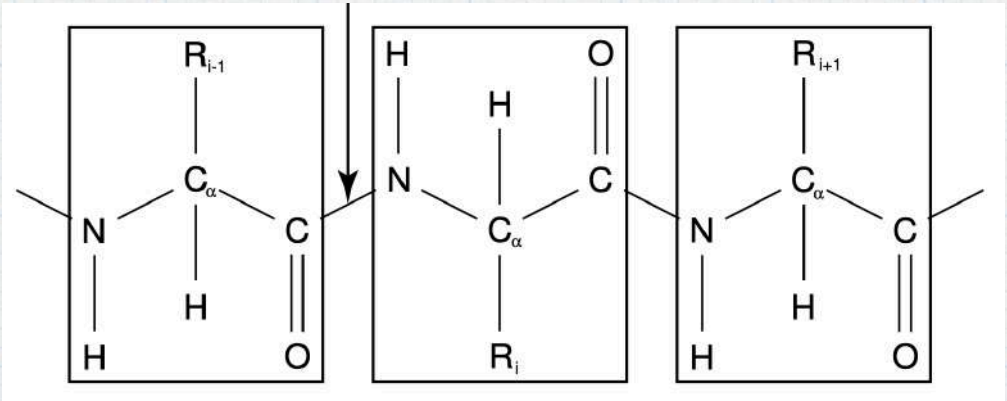
* **Luận thuyết trung tâm (Central Dogma) là một học thuyết được đề xuất bởi Francis Crick vào 1955, khi phát hiện ra mô hình DNA. Luận thuyết cho rằng:**

  - **DNA là vật chất mang thông tin di truyền của mọi sinh vật, từ đó quy định các tính trạng của cơ thể thông qua vai trò trung gian của RNA để tổng hợp nên protein.**



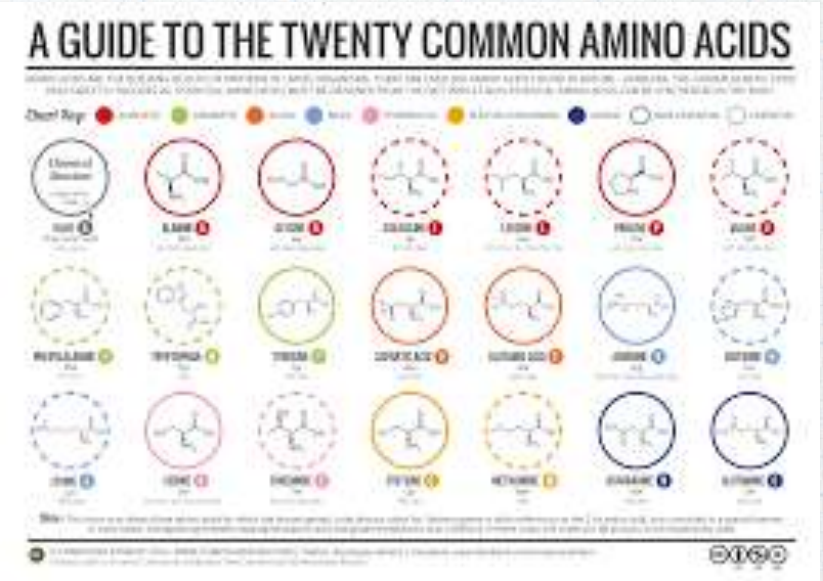
Về protein

* Protein có các đơn phân tử là các amino acid.
* Mỗi amino acid có cấu trúc cơ bản giống nhau (như hộp trong hình), chỉ khác nhau ở các nguyên tử tạo thành (là R_i)
* Nguyên tử carbon để gắn nhóm amino, nhóm carboxyl và R_i được gọi là alpha carbon (C_α).

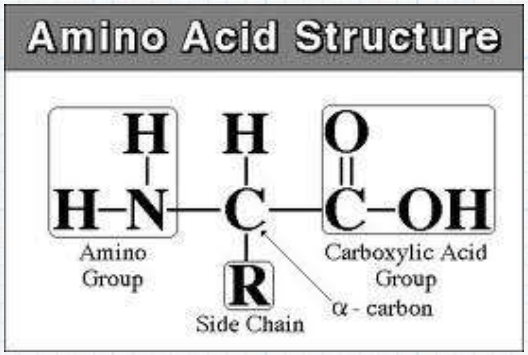


* Hiện nay biết được 20 amino acid (aa) khác nhau (mỗi ac được mã hoá theo bảng mã 1 ký tự hoặc 3 ký tự) hình thành protein trong tự nhiên.

* Liên kết được tạo thành giữa amino acid thứ i-1 và thứ i bằng cách liên kết nhóm carboxyl của amino acid thứ i-1 và nhóm amino của amino acid thứ i (loại bỏ đi một phân tử nước), được gọi là liên kết peptide



* Protein đóng vai trò quan trọng trong hầu hết các tiến trình sinh học (Biological Processes), là nền tảng của sự sống.
* Protein thực hiện một khối lượng lớn các công việc quan trọng mà một tế bào cần phải làm. Chẳng hạn,
  - Tạo nên cơ bắp, tóc và móng tay của bạn, làm cho các phản ứng hóa học xảy ra, tiêu hóa thức ăn
  - Duy trì tính toàn vẹn của cấu trúc tế bào, vận chuyển tế bào, làm xúc tác, truyền dẫn tín hiệu, tham gia vào hệ thống miễn dịch, v.v...

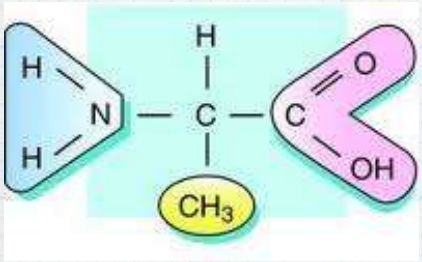


* Như vậy có thể nói: một phân tử protein được hình thành từ một mắt xích (chain) của các amino acid.

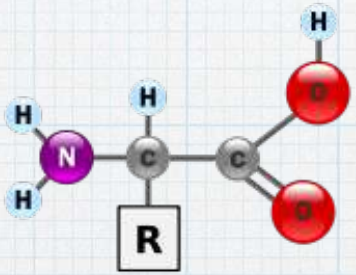
* Mỗi amino acid bao gồm một nguyên tử Carbon ở trung tâm (C_α), gắn với carbon này là nguyên tử hydro, nhóm amino (NH_2), nhóm carboxyl (COOH) và một mắt xích đặc trưng là amino acid này.

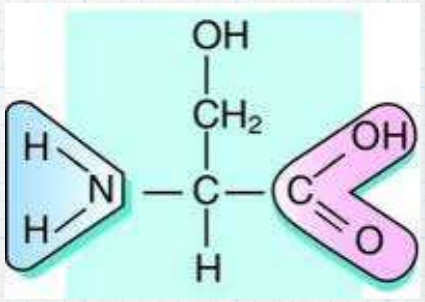
* C_α gắn với NH_2 gọi là liên kết N-C_α, C_α gắn với COOH gọi là liên kết C_α-C

* Chẳng hạn, với amino acid: Alanine, Serine



Alanine



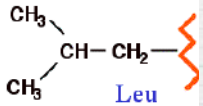
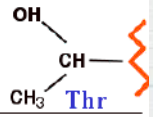
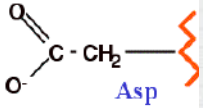
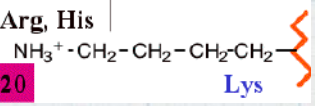


Serine

* Mỗi amino acid khác nhau các thành phần nguyên tử nên có tính chất hoá học khác nhau, từ đó có những đặc trưng khác nhau.

* Có thể có loại amino acid không phân cực (non-polar) hay kỵ nước (hydrophobic), phân cực (không tích điện), có tính acid (tích điện âm), có tính base (tích điện dương)

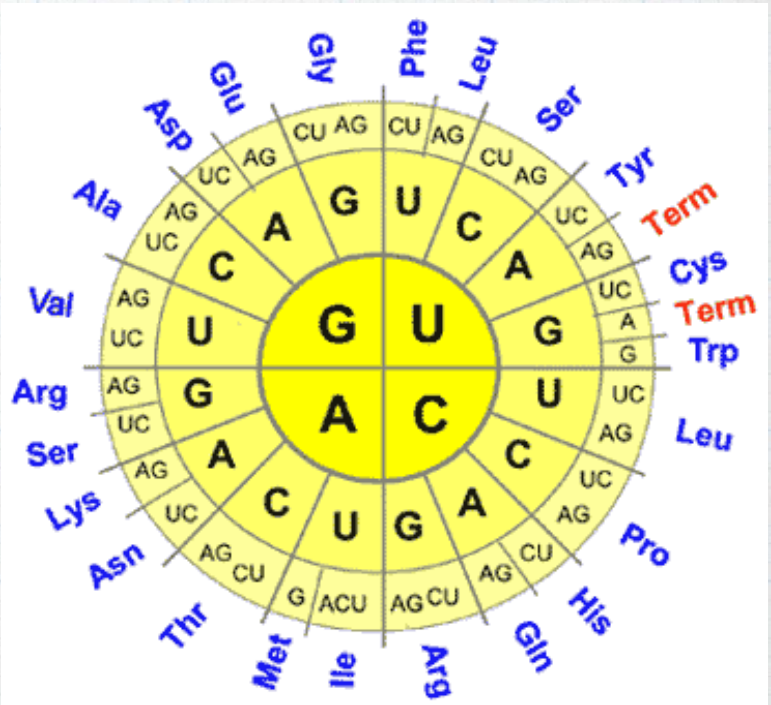
Amino acids groups

Group	Characteristics	Names	Example (-Rx)
non-polar	hydrophobic	Ala, Val, Leu, Ile, Pro, Phe Trp, Met	 Leu
polar	hydrophilic (non-charged)	Gly , Ser, Thr, Cys, Tyr, Asn Gln	 Thr
acidic	negatively charged	Asp, Glu	 Asp
basic	positively charged	Lys, Arg, His	 Lys
		Total = 20	

* Các amino acid ky nước bao gồm Isoleucine (I), Leucine (L), Methionine (M), Phenylalanine (F), Valine (V).
* Arginine (R) và Lysine (K) tích điện dương; Aspartic (D), Glutamic (D) tích điện âm.
* Amino acid phân cực bao gồm Asparagine (N), Glutamine (Q) Histidine (H), Serine (S), Threonine (T)

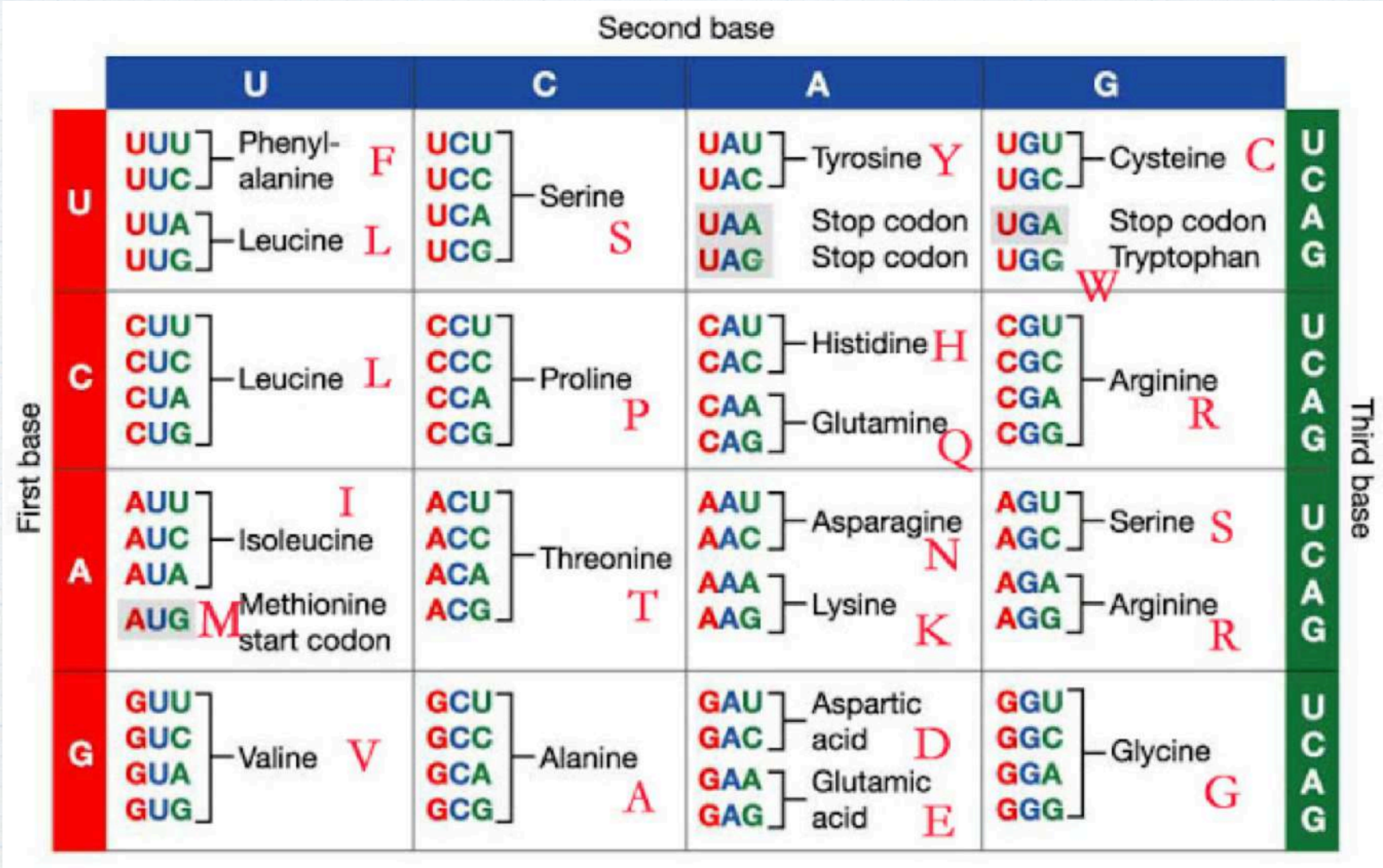
* Alanine (A) là một amino acid nhỏ không phân cực.
* Glycine (G) là amino acid nhỏ nhất, chỉ bằng hydro.
* Cysteine (C) là một phần trong liên kết disulfide.
* Proline (P), Tryptophan (W), Tyrosine (Y) là amino acid có dạng hình vòng lớn.

* Trình tự các base trên DNA quyết định trình tự amino acid trên protein tương ứng.
* Mỗi amino acid có 3 base, nên với 4 base A, C, G, T sẽ có số lượng amino acid lý thuyết là $4^3 = 64$ (gọi là 64 codon)
* Tuy nhiên, hiện nay chỉ phát hiện được 20 amino acid.



* Margaret Oakley Dayhoff (American Physical Chemist, pioneer in Bioinformatics) đề xuất dùng one-letter code để mã hóa 20 amino acid này

Amino acid	Three letter code	One letter code
alanine	ala	A
arginine	arg	R
asparagine	asn	N
aspartic acid	asp	D
asparagine or aspartic acid	asx	B
cysteine	cys	C
glutamic acid	glu	E
glutamine	gln	Q
glutamine or glutamic acid	glx	Z
glycine	gly	G
histidine	his	H
isoleucine	ile	I
leucine	leu	L
lysine	lys	K
methionine	met	M
phenylalanine	phe	F
proline	pro	P
serine	ser	S
threonine	thr	T
tryptophan	trp	W
tyrosine	tyr	Y
valine	val	V


Second base
U
C
A
G
First base
Third base
U
UUU Phenyl-alanine F
UUC
UUA Leucine L
UUG
UCU Serine S
UCC
UCA
UCG
UAU Tyrosine Y
UAC
UAA Stop codon
UAG Stop codon
UGU Cysteine C
UGC
UGA Stop codon
UGG Tryptophan W
C
CUU Leucine L
CUC
CUA
CUG
CCU Proline P
CCC
CCA
CCG
CAU Histidine H
CAC
CAA Glutamine Q
CAG
CGU Arginine R
CGC
CGA
CGG
A
AUU Isoleucine I
AUC
AUA
AUG Methionine start codon M
ACU Threonine T
ACC
ACA
ACG
AAU Asparagine N
AAC
AAA Lysine K
AAG
AGU Serine S
AGC
AGA Arginine R
AGG
G
GUU Valine V
GUC
GUA
GUG
GCU Alanine A
GCC
GCA
GCG
GAU Aspartic acid D
GAC
GAA Glutamic acid E
GAG
GGU Glycine G
GGC
GGA
GGG

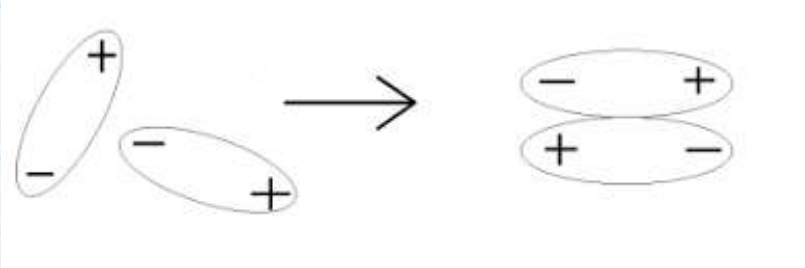
- Chỉ có 61 codon chứa thông tin (mã hóa amino acid cụ thể)
- 3 codon: UAA, UAG, UGA là dấu hiệu kết thúc
- Codon AUG vừa là amino acid có tên Methionine (Met) vừa là dấu hiệu bắt đầu

* Lưu ý: Trong 20 amino acid này có 9 amino acid gọi là thiết yếu; bởi nó không thể được tạo ra trực tiếp từ cơ thể con người, mà được cung cấp thông qua nguồn thực phẩm dinh dưỡng từ bên ngoài, Đó là:

  - histidine, isoliucine, leucine, lysine, methionine, phenylalanine, threonine tryptophan, valine

* Với các amino acid trong mắt xích có tính chất hoá lý khác nhau, nên có nhiều lực khác nhau tạo nên nếp gấp protein

* Chẳng hạn lực liên kết hydro, tương tác tĩnh điện, lực Van der Waals.



Cấu trúc protein

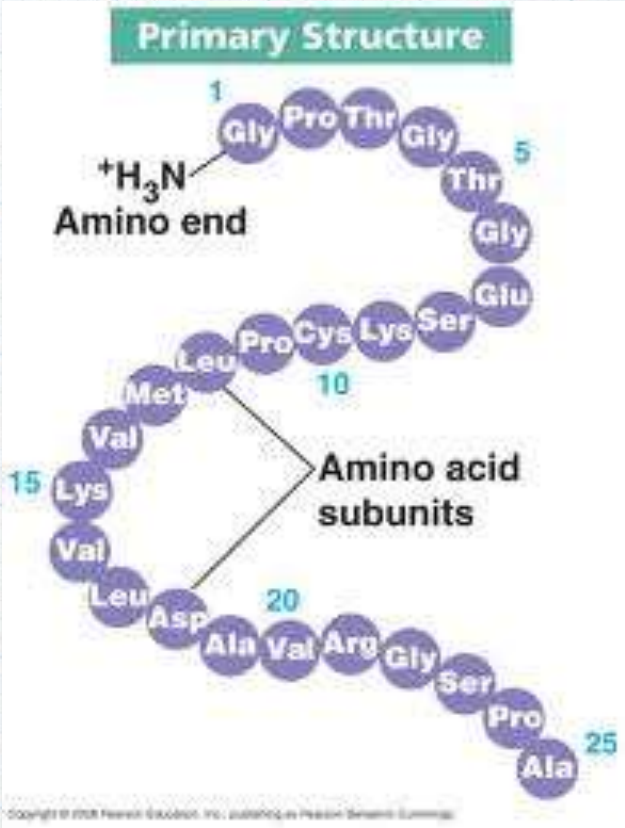
* Cấu trúc protein rất cần thiết cho sự hiểu biết về chức năng protein.
  - Các phân tử protein tuyến tính gấp lại thành các cấu trúc ba chiều (Three-Dimensional Structures - 3D) và các tính chất chức năng của chúng phụ thuộc rất nhiều vào cấu trúc 3D này.
  - Để nhận biết chức năng của protein ở cấp độ phân tử, cần phải xác định cấu trúc 3D của chúng.

* Ở góc độ cấu trúc, thường coi trình tự protein bao gồm các đơn vị peptide (peptide unit)
* Đơn vị này bao gồm các nguyên tử amino acid chính ở giữa các nguyên tử C_α liên tiếp.
* Trong cấu trúc protein, các nguyên tử trong một đơn vị peptide nằm trong cùng một mặt phẳng có chiều dài liên kết và góc quay tương tự nhau.

* Mỗi đơn vị peptide về cơ bản chỉ có 2 bậc tự do quay xung quanh liên kết N-C_α và C_α-C của nó.
  - Góc ϕ chỉ góc quay xung quanh liên kết N-C_α
  - Góc ψ chỉ góc quay xung quanh liên kết C_α-C
* Như vậy hình dáng xương sống (backbone conformation) của protein được chỉ định bởi một dãy các góc ϕ và ψ.

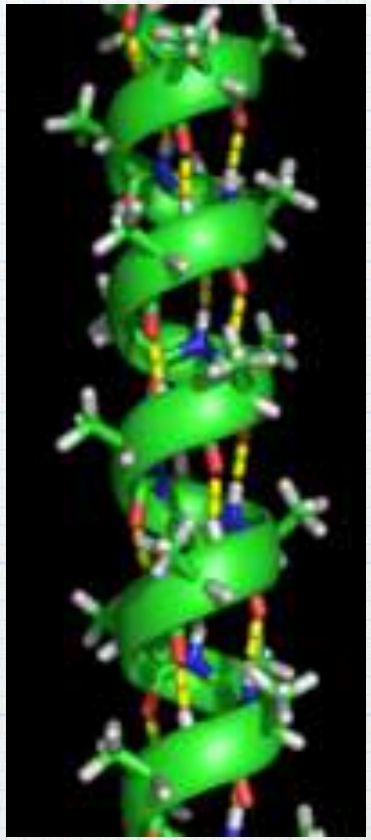
* Do amino acid là các đơn phân tử cấu thành nên protein, nên chuỗi peptid hay polypeptide là trình tự protein.

* Trong trường hợp chỉ quan tâm đến các liên kết peptide trong chuỗi này, ta có cấu trúc bậc I của protein (protein primary structure) hay còn gọi là cấu trúc sơ cấp.

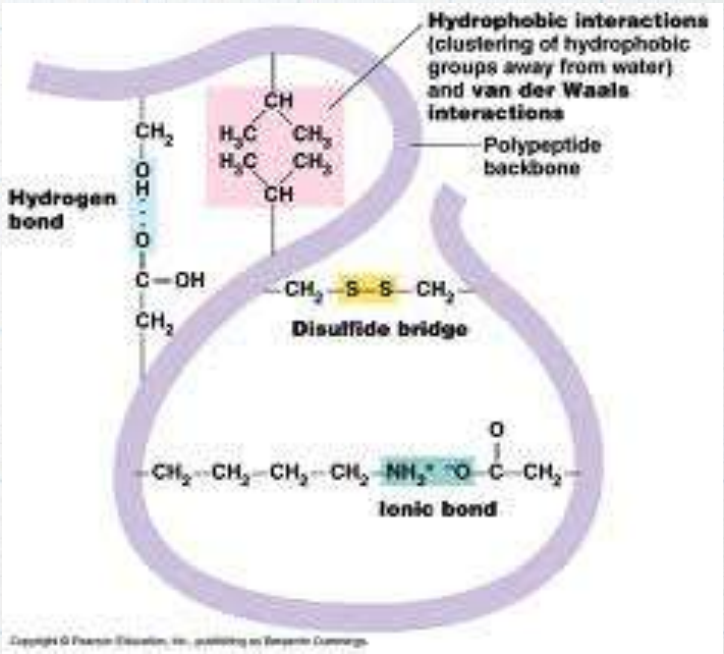


* Khi các amino acid gần nhau liên kết với nhau thông qua liên kết hydro giữa nhóm amin (NH) của amino acid này với nguyên tử Oxy của amino acid khác sẽ tạo nên vòng xoắn của chuỗi polypeptide.

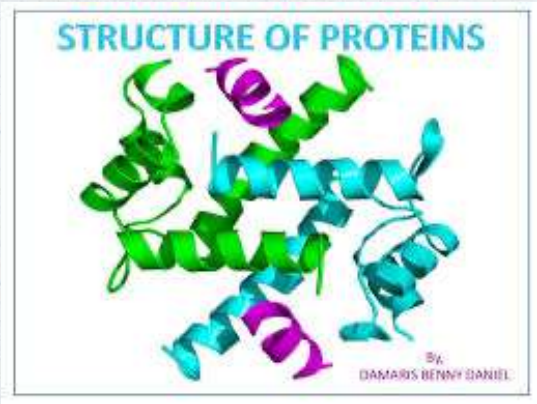
* Khi đó có cấu trúc bậc II của protein (protein secondary structure) hay cấu trúc thứ cấp.

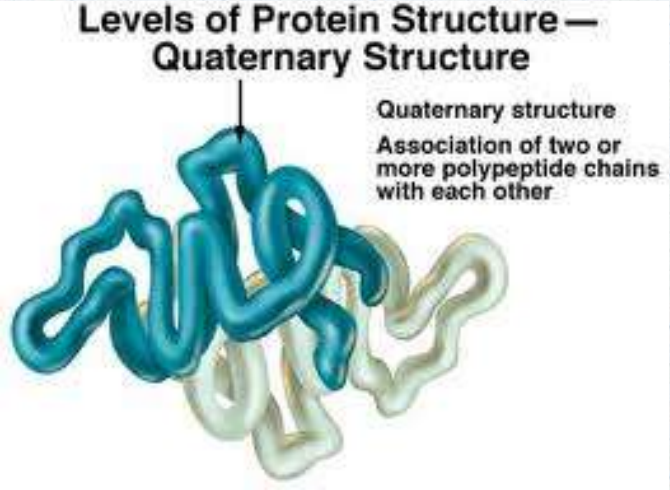


* Ngoài các liên kết hydro để tạo ra cấu trúc bậc II, các nhóm amino acid trên chuỗi polypeptide còn liên kết lại cùng nhau.
  - Chẳng hạn, các Cystein sẽ liên kết với nhau, hoặc các Proline liên kết với nhau để hình thành nên các nhóm riêng.
* Khi đó tạo nên cấu trúc bậc III (Protein tertiary structure)



* Khi có nhiều hơn một chuỗi polypeptite với các cấu trúc bậc III được liên kết với nhau, sẽ tạo nên cấu trúc protein bậc IV (Quaternary structure)

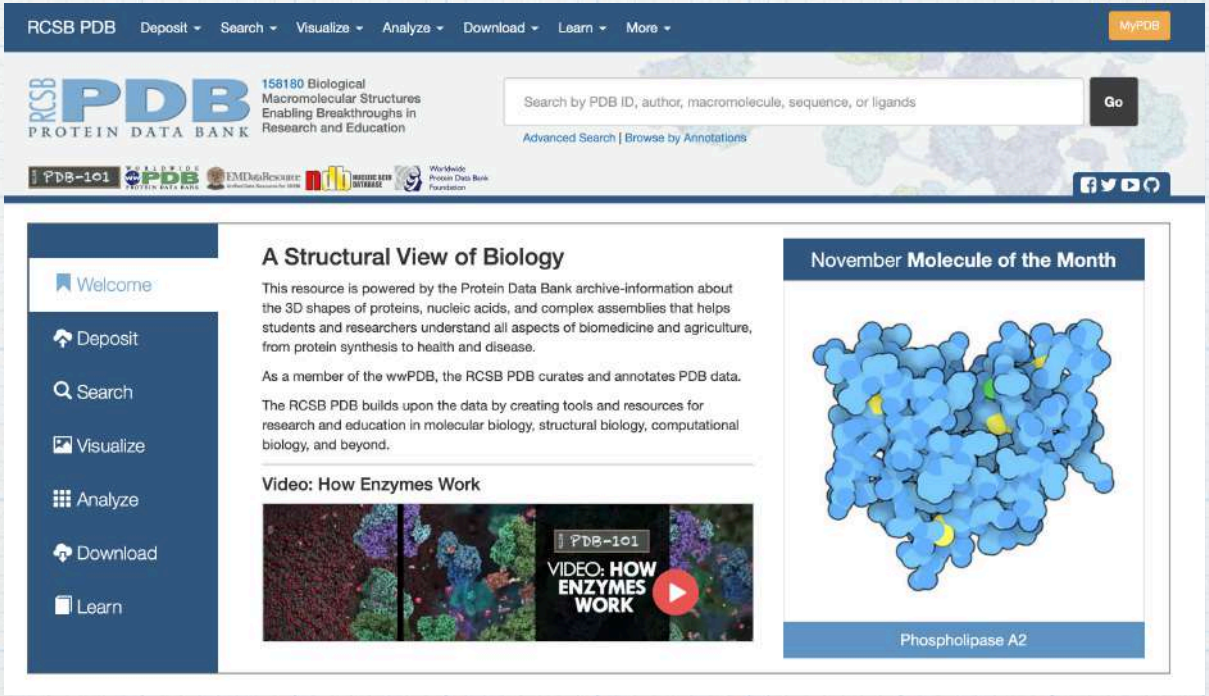




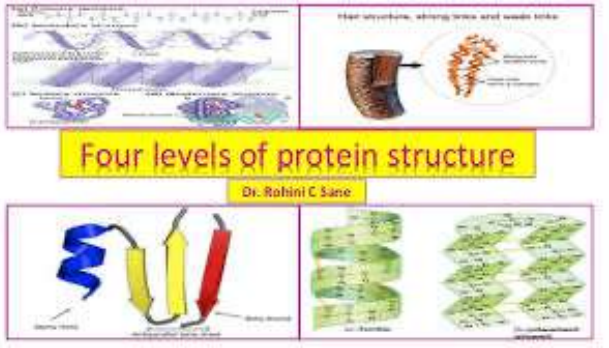
Dự đoán cấu trúc protein

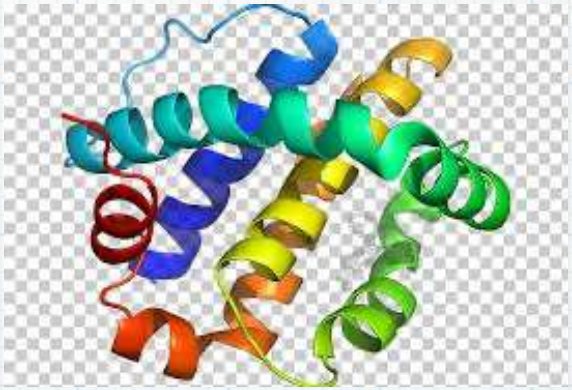
* Việc dự đoán chính xác và đáng tin cậy về cấu trúc của một trình tự protein là một trong những nhiệm vụ khó khăn nhất (most challenging tasks) trong sinh học tính toán (Computational Biology)

http://www.rcsb.org



- Cấu trúc protein được xác định bằng thực nghiệm như sử dụng phương pháp nhận dạng tinh thể bằng tia X, quang phổ thông qua cộng hưởng từ hạt nhân (NMR).
- Tinh thể học tia X bị hạn chế bởi sự khó khăn trong việc tạo ra một số protein để tạo thành tinh thể.
- NMR chỉ có thể được áp dụng cho các phân tử protein tương đối nhỏ





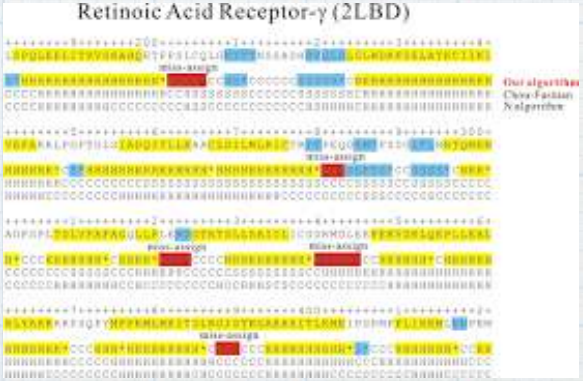
* Qua những thành tựu giải trình tự toàn bộ bộ gen đã dẫn đến một số lượng lớn các trình tự DNA đã biết được.

* Nhưng cấu trúc protein tương ứng của các trình tự được tìm kiếm với tốc độ chậm hơn rất đáng kể.

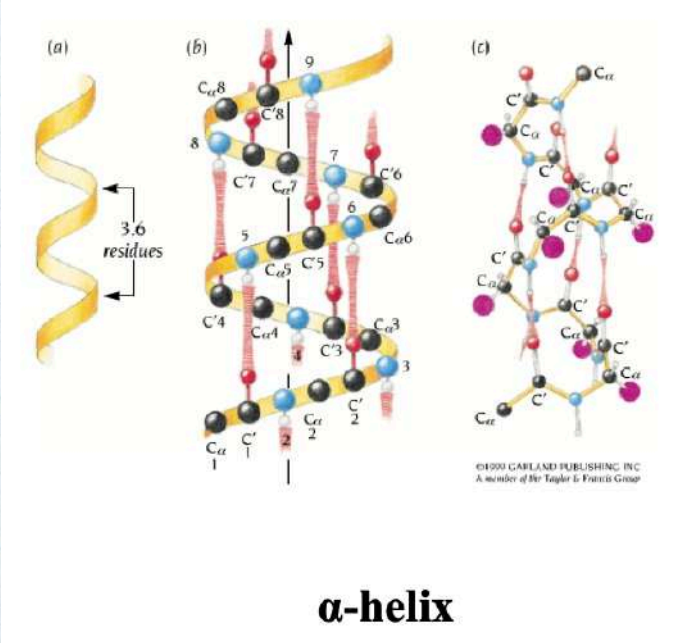
Cấu trúc bậc hai

* Việc dự đoán cấu trúc thứ cấp hay cấu trúc bậc II (secondary structure prediction) cung cấp bước đầu tiên đầy ý nghĩa (significant first step) để hướng đến dự đoán cấu trúc bậc ba (tertiary structure prediction)

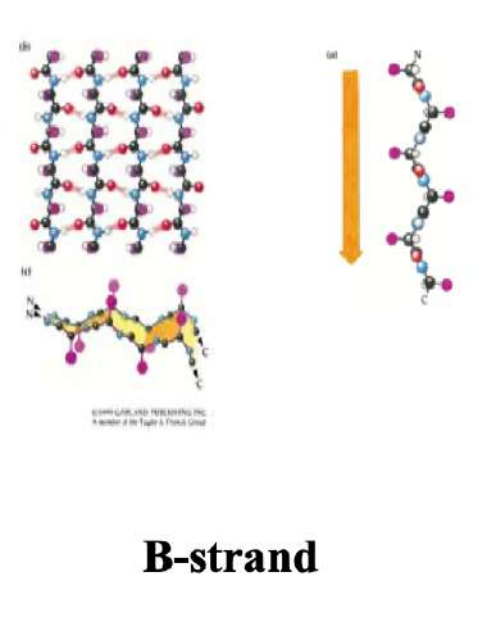
* Qua đó cung cấp thông tin về hoạt động, mối quan hệ và chức năng của protein.



* Có hai trạng thái cấu trúc đều đặn là α-helix (H) và β-strand hay β-sheet (E), và một trạng thái không đều là coil region (C).



α-helix

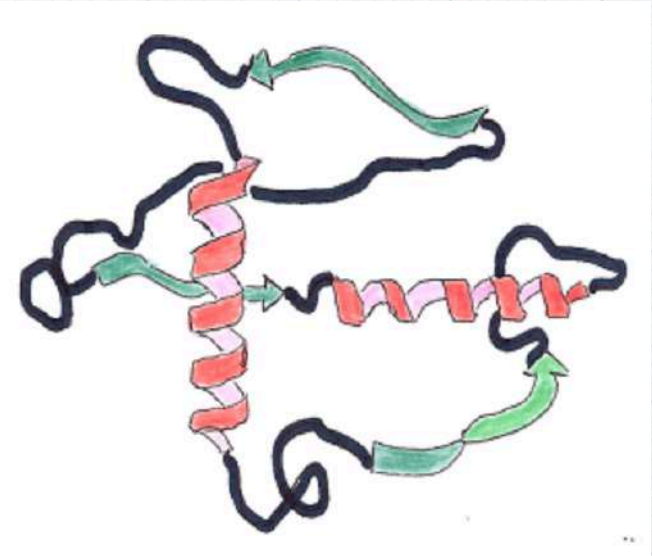


B-strand

* Cấu trúc thứ cấp là cấu trúc sắp xếp cục bộ thông qua liên kết hydrogen trong trình tự peptide.

* Cấu trúc này thông thường bao gồm các α-helices và β-sheets. Một α-helix chính tắc có khoảng 3,6 residues trên một vòng xoắn và được xây dựng từ một đoạn amino acid liên tục thông qua sự hình thành liên kết hydrogen giữa amino acids ở vị trí thứ i và vị trí thứ i + 4.

* α-helix có dạng như màu đỏ hình bên.
* Residue của amino acid là những gì còn lại sau khi lấy đi tất cả những phần giống nhau của amino acid.



* Các residue tham gia trong một α-helix có góc ϕ khoảng -60^{o} và góc ψ khoảng -50^{o}
* α-helix khác nhau đáng kể về chiều dài, có thể từ 4 hoặc 5 đến vài trăm amino acid được tìm thấy trong một trình tự protein.
* β-strand là một cấu trúc được mở rộng hơn.

* β-strand là một cấu trúc mở rộng hơn với 2,0 residue mỗi vòng xoắn. Góc ϕ và ψ tương ứng là -140^0 và 130^0.
* Một β-strand tương tác thông qua liên kết hydrogen với β-strands ở khoảng cách xa trong trình tự để tạo nên β-sheet (tấm β)
* Trong các sheet song song với nhau, các strand chạy theo cùng một hướng; trong khi ở các tấm phản song song, chúng chạy theo các hướng xen kẽ.
* Một strand thường có trung bình 5 - 10 residue, và có sáu strand trên mỗi sheet.

Dự đoán cấu trúc bậc II

* Cấu trúc protein tự nhiên tồn tại trong các trạng thái cấu trúc đều đặn là α-helix (H), β-strand hay β-sheet (E), và không đều là coil region (C).
* Cho một trình tự protein (protein sequence) với các amino acid $r_1r_2...r_n$.
* Bài toán dự đoán cấu trúc bậc II được mô tả như sau:
  - Cho một trình tự protein X với các amino acid $r_1r_2...r_n$, hãy dự đoán xem mỗi amino acid r_i có phải là α–helix (H), β–strand (E), hay là (C)

* Độ chính xác của dự đoán cấu trúc bậc II thường được đánh giá theo Q3 (gọi là 3-state accuracy); là tỷ lệ phần trăm residue mà dự đoán H, E, C là chính xác. Q3 được tính như sau:

$$Q3 = \frac{N_H + N_E + N_C}{N} 100$$

* Trong đó, N_H, N_E, N_C là số cấu trúc helix, strand và coil được dự đoán chính xác, N là tổng số residues (amino acid).

* Vì residues được biết trong một cấu trúc protein thường xấp xỉ khoảng 30% trong helices, 20% trong strands và 50% trong cả hai, nên một thuật giải tầm thường luôn dự đoán C theo Q3 là 50%.

* Độ đo Q3 không truyền tải nhiều loại thông tin hữu ích. Tuy nhiên, hiện nay vẫn là một biện pháp ngắn gọn, hữu ích thường được sử dụng để so sánh hiệu quả của các phương pháp dự đoán khác nhau.

* Hiện nay, dự đoán được với độ chính xác khoảng 86%, cộng đồng đang mong muốn đạt được 88 - 90% để kết thúc một chặng đường dài trong việc nghiên cứu dự đoán cấu trúc bậc hai của protein.

* Có thể dùng độ chính xác trung bình trên toàn bộ tập dữ liệu thử nghiệm, khi đó sử dụng Q3 trung bình để đánh giá hiệu suất của mô hình. Q3 trung bình được xác định là:

$$Average\ Q3 = \frac{\sum_{i=1}^{n} Q3(X_i)}{n}$$

* Trong đó n là số trình tự protein có kết quả dự đoán tốt trên tập dữ liệu thử nghiệm, X_i là trình tự protein thứ i, và $Q3(X_i)$ là độ chính xác Q3 của X_i.

* Hiện nay vẫn dử dụng phương pháp DSSP (Dictionary of Secondary Structure of Proteins) để thể hiện 8 trạng thái của cấu trúc bậc II dựa trên liên kết hydrogen là H, E, B, T, S, L, G, và I. Trong đó,
  - G là liên kết hydrogen giữa amino acid thứ i và i+3
  - I hay π-helix là liên kết hydrogen giữa amino acid thứ i và i+5
  - B là cầu nối , là residue của β-strand
  - S là chỗ thắt nút (bend)
  - T là cuộn liên lết hydrogen

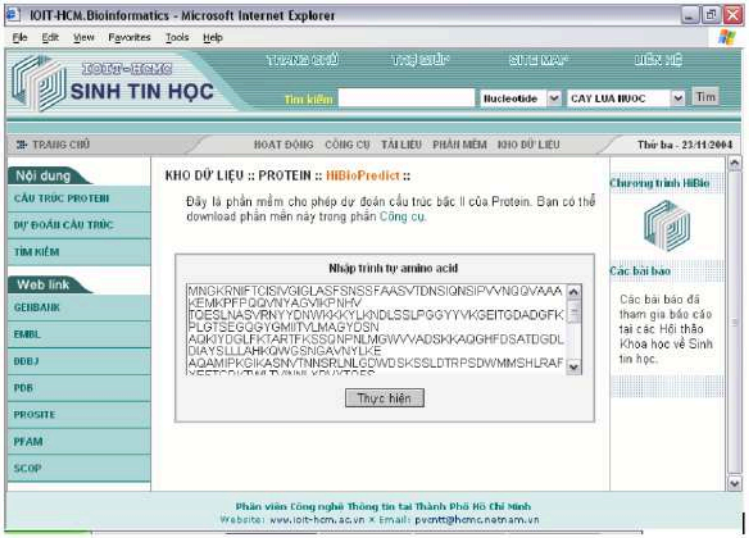
* Tám trạng thái này được gom vào 3 loại:
  - helix được coi là G, H và I (gộp thành H)
  - sheet hay strand là B và E (gộp thành E)
  - tất cả các trạng thái khác được coi là coil (gộp thành C)
* Nếu chính xác phải dự đoán theo 8 nhóm (độ đo Q8 - đây là vấn đề thời sự của dự đoán cấu trúc bậc II)

- Dự đoán cấu trúc protein bậc II bắt đầu vào năm 1951 khi Pauling và Corey dự đoán helix và sheet cho trình tự polypeptide, ngay cả trước khi cấu trúc protein đầu tiên được xác định*
- Phương pháp thống kê và học máy đã được dùng để dự đoán cấu trúc thức cấp của protein

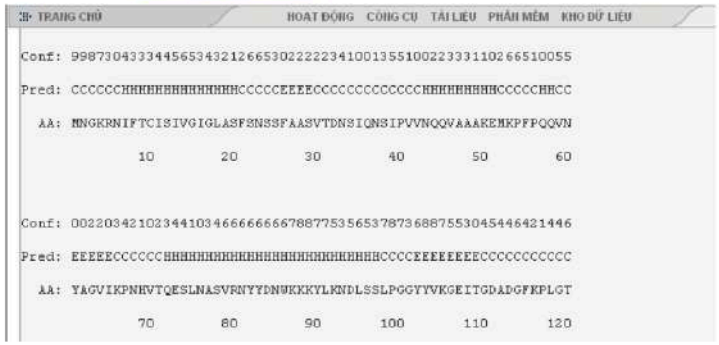
*Sixty-five years of the long march in protein secondary prediction: the final stretch, Brief Bioinform. 2018 May 1;19(3):482-494. doi: 10.1093/bib/bbw129]

* Cách tiếp cận đầu tiên của việc dự đoán cấu trúc bậc II protein là tổ hợp các quy tắc thống kê và heuristic.
* Phương pháp học máy tỏ ra hữu hiệu do việc khai thác thông tin tiến hóa, cũng như thông tin thống kê về các amino acid.
* Chẳng hạn Neural Network, Hidden Markov Model, Support Vector Machines, K-nearest Neighbors đã có những thành công với Q3 đạt đến 80%

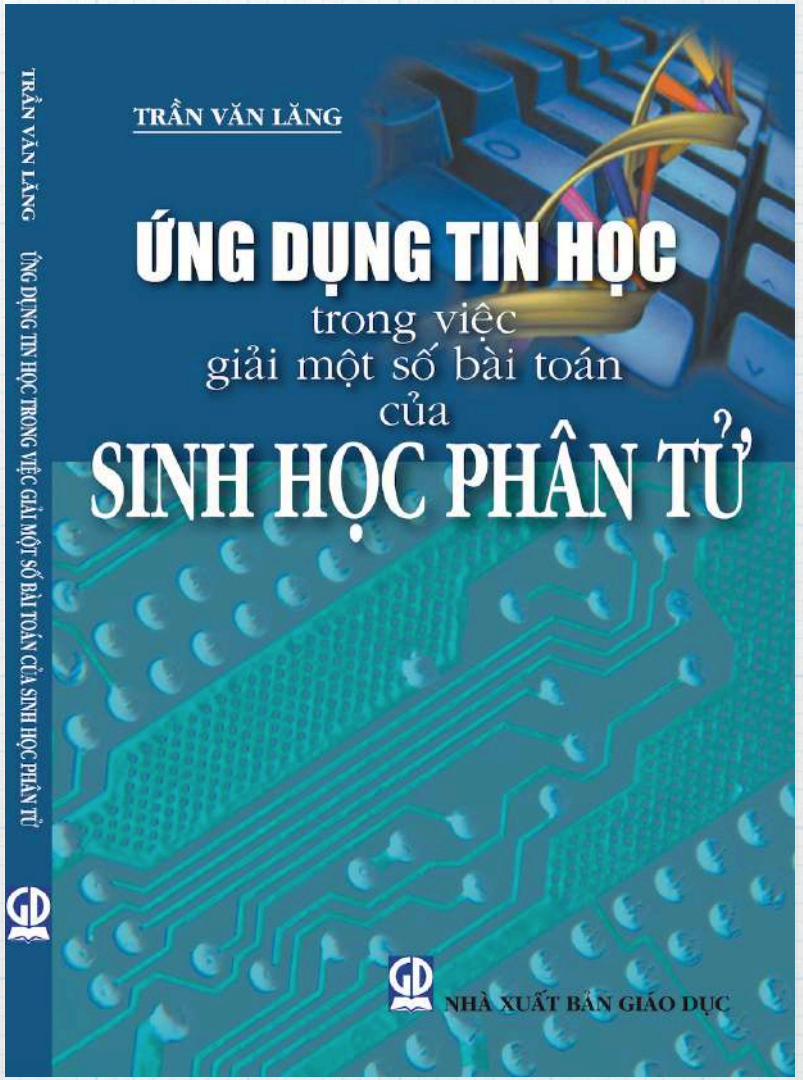
* Nhóm chúng tôi khi thực hiện vấn đề liên quan vào 2004, sau đó biên soạn quyển sách cũng đã có kết quả dự đoán bằng Neural Network là 78%.



Hình 120. *Dự đoán cấu trúc bậc 2 protein - Nhập dữ liệu đầu vào*



Hình 121. *Dự đoán cấu trúc bậc 2 - Kết quả dự đoán*



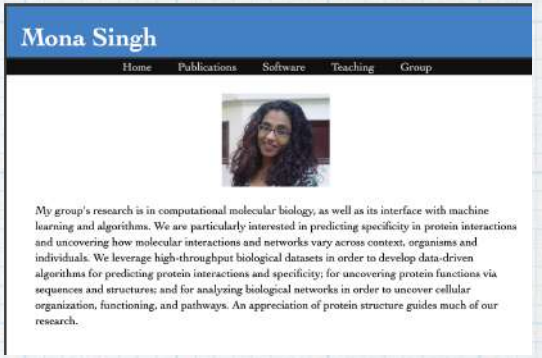
* Gần đây, một số công trình đã sử dụng Deep Learning (DL) với độ chính xác Q3 lên đến 84%. DeepCNF là một minh chứng; bằng cách phần mở rộng của DL với Conditional Neural Field - CNF, tích hợp các trường ngẫu nhiên có điều kiện vào Shallow Neural Networks.

* SPIDER3 đã cải thiện dự đoán bằng cách capture các tương tác không cục bộ với LSTM (Long Short-Term Memory) cũng đã có những kết quả vượt lên DeepCNF.

* Partition and semi-random subspace method (PSRSM) cũng là cách tiếp cận nâng cao độ chính xác lên 85,89% bằng cách:

  - Phân vùng tập dữ liệu huấn luyện thành nhiều tập con căn cứ trên chiều dài trình tự proteins.
  - Sau đó sinh ra các không gian con bằng cách bán ngẫu nhiên, phân lớp huấn luyện trên các không gian con này
  - Cuối cùng chọn bằng cách tổ hợp dựa trên quy tắc bình bầu đa số.

Tài liệu tham khảo



* https://www.cs.princeton.edu/~mona/
* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6026213/
* https://github.com/LucaAngioloni/ProteinSecondaryStructure-CNN
* https://www.ncbi.nlm.nih.gov/pubmed/28040746

Trân trọng,

Trần Văn Lăng
langtv@vast.vn