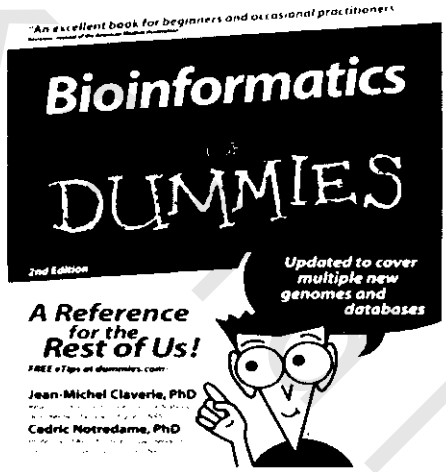
الفصل الأول

أساسيات هامة للبدء فى تعلم المعلوماتية الحيوية Bioinformatics

يعتبر علم المعلوماتية الحيوية دمج بين علوم البيولوجيا وعلوم الحاسب الألى . ويختص هذا العلم بدراسة كيفية استخدام إمكانيات الحاسب الآلي في تحليل وتفسير النتائج البيولوجية المختلفة. ومن المراجع السهلة التى ينصح بها للمبتدئين فى هذا العلم كتاب (Bioinformatics for Dummies) هو أحد السلاسل الشهيرة التى يهتم بتبسيط العلوم والمفاهيم المختلفة للمبتدئين وقد تم الأستعانة ببعض الصور التوضيحة منه لتوضيح بعض المفاهيم الخاصة بهذه العلم



شكل٤ : صورة توضح شكل غلاف الكتاب الخاص بتعليم علم المعلوماتية الحيوية للمبتدئين الذى تم الأستعانه ببعض الصور التوضيحية منه لتوضيح بعض المفاهيم

والراغب فى دراسة وفهم هذا العلم يجب أن تتوفر لدية مجموعة من الأدوات والمعلومات الخاصة بهذا العلم مثل:-

١- امتلاك جهاز كمبيوتر به مايكروسوفت ويندوز.

٢- أن يكون لديك إنترنت متصل بحاسوبك ويفضل أن يكون سريع.

٣- أن تكون لديك بعض المعلومات الأساسية في البيولوجيا الجزيئية.

٤- أن تجيد استخدام شبكة الإنترنت ولكن ليس بالضرورة أن تجيد استخدام كل البرامج العادية للحاسب.

٥- معرفة ما هي البرامج والوسائل التي تحتاجها البحث الخاص بك الخاصة بالمعلوماتية الحيوية Bioinformatics.

ويجب ملاحظة أن معظم شركات البيوتكنولوجى التجارية الخاصة تعتبر أن الإنترنت وسيلة غير أمنة لنقل البيانات لذلك لها برامجها الخاصة وشبكاتها الخاصة لنقل بياناتها وتحليلها حيث تعتبرهذه البيانات من من الأسرار الهامة لدى هذه الشركة لأغراض المنافسة التجارية . بينما عندما يستخدم الباحث العادى البرامج التحليلية الخاصة بالمعلوماتية الحيوية الوجوده مجانا على الأنترنت لا يعتبر هذا سراً من هذه الأسرارحيث أنه سيقوم فى النهاية بنشر نتائجة فى المجلات العلمية. وعلى الرغم من أن هناك الكثير من البرامج التحليلية المجانية الخاصة بالمعلوماتية الحيوية متاحة على شبكة الإنترنت إلا أن هناك كثير من المعامل الخاصة والمؤسسات العلمية الكبيرة تفضل تطوير برامج تحليلية خاصة بها تتناسب مع طبيعة الأبحاث بها. لذلك تحرص هذه المعامل البيولوجية على وجود مبرمج للحاسب الألى واحد على الأقل ضمن فريقها البحثى لتطوير تلك البرامج الخاصة.

ولكن ما هي المعلوماتية الحيوية؟

١ - التعريف الضيق:

* *المعلوماتية الحيوية الكلاسيكية:* هو علم يستخدم في المقام الأول لتحليل التتابعات سواء بروتينية اوجينومية.

* *المعلوماتية الحيوية الحديثة:* (ما بعد علوم الجينوم) هوعلم يشتمل على دراسة المقارنة للجينومات المختلفة والحامض النووي وتحليل وظائف الجينومات وتحليل المجموعة البروتينية الكاملة للجينوم بالأضافة الى المعلوماتية الطبية.

٢- التعريف الواسع للمعلوماتية الحيوية:

هو عبارة عن تحليل كمية كبيرة من المعلومات البيولوجية سواء كانت تتابعات وراثية أو غيرها عن طريق الحاسب الآلي.

قبل ظهور علم المعلوماتية الحيوية ،كان هناك مصطلحان علميان لوصف مكان إجراء التجارب البيولوجية و هما:

١ - In vivo: وهي إجراء التجارب على الكائن الحي مباشرة .

٢ - In vitro: وهي إجراء التجارب داخل المعمل .

حديثا وبعد أستخدام علم المعلوماتية الحيوية تم إدخال مصطلح In silico وهي إجراء التجارب نظرياً على الحاسب الآلي قبل إجرائها على النظم الحيوية أو فى المعامل.

التتابعات البيولوجية:

أولا التتابعات البروتينية (Protein sequences)

تتكون البروتينات عادة من ٢٠ وحدة بنائية مختلفة يطلق عليها الأحماض الأمينية. وتختلف البروتينات فى تركيبها ووظائفها بأختلاف عدد ونوع الأحماض الأمينية الداخلة فى تركيبها كما يوضح الجدول التالى:-

(جدول١): قائمة بـ ٢٠ حمض أميني مع ذكر أسمائها الكاملة والشفرة الثلاثية (المكونة من ثلاثة أحرف) والشفرة الأحادية (المكونة من حرف واحد) لهذه الأحماض والتى يطلق عليها عالميا شفرات الاتحاد الدولي للكيماء البحتة والتطبيقية (الأيوباك IUPAC).

The 20 Amino Acids and Their Official Codes

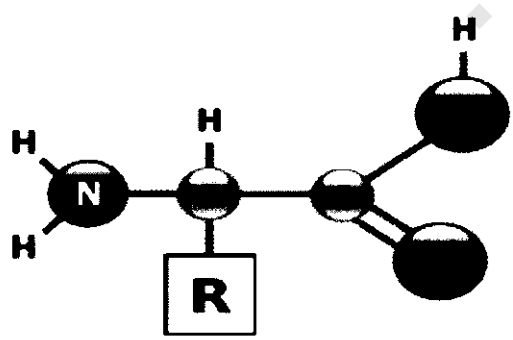
#	1-Letter Code	3-Letter Code	Name
1	A	Ala	Alanine
2	R	Arg	Arginine
3	N	Asn	Asparagine
4	D	Asp	Aspartic acid
5	C	Cys	Cysteine
6	Q	Gln	Glutamine
7	E	Glu	Glutamic acid
8	G	Gly	Glycine
9	H	His	Histidine
10	I	Ile	Isoleucine
11	L	Leu	Leucine
12	K	Lys	Lysine
13	M	Met	Methionine
14	F	Phe	Phenylalanine
15	P	Pro	Proline
16	S	Ser	Serine
17	T	Thr	Threonine
18	W	Trp	Tryptophan
19	Y	Tyr	Tyrosine
20	V	Val	Valine

أكواد الأحماض الأمينية السبعة الإضافية:

تستخدم هذه الأكواد السبعة الأضافية اما لوصف بعض الأحماض الأمينية الجديدة التى تم أكتشافه حديثا فى بعض أنواع البكتريا أو للتعبيرعن عدم التأكد من صحة وجود بعض الأحماض الأمينية.

(جدول٢): قائمة أكواد الأحماض الأمينية السبعة الإضافية (شفرات الاتحاد الدولي للكيماء البحتة والتطبيقية (الأيوباك IUPAC).

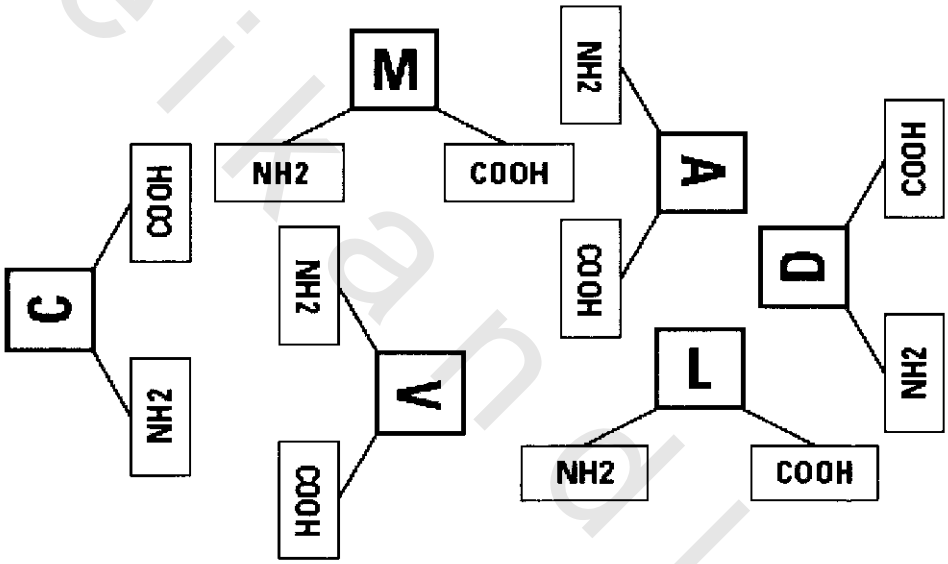
1-Letter Code	3-Letter Code	Meaning
B	Asn or Asp	Asparagine or aspartic acid
J	Xle	Isoleucine or leucine
O (letter)	Pyl	Pyrrolysine
U	Sec	Selenocysteine
Z	Gln or Glu	Glutamine or glutamic acid
X	Xaa	Any residue
--	-----	No corresponding residue (gap)



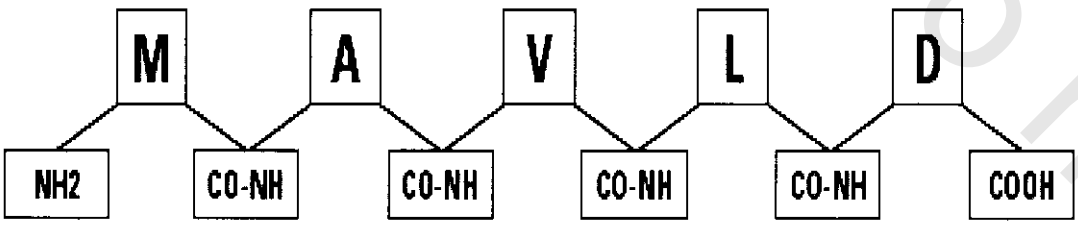
شكل ٥: يوضح البنية الكيميائية لحمض أميني في الكربون ألفا، لاحظ جذر الأمين NH٢ إلى اليسار وجذر الكربوكسيل COOH إلى اليمين

كيفية قراءة التتابع البروتيني:

تكوين السلسلة البيتيرية : الأحماض الأمينية (Amino Acid) هي لبنات البناء الرئيسية لبناء البروتين والببتيد في الجسم. يمكن ملاحظتها بسهولة بعد هضم البروتين (شكل٥ و٦). على الرغم من أختلاف تركيب وشحنة الأحماض الأمينية ألا أنها جميعا تتشابه فى وجود طرف أمينى(NH2) وطرف كربوكسيلى(COOH) . ترتبط الأحماض الأمينية مع بعضها عن طريق ارتباط الطرف الأميني(NH2) مع الطرف الكربوكسيلى(COOH) للحمض الأميني الآخر مكوناً رابطة ببتيدية مع خروج جزيء ماء (H2O)كما هو موضح بهذا الشكل٥ .



شكل ٦: صورة توضح التركيب العام للأحماض الأمينية المختلفة. يجب ملاحظة الطرف الأميني(NH2) والطرف الكربوكسيلى(COOH) لكل حمض الأميني.



شكل ٧: شكل يوضح ارتباط الطرف الأميني(NH2) مع الطرف الكربوكسيلى(COOH) للحمض الأميني الآخر مكوناً رابطة ببتيدية مع خروج جزيء ماء(H2O).

ولكن كيف يمكن ان نقرأ السلسلة الببتيدية؟

فى الغالب يتم قراءة السلسلة الببتدية بدأ من الطرف الأمينى أقصى اليسار الى الطرف الكربوكسيلى اقصى اليمين.

مثال لقرأة سلسلة ببتيدية قصيرة مكونة من خمس أحماض أمينية بالشفرة أحادية الكود أو بالأسم الكامل (M إلى D) :

MAVLD = Methanine – Alanine – Valeine – Leusine – Aspartic

كما يمكن قرأه نفس الأحماض الأمينية بإستخدام الشفرة الثلاثية

MAVLD = Met – Ala – Val – 1eu – ASP

مثال: الأنسولين Insulin يتكون من ١١٠ حمض أميني والذى يمكن كتابته على الصورة التالية (٣٠ جليسين Glycines + ٤٤ ألانين Alanines + ٥ تيروزين Tyrosines + ٤ جلوتامينGlutamines + ...)

ولكن هذه الطريقة تستهلك كميه كبيرة من المساحة التخزينية لجهاز الحاسب الآلي وخاصة عن التعامل مع كمية كبيرة من التتابعات البروتينية.

ولذلك تظهر فائدة الشفرات الأحادية والتى يمثل فيها كل حمض أمينى عن طريق حرف واحد وذلك لتوفير المساحة التخزينية لجهاز الحاسب الآلي وأيضا لزيادة سرعة التعامل مع اكبر قدرمن هذة البينات فى أقل وقت ممكن. وبناء على ذلك فيمكن مثلا كتابة جزىء الأنسولين الكلمل بإستخدام الشفرات الأحادية فى عدد أقل من السطور كما يلى :

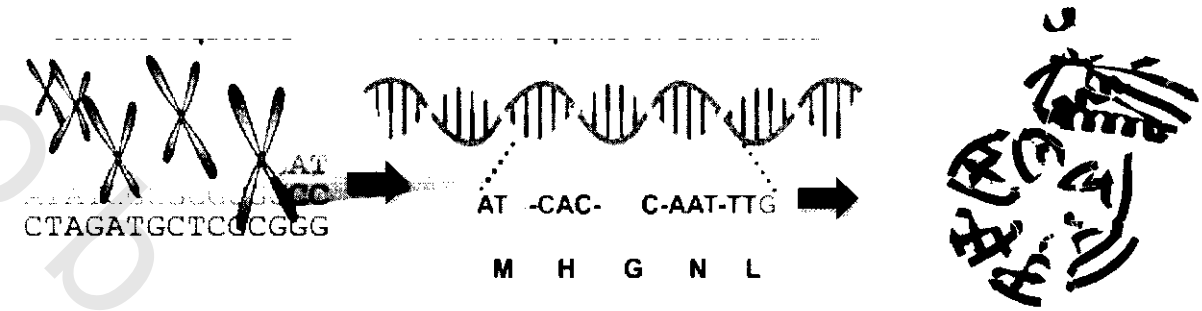
Insulin =
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLV
CGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALE
GSLQKRGIVEQCCTSICSLYQLENYCN

الشكل الثلاثي الأبعاد للبروتين Protein 3-D structures

إذا حدث تغير في أحد التتابعات الشفرية المكونة للأحماض الأمينية يؤدى ذلك إلى التغير في التركيب البروتينى الناتج وبالتالي التغير في الوظيفة وافضل مثال على ذلك هومرض أنيميا خلايا الدم المنجلية حيث أن التغير الحادث فى أحد الشفرات الوراثية لأحد الأحماض الأمنية يؤدى الى تغير التركيب البروتينى الناتج وبالتالى قدرتة على أداء وظيفته.

obeikandl.com

التركيب ثلاثى الأبعاد للبروتين ← التركيب البروتينى الناتج ← التتابع الوراثى داخل الجينوم



شكل ٨: شكل يوضح التركيب ثلاثى الأبعاد للبروتين والذى يؤثر فى نشاط ووظيفة البروتين.

نبذة تاريخية عن الشكل ثلاثي الأبعاد(3D) للبروتين :-

أول شكل ثلاثي الأبعاد(3D) للبروتين تم تحديده عام ١٩٨٥ بواسطة العالمان Kendraw and Perutz باستخدام الأشعة السينية وعلم البلوريات. وقد حصل كلاهما على جائزة نوبل في العلوم وبالإضافة إلى فوز أحدهما بجائزة نوبل في مجال البيولوجيا الجزيئية. وفى ذلك السياق وقد لوحظ أن :

١- البروتينات المتشابه فى التركيب ثلاثي الأبعاد تؤدي وظائف متشابهة. من المتوقع أن تكون البروتينات التي تحتوى على تتابعات ببتيدية متشابهة أن تتشابه فى الشكل ثلاثي الأبعاد الخاص بها وبالتالي تؤدي نفس الوظائف.

- وظيفة البروتين تعتمد اعتماد مباشر على الشكل الثلاثي الأبعاد لهذا البروتين.

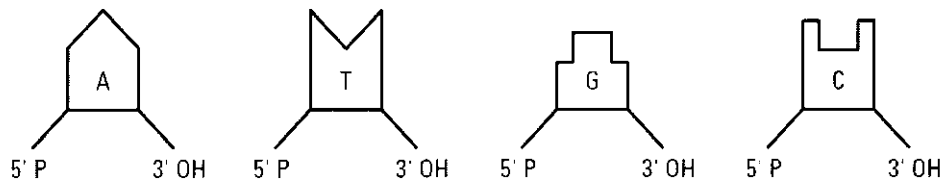
الوظيفة ← التركيب ← التتابعات

Function→Sequence → Structure

ثانيا تتابعات الـ DNA (Sequences DNA)

كيفية قراءة تتابعات الـ DNA بالطريقة الصحيحة؟

- يتكون الـ DNA من وحدات تسمى النيوكليوتيدات(Nucleotides) والتى يرمز لها بـ C (سيتوزين) ، A(أدنين)،T (ثيامين)، G(جوانين). ولكل قاعدة من القواعد النتروجينية السابقة طرف هيدروكسيلي(3`OH ) و طرف فوسفاتي(5` Po4) وارتباطها معا يكون روابط فوسفودای أستر.



شكل ٩: شكل يوضح القواعد النتروجينية الأربعة التى يتكون منها الـ DNA.

- الأحماض النووية أسهل في قرائتها وتخزينها على الحاسب عن الأحماض الأمينية لأنها تحتوي على ٤ قواعد فقط أي أربع رموز فقط وهي A – T – G – C.

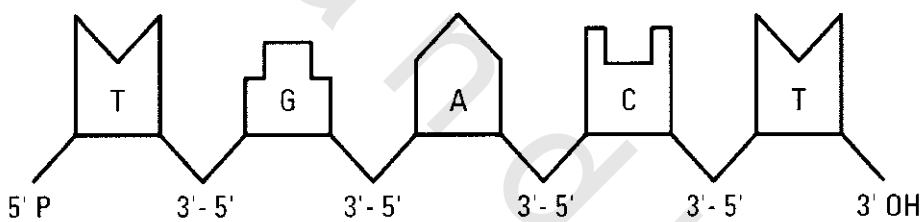
وبالتالى يمكن قراءة التتابع التالى كما يلى:

TGACT = Thymine-Guanine-Adenine-Cytosine-Thymine

أو

T (ثيامين) – C (سيتوزين) – A(أدنين) – G (جوانين) – T (ثيامين)

TGACT =



شكل ١٠: شكل يوضح أرتباط القواعد النتروجينية الأربعة التى يتكون منها الـ DNA بروابط فوسفودای أستر

شفرات الأيوباك لتتابعات الـ DNA:

شفرات الأيوباك هى الشفرات المتعارف عليها عالميا للتشفير عن النيوكليوتيدات والتى تظهر فى الجدول التالى:

جدول٣ : قائمة أكواد النيكليوتيدات لتتابعات الـ DNA (شفرات الاتحاد الدولي للكيماء البحتة والتطبيقية (الأيوباك IUPAC)).

1-Letter Code	*Nucleotide Name*	*Category*
A	Adenine	Purine
C	Cytosine	Pyrimidine
G	Guanine	Purine
T	Thymine	Pyrimidine
N	Any nucleotide (any base)	(n/a)
R	A or G	Purine
Y	C or T	Pyrimidine
--	-----	None (gap)

ملحوظات:

R يرمز به عند الاشتباه في G أو A

Y يرمز به عند الاشتباه في T أو C

N يرمز به عند الاشتباه في أي نيوكليوتيده من الأربع (A,T,C,G)

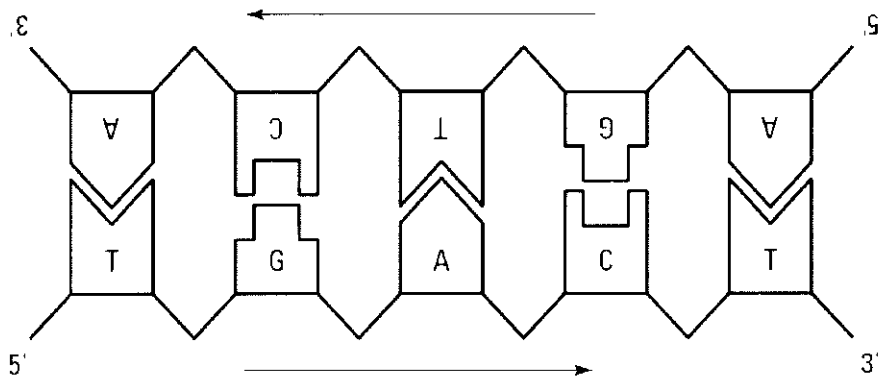
يوضع الرمز (--) فى التتابع عندما لاتوجد اى نيوكليوتيدة

- Palindromes التتابعات المتعاكسة (البالندرومية)

أجزاء من الـ DNA إذا تم قرأتها فى أتجاه معين (٥-٣ مثلا) من أحد الخيطين تعطى نفس القراءة عند قرائتها من الطرف الآخر فى الأتجاه (٥-٣) . أى أن هذه التتابعات هى صور متشابه ومتكاملة ومتعاكسة فى الأتجاه لمقطع من الـ DNA كما هو موضح فى المثال التالى.

ATGCTGATCTTGGCCATCAATG and

CATTGATGGCCAAGATCAGCAT



شكل ١١ : شكل يوضح التتابعات المتعاكسة المتكاملة (البالندرومية)

الأهمية البيولوجية للتتابعات المتعاكسة (البالندرومية) الـ Palindromes

التتابعات المتعاكسة ذات أهمية كبيرة داخل الجينوم وتؤدى وظائف هامة. فعلى سبيل المثال:

- معظم إنزيمات قطع (القصر) الـ DNA (Restriction enzyme) تقطع عند هذه المناطق وتسمى فى هذه الحالة "مناطق القصر".
- هذه المناطق تستخدم كمناطق ارتباط بروتينات التحفيز الخاصة ببدء نسخ الجينات لزيادة نشاطها أو إيقافه.
- لها تأثير قوي ومباشر على الشكل الثلاثي الأبعاد لجزيئات الـ DNA.

ثالثا تتابعات الـ RNA (Sequences DNA)

جزيئات الحمض النووى الريبوزى (الـ RNA)هي أكثر أنواع الأحماض النووية نشاطاً حيث يتم تخليقها وتحللها بشكل مستمر داخل الخلية الحية.

ويلخص علماء المعلوماتية الحيوية (Bioinformatics) الفرق بين تتابعات DNA و RNA كما يلى:-

١ - الـRNA يختلف عن الـ DNA بقاعدة واحدة هي (U بدلاً من T).

٢-الـ RNA خيط مفرد وليس حلزون مزدوج مثل DNA.

وفيما يلى جدول يوضح قائمة الأحماض النووية (النيوكليوتدات) التى تكون الـRNA وأختصاراتها المعروفة فى مجال المعلوماتية الحيوية.

جدول ٤ : قائمة أكواد النيكليوتيدات لتتابعات الـ RNA (شفرات الاتحاد الدولي للكيماء البحتة والتطبيقية (الأيوباك IUPAC)).

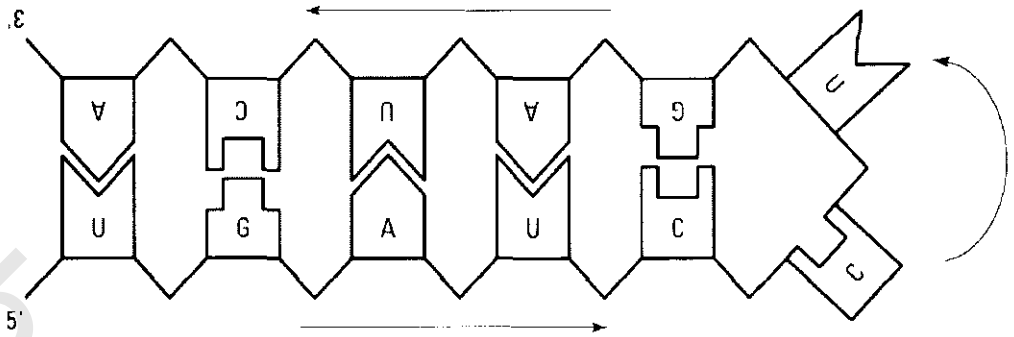
1-Letter Code	*Nucleotide Base Name*	*Category*
A	Adenine	Purine
C	Cytosine	Pyrimidine
G	Guanine	Purine
U	Uracil	Pyrimidine
N	Any nucleotide	Purine or Pyrimidine
R	A or G	Purine
Y	C or U	Pyrimidine
--	-------	None (gap)

هناك أكثر من نوع من أنواع الـRNA:

- الـRNA الرسول mRNA والذى ينتج عن الترجمة المباشرة للجينات.
- الـRNA الناقل tRNA والذى يقوم بنقل الأحماض الأمينية إلى الريبوسوم لتخليقها.
- الـRNA الريبوسومى rRNA يدخل في تركيب الريبوسوم.
- جزيئات الـRNA الصغيرة (Small RNAs).

- عند وجود مناطق متشابهة في جزيئات الـRNA يؤدى ذلك عمل التفاف بعض المناطق على بعضها وتظهر كأنها مناطق مزدوجة الخيط.
- الوظيفة البيولوجية للجزيئات الـRNA تأتي من الشكل ثلاثي الأبعاد أو من التتابعات المتكاملة مع الجين المتخصص.
- جزيئات الـRNA الصغيرة (Small RNAs) والتي ،تم التعرف عليها حديثا، تقوم بوظيفة بيولوجية هامة في تنظيم نشاط بعض الجينات.

obeikandi.com



شكل ١٢ : شكل التفاف خيط الـ RNA على بعضها لتكوين خيط مزدوج

الشفرات الجينية

معلومات هامه يجب ذكرها:

فى لغة المعلوماتية الحيوية نقول ان الجين يتكون من عدد من المواقع (Bases) بدلاً من عدد من النيكلوتيدات المزدوجة. فمثلا جزىء الـ DNA الذى طوله ٤٠٠ نيكلوتيده يتضمن فى الواقع ضعف عدد القواعد (٨٠٠ قاعدة) لأن كل موقع يحتوي زوج من النيكلوتيدات.لذللك يتم قياس طول الـ DNA بعدد أزواج القواعد واختصارها Base Pair (bp)

وهناك وحدات أكبر مثل kb أو Mb هذه الوحدات (Mega – bp) تستخدم أيضاً.

- الف قاعدة Kb (Kilo base) = 1000 bp
- مليون قاعدة Mb (Mega base) = ١٠٠٠ kp

ومن الجدير بالذكر بأن الشفرات الوراثية هى شفرة عامة لجميع الكائنات الحية ولكن هناك بعض الاستثناءات. النيكلوتيدات الأربع تكون ٦٤ شفرة وراثية (Codes) والتى بدورها تشفرلأنتاج ٢٠ حمض أميني مختلف. كل حمض أمينى له أكثر من شفرة واحدة وكل شفرة تتكون من ثلاثة أحرف(ثلاث نيوكليوتيدات) فقط. توجد أيضا شفرات عامة لبدء أو أنهاء عملية النسخ الجينى فى بداية ونهاية الجين. ومن الجدير بالذكر أنه تم اكتشاف هذه الشفرات الوراثية في الستينات من القرن العشرين.

ولكن كيف يمكن التعرف على الشفرات الوراثية فى أى تتابع وراثى؟

كما نرى في المثال التالي:

١) قراءة تتابع جزئ من الـ DNA:

ATGGAAGTATTTAAAGCGCCACCTATTGGGATATAAG

٢) تتقسيم تتابع جزئ الـ DNA إلى ٣ قواعد متتالية:

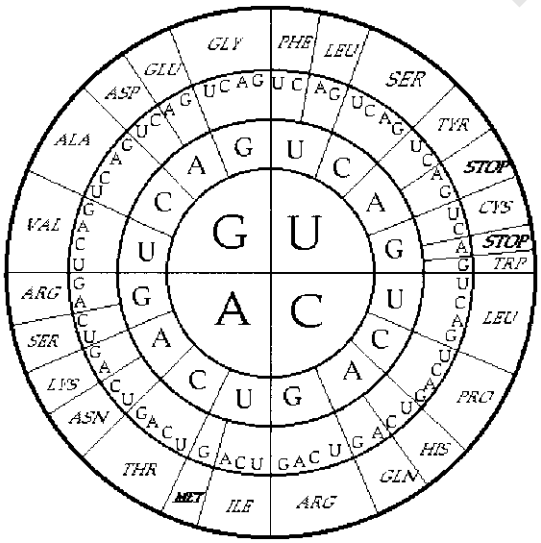
ATG- GAA- GTA- TTT –AAA- GCG- CCA- CCT- ATT- GGG-
ATA- TAA- G..

٣) تترجم كل شفرة ثلاثية إلى الحمض الأميني المقابل:

M E V F K A P P I G I STOP

وذلك مع ملاحظة أنه إذا كان تركيب الـDNA من الطرف `5 إلى الطرف `3 فإن السلسلة الببتيدية تتكون من الطرف الأمينى N إلى الطرف الكربوكسيلى C على التوالي.

ولذلك فإننا إذا علمنا من أين نبدأ قراءة الجين فإننا سوف نتوقع التركيب البروتينى الناتج من هذا الجين. لذلك فيمكن برمجة الحاسب الألى لكى يقوم بترجمة تتابعات الـ DNA ويحاكى الحاسب الألى الخلية في ترجمة الجينات الوراثية.



A

2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

B

شكل ١٣ : تركيب الـ٦٤ شفرة وراثية(Codes) والأحماض الأمينية المتوقعة منها. A) الشكل الحلقى لقراءة الشفرات الوراثية.
B) جدول الشفرات الثلاثية والأحماض الأمينية المتوقعة منها.

تغيير قراءة الإطارللتتابعات الوراثية
(Changing the reading frames)

كما أتضح مسبقا فأن قراءة التتابعات الوراثية يعتمد على تقسيم التتابع المدروس الى أكواد مكونه من ٣ قواعد كما في المثال السابق .ولكن أذا تم استبعاد النيكلوتيده الأولى من بداية التتابع المرغوب فإن أطار قراءه سوف يتغير القراءة ومما يؤدي إلى تغير كبير في قراءة البروتين الناتج كما يلي:

تتابع وراثى

ATGGAAGTATTTAAAGCGCCACCTATTGGGATATAAG

- تغيير أطار قراءه بعد خذف القاعدة الأولى

A-TGG-AAG-TAT-TTA-AAG-CGC-CAC-CTA-TTG-GGA-TAT-AAG

- الأحماض الأمينية الناتجة

W K Y L K R H L L G Y K

من الملاحظ أن تجاهل قراءة القاعدة الأولى أدى الى تغيرالأحماض الأمينية الناتجة. وبناء علية فأن تجاهل قراءة القاعدة الثانية أو الثالثة لنفس التتابع سوف يؤدى أيضا الى تغيرالأحماض الأمينية الناتجة من نفس التتابع وبالتالى ظهور ثلاث تتابعات ببيتدية محتملة من تتابع وراثى واحد من الـ DNA

الستة طرق الممكنة لقراءة إطارالتتابع
(Six possible reading frames of DNA)

- هناك ٣ طرق مختلفة لقراءة تتابع وراثى واحد من الـDNA الواحد ، حيث أن الشفرة الوراثية شفرة ثلاثية القواعد وهناك خيطان من الـ DNA ولذا يكون لدينا ٦ طرق للقراءة وهذا ما يسمى القرأت الستة المحتملة للتابع الوراثى Six possible reading frames of DNA وبالتالي لكل جين ٦ طرق مختلفة محتملة يمكن أن ينتج بها بروتينات مختلفة يمكن أن يتنبئ بها الحاسب الألى .يستثنى من ذلك بعض الفيروسات التي تحتوي على خيط مفرد وبالتالي يستثنى من هذه القاعدة. ويستثنى من ذلك أيضاً أن هناك بعض التتابعات تسمى Stop condons (إشارات وقف القراءة TAA – TGA – TAG) . أما دون ذلك يسمى أطارالقراءة المفتوحة (Open Reading Frame ORF).

مناطق الـ DNA الغير مشفر (Non coding DNA):

هناك مناطق كبيرة من الـ DNA غير مشفرة داخل الجينومات المختلفة وتسمى الـDNA الغير مشفر (Non coding DNA) .هذه المناطق قد زادت من تعقيد الـ DNA فى الكائنات الراقية لأن الكائنات الراقية تحتوي على مناطق كثيرة غير مشفرة والتى لاتنتج اى بروتينات .وتوجد أيضا بعض المناطق الغير المشفرة داخل الجينات وفي هذه الحالة تسمى أنترونات (Introns) وذلك غالبا فى جينات الكائنات حقيقيات النواه.

هناك برامج حوسبية كثيرة داخل تطبيقات المعلوماتية الحيوية تعمل على اكتشاف المناطق الغير مشفرة فى تتابعات الـ DNA والتي تحدد بدقة أين تبدأ وتنتهي الجينات وأين توجد أنترونات (تتابعات غير مشفرة) وأكسونات (تتابعات مشفرة) داخل هذه الجينات.

العمل مع جينومات بأكملها (ويحكي التاريخ):

- تم اكتشاف أول تقنية فعالة لتعرف على تتابعات الحمض النووي الـ DNA سنة ١٩٧٧ وأطلق عليها "طريقة سانجر" للسلسة نسبة الى مكتشفها. أما فى سنة ١٩٩٥ تم انتهاء من سلسة تتابعات أول جينوم كامل (وهو فيروس الإنفلوانزا Hemphilus influezae.

وخلال هذه الفترة كان علماء البيولوجي يعملون على قطع صغيرة من DNA وكان طولها بضعة آلاف من النيكلوتيدات. وذلك لأنهم كانوا يبحثون عن جينات محددة وكانوا يعملون عليها من قبل. معظم أدوات وبرامج المعلوماتية الحيوية المتاحة الآن أنشئت منذ تلك الفترة مثل

- برامج أيجاد التشابه بين التتابعات BLAST .
- برامج تحديد القرابة والتقسيم Phylogeny .
- طرق التصور النوعي والتصميم للتتابعات الصغيرة Design.
- وسائل عرض مختلفة للتتابعات الصغيرة(مثل التتابعات البروتينية التى لا تتعدى بضعة الآلاف من الحروف) Visualization.

الجينومية: الحصول على جميع الجينات في آن واحد
Genomics: Getting all the gene

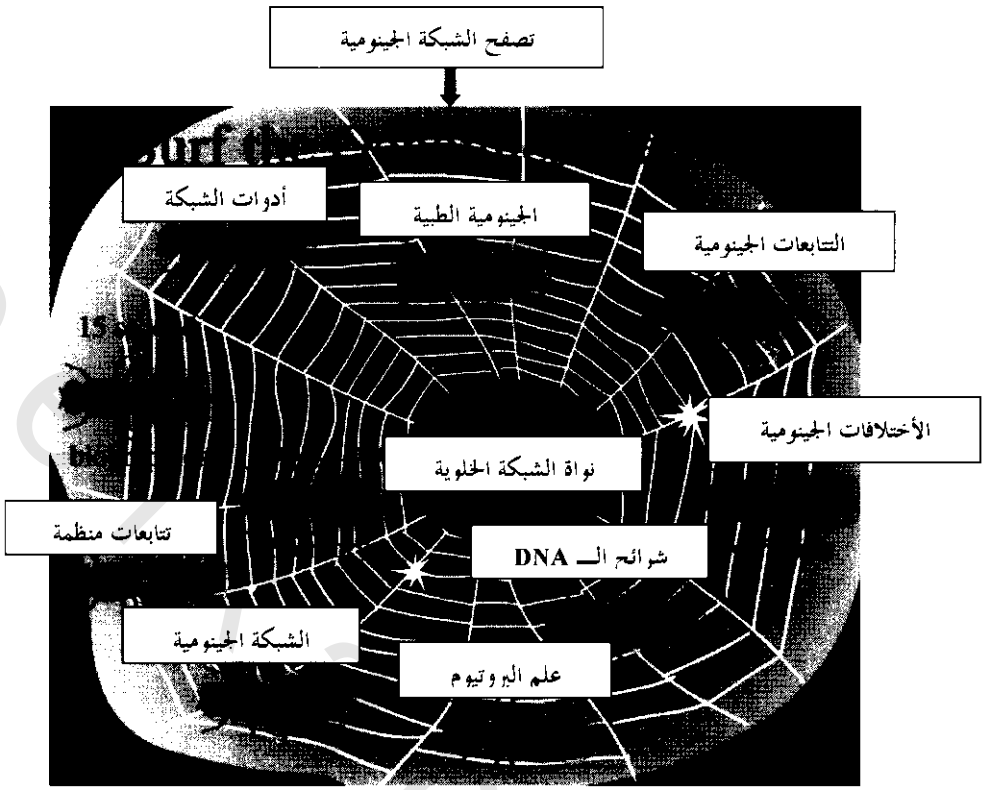
بعد الأنتهاء من معرفة التتابع الكامل لأول جينوم كان ذلك بداية "عصر الجينومية". هذا الفرع الجديد من العلم وهذه الثورة الجديدة دعت العلماء إلى تصميم أدوات جديدة من أدوات "المعلوماتية الحيوية" تكون قادرة على تخزين قواعد البيانات والبحث والتحليل وعرض هذه القطع الضخمة من الـ DNA فيماسمى بعد ذلك بنوك الجينات (Gene Banks) .

ودفع هذا التطور تطور ظهور علم "المعلوماتية الحيوية" بأدواته الجديده والذى ركز فى البداية على تحليل تتابعات الأحماض النووية الكبيرة وتحليلها إلى مكوناتها الأساسية (تتابع الجينات، ترجمة الجينات ، مناطق الشفرات البروتينية، مناطق الـ DNA الغير مشفرة، العناصر التنظيمية وغيرها).

وهذه المرحلة يتبعها مرحلة أكثر طولاً وأهمية وهي مرحلة تحليل الجينوم للتعرف على الوظائف البيولوجية المختلفة. وتشتمل المرحلة الثانية على:-

١ - البحث فى تتابعات الجينومات المتاحة.
٢ - تحليل جينومات محددة إلى تتابعات متخصصة.
٣- عرض الشكل الكلى للجينومات المختلفة وتحديد التشابه والأختلاف بينها.
٤ - تحليل التتابعات الجينومية المشفرة (ORF).
٥ - تحليل التتابعات الجينومية للكائنات الراقية Gene Scan.
٦ - إيجاد الجينات المتشابهة والمتناظرة Orthologous و Paralogous.
٧- إيجاد التكرارات (التعرف على المناطق والتتابعات المتكررة داخل الجينوم).

ويظهر الشكل الشبكى التالى بعض الأفكار الحديثة فى هذا العلم



شكل١٤: شكل الشبكى يوضح بعض النقاط البحثية الحديثة فى علم المعلوماتية الحيوية بعد تطور عوم الجينوم

التعرف على البروتينات بواسطة أسمائها

عند محاولة البحث عن تتابع بروتينى عن طريق الأسم فقط فى قاعدة البيانات ستجد العديد من التتابعات البروتينية التى تشبه أوتقارب أسم التتابع المرغوب فى قاعدة البيانات. لذلك يجب أن لاتنزعج أوتحتار فربما تكون هذه هى بداية الطريق لعمل بحث جيد والتوصل لنتائج مزهلة.

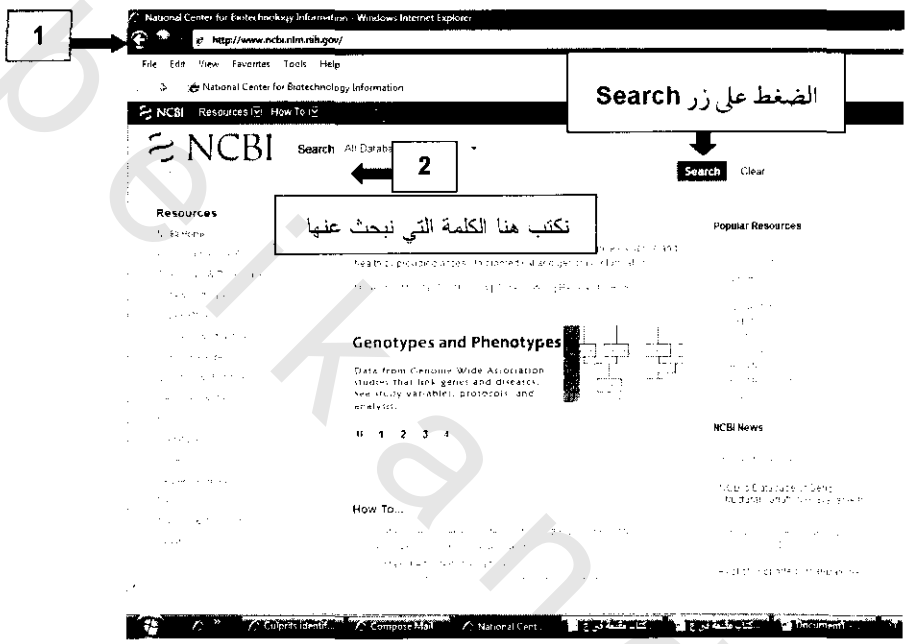
فمثلا إن كنت لم تسمع عن كلمة "HSF" قط في حياتك فأين تذهب من هذا؟

كيف يمكن معرفة المزيد عن هذا المصطلح الذي يبدو غامضاً مع هجاء غريب؟

كيف يمكنك أن تقرر بسرعة ما إذا كانت هذا البحث يمكن الأستمرار فيه فى سياق ما سبق لك معرفته عن الجينات الخاصة بك؟

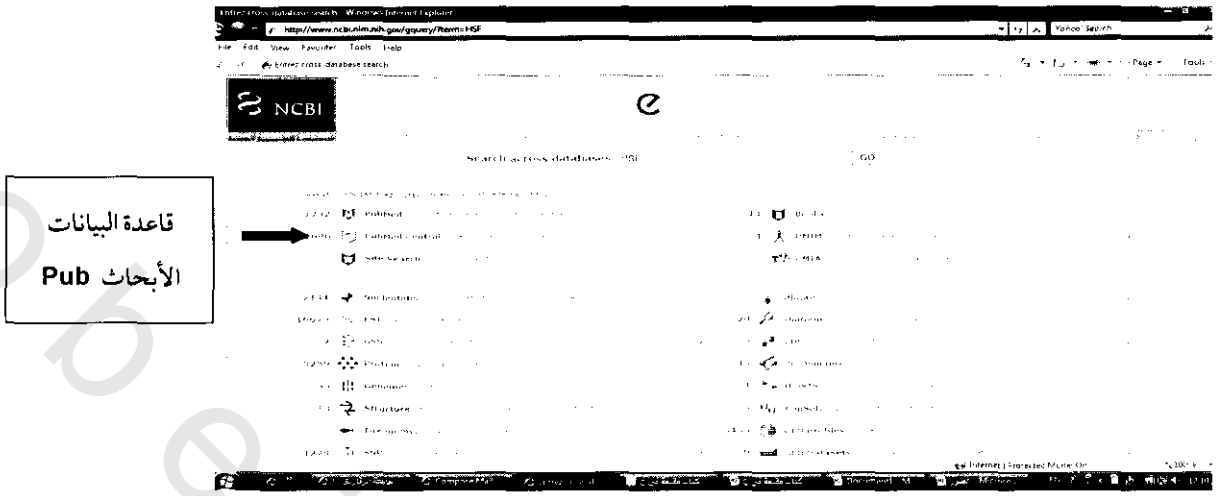
سوف يتم الحصول على الجواب من زيارة الصفحة الرئيسية للمركز الوطنى للبيوتكنولوجى National Center or Biotechnology Information (NCBI)

المركز الوطنى للبيوتكنولوجى (NCBI Home page)



شكل ١٥ : صورة توضح شكل الصفحة الرئيسية للمركز الوطنى للبيوتكنولوجى.هذه الصورة توضح الصفحة الرئيسية لـ NCBI السهم رقم ١ يوضح رابط الموقع وهو http://www.ncbi.nlm.nih.gov/ . السهم رقم٢ يوضح المكان الذي يكتب به عنوان البحث المراد البحث عنه. عند كتابة كلمة للبحث عنها تفتح لنا الصفحة التالية.

- عند الدخول على الصفحة الرئيسية للمركز الوطنى للبيوتكنولوجى (NCBI) نجد كلمة Search (All Databases) . نكتب هنا الكلمة التي نبحث عنها ثم نقوم بالضغط على زر Search تظهر لنا صفحة بها جميع قواعد البيانات المتاحة وما تحتويه عن هذه الكلمة المستخدمة في البحث كما فى شكل ١٥.



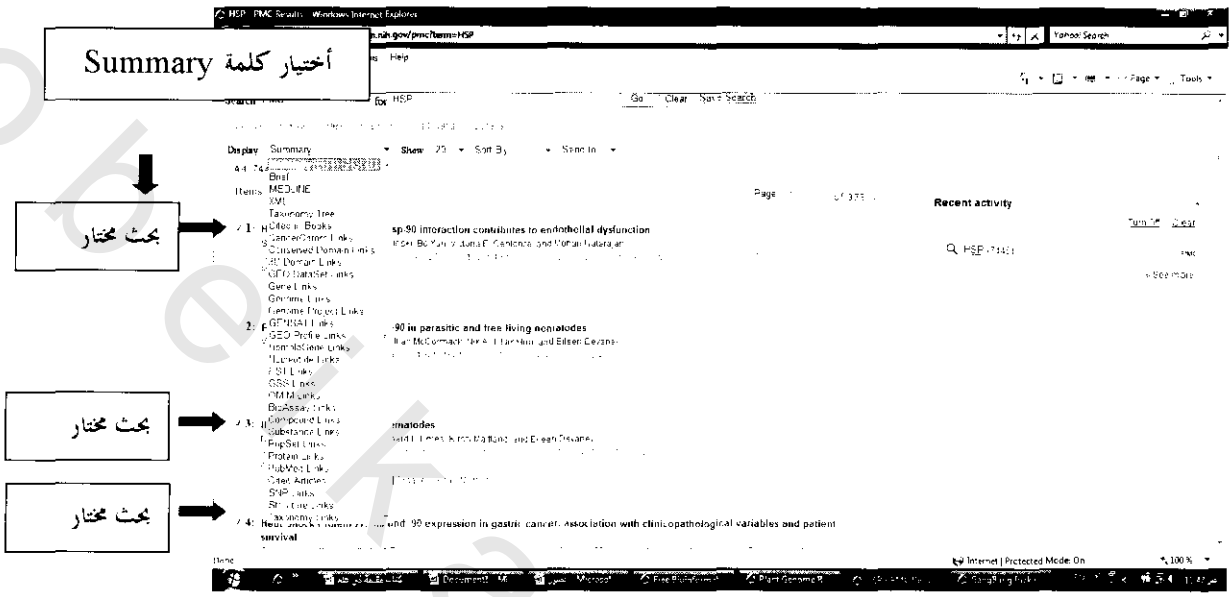
شكل ١٦ : هذه الصورة توضح جميع قواعد البيانات التي تحتوي معلومات عن هذا بروتين"HSF".

- عند الضغط على قاعدة البيانات Pub Med central نذهب إلى صفحة معلومات جديدة بها العديد من المقالات العلمية المحكمة والمنشورة فى مجلات عالمية عن كلمة البحث "HSP" التي نبحث عنها. هذه الصورة هي صورة الصفحة الرئيسية للـ PubMed وبها مجموعة من الأبحاث المراد البحث فيها فعند الرغبة في تحويلها إلى Text نضغط على Send to ونختار منها Text كما سيوضح في الصورة التالية شكل ١٧.

شكل ١٧ : هذه الصورة توضح شكل نتائج البحث فى قاعدة البيانات **Pub Med central**

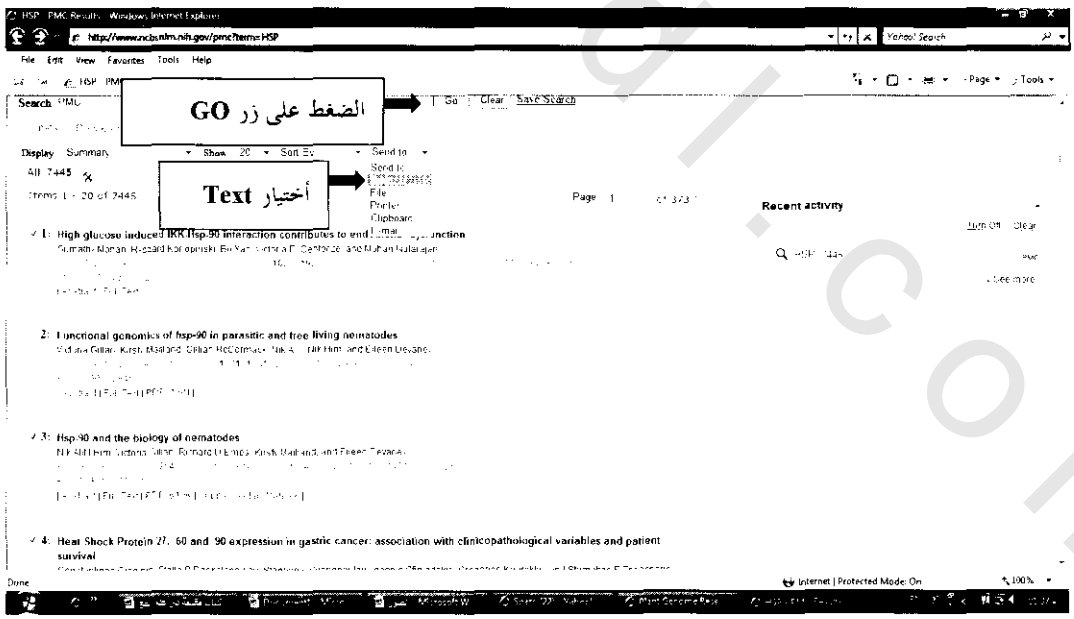
بعد البحث عن اى معلومات عن بروتين"HSF".

شكل ١٧- يمكن أختيار أى عدد من الأبحاث الناتجة ثم أختيار كلمة Summary للحصول على ملخصات الأبحاث المرغوبة كما فى (شكل ١٧الى شكل ٢٢)



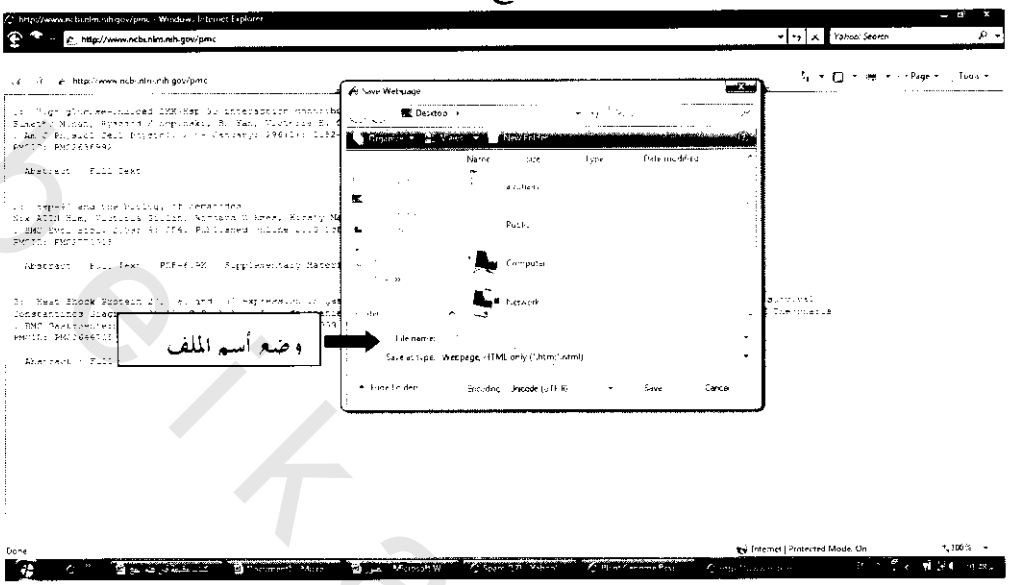
شكل ١٨ : أختيار أى عدد من الأبحاث الناتجة ثم أختيار كلمة Summary

- عند الضغط على Send to واختيار كلمة Text تحول الموضوع إلى كتابةText



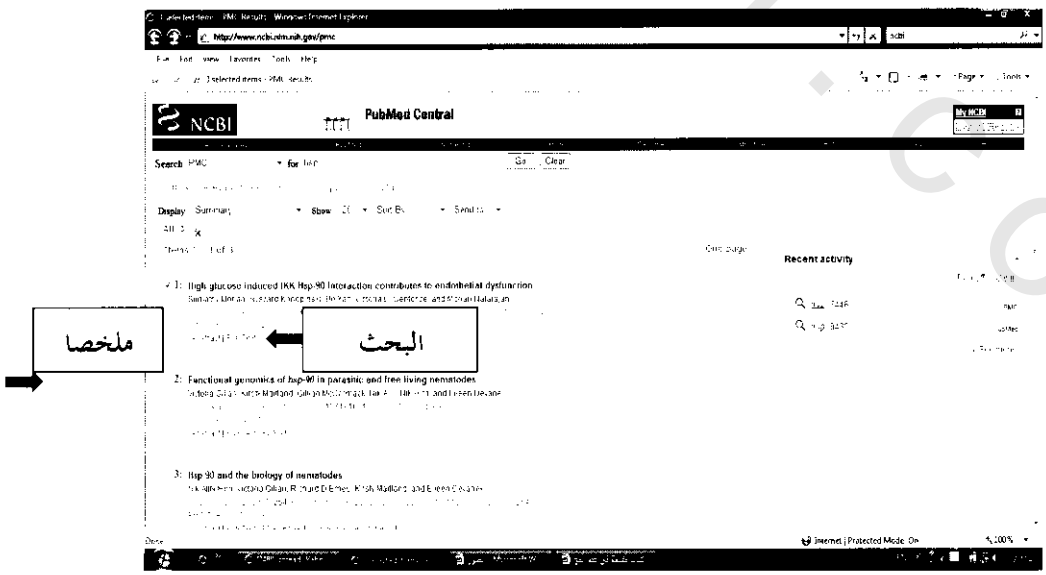
شكل ١٩ :أختيار أى عدد من الأبحاث وأظهارها على صورة ماف نصى Text مع ملاحظة أن الأبحاث الناتجة ستكون على شكل أسم البحث واسماء المؤلفين وأسم المجلة المنشور بها البحث وتاريخ النشر .

- كما يمكننا أن نختار المواضيع المراد فحصها عن طريق الصناديق المقابلة نضع علامة صح في الصندوق المقابل للمواضيع التي نريد تصفحها ثم حفظها أو طباعتها. حفظ النتائج عن طريق اختيار مكان الحفظ File → Save as.



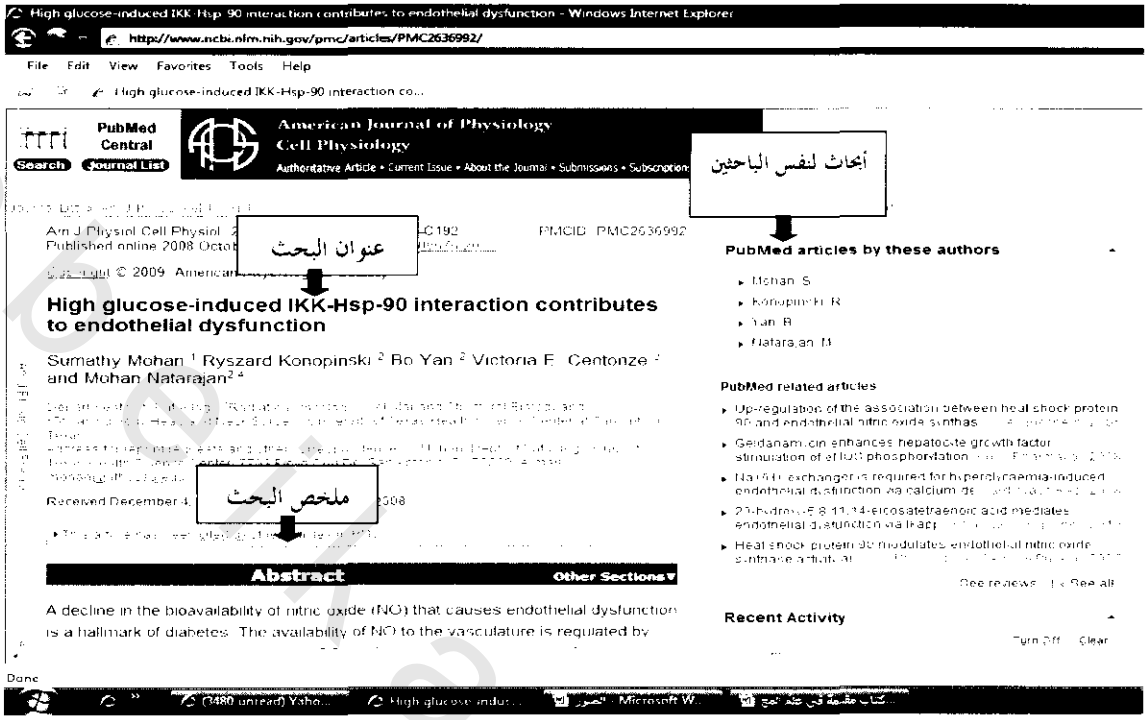
شكل ٢٠: أختيار حفظ النتائج عن طريق اختيار مكان الحفظ

- أما عند الرغبة في إظهار ملخصات الأبحاث الموجودة (Abstracts) أو رؤية الأطلاع على البحث بالكامل (Full text) (إذا كان موجودا) فيمكن الضغط على زر Abstracts أو Full text الظاهر أسفل عنوان البحث كما في الصورة (شكل ٢١).



شكل ٢١: الضغط على زر Abstracts أو Full text الظاهر أسفل عنوان البحث

وسوف تظهر لنا الملخصات أو الأبحاث الكاملة كما في الصورة (شكل ٢٢).



شكل ٢٢: أظهار الملخصات أو الأبحاث الكاملة

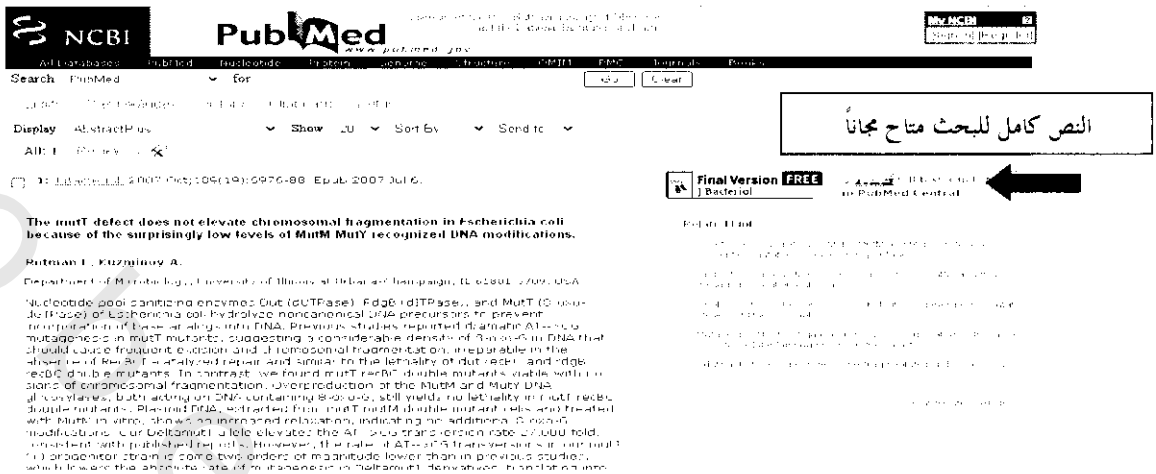
فعند الدخول على بحث معين سوف يفتح لنا صفحة البيانات الخاصة به ويظهر الموضوع كامل وعند الرغبة في حفظ هذا البحث نقوم بالضغط على كلمة File ثم Save as ثم نحدد المكان المراد الحفظ فيه كما في شكل ٢٠.

طرق متقدمة فى البحث داخل قاعدة الأبحاث Pub Med

يمكن البحث داخل قاعدة البيانات الخاصة بالأبحاث عن طريق:-

- البحث في Pub Med باستخدام اسم البروتين أو الجين.
- البحث في Pub Med باستخدام أسماء المؤلفين.
- البحث في Pub Med باستخدام أسماء البروتينات أو الجينات وأسماء المؤلفين.

كما يمكن أن نحصل على النص كامل للبحث مجاناً من (كما يتبين هذا من مستطيلات ملونة كبيرة تظهر أعلى العنوان) والرسم التالي سوف يوضح ذلك.



شكل ٢٣ : أمكانية الحصول على الأبحاث الكاملة مجانا عن ظهور المربع المشار اليه بالسهم.

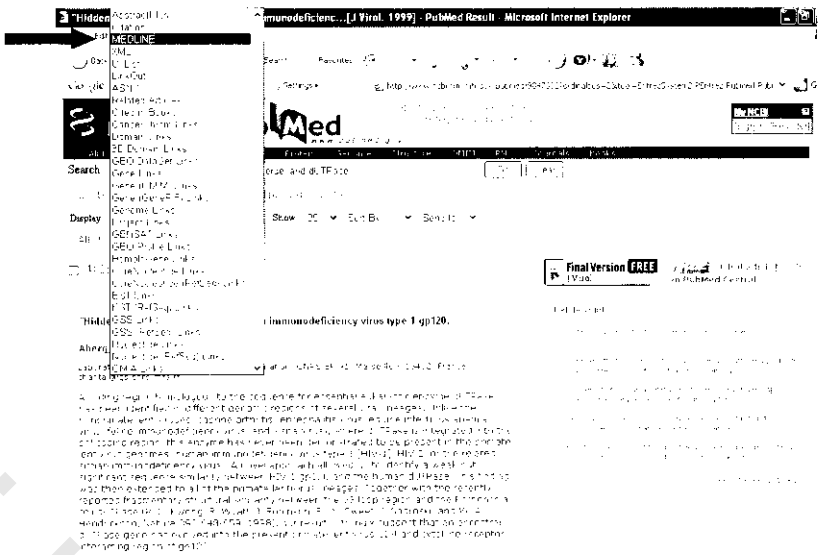
ولكن أحينا يكون البحث الكامل متاح ولكن يحتاج الى دفع أشتراك فى قاعدة البيانات المحتوية عليه كما فى الشكل التالى:

شكل ٢٤ : أمكانية الحصول على الأبحاث الكاملة ولكن عن طريق الأشتراك بقاعدة البينات.

الأستفادة من قاعدة بيانات الـ Medline

البحث عن البحث المطلوب، يمكن الأستعانة بقاعد بيانات الـ Medline للحصول على بعض الأختصارات الهامة كما يظهر فى الشكل التالى (شكل ٢٥):

بيانات الـ **Medline**

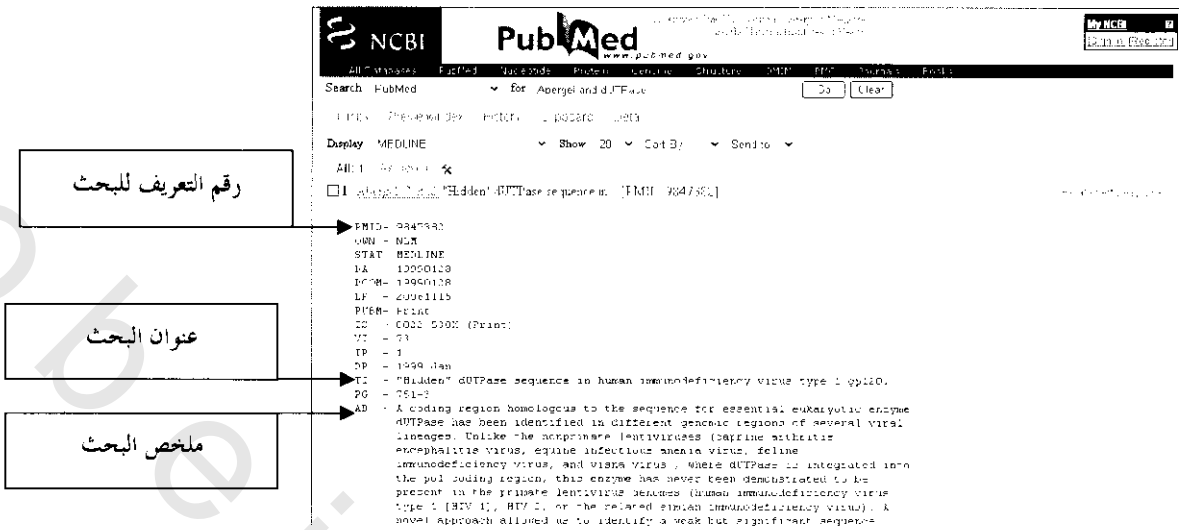


شكل ٢٥: كيفية الوصول الى قاعدة بيانات الـ Medline.

بعد اختيار كلمة الـMedline يظهر لنا بعض الأختصارات الهامة التى يمكن أستخدامها فى البحث المتقدم كمايلى بعض الأمثلة:-

رقم التعريف للبحث	PMID
عنوان البحث	TI
عنوان المعمل الذى أجرى البحث	AD
أسم الباحث	AU
مصدر البحث (أسم المجلة)	SO
تاريخ النشر	DP
ملخص البحث	AB

obeikandl.com



شكل ٢٦: شكل يوضح جزء من الى قاعدة بيانات الـ Medline.

كيفية أستخدام هذه الأختصارات فى البحث

على سبيل المثال:

يمكن أستخدام أي من هذه الأختصارات فى الأبحاث بعد وضعها بين علامتين [.....] كما يلى:-

- Down [AU] – أي أن أسم الباحث يسمى (Down).
- Down [TI] – أي أن الأبحاث تحتوى في عنوانها على كلمة (Down).
- Down [AD] – أي أن عنوان الباحث يحتوى على كلمة (Down).
- Down [AB] – أي أن الأبحاث تحتوى في ملخصها على كلمة (Down).

كيف يمكن تطبيق الأدوات السابقة لمعرفة باحث بعمل في نفس مجالك البحثى قريب منك جغرافيا

مثال عند وصولك إلى بلد ما جديد (على سبيل المثال الزقازيق) وتريد معرفة هل يعمل أحد فيها على "أبحاث النبات" فيجب أن نتبع الخطوات التالية كما يلى:

١- ندخل إلى www.ncbi.nlm.nih.gov/entrez/.

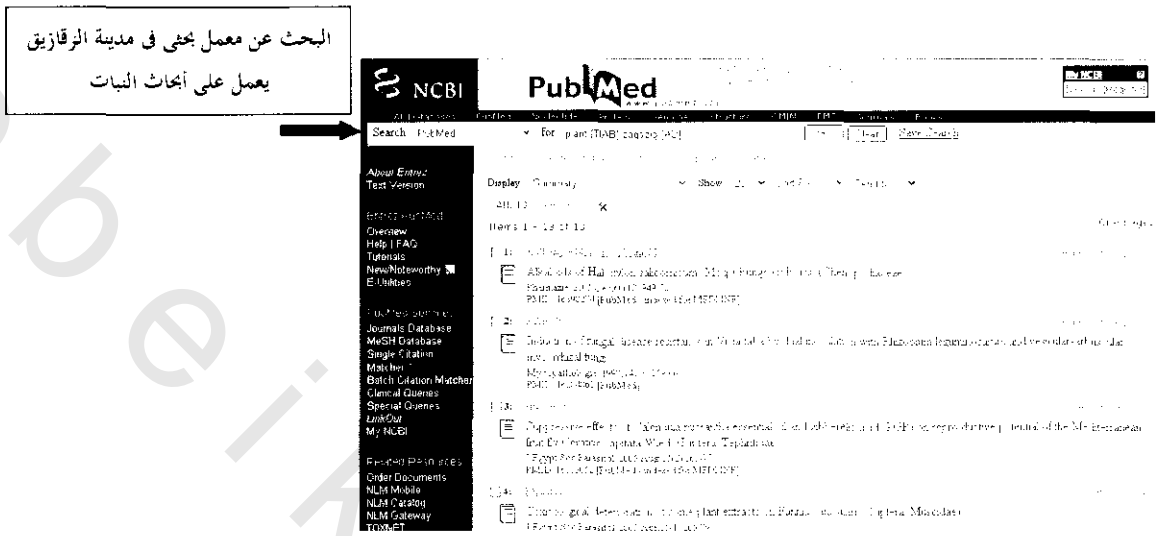
٢- نكتب في مربع For كلمة Plant [TiAB] Zagazig [AD]. ثم نضغط على كلمة Go ومعناها ابحث عن معمل بحثى فى مدينة الزقازيق يعمل على أبحاث النبات.

٣- اضغط على قائمة الباحثين الزرقاء والتي تحتها خط.

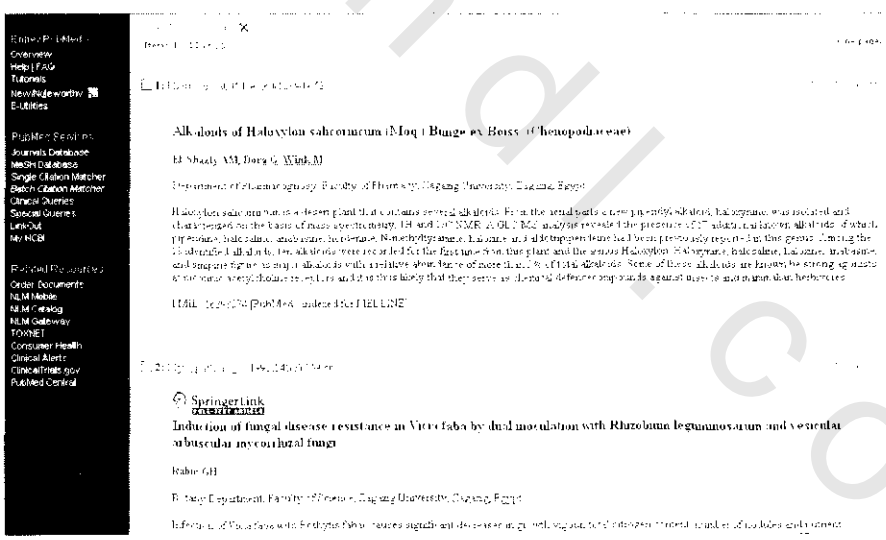
٤- العودة إلى العنوان السابق (بعد الخطوة ٢) باستخدام زر العودة في المتصفح.

٥- في العرض القائمة المنسدلة، وتغير الخيار من Summary إلى Abstract.

وفيما يلي شرح ذلك بالصور.



شكل ٢٧: البحث عن **Plant [TIAB] Zagazig [AD]** ومعناها هذا أي ابحث عن كلمة **Plant** في العنوان أو الاختصار ويكون مكان العمل في الزقازيق وهذه هو عملها بالصور.



شكل ٢٨: ملخص أحد الأبحاث الناتجة عن البحث عن أبحاث البنات في مدينة الزقازيق

البحث في الـ Pub Med باستخدام محددات البحث المتقدم (Limits)

Searching Pub Med using limits

عند الضغط على Limits أي بحث متقدم نجد الأختيارات التالية:-

١- "Links to full text" أي يعمل على إظهار ملخصات الأبحاث التي لها نص كامل.

٢- "Links to free full text" أي إظهار الأبحاث المجانية فقط.

٣- "Abstract" أي عرض ملخصات الأبحاث فقط.

٤- "Published in the last" تحديد الفترة السابقة التي تم خلالها النشر.

٥- "Added to PubMed in the last" تحديد تاريخ أضافة البحث على الـ PubMed.

٦-"Humans or Animal"يمكن هذا الأختيار من تحديد نوع البحث في نوع الكائن إنسان أو حيوان

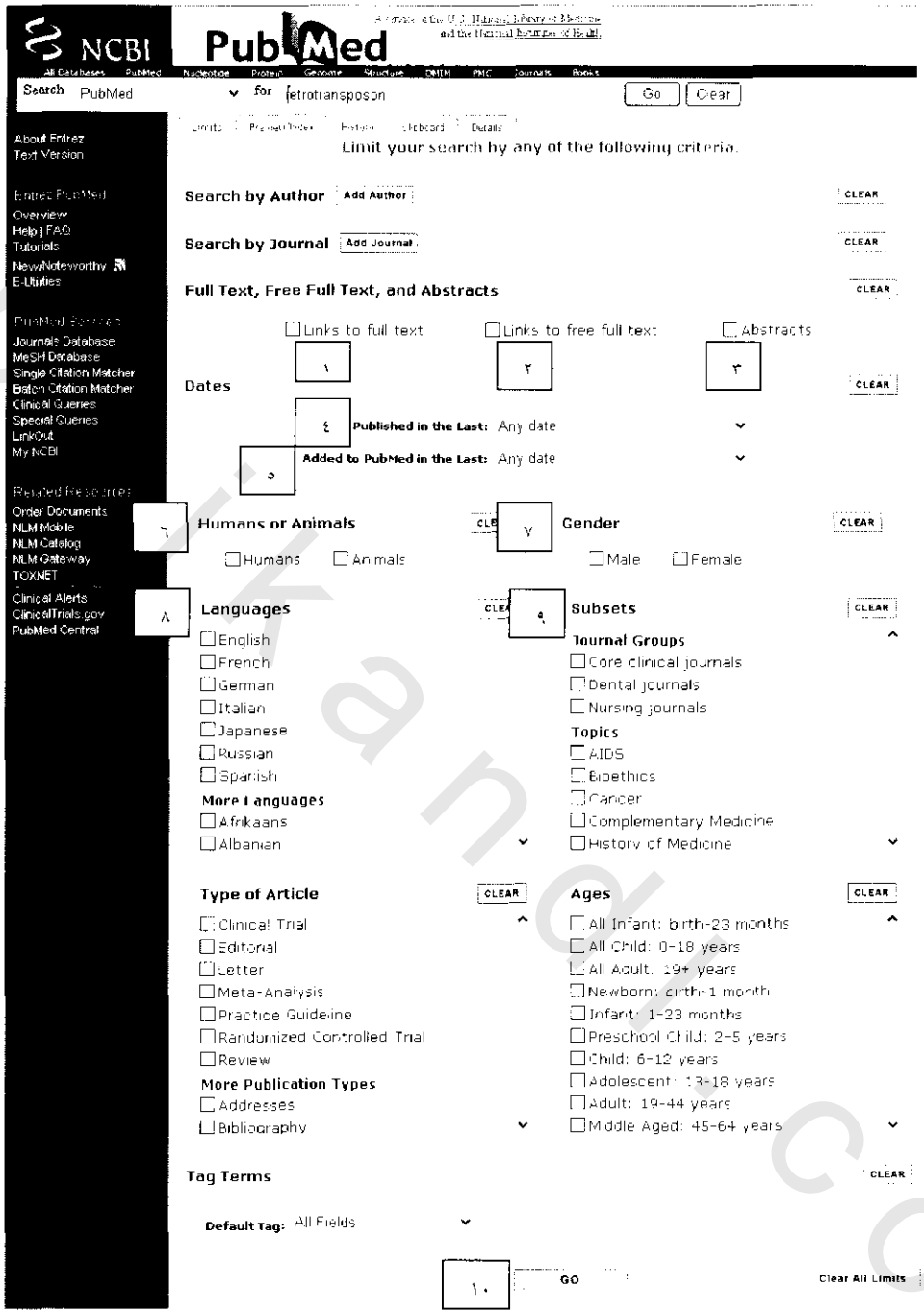
٧- كما يمكن تخصيص الجنس ذكر أم أنثى "Male or Female".

٨- "Languages "اللغة ومجال المجلة العلمية التي تم النشر فيها وكذلك نوع البحث والعمر.

٩- ثم تحديد حقل قواعد البيانات أي مكان البحث.

١٠- ثم الضغط على كلمة Go كما فى الشكل ٢٩.

obeikandl.com



شكل ٢٩: يوضح كيفية أستخدام الأختيارات المختلفة لمحددات البحث.

هناك بعض النصائح الإضافية عند البحث فى قاعدة أبحاث الـ PubMed وهي:

- عند البحث عن كلمتين أوأكثر معا مثل "Down Syndrome " يجب أن نضعهما بين علامات التنصيص حتى يتم البحث على هاتين الكلمتين معا ككلمة واحدة.
- الروابط المنطقية مثل AND أو OR أوNOT لابد من كتابتها بحروف Capital وهذه يعني أنها روابط وليس كلمة مكونة من هذه الحروف.
- مثلا عند كتابة dulpase [TI] Pyrophosphatase [TI] NOT Smith [AU] ومعناها ابحث في كلمة dutpase في العنوان أو pyrophosphate في العنوان فيما عدا الأبحاث الخاصة بالعالم Smith.
- إضافة حروف Capital في نهاية الاسم معناها أنه اسم مختصر مثل "Abergel C"
- يمكن البحث عن اى مقال علمى عن طريق رقم الـ PMID أي كل بحث له رقم خاص به والذى يوجد بجانب البحث.

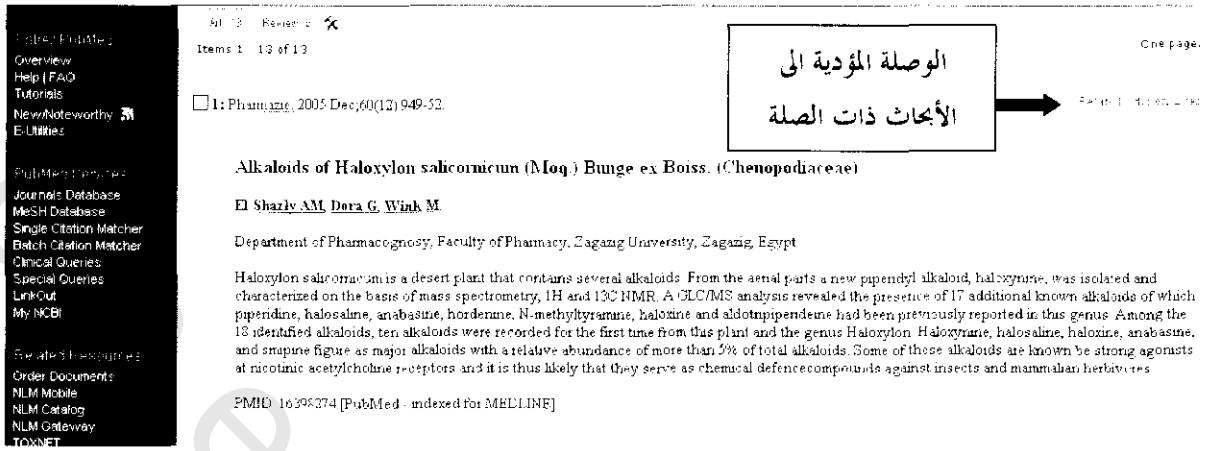
ملحوظة هامة جداً: في حالة استخدام أختيارات محددات البحث المتقدم (Limits) ، يجب بعد الانتهاء إزالتها وألا أنها سوف تظل سارية في كل الأبحاث القادمة.

أهم الأسباب التي تؤدي إلى فشل عملية البحث:

- أخطاء في الهجاء.
- استخدام محددات البحث المتقدم (Limits) غير السليم.
- استخدام اختصار خاطئ.

نصائح هامة

- قراءة أكثر من ملخص مختصر من الأبحاث في مجال التخصص الذي تبحث عنه لمعرفة أهم الكلمات والمصطلحات التي يتم استخدمها أثناء عملية البحث والتى في حالة عدم معرفتها وأستخدامها سوف نفقد الكثير من هذه الأبحاث.
- أستخدام الوصلة المؤدية الى الأبحاث ذات الصلة والموجودة تحت عنوان (Related Articles) والتى توجد فى أقصى اليمين فى صفحة نتائج البحث شكل ٣٠:



شكل ٣٠: يوضح كيفية أستخدام أستخدام الوصلة المؤدية الى الأبحاث ذات الصلة.

أشياء يمكن أن لا تجدها في قاعدة أبحاث الـ PubMed :

١- أسماء الباحثين بعد الاسم العاشر (رقم عشرة) لا يمكن أن تجدها في الأبحاث التي نشرت قبل سنة ١٩٩٥. وهذه تعتبر مشكلة كبيرة على سبيل المثال عند الرغبة فى معرفة الباحثين الأوائل فى مجال في مجال الجينومة.

٢- لا توجد الأبحاث التي نشرت قبل ١٩٦٥ على الـ PubMed أي لا تستطيع الاعتماد على الـ Pub Med في كتابة مقال تاريخي أو عن تاريخ العلم فى أي مجال.

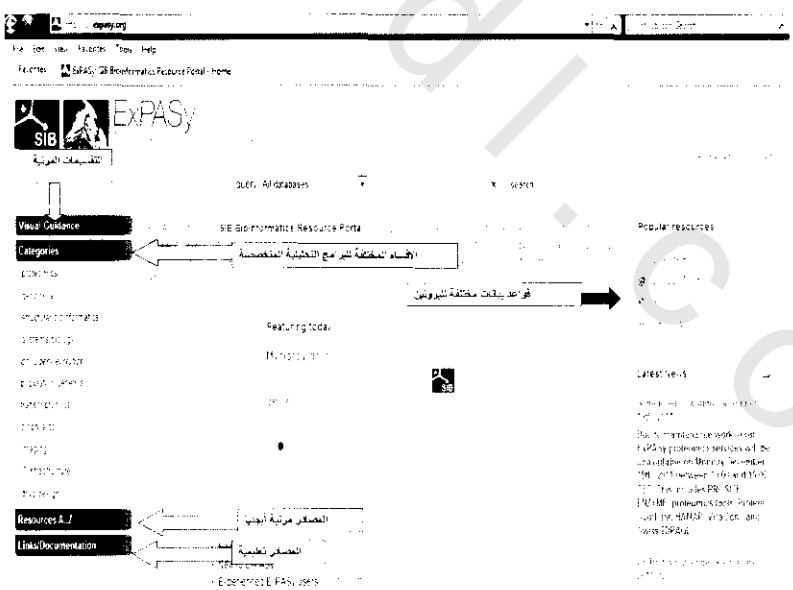
٣- لا توجد ملخصات معظم الأبحاث التي نشرت قبل سنة ١٩٧٦.

موقع الأكسبسي: الموقع المتميز فى المعلومات البروتينية
prime internet site for protein information EXPASY: A

موقع الأكسبسى EXPASY هو قاعدة بيانات بروتينية متميزة تختص بتجميع تراكيب وأشكال ووظائف البروتينات كما يضم مجموعة كبيرة من الروابط لمواقع وبرامج متخصصة فى تحليل التتابعات البروتينية . قد تم إنشاءه وإدارته عن طريق أحد العلماء المشهورين في علم المعلوماتية الحيوية للبروتينات وهو البوفيسور Amos Bairoch. يضم هذا الموقع العديد من الروابط وقواعد بيانات أخرى.

EXPASY اختصارلعبارة Expert Protein Analysis System ويحتوي هذا الموقع المتقدم على قاعدتين هامتان جداً وهاتان القاعدتان هما Swiss port protein و TrEMBI مرتبطتين في قاعدة واحدة تسمى Uniport knowledgebase.

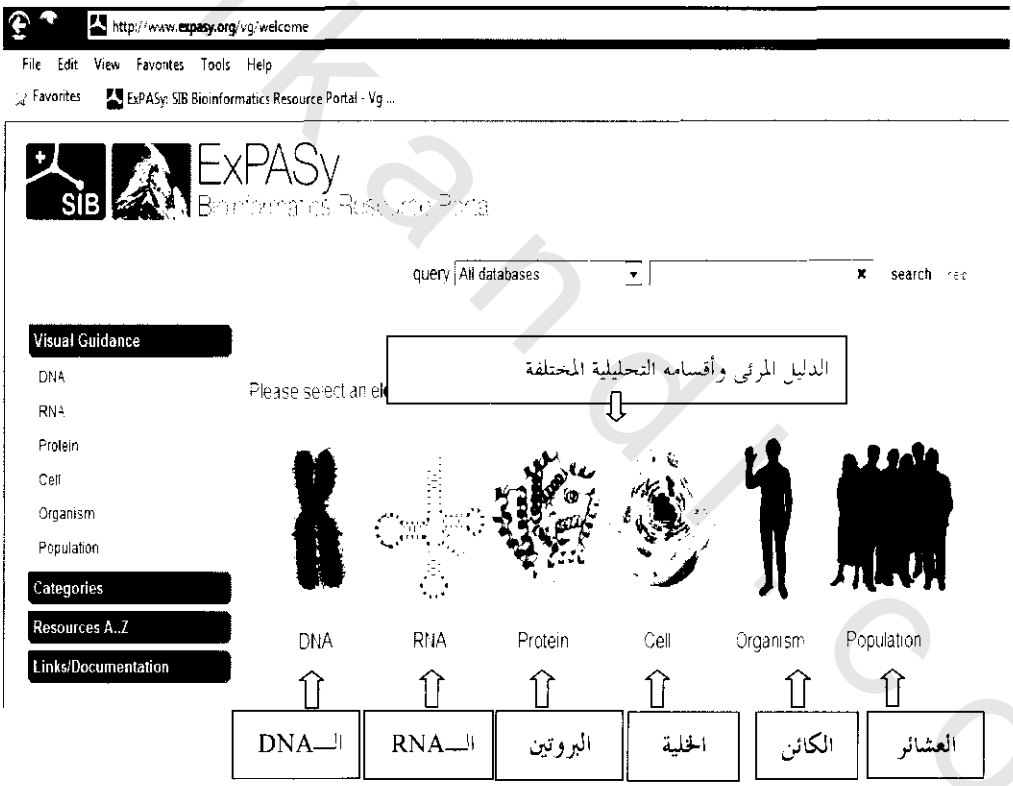
قاعدة الـ Swiss Port تحتوي على التتابعات الحقيقية للبروتينات الناتجة من الأبحاث مع وصف كامل لتركيبها وأجزائها الفعالة والتغيرات التى تحدث فى تركيبها بعد عملية الترجمة وغير ذلك من المعلومات الهامة.أما قاعدة الـ TrEMBL فهى تختص بالتنبؤ بالتتابعات الخاصة بروتين م وبرامج المعلوماتية الحيوية الخاصة بتحليل هذه التتابعات البروتينية شكل ٣١.



شكل ٣١: الصفحة الرئيسية لموقع لموقع الـ EXPASY. يحتوى الموقع على وصلات لقواعد بيانات Uniport knowledgebase ناحية اليمين وعدد كبير لبرامج تحليلية أخرى ناحية اليسار مقسمة الى عدة أقسام.

ويحتوى الموقع على عدة أقسام مختلفة ويحتوى كل منها على برامج متخصصة فى الفروع المختلفة من المعلوماتية الحيوية و تضم الأدوات والبرامج الخاصة بالتحليل والتى تبدأ بقسم الدليل المرئى وتنتهى بالقسم الخاص المحتوى على وصلات تعليمية وخدمات Links/Documentation.

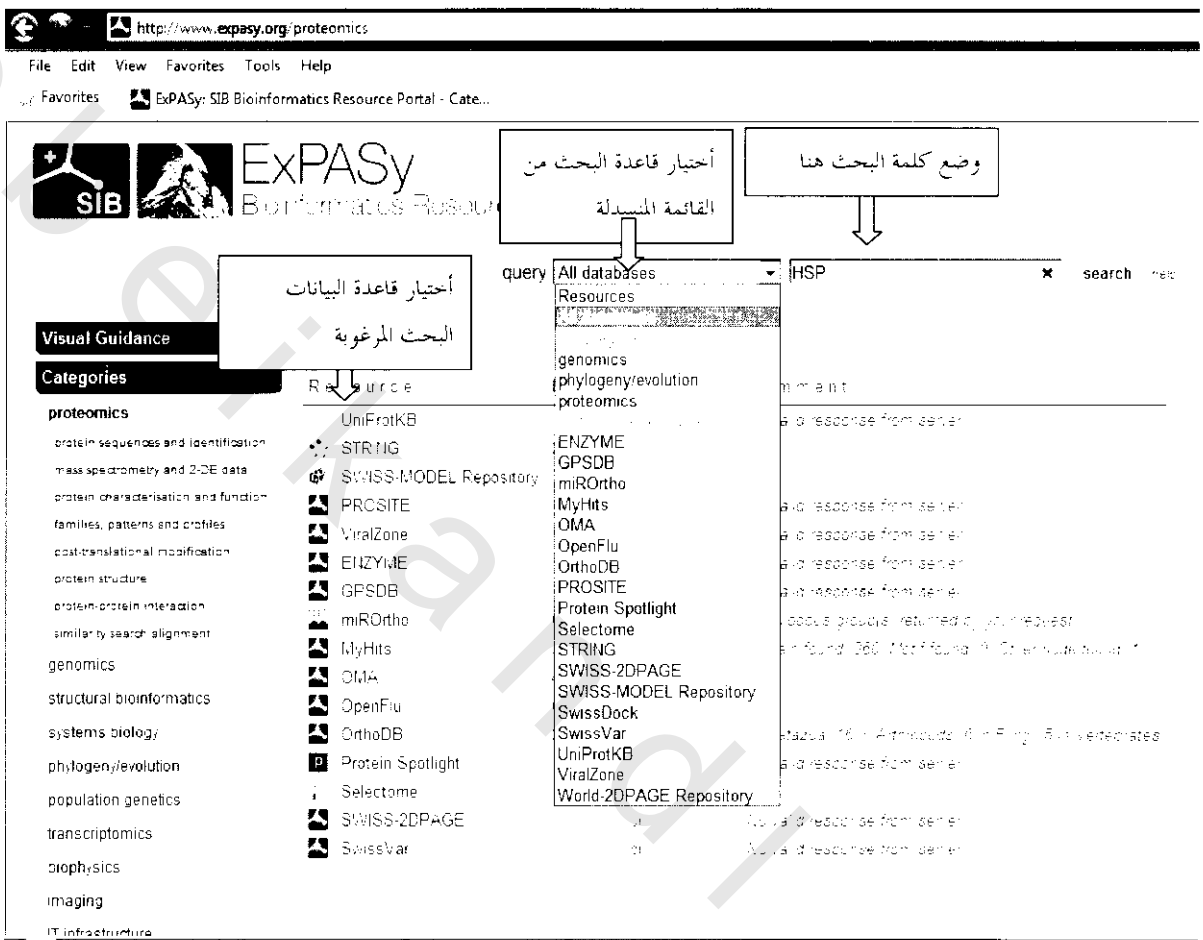
عند الدخول على أول الوصلات والتى تسمى (Visual Guidance) الموجودة فى القائمة اليسرى سوف تظهر مجموعة من الأقسام التى تشمل أقسام الـDNA والـRNA و الـبروتين و الـخلية و الـكائن و الـعشائر. كل قسم منهم يحتوى على مجموعة من البرامج التحليلية وقواعد البيانات الخاصة بكل قسم على حده كما يظهر فى شكل ٣٢.



شكل ٣٢ : الصفحة الخاصة بالدليل المرئى لقواعد بيانات وتحتوى على عدد كبير من البرامج التحليلية مقسمة الى عدة أقسام وهى قسم الـDNA و الـRNA و الـبروتين و الـخلية و الـكائن و الـعشائر.

فمثلا عند البحث عن بروتين الصدمة الحرارية والمسمى HSP نكتب كلمة البحث فى المكان المخصص للبحث والموجود ناحية اليمين ثم الضغط على كلمة البحث Search . سوف يقوم محرك البحث الخاص بالموقع على البحث فى جميع

قوائم البيانات المدرجة وسوف يظهر لنا ملخص نتيجة البحث كما فى الشكل ٣٠. يجب الأشارة الا أن كل قاعدة بيانات تعطى معلومات بيولوجيه أضافية هامة مرتبطة بكلمة البحث لذلك يجب على البحث التعرف عليها وأختيار ماينا سبه.

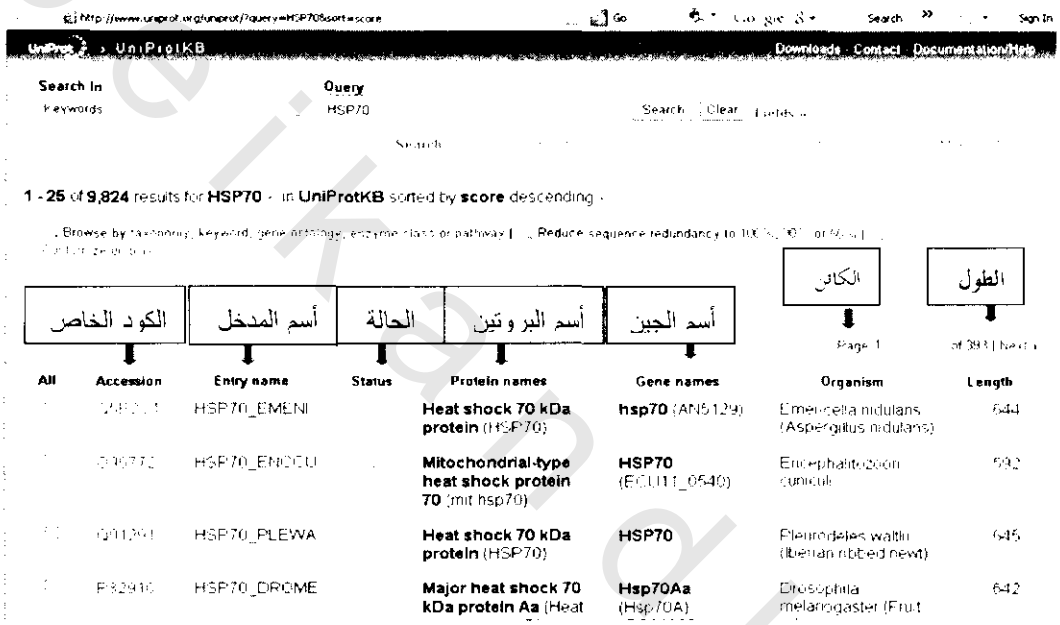


شكل ٣٣: عند البحث عن بروتين الصدمة الحرارية والمسمى HSP فى جميع قوائم البيانات المدرجة.

عند الدخول على قاعدة البيانات المعروفة بأسم UniportKB للبحث على بروتين HSP70 سوف تظهر النتائج كما فى جدول الشكل ٣٤. العمود الأول يحتوى على الكود الخاص(Accession) بأسم البروتين فى قاعدة البيانات. أما العمود الثانى يحتوى الأسم الخاص بأسم المدخل فى قاعدة البيانات أما بقية الأعمدة فهى كالتالى:

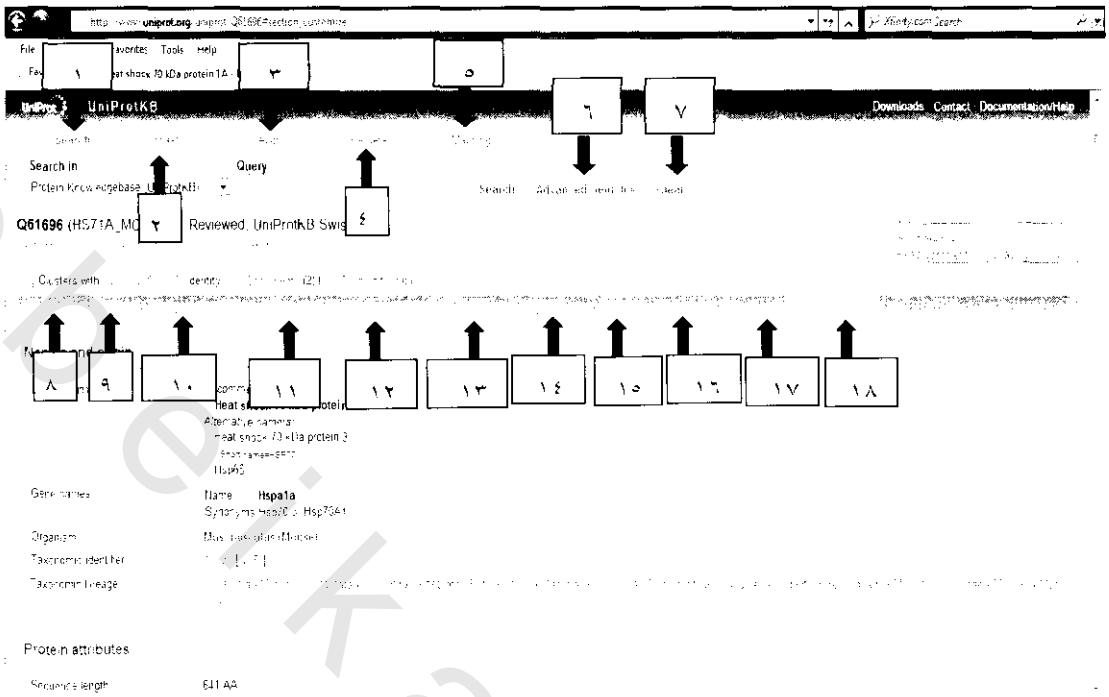
- عمود الحالة Status احتوائها على نجمة صفراء أي أن تتابع هذه البروتينات يوجد في قاعدتي البيانات , Swiss-protein و Tr EMBL وتمت مراجعة هذا البروتين.

- عمود البروتين Protein يحتوى على اسم البروتين (HSP 70) ويجب ملاحظة كيفية كتابة الأسم سواء كان بأحرف كبيرة أو بأحرف صغيرة.
- عمود أسم الجين Gene name يحتوى على اسم الجين (ANS129 (hsp 70.
- عمود أسم الكائن Organism يحتوى على أسم الكائن الذى تم أستخلاص البروتين منه.
- عمود الطول Length يحتوى على طول البروتين "أي كم عدد الأحماض الأمينية التى تكون البروتين.



شكل ٣٤ : عند البحث عن بروتين الصدمة الحرارية والمسمى HSP70 فى جميع قوائم البيانات المدرجة.

عند الضغط على الدخول على العمود الأول الذى يحتوى على الكود الخاص(Accession) بالبروتين سوف يتم فتح صفحة جديدة تحتوى على كثير من المعلومات الخاصة بهذا البروتين.



شكل ٣٥ : نتيجة البحث عن بروتين الصدمة الحرارية (hsp70) ANS129 .

ويمكن تلخيص نتائج الصفحة التى تظهر فى شكل ٣٥ كما يلى:

١. الزر(Search) الخاص بمحرك البحث داخل الصفحة للبحث عن أي بروتين أوجين.

٢. الزر(Blast) الخاص بعمل مقارنة لتتابعات البروتين المراد دراسة مع مجموعة البروتينات الموجودة فى قاعدة البيانات.

٣. الزر (Align)الخاص بقارنة التتابع المراد دراسة مع بروتين أخر أومجموعة بروتينات أخرى مختارة.

٤. الزر (Retrieve) الخاص بالحصول على التتابع المرغوب عن طريق كتابة الرقم الكودى.

٥. معرفة مكان التتابع المرغوب فى قاعدة البيانات (ID Mapping).

٦. خيارات البحث المتقدم (Advanced Search).

٧. زر(Clear) خاص بمسح محتوى البحث لعمل بحث أخر.

٨. زر الأسماء (Names)المعلومات الأولية الخاصة بأسم البروتين والجين المسؤول عنه والرقم التقسيمى وأسم الكائن الذى تم عزل البروتين منه

٩. زر (Attributes) خاص بأعطاء معلومات عن طول التتابع وحالتة (كامل أوغيركامل) ومعلومات تؤيد وجود هذا البروتين من عدمه.

١٠. زر التعليقات العامة (General Attributes) ويحتوى على وظيفة البروتين والوحدات المكونة للبروتين (Subunits) ومكان التواجد (location) ودرجة التشابه مع مجموعات أوعائلات بروتينية أخرى(Sequence Similarity).

١١. تعريفات وظيفية خاصة بالبروتين (Ontologies) مثل كلمات البحث (Key words) ومعلومات وظيفية للجين(GO)(Gene Ontologies) وتشمل العمليات البيولوجية والمحتوى السيتوبلازمى والوظيفة الجزيئية والمصطلحات التقنية لهذا البروتين.

١٢. خصائص التتابع (Sequence Features) وتشمل شرح تفصيلى لمحتويات التتابع والأماكن الفعالة داخله ووظيفتها.

١٣. التتابع (Sequence) ويحتوى هذا القسم على التتابع الخاص بهذا البروتين.

١٤. قسم المراجع (Refrences) ويحتوى على المراجع الخاصة بالمعلومات الموجودة بالبحث والتى يمكن الرجوع اليها.

١٥. قسم المراجع المرتبطة (Cross-Refrence) ويحتوى على وصلات للتتابع المدروس فى قواعد البيانات المختلفة مثل قواعد بيانات التتابعات أو الأشكال ثلاثية الأبعاد أو البروتوكولات أوغيرها من قواعد البيانات الهامة لتركيب أو وظيفة البروتين.

١٦. قسم معلومات المدخل (Entry Information) ويحتوى على أسم وحالة وتاريخ المعلومات المدخلة للبروتين الذى تم البحث عنه.

١٧. قسم الوثائق (Relevant Documents) ويحتوى على كل الوثائق ذات الصلة بالبروتين المطلوب.

١٨. زر تنسيق الترتيب (Customize order) لأمكانية تعديل ترتيب الأقسام السابقة.

تطبيق عملى على كيفية دراسة بروتين معين بأستخدام قاعدة البيانات UniportKB

١. الدخول على موقع http://www.uniprot.org/ وكتابة أسم البروتين المطلوب دراسته بعد الأطلاع على المراجع والأبحاث فى الـPubMed (شكل ٣٦).

٢. تظهر النتيجه بوجود ١٠ تتابعات منهم ٩ غير مراجعين (Unreviewed) وواحد فقط تمت مراجعته (Reviewed). (شكل ٣٧).

٣. يتم أختيار التتابع المراجع (ذو النجمة الصفراء) بالضغط عليه فتظهر لنا جميع البيانات والمعلومات الخاصة بهذا البروتين(شكل ٣٨).

٤. نضغط على المعلومات الخاصة بالتتابع لرؤية وحفظ التتابع المرغوب فى ملف بصيغة الـFASTA عن طريق الضغط على كلمة FASTA على اليمين(شكل ٣٩).

٥. ثم نضغط على الزر (Blast) فى أعلى الصفحة والذى سوف يقوم بمقارنة التتابع المدروس مع جميع تتابعات البروتينات الموجودة فى قاعدة البيانات لأيجاد الأكثر تشابها(شكل ٤٠).

٦. تظهر النتيجة بظهور عدد من التتابعات المتشابه مع التتابع المدروس فى التركيب الأولى والتى يتم أختيار بعض منها للدراسة (شكل ٤١).

٧. تظهر المقارنة فى شكل توازى (Alignment) أوشكل شجرة قرابه (Tree) مع أمكانية تلوين بعض نقاط التشابه أو الأختلاف فى التتابعات المدروسة من الناحية الكيميائية والفزيائية(شكل ٤٢).

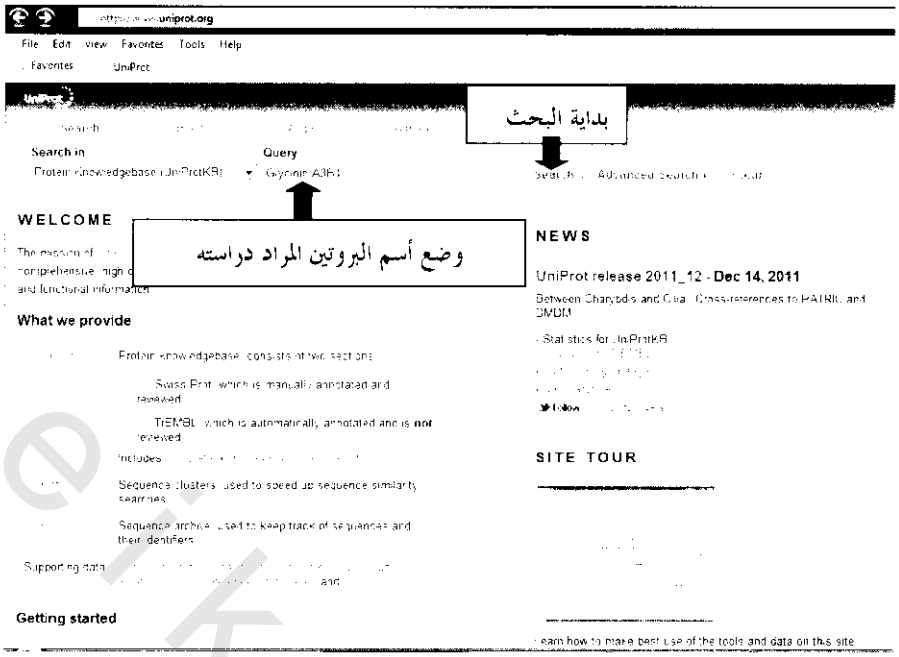
٨. يمكن حفظ ملف توازى التتابعات فى أكثر من صورة لأجراء تعديلات علية أو تحليله بواسطة برنامج أحر مثل Text, Tree, Fasta, Jalview (شكل ٣٩).

٩. يمكن الحصول وحفظ التتابعات المدروسة فى أى صورة عن طريق الأمر (Retrieve) (شكل ٤٣).

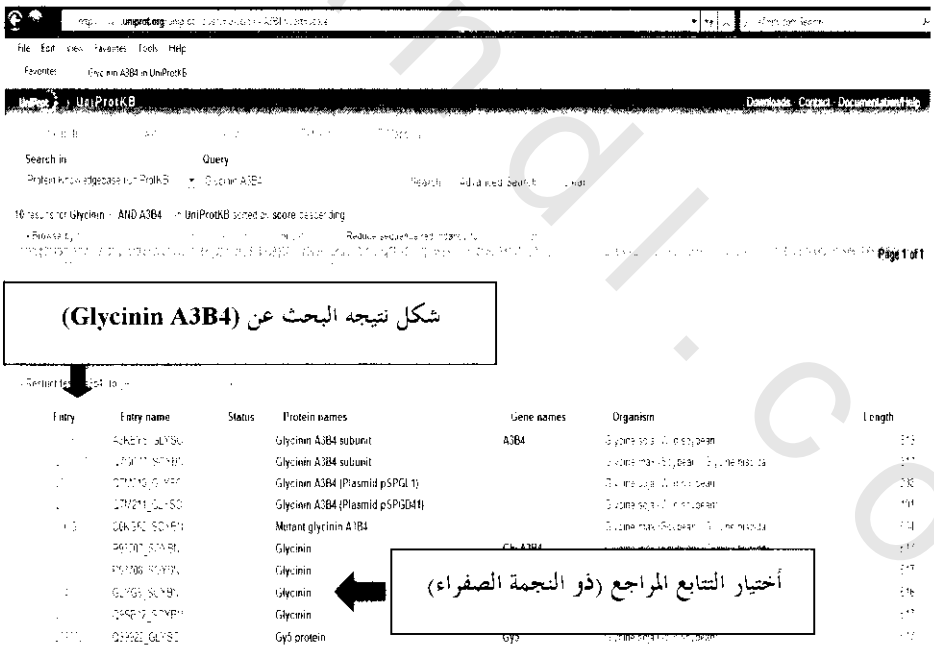
١٠. يمكن الحصول على معلومات أضافية من قواعد البيانات الأخرى عن طريق زر ID Mapping (شكل ٤٤).

١١. بالضغط على زر Swap يتم الأنتقال الى قاعدة بيانات أخرى تحتوى على معلومات أضافية عن البروتين المطلوب (شكل ٤٤).

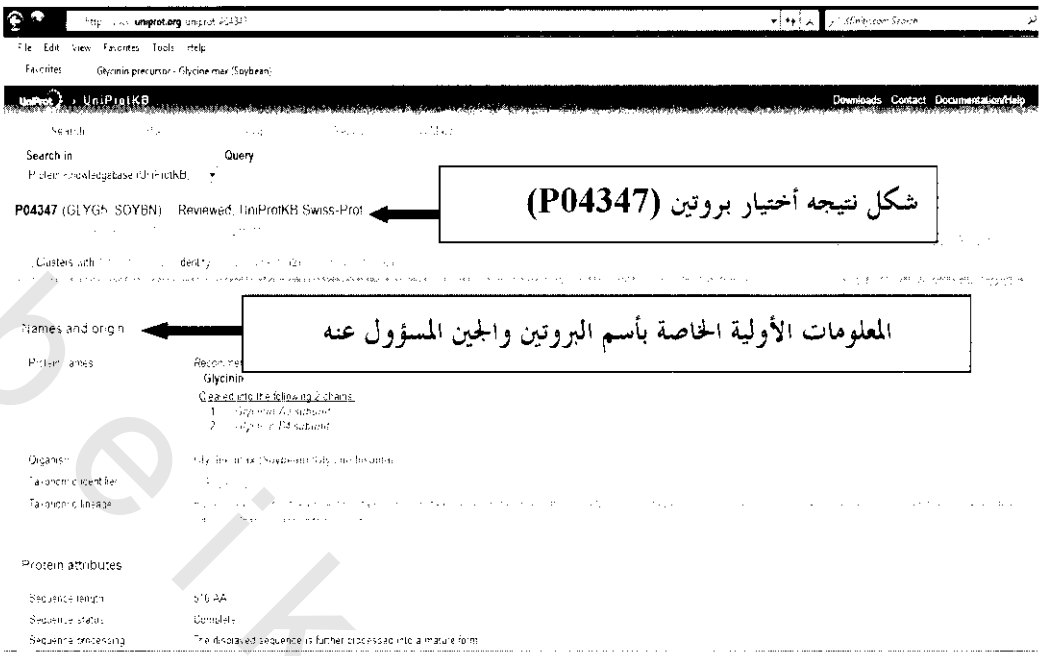
١٢. الأنتقال الى قاعدة الـ PDBe على سبيل المثال للبحث عن معلومات جديدة عن البروتين المرغوب (شكل ٤٥).



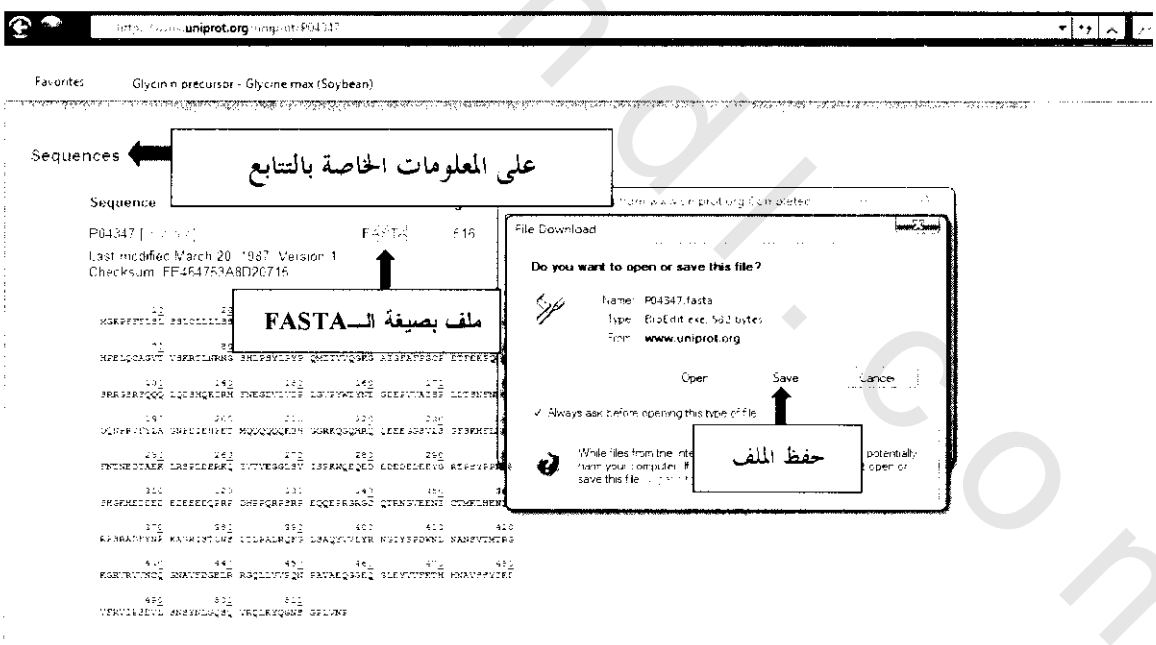
شكل ٣٦:الدخول على موقع http://www.uniprot.org/ وكتابة أسم البروتين المراد دراسة.



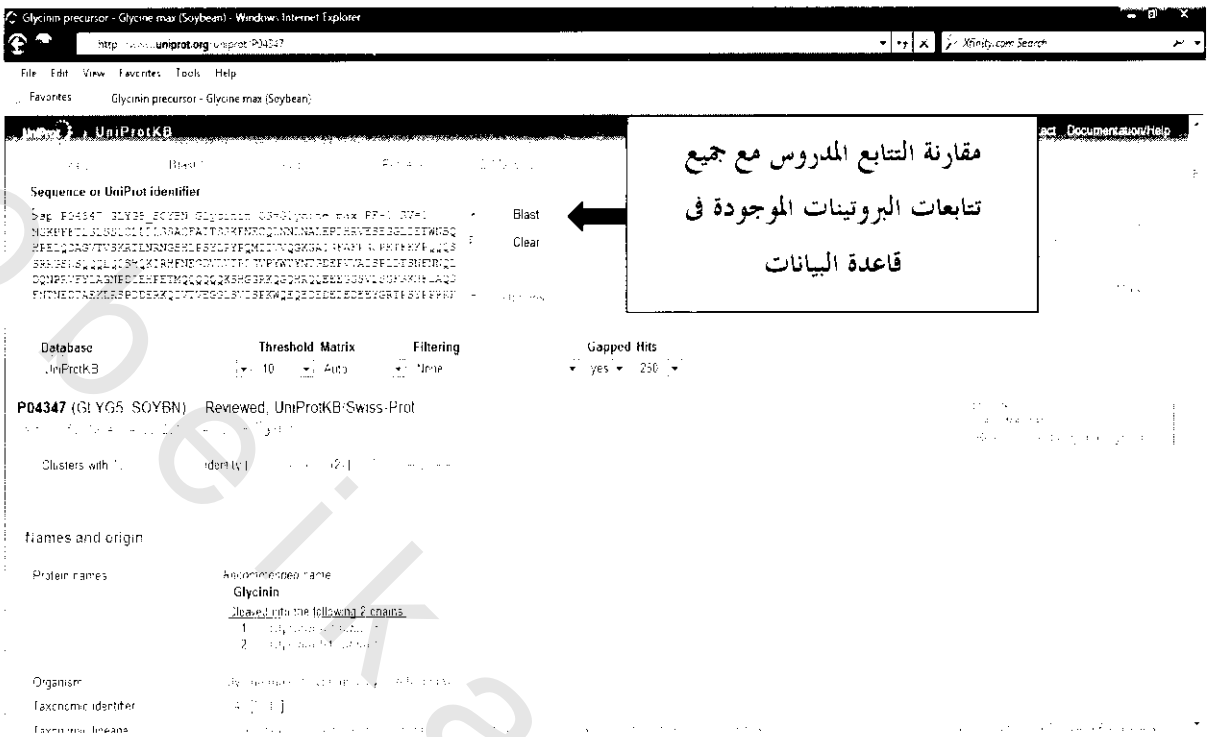
شكل ٣٧: شكل نتيجه البحث عن بروتين (Glycinin A3B4) والتى أسفرت عن وجود ١٠ تتابعات منهم ٩ غير مراجعين (Unreviewed) وواحد فقط تمت مراجعته (Reviewed)واحد فقط تمت مراجعته (Reviewed)



شكل ٣٨: شكل نتيجه أختيار بروتين (P04347) واحد فقط تمت مراجعته (Reviewed) والمعلومات الخاصة بهذا البروتين

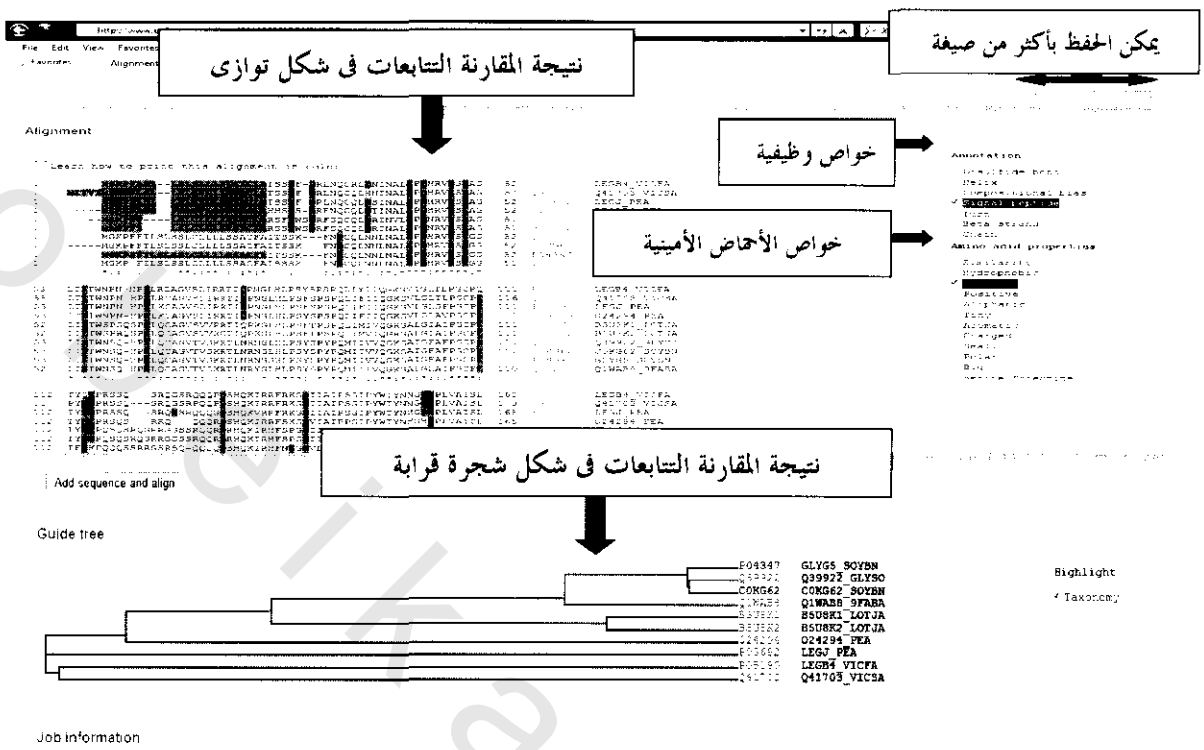


شكل ٣٩: على المعلومات الخاصة بالتتابع (Sequence) لرؤية وحفظ التتابع المرغوب فى ملف بصيغة الـFASTA

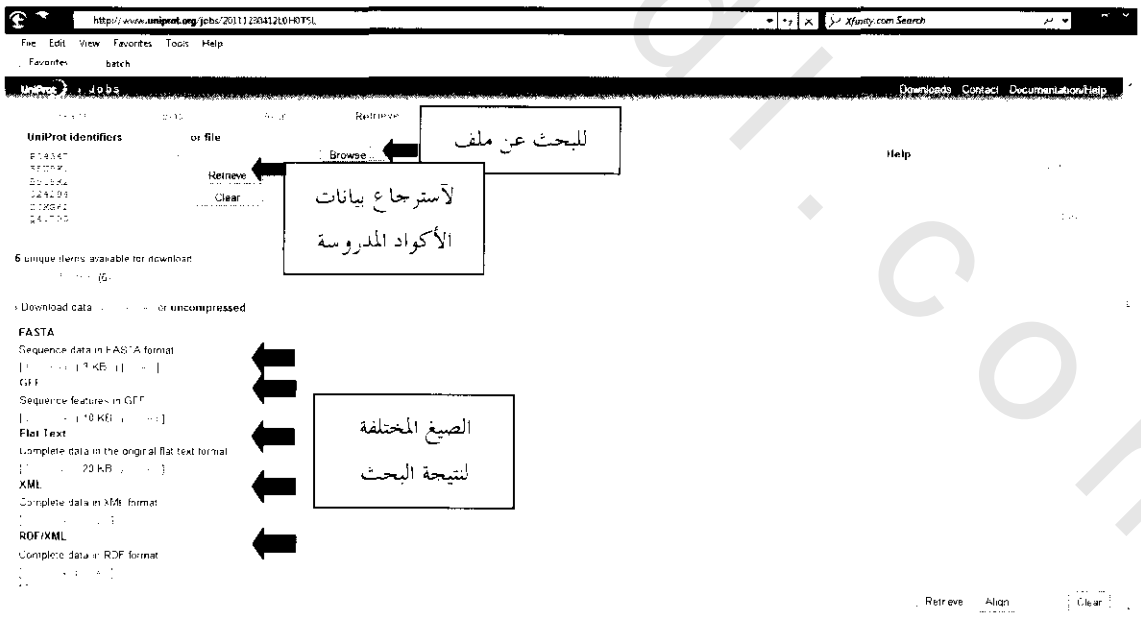


شكل ٤٠ : مقارنة التتابع المدروس مع جميع تتابعات البروتينات الموجودة فى قاعدة البيانات عن طريق الضغط على الزر (Blast) فى أعلى الصفحة.

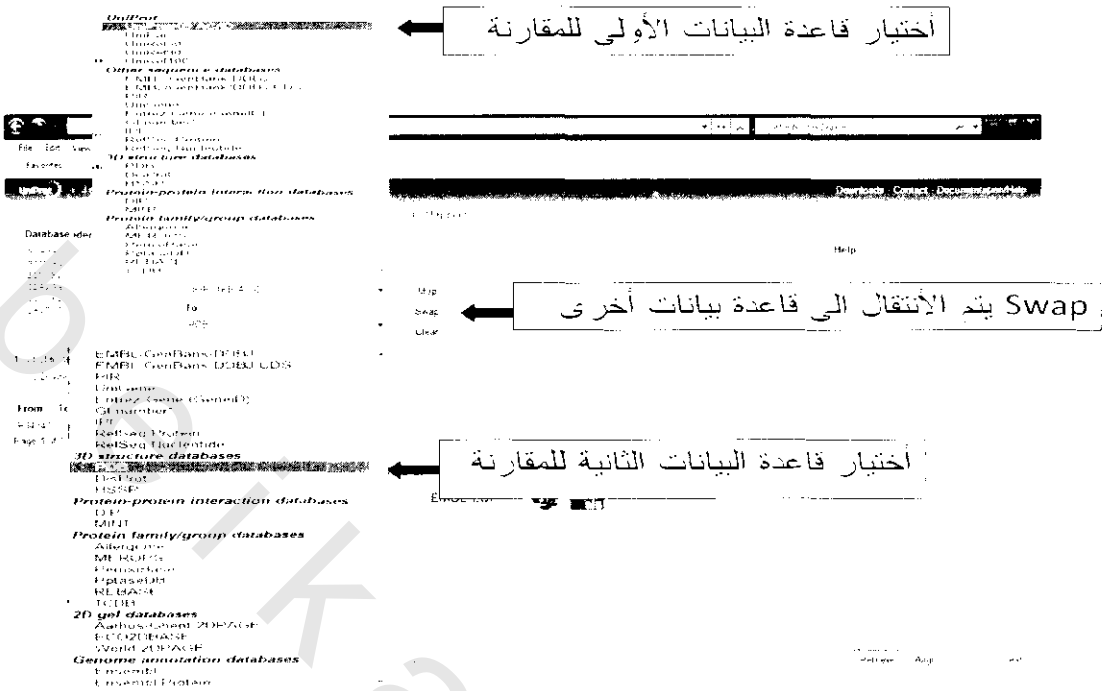
شكل ٤١ : ظهور مجموعة من التتابعات المشابهة للتتابع المدروس مع جميع تتابعات البروتينات الموجودة فى قاعدة البيانات.



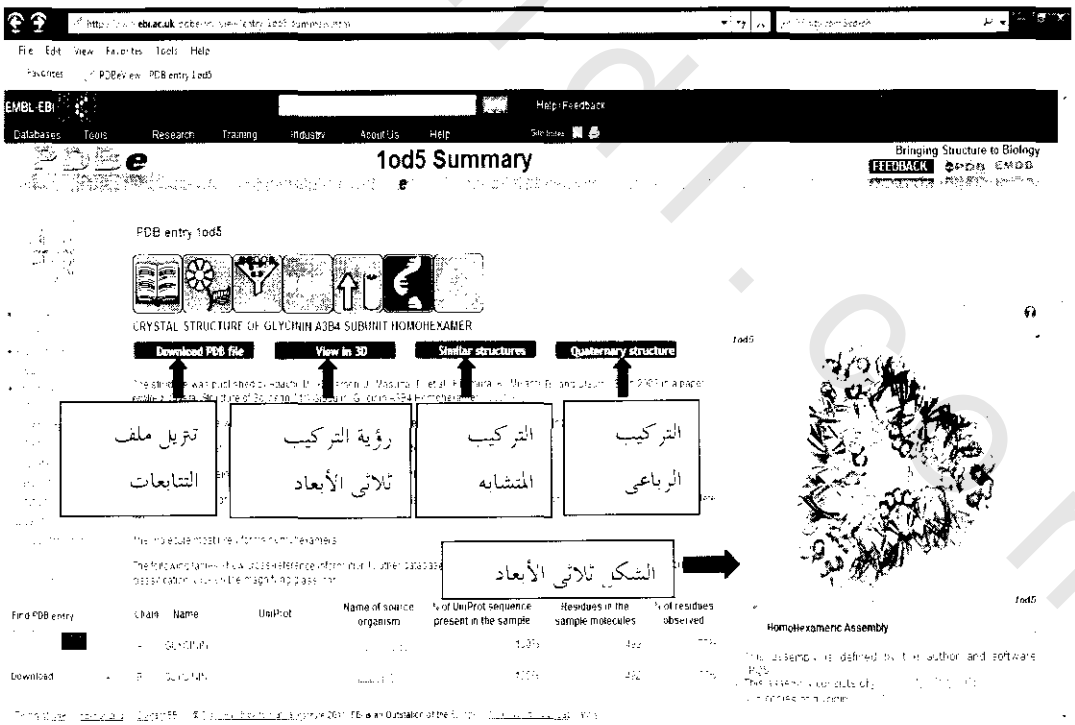
شكل ٤٢ : تظهرنتيجة المقارنة فى شكل توازى (Alignment) أوشكل شجرة قرابه (Tree) مع أمكانية تلوين بعض نقاط التشابه أو الأختلاف فى التتابعات المدروسة من الناحية الكيميائية والفزيائية.



شكل ٤٣ : يمكن الحصول وحفظ التتابعات المدروسة فى أى صورة عن طريق الأمر (Retrieve).



شكل ٤٤ : الحصول على معلومات أضافية من قواعد البيانات الأخرى عن طريق زر ID Mapping



شكل ٤٥ : الحصول على معلومات جديدة عن البروتين المرغوب من قاعدة الـ PDB والتى تحتوى على الشكل ثلاثى الأبعاد.

تحليل تتابعات المادة الوراثية الـ DNA

أولا: الحصول على تتابعات DNA

أشياء يجب تذكرها عن التتابعات البروتينية:

- التتابعات البروتينية هي تتابعات بسيطة يتراوح طولها فى المتوسط ٤٠٠ حمض أميني وقد تزيد أو تقل عن ذلك بحوالى ٢٠٠ حمض. ويستثنى من ذلك بعض التتابعات البروتينية الكبيرة.
- البروتينات الميكروبية أو حقيقيات النواة (النباتات أو الحيوانات) لها نفس الخصائص تقريباً.

أن تركيب الجينات المسئولة عن أنتاج البروتينات تكون أكثر تعقيداً في الحيوانات الراقية حيث تختلف أطوال الجينات في الإنسان عنها في الميكروبات (طولها في الميكروبات بضع مئات أوالاف من القواعد بينما في الأنسان قد تتعدى عشرات أومئات الآلاف من القواعد).

ملحوظة: ليس كل تتابعات الـ DNA ينتج عنها بروتين وداخل الجين نفسه هناك مناطق كثيرة عديمة الفائدة تسمى (junk DNA) ولا تنتج أي بروتين.

وتجدر الأشارة الى أن هناك ٣ أنواع من تتابعات الـ DNA تدخل في تركيب الجين:

١. مناطق مشفرة للبروتينات (The protein coding region)
٢. مناطق منظمة لعمل الجينات في العادة هي مناطق تسبق المناطق المشفرة للبروتينات(Requlatory region) .
٣. مناطق غير مشفرة للبروتينات(Untranslated regions) وهى مناطق لاتنتج بروتينات وتسمى (INTRONS) وهي توجد قبل أوبعد المناطق المشفرة للبروتين (EXONS) . وبناءاً على ذلك فالتعامل مع تتابعات الـ DNA تكون أكثر تعقيدا من التعامل مع تتابعات البروتين.

صيغ (Format) تتابعات الـDNA

هناك أكثر من صيغة أوطريقة تكتب بها تتابعات الـDNA أوالبروتين وتختلف البرامج المختلفة فى تقبلها لنوع الصيغة المستخدمة. وهناك نوعين مشهورين من هذه الصيغ وهى صيغة الـ FASTAو RAW.

س- ما الفرق بين FASTA format و RAW؟

- الـ FASTA format يبدأ بعلامة > ثم يلى ذلك فى نفس السطرأسم أو تعريف عن التتابع ثم يلى ذلك التتابع نفسة بدون أى فواصل أو مسافات أوحروف غريبة وأيضاً هناك FASTA للبروتين وهناك أخرى للجين (شكل ٤٦).

```
>gi|184402|gb|M64673.1|HUMHSF1 Human heat shock factor 1 (TCF5) mRNA, complete cds
CGGGCCCGTTGCAAGATGGCGGCGGCCATGCTGGGCCCCGGGGCTGTGTGTGCGCAGCGGGCGGCGGCGC
GGCCCGGAAGGCTGGCGCGGCGACGGCGTTAGCCCGGCCCTCGGCCCCTCTTTGCGGCCGCTCCCTCCGC
CTATTCCCTCCTTGCTCGAGATGGATCTGCCCGTGGGCCCCGGCGCGGCGGGGCCCAGCAACGTCCCGGC
CTTCCTGACCAAGCTGTGGACCCTCGTGAGCGACCCGGACACCGACGCGCTCATCTGCTGGAGCCCGAGC
GGGAACAGCTTCCACGTGTTCGACCAGGGCCAGTTTGCCAAGGAGGTGCTGCCCAAGTACTTCAAGCACA
ACAACATGGCCAGCTTCGTGCGGCAGCTCAACATGTATGGCTTCCGGAAAGTGGTCCACATCGAGCAGGG
CGGCCTGGTCAAGCCAGAGAGAGACGACACGGAGTTCCAGCACCCATGCTTCCTGCGTGGCCAGGAGCAG
CTCCTTGAGAACATCAAGAGGAAAGTGACCAGTGTGTCCACCCTGAAGAGTGAAGACATAAAGATCCGCC
AGGACAGCGTCACCAAGCTGCTGACGGACGTGCAGCTGATGAAGGGGAAGCAGGAGTGCATGGACTCCAA
GCTCCTGGCCATGAAGCATGAGAATGAGGCTCTGTGGCGGGAGGTGGCCAGCCTTCGGCAGAAGCATGCC
CAGCAACAGAAAGTCGTCAACAAGCTCATTCAGTTCCTGATCTCACTGGTGCAGTCAAACCGGATCCTGG
GGGTGAAGAGAAAGATCCCCCTGATGCTGAACGACAGTGGCTCAGCACATTCCATGCCCAAGTATAGCCG
GCAGTTCTCCCTGGAGCACGTCCACGGCTCGGGCCCCTACTCGGCCCCCTCCCCAGCCTACAGCAGCTCC
AGCCTCTACGCCCCTGATGCTGTGGCCAGCTCTGGACCCATCATCTCCGACATCACCGAGCTGGCTCCTG
```

شكل ٤٦ : شكل التتابعات بصيغة أوطريقة الـ FASTA format.

- صيغة الـRAW وتعتمد على وضع التتابع نفسة بدون أي علامات ولا أي تعريف ولا فواصل أو مسافات أوحروف غريبة (التتابع فقط) (شكل ٤٧).

```
CGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCTAG
AGGCCCCGGAGGAGCTGGGCCAGCCGCCCACCCCCACCCCCAGTGCAGGGCTGGTCTTGGGGAGGCAGGG
CAGCCTCGCGGTCTTGGGCACTGGTGGGTCGGCCGCCATAGCCCCAGTAGGACAAACGGGCTCGGGTCTG
GGCAGCACCTCTGGTCAGGAGGGTCACCCTGGCCTGCCAGTCTGCCTTCCCCCAACCCCGTGTCCTGTGG
TTTGGTTGGGGCTTCACAGCCACACCTGGACTGACCCTGCAGGTTGTTCATAGTCAGAATTGTATTTTGG
ATTTTTACACAACTGTCCCGTTCCCCGCTCCACAGAGATACACAGATATATACACACAGTGGATGGACGG
ACAAGACAGGCAGAGATCTATAAACAGACAGGCTCTAAAAAAAAAAAAAAAAAAAAA
```

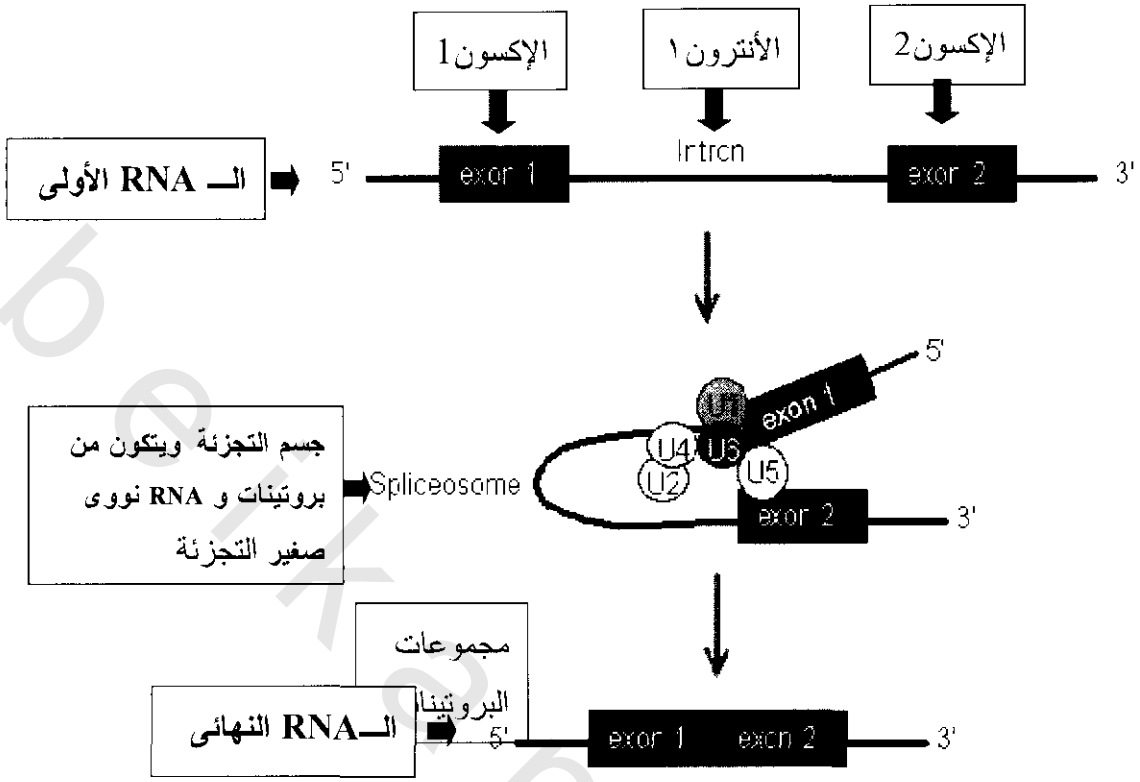
شكل ٤٧ : شكل التتابعات بصيغة أوطريقة الـ RAW.

ملحوظة: يجب الأخذ في الاعتبار الأبتعاد عن الأحرف الأخرى الغريبة (Parasite characters) عند تحميل ملفات التتابع الـ DNA .

مصطلحات هامه:

الـ RNA الأولى (PreMature RNA) : هو عبارة الـ RNA الناتج من نسخ الجين بالكامل من البداية حتى النهاية ولذلك فهو يشمل الإكسونات والإنترونات.

جزيء الـRNA النهائى (Mature RNA): ينتج من RNA الأولى بعد إزالة مناطق الإنترونات في عملية تسمى (RNA splicing). مجموعات البروتينات المستخدمة فى هذه العملية تكون مايسمى بجسم التجزئة (Spliesomes)

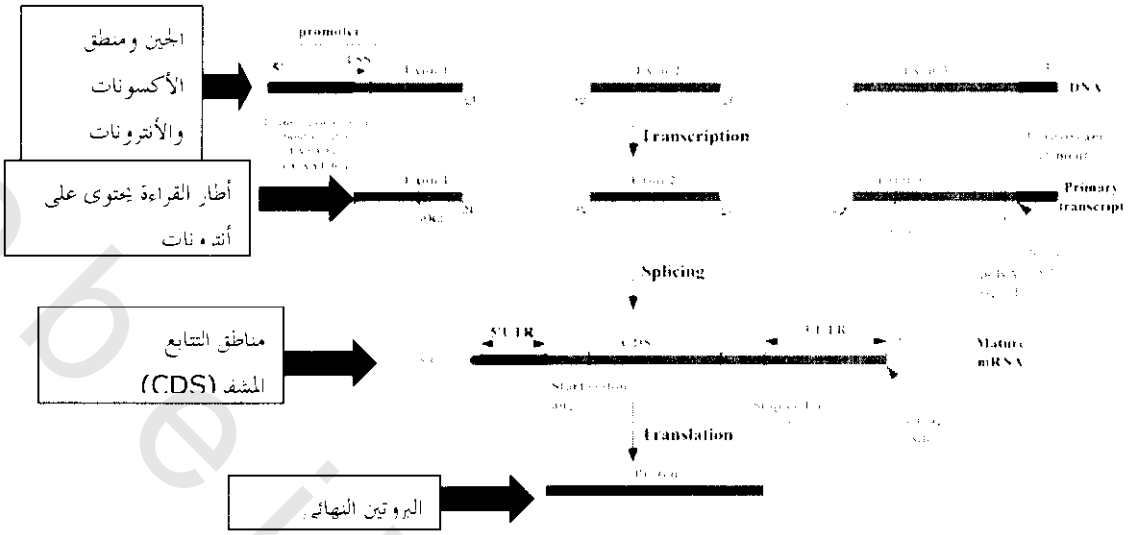


شكل ٤٨ : شكل تخطيطى يوضح إزالة مناطق الإنترونات للحصول على جزيء الـRNA النهائى في عملية تسمى (RNA splicing).

ماهو التتابع المشفر(Coding Sequence(CDS ؟

وكيف يختلف أطار القراءة(Open reading frame (ORF؟

التتابع المشفر هو المنطقة الحقيقية من الـDNA التى تترجم لتكون البروتينات. يحتوى أطار القراءة على أنترونات فى حين أن التتابع المشفر يحتوى فقط على مناطق الأكسونات المشفرة فقط مجمعة بشكل ممتالى ويمكن تميزها الى شفرات ثلاثية لأحماض أمينية فى الريبوسوم. فى أوليات النواه مناطق التتابع المشفر(CDS) ومناطق أطار القراءة (ORF) متشابه لعدم وجود أنترونات. شكل ٤٩



شكل ٤٩ : شكل تخطيطى يوضح الفرق بين التتابع المشفر وأطار القراءة

الحصول على التتابعات DNA من تتابعات البروتين: لماذا يجب الرجوع للخلف؟

هناك حالات يستوجب فيها البحث أن نحصل فيها على تركيب الـ DNA الخاص بجين ينتج بروتين معين فعلى سبيل المثال:-

عندما نرغب في كلونة الجين المسئول عن تكوين بروتين معين لنقله من كائن إلى كائن حتى يستطيع إنتاج هذا البروتين بشكله الطبيعي أو حتى بشكل محور بعد إحداث طفرة متخصصة به. فى هذه الحالة يجب أن نتبع الخطوات التالية:-

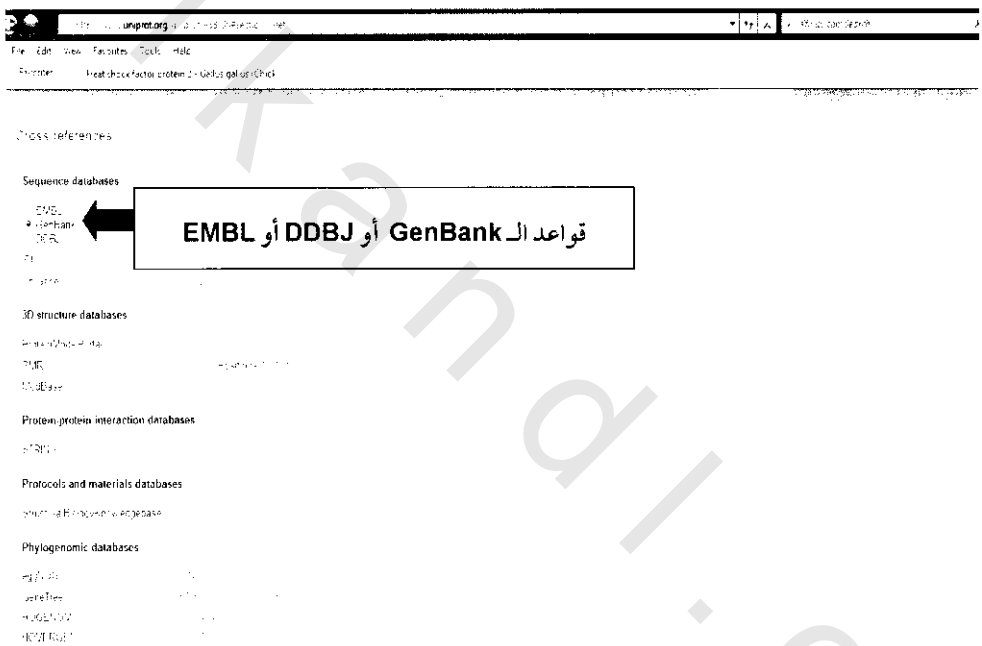
١- الذهاب لموقع EXPASY (http://www.expasy.org/) ويتم البحث عن بروتين معين وليكن (HSF) ونكتب الاسم ثم نضغط Go (شكل ٢٧).

- سوف تظهر لنا قاعدة بيانات UniportKB (القاعدة الخاصة بتتابعات البروتين) والتى تحتوى على مجموعة التتابعات البروتينية المرتبطة بهذا الاسم كل واحدة من هذه التتابعات لها رقم محدد (كود) خاص بها مثل Q00613. كل واحد من هذه الأرقام لا تكرر سوى مرة واحدة كما (شكل ٣٦-٣٧).

٢- عند الضغط على الرقم الخاص بالبروتين تظهر لنا المعلومات المطلوبة عن هذا البروتين ثم نضغط على Cross-refrence لكي نحصل على الجين المسئول عنه. وتظهر التتابعات الموازية لهذا البروتين في GenBank أو DDBJ أو EMBL(شكل ٥٠).

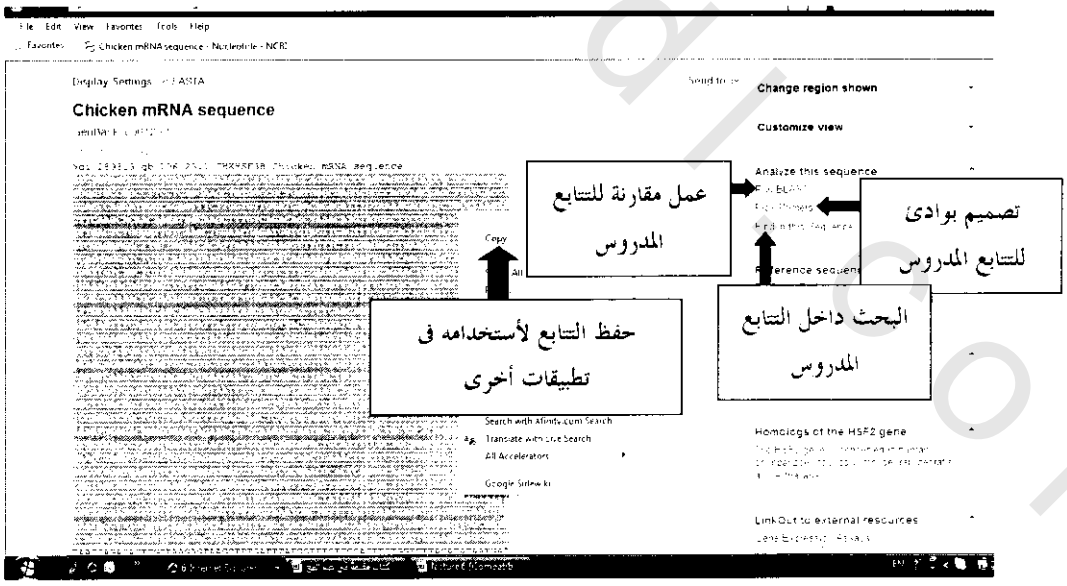
٣- عند اختيار GenBank تظهر لنا معلومات كثيرة عن هذا الجين من قاعدة الـ NCBI ثم في النهاية يظهر لنا التتابع البروتيني الخاص بهذا الجين ثم التتابع الجيني في NCBI فى شكلين هما شكل تخطيطى(Graphics)و شكل صيغة FASTA (شكل ٥١).

٤- ثم نقوم بعمل نسخ لهذا التتابع وحفظه بالطريقة العادية المعروفة وذلك لاستخدامها مرة وتحليها فى تطبيقات أخري مثل التعرف على منطق القصر أو تصميم بوادئ ت Primers أو لعزل الجين أو مقارنة التتابع مع تتابعات أخرى لمعرفة صلة القرابة.



شكل ٥٠: شكل يوضح نتيجة الضغط على Cross-refrence لكي نحصل على الجين المسئول عن البروتين. وتظهر التتابعات الموازية لهذا البروتين في قواعد الـ GenBank أو DDBJ أو EMBL

شكل ٥١: شكل يوضح أختيار الشكل الذى يظهر به التتابع الجينى في الـNCBI ، هما شكل تخطيطى(Graphics) أو شكل صيغة FASTA



شكل ٥٢: بعمل نسخ لهذا التتابع وحفظه بالطريقة العادية المعروفة وذلك لاستخدامها وتحليها فى تطبيقات أخري مثل التعرف على منطق القصر أو تصميم بوادئ ت Primers أو لعزل الجين أو مقارنة التتابع مع تتابعات أخرى لمعرفة صلة القرابة

بعض التطبيقات التى يمكن عملها على تتابع الـجين

أولا:- عمل مقارنة بين التتابع الجينى المدروس والتتابعات الجينات المشابهة على قاعدة البيانات:

يمكن أستخدام أمر BLAST وهو أختصار لعبارة Basic Local Alignment Search Toolوالموجود فى أقصى يمين الصفحة شكل (٥٢). ويستخدم هذا الأمر بكثرة وبكفاءة كبيرة في البحث عن نسبة التشابه أو الأختلاف بين التتابعات المختلفة ذات الصلة بالجين المدروس فى قاعدة البيانات. ويمكن تلخيصها كما يلى:

١- هذه الطريقة تم تصميمها بواسطة العالم Smith Waterman.

٢- تظهر هذه الطريقة أفضل المناطق المتشابهة داخل التتابعات المختلفة.

٣- تعطي المعنوية الإحصائية لدرجات التشابه بين التتابعات المختلفة.

٤- يمكن أن نستخدم هذه الطريقة في البحث عن تتابعات الـ DNA والبروتينات وجميع التوافيق والتباديل الممكنة بينهم مثل:

- مقارنة تتابع الـDNA مع قاعدة التتابعات DNA عن طريق برنامج Blastn.
- مقارنة ترجمة تتابع الـDNA مع تتابعات البروتينات عن طريق برنامج Blastx.
- مقارنة تتابع بروتينى مع قاعدة تتابعات البروتينات عن طريق برنامج Blastp.
- مقارنة ترجمة تتابع بروتينى مع قاعدة تتابعات عن طريق برنامج tBlastn DNA.
- مقارنة ترجمة تتابع DNA أمام قاعدة ترجمة تتابعات DNA عن طريق برنامج tBlastx.
- ويوجد هذا برنامج الـBLAST على الإنترنت أو كبرنامج مستقل أو داخل شبكات المواقع المتخصصة لقواعد البيانات مثل NCBI أو EMBL .

ماهى النتائج التى يمكن أن نحصل عليها من برامج الـBLAST ؟

تعطينا نتائج الـ Blast درجة التشابه أو الأختلاف بين التتابع المدروس والتتابعات الموجودة على قاعدة البيانات مع حساب درجة التشابه نتيجة الصدفة. مع الأخذ فى الأعتبار أن التتابعات المستخدمة فى المقارنة:

١- تتابعات عشوائية ٢- ذات تركيب ثابت.

obeikandl.com

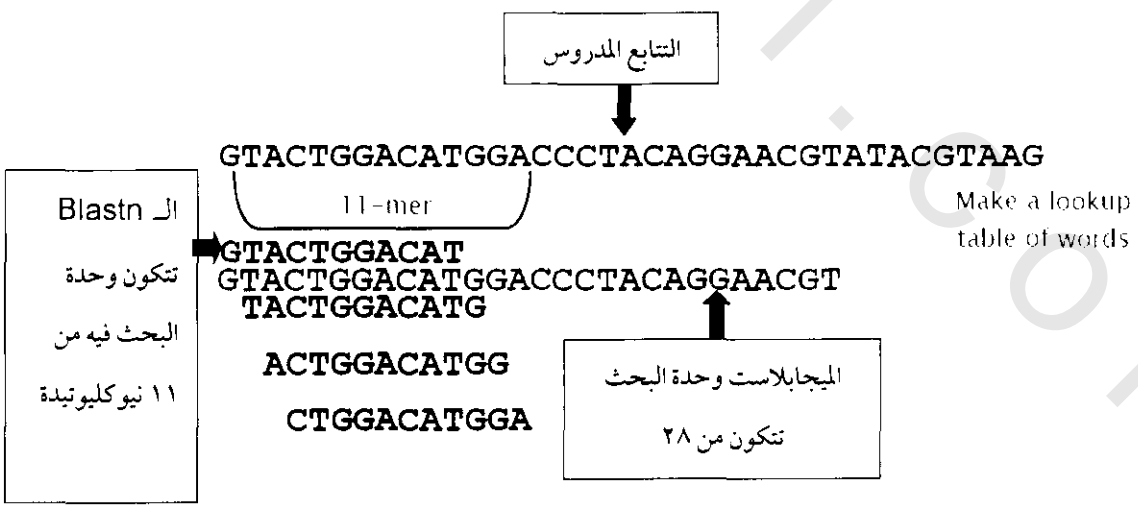
وتكون النتيجة المتوقعة بالتالى أن التشابهات التي توجد بين هذه التتابعات توضح تشابه تطوري، أي هذه التتابعات قادمة من أصل وراثي واحد ، وذلك لا يعني أن تكون لها نفس الوظيفة.

أنواع أخرى للبحث المتقدم عن طريق الـ Blast

١- ميجابلاست Megablast : هو أحد أنواع البحث بأستخدام الـ Blast لمقارنة تتابع من الـ DNA مع قاعدة بيانات تتابعات الـ DNA مثل نظيره الـ Blastn.ولكن يتميز الميجابلاست فى أن وحدة البحث تتكون من ٢٨ نيكلوتيده وبحد أدنى نيكلوتيده ١٢ فى حين أن نظيره الـ Blastn تتكون وحدة البحث فيه من ١١ نيوكليوتيدة وبحد أدنى ٧ نيكلوتيده.كما يظهر فى الجدول التالى مقارنة بين النوعين

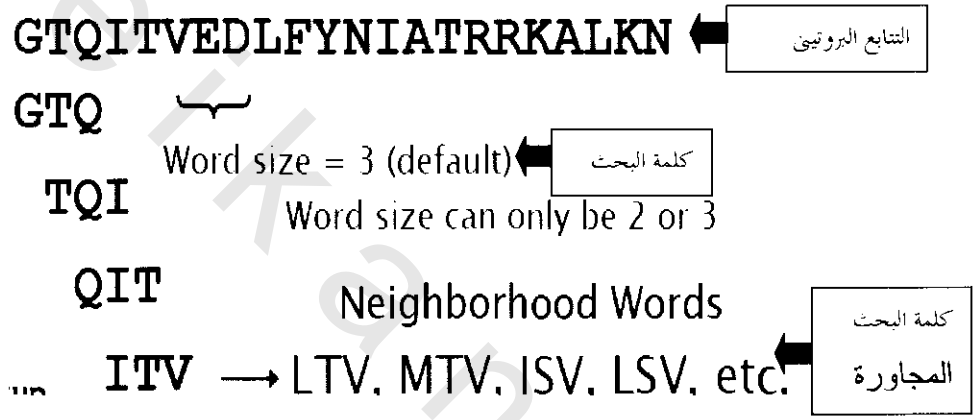
وحدة البحث	الوحدة المستخدمة	بحد أدنى
Blastn	١١	٧
Megablast ميجابلاست	٢٨	١٢

يعنى ذلك أن الميجابلاست سوف يظهر التتابعات الأكثر قرابة للتتابع المدروس وبالتالى يظهر عدد أقل من نتائج البحث. ولذلك هو اكثر دقة فى البحث عن التتابعات الأكثر قرابة للتتابع المدروس كما فى الشكل ٥٣.



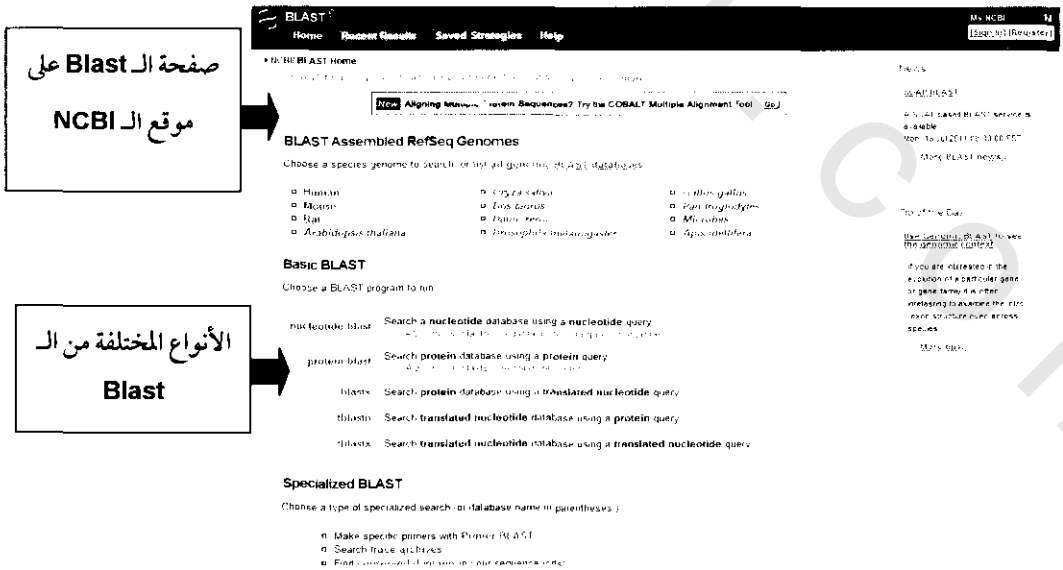
شكل ٥٣: يتميز الميجابلاست فى أن وحدة البحث تتكون من ٢٨ نيكلوتيده وبحد أدنى نيكلوتيده ١٢ فى حين أن نظيره الـ Blastn تتكون وحدة البحث فيه من ١١ نيوكليوتيدة وبحد أدنى ٧

أما فى حالة Blastp لمقارنة تتابع بروتينى مع قاعدة تتابعات البروتينات فأن وحدة البحث تتكون من٣ حروف أي ٣ أحماض أمينية ويعنى ذلك أنه عندما تكون هناك ٣ أحماض أمينية متشابهة تتكون بداية التشابه.كما يتطلب أيضا أن تكون هناك وحدتين متشابهتين للبحث على الأقل فى التتابع ولا تزيد المسافة بينهما عن ٤٠ حمض أميني. وهناك أيضا مايسمى "بوحدات البحث المجاورة" وهى عبارة عن وحدات بحث مكونة من ثلاث أحماض أمينية أيضا ولكن تكون مشابه للأحماض الأمينية المدروسة فى خواصها الكيميائية والفيزيائية ولكن غير متطابقة معها كما فى الشكل ٥٤.



شكل ٥٤: مقارنة تتابع بروتينى مع قاعدة تتابعات البروتينات

فأن وحدة البحث تتكون من٣ حروف أي ٣ أحماض أمينية والتى قد تكون متطابقة أوغير متطابقة "متجاوره".



شكل ٥٥. شكل صفحة الـ Blast على موقع الـ NCBI. توجد الأنواع المختلفة من الـ Blast فى الجانب الأيمن من الصفحة.

كيفية الوصول إلى BLAST:

١- فتح صفحة NCBI.(http://www.ncbi.nlm.nih.gov /) (شكل ٥٦).

٢- الضغط على كلمة BLAST ناحية اليمين (مشار إليها بسهم أحمر في الصور التوضيحية (شكل ٥٦).

٣- في أسفل الصفحة على اليسار نرى كلمة Nuclotide plast نضغط عليها (شكل ٥٥).

٤- في المستطيل الفارغ في الأعلى يتم وضع التتابع المراد مقارنته أو الرقم الكودى للتابع(شكل٥٧).

٥- نختار قاعدة البيانات التي نريد البحث بها من Choose search set (شكل٥٧).

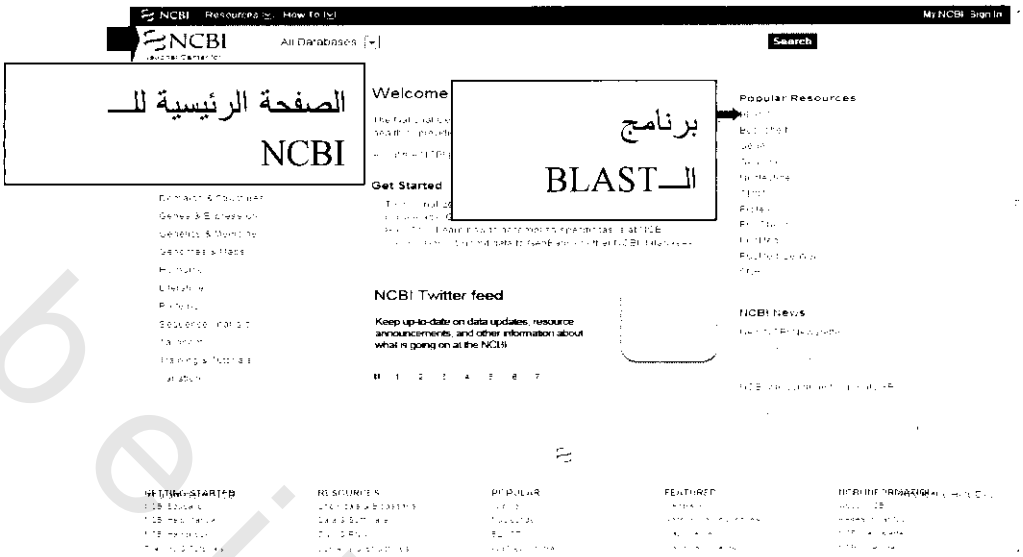
٦- نختار نوع الـ BLAST المراد تطبيقه في الأسفل من (Program Seletion)

٧- فيقوم بعمل مقارنة ويقوم بإعطاء نتيجة عبارة عن صفحة طويلة مكونة من ثلاث أقسام. القسم الأول يتكون من الملخص البيانى (Graphic Summary) بين التتابعات القريبة من التتابع المدروس مرتبة حسب درجة التشابه أو القرابة(شكل أ٥٨) . وبالنزول للأسفل نجد القسم الثانى والخاص بوصف هذه التتابعات (Descriptions) (شكل ب٥٨). وبالنزول الى أسفل نجد القسم الثالث الخاص بالتوازى والمقارنة (Alignment) والذى يختص بقارنة التتابع المدروس مع التتابعات القريبة (شكل ج٥٨).

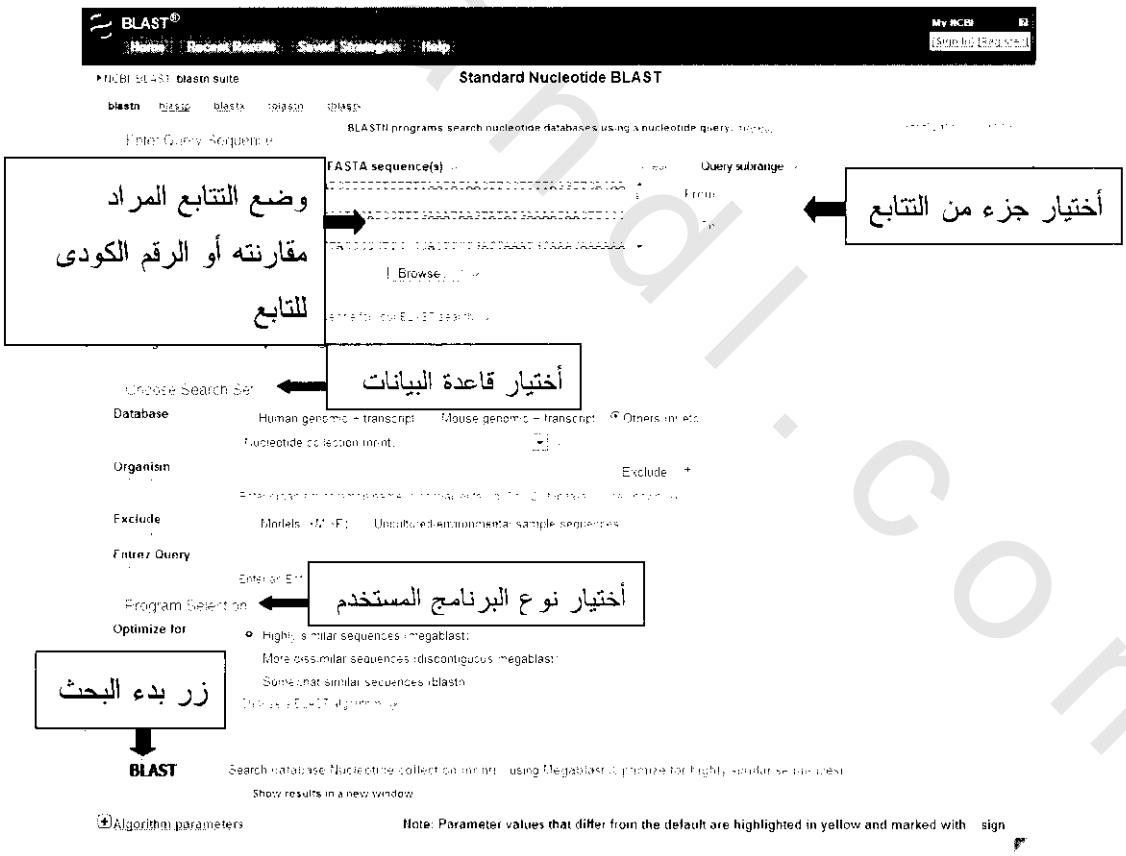
٨- يمكن الحصول على بعض التتابعات المرغوبة وحفظها لأستعمالها فى برامج أخري عن طريق الضغط على (Get selected sequences) .كما يمكن أيضا عمل شجرة قرابة بين التتابعات المختارة عن طريق أختيار الأمر (Distance tree of results) وذلك بعد أختيار التتابعات المرغوبة للمقارنة (شكل ج٥٨).

٩- عند الضغط على (Distance tree of results) نحصل على شكل تخطيطى يوضح درجة التشابه والأختلاف بين التتابعات المدروسة. يظهر لنا هذا الشكل الموضح في الصور(شكل ٥٩).

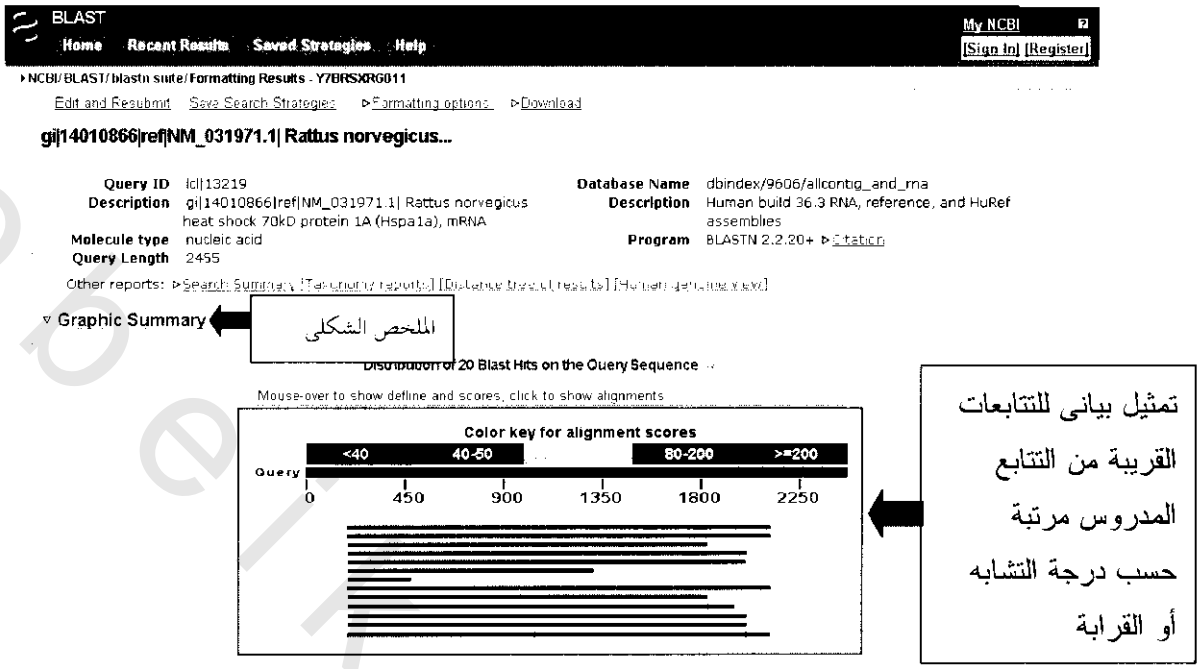
١٠- ويمكن أختيار أى نوع من الأنواع مختلفة من أشكال شجرة القرابة(شكل ٥٩).



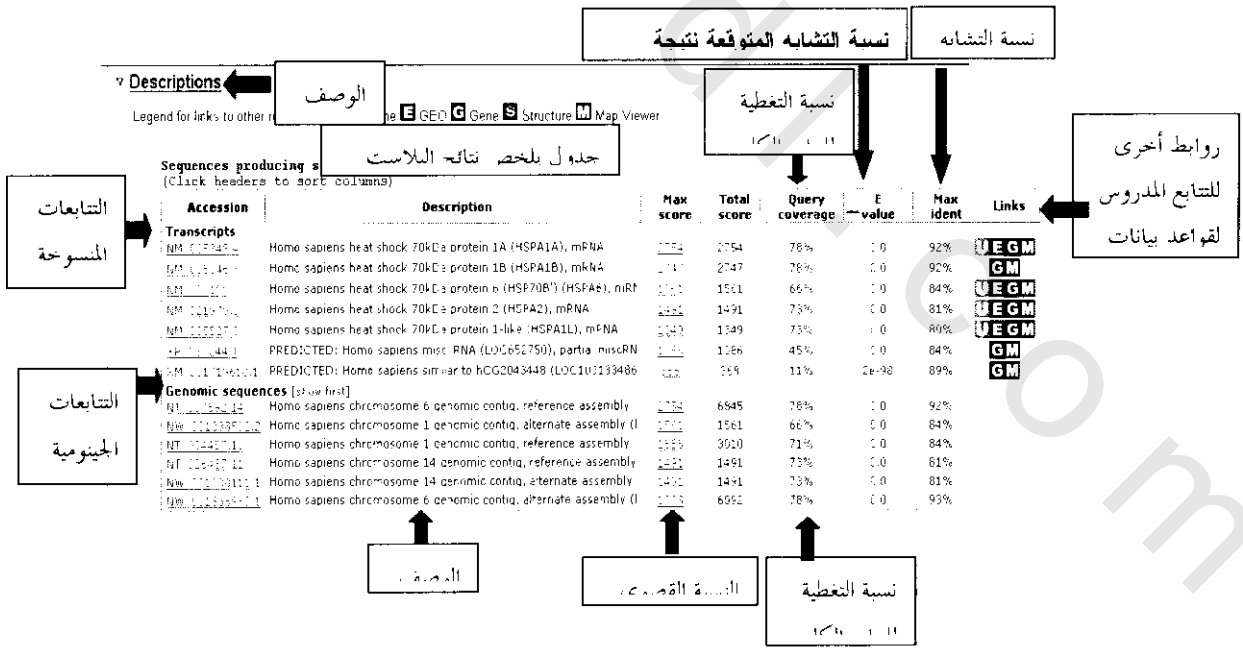
شكل ٥٦ : فتح الصفحة الرئيسية للمركز الوطنى لمعلومات التكنولوجيا الحيوية NCBI.



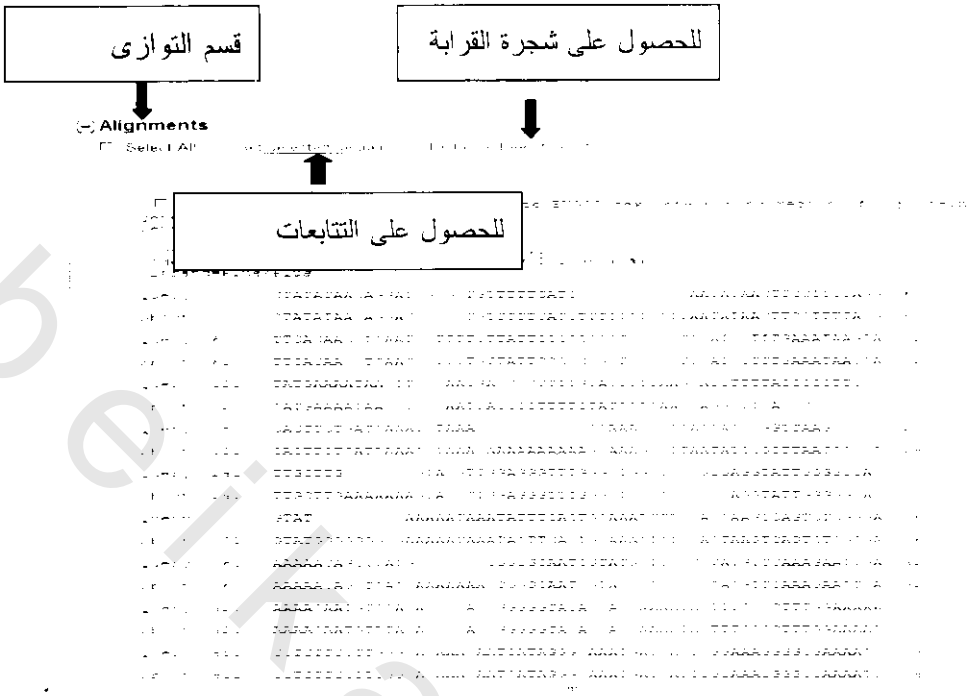
شكل ٥٧ : فتح الصفحة الرئيسية لبرنامج الـ Blast الخاص بمقارنة تتابعات الـ DNA.



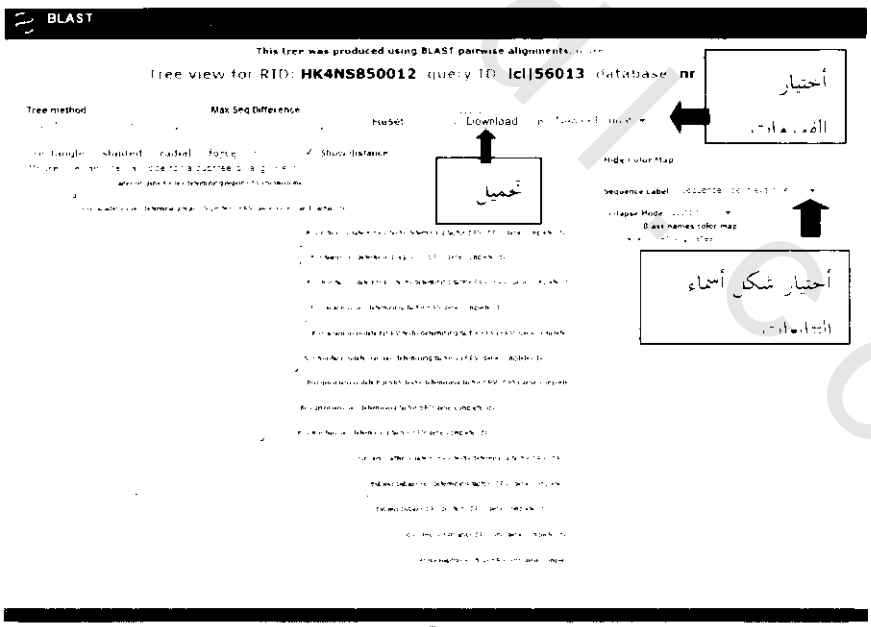
شكل٥٨ أ: الملخص الشكلى (Graphic Summary) بين التتابعات القريبة من التتابع المدروس مرتبة حسب درجة التشابه أو القرابة.



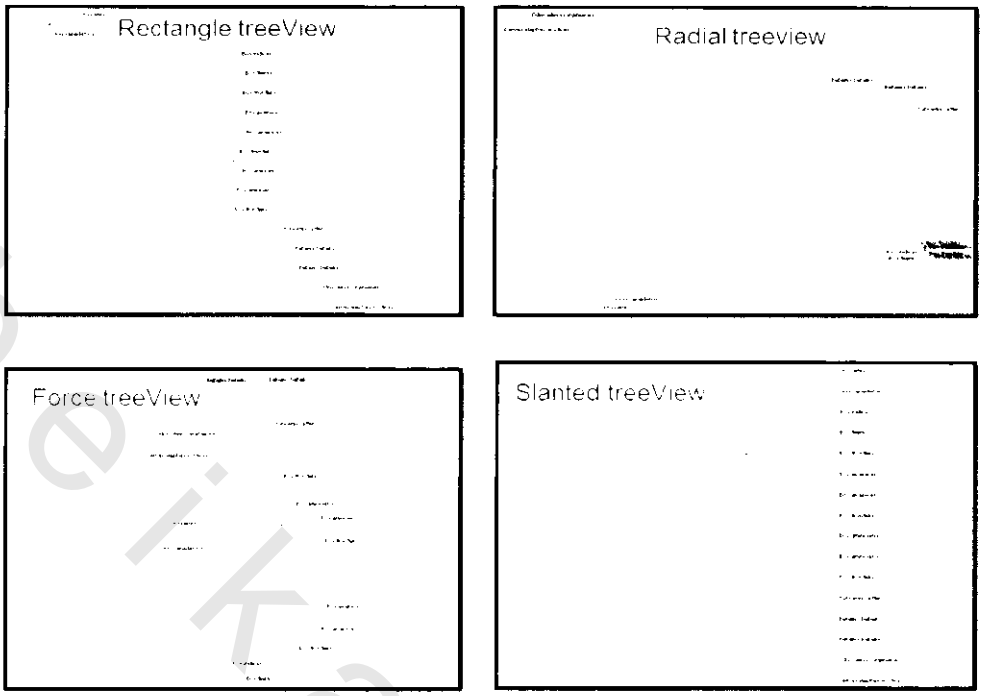
شكل ٥٨ ب: القسم الثانى والخاص بوصف هذه التتابعات (Descriptions).



شكل ٥٨ ج: القسم الثالث الخاص بالتوازى والمقارنة (Alignment) والذى يختص بمقارنة التتابع المدروس مع التتابعات القريبة.



شكل ٥٩: شجرة القرابة بين التتابعات المختارة.



شكل ٦٠: ويمكن أختيار أى نوع من الأنواع مختلفة من أشكال شجرة القرابة.

أنواع الـ **BLAST** الخاصة بالبروتين فقط:

النوع الأول: بلاست التتابعات البرتينية

blastp (protein-protein BLAST)

ويعمل هذا النوع على مقارنة التتابعات البروتينية بقاعدة بيانات البروتينات للبحث عن أكثر التتابعات البروتينية قرابة و تشابها من التتابع المدروس فى الكائنات الكائنات -المختلفة الموجودة فى قاعدة البيانات.

*النوع الثانى:- بلاست الـ **PSI** بلاست المناطق المتشابهة المتكررة*

PSI-BLAST (Position-Specific Iterated BLAST)

النوع الثالث:- بلاست النطاقات الوظيفية المتشابهة ***PHI-BLAST (Pattern Hit Initiated BLAST)*** *والذى يكرر البحث عن المناطق المتشابهة.*

ولكل واحد من هذه الأنواع مجموعة كبيرة من التطبيقات . ويعتبر النوع الأول هو الأكثر أستخداما.

obeikandl.com

كيف تقرأ نتائج التوازى بين التتابعات الناتجة من الـ Blast

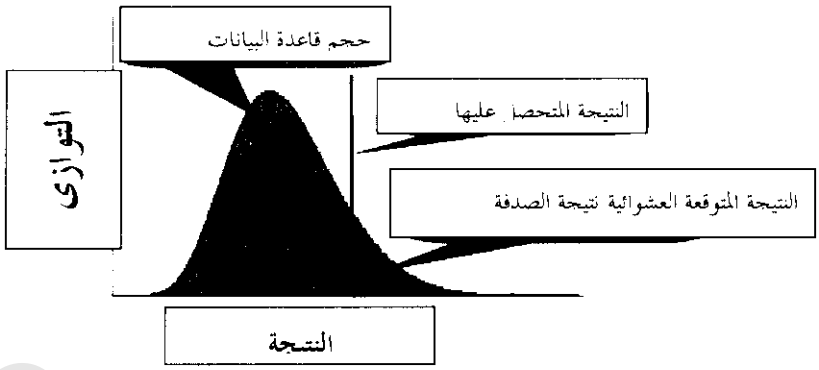
لكى نعرف كيف نقرأ نتائج التوازى بين التتابعات الناتجة من برنامج الـ Blast يجب أن نتعرف على بعض المصطلحات الملخصة فى الجدول التالى:

مصطلحات هامة لقراءة نتائج التوازى والتشابه بين التتابعات

الرقم	المصطلح	تعريف	التفسير
١	Expected value أو E.Value نسبة الصدفة	نسبة التشابه المتوقعة نتيجة الصدفة	كلما قلت هذه النسبة كلما زادت معنوية النتائج والعكس صحيح
٢	تطابق (*) Identical match	هو التشابه وتطابق ويعبر عنها برمز النجمة	يعبر عن التطابق الكامل بين الوحدة المدرسة والمقارنة
٣	أشارة موجبة Positive score (Conservative) (:) (.)	يعبر عن وجود تشابه غير متطابق ويرمز له بنقطة أوأثنين فمثلا إذا كان حدث تغير ولكن داخل نفس المجموعة (أي قاعدة بيورين بقاعدة بيورين أو بيريميدين بقاعدة بيريميدين أو لبوسين بأيزوليوسين) يأخذ رقم موجب.	يعبر عن تشابه فى الخواص بين الغالبية المقارنة مع وجود أحد الأختلافات أى أن التغير الناتج لاينتج عنه أختلاف معنوى فى شكل ووظيفة البروتين النهائى
٤	أشارة سالبة Negative score (-)	في هذه الحالة يكون التغير بين بروتين وبيرميدين أو العكس لذلك لا يكون Score إشارة سالبة	أى أن التغير الناتج ينتج عنه أختلاف معنوى فى شكل ووظيفة البروتين النهائى
٥	فراغ ( ) Gap	وهي منطقة فارغة أي هناك أحماض أمينية لا تقابل أي حمض آخر أو نيوكليوتيدات لا تقابل أي نيوكليوتيدات أخرى	ويعنى ذلك وجود طفرات حذف أو أضافة فى التتابع المدروس بمقارنة بتتابعات أخرى

الأسس الأحصائية لحساب نسبة الصدفة
Expected value أو E.Value

الرقم الأعلى للرقم (High Score) لنتيجة التوازى بين تتابعين عشوائيين يتبع الحد الأقصى للتوزيع العشوائى (Exterm Value distribution). ويجب معرفة ان نسبة الصدفة تعبر عن عدد المرات التى تحدث نتيجة الصدفة. وينطبق المثال التالى على التوازى بدون وجود فراغات بينية (شكل ٦١).



شكل٦١: شكل يوضح الأساس الأحصائى لحساب نسبة الصدفة .E. Value

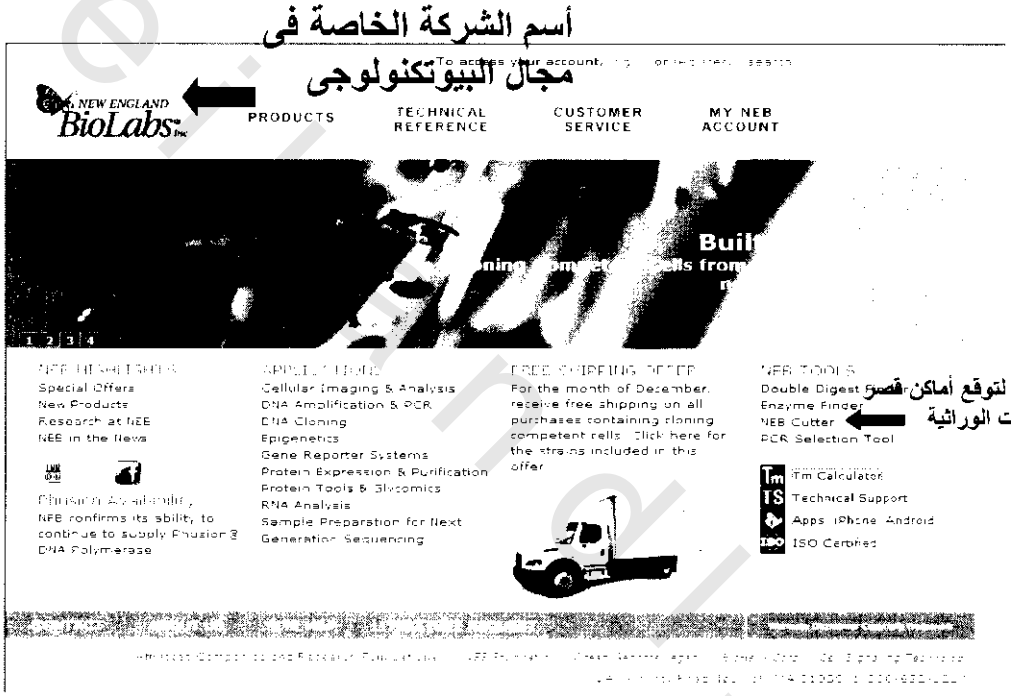
ويوضح الشكل التالى بعض معانى الأشارات الهامة المستخدمة للتعبيرعن نتائج التوازى بين التتابعات الناتجة من الـ Blast (شكل ٦٢)

```
>gi|127552|sp|P23367|MUTL_ECOLI  DNA mismatch repair protein mutL
          Length = 615

 Score = 42.0 bits (97),  Expect = 3e-04
 Identities = 26/59 (44%), Positives = 33/59 (55%), Gaps = 9/59 (15%)

Query  9    LPKNTHPFLYLSLEISPQNVDVNVHPTKHEVHF-----LHE---ESILEV-QQHIESKL  58
            L  +  P   L LEI P  VDVNVHP KHEV F     +H+    + +L V QQ +E+ L
Sbjct  280  LGADQQPAFVLYLEIDPHQVDVNVHPAKHEVRFHQSRLVHDFIYQGVLSVLQQQLETPL  338
```


شكل٦٢: شكل يوضح مصطلحات هامة لقراءة نتائج التوازى والتشابه بين التتابعات.

رسم خريطة لمواقع أنزيمات القصر تتابع جينى

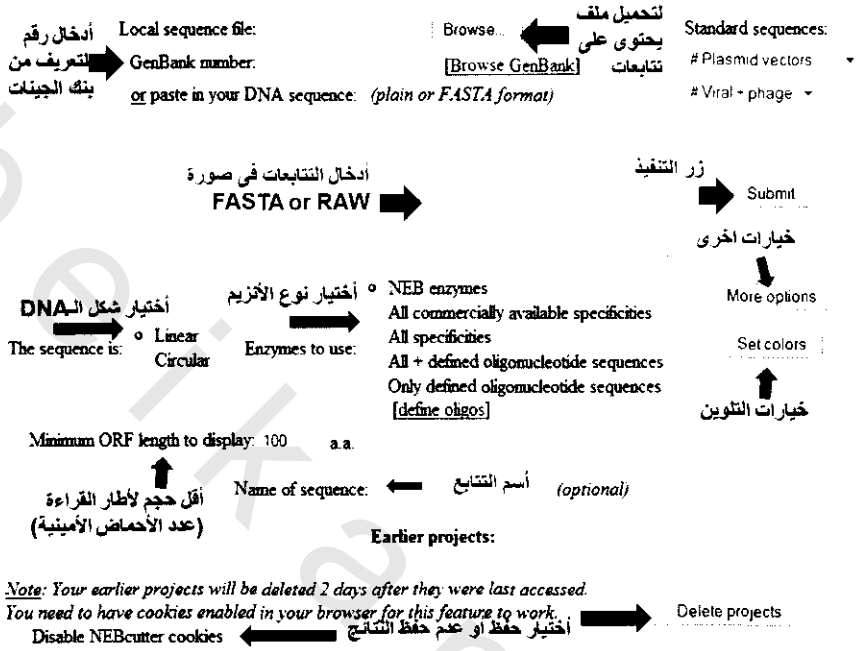
تتطلب بعض تجارب الهندسة الوراثية عمل خريطة كاملة لمواقع أنزيمات القصر فى تتابع جينى معين. وتفيد مثل هذه الخرائط العديد من تجارب التطويع الجينى. وهناك العديد من البرامج المجانية التى تساعد على رسم وتوقع شكل هذه الخرائط. بعض هذه البرامج توفره الشركات المنتجة لهذه الأنزيمات على مواقعها. ومن أشهر البرامج المستخدمة هو برنامج NEB cutter والموجود على موقع www.neb.com والخاص بشركة New England Biolabs الشهيرة فى منتجات البيوتكنولوجى.



شكل٦٣: شكل يوضح الصفحة الرئيسية لموقع شركة البيوتكنولوجى الخاصة New England Biolabs .

عند وضع التتابع المرغوب دراسته فى برنامج الـ NEB cutter يعمل البرنامج على التعرف على مناطق أطار القراءة فى الجين المرغوب بناء على الأكواد والشفرات الوراثية المعروفة فى بكتريا الـ E.coli كما يقوم البرنامج أيضا بالتعرف على مناطق القصر (القطع) فى التتابع المدروس الخاصة بأنزيمات القصر من النوع الثانى TypeII وبعض أنواع TypeIII المتوفرة تجاريا والتى تقطع فى أماكن متفردة. فى البداية يستخدم البرنامج الأنزيمات المتوفرة من الشركة ولكن يمكنك اختيار اى نوع من الأنزيمات الأخرى المتوفرة تجاريا كما يظهر فى(شكل ٦٤) وهناك أختيارات أخرى سوف تظهر مع صفحة النتائج. ويجب الأخذ فى الأعتبار أن الحد الأقصى لحجم الملف

الذى يمكن تحميله هو واحد ميجابايت كما أن أقصى طول للتتابع الذى يمكن دراسته هو ٣٠٠ الف قاعدة. ويوضح الشكل التالى صفحة الأدخال فى البرنامج مع وجود عدد من الأختيارات المختلفة.

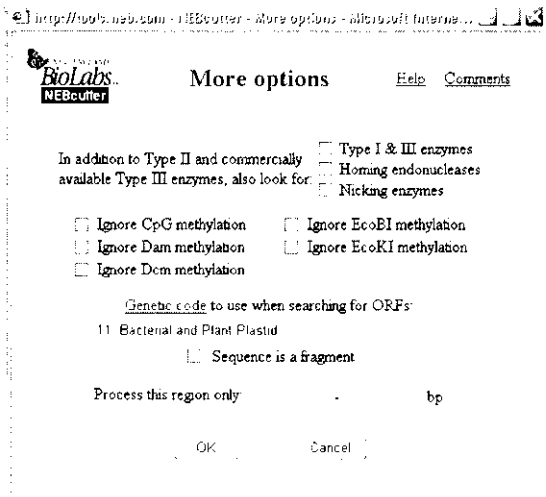


شكل٦٤: شكل يوضح صفحة الأدخال الرئيسية لبرنامج الـ NEBcutter والتى تحتوى على العديد من الأختيارات .

وبالضغط على زر خيارات أخرى (More options) تظهر لنا مجموعة جديدة من الأختيارات الهامة مثل أختيار مجموعة أخرى من أنزيمات القصر أو تجاهل بعض المناطق المعينة أثناء عملية القطع او تحديد منطقة معينة من التتابع المدروس لأجراء عملية القطع (شكل٦٤).

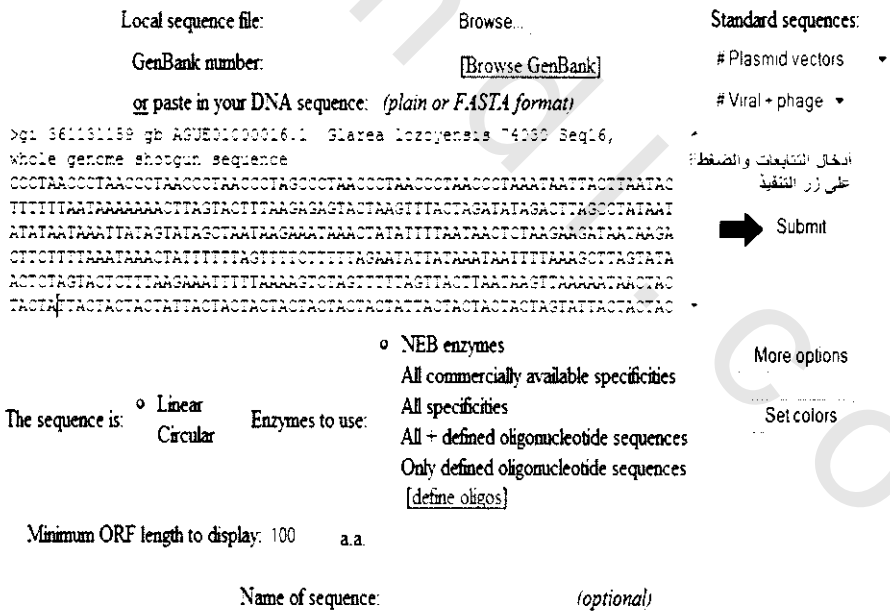
خيارات أضافية فى برنامج الـ NEBcutter

- تجاهل تحويرات مناطق الميثلة (أضافة مجموعة ميثل) المختلفة (مثل مناطق CpG, Dam, Dcm, EcoBI and EcoKI).
- أستخدام أنزيمات النوع الأول والثالث من أنزيمات القصر (type I & III restriction enzymes).
- أختيار جدول الأكواد الوراثية المختلفة لأحد الأنواع.
- أمكانية البحث فى جزء محدد من التتابع المدخل.
- أمكانية تخزين النتائج وأمكانية أسترجاعها خلال بضعة أيام.



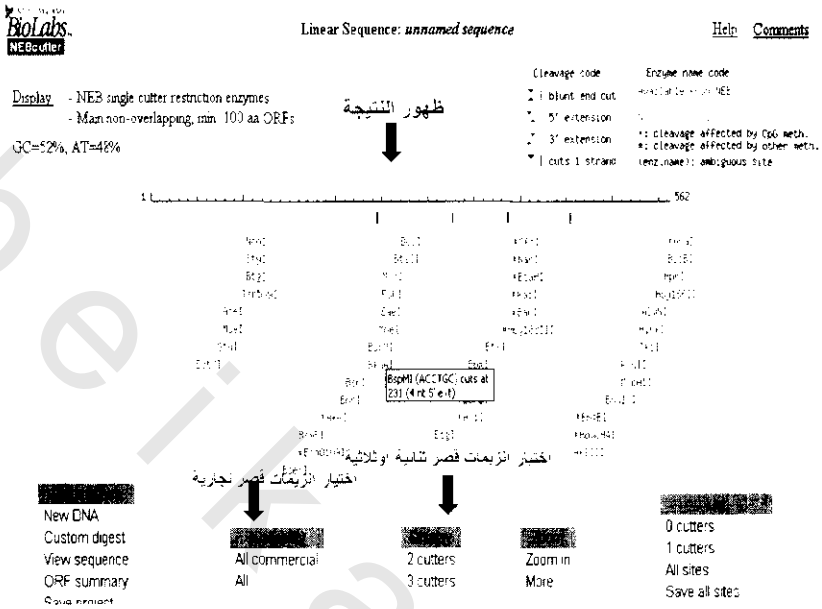
شكل٦٥: شكل يوضح الخيارات الأضافية لبرنامج الـ **NEBcutter** والتى تحتوى على العديد من الأختيارات المكملة.

ثم يتم أدخال التتابعات المرغوبة فى المكان المخصص والضغط على زر تنفيذ كما فى (الشكل ٦٦).



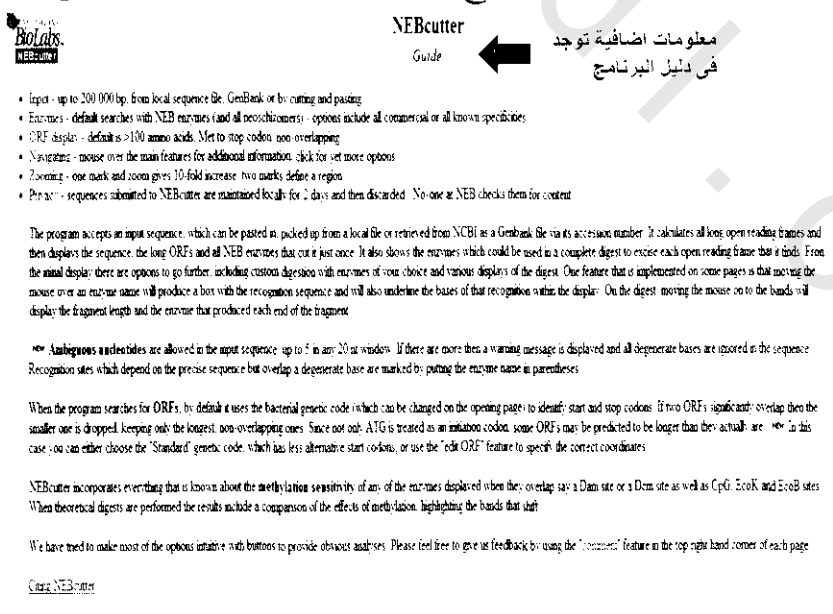
شكل٦٦: أدخال التتابعات فى صورة صيغة الـ **FASTA** والكبس على زر التنفيذ.

تخرج النتيجة من البرنامج كما فى (شكل٦٧) بحيث تظهر أماكن القصر المختلفة بأستخدام أنزيمات القصر المختلفة على طول التتابع المدروس.



شكل ٦٧ : شكل صورة النتائج المخرجة من لبرنامج الـ **NEBcutter** .

وعند الضغط على كلمة دليل فى أعلى يمين صفحة الأدخال نحصل على معلومات حول البرنامج وخواصه وكيفية أستخدامه كمافى شكل ٦٨ .



شكل ٦٨ : صورة توضح شكل صفحة الدليل لبرنامج الـ **NEBcutter** .

التناظر بين التتابعات البيولوجية

Sequence Alingment between biological sequences

أن أداة التناظر أو التوارى هى أداه هامة جداً في الدراسات البيولوجية ولذلك تستخدم هذه الأداة في علم Bioinformatics فى مقارنة البينة بين التتابعات البيولوجية مثل :

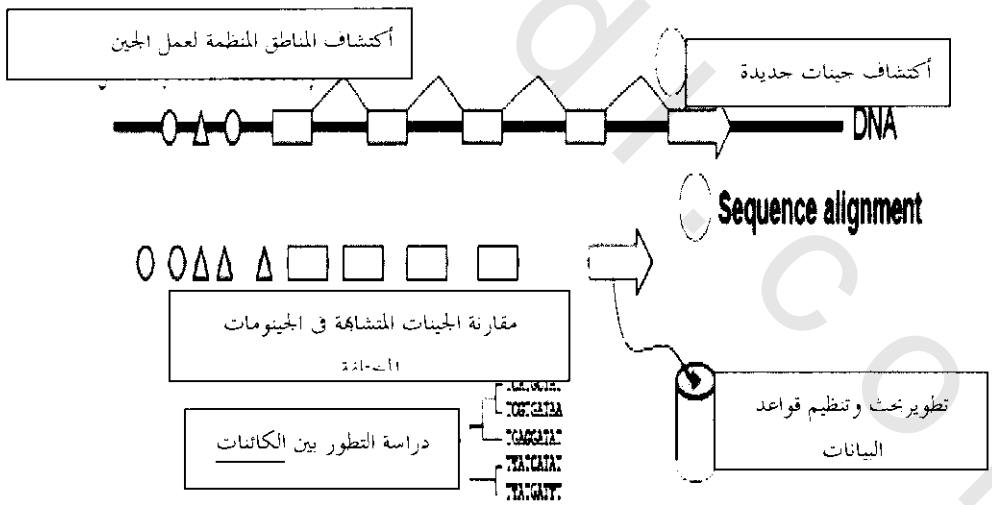
١ - مقارنة تتابعات الـ DNA.

٢- مقارنة تتابعات RNA.

٣- مقارنة تتابعات بروتينية.

وتتم هذه المقارنة بغرض إيجاد التشابهات أو الأختلافات بين التتابعات المختلفة. ويمكن عن طريق نتائج التناظر مساعدة الكثير من الأبحاث البيولوجية على سبيل المثال:

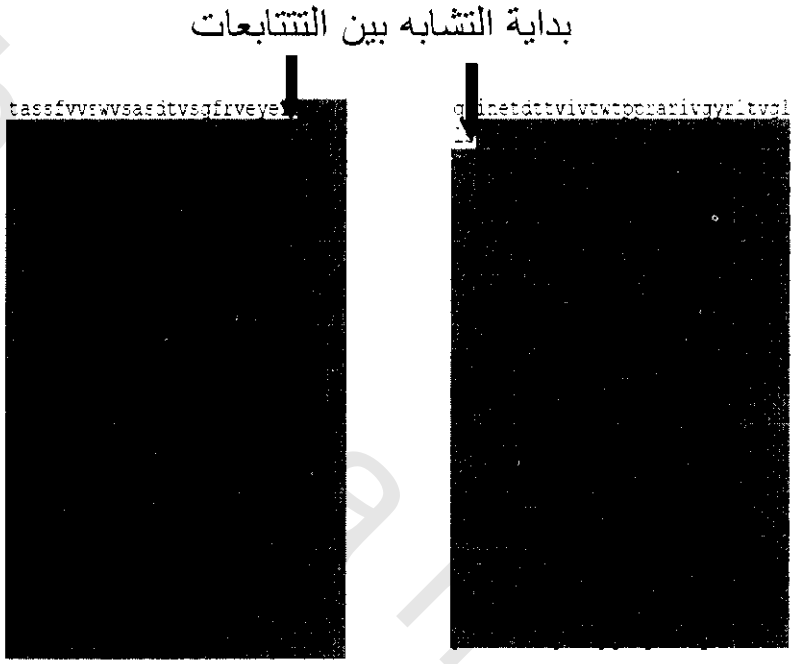
- اكتشاف الأصل المشترك لمجموعة جينات أو بروتينات.
- التنبؤ بالوظيفة والتركيب للجين او البروتين المدروس.
- تحديد التتابعات التحتية Sub sequence المتشابه في الجينات والبروتينات.
- تحديد المناطق النشطة المتشابهة.
- تحديد التتابعات المتداخلة.



شكل ٦٩ : صورة توضح بعض التطبيقات الناتجة من تناظر التتابعات البيولوجية

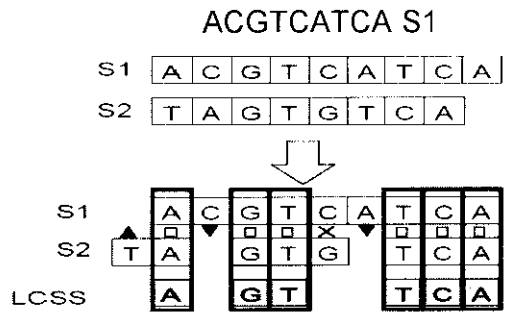
Sequence similarity

من الصعب بالعين المجردة إيجاد التشابه بين تتابعين أو جينات بكثرة عدد الحروف لذلك كان لابد من وجود برامج الحاسب الآلي لإيجاد هذه الفروق كما يظهر فى الشكل التالى:



شكل ٧٠: صورة توضح بعض صعوبة إيجاد التشابه بين تتابعين بالعين المجردة بدون عمل تناظر للتتابعات البيولوجية.

فعلى سبيل المثال: كيف يمكن المقارنة بين تتابعين من DNA لو أخذنا خيطين أحدهما S1 والآخر S2 للبحث عن أطول مقطع متشابه بين التتابعين سوف نحاول أن نحرك بعض التتابعات وليكن في الخيط الثاني ووضع بعض الفجوات بهذا التتابع لنحصل على أكبر قدر من التشابه بين الخيطين.



شكل ٧١: صورة توضح تحريك التتابعات لنحصل على أكبر قدر من التشابه بين الخيطين.

كيف يمكن أن نحسب أفضل توازي بين تتابعين؟

الناتج النهائي للتحوير في الخيطين لحساب أفضل توازى يشتمل على:

١- الدرجة المعطاه للطفور الناتج من التبديل أو التبادل في القواعد Substitution.

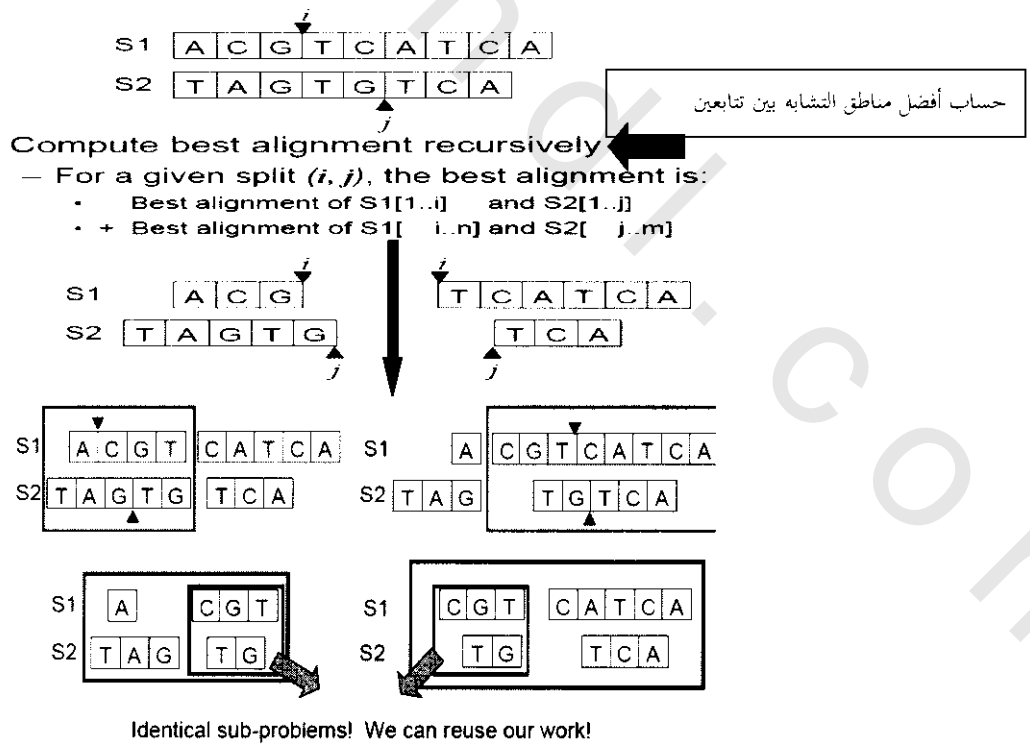
٢- الدرجة المعطاه للطفرة الناتجة من الإضافة أو الحزف تسمى Indel.

٣- الدرجة المعطاه للتشابه Similarirty.

ونحتاج إلى معادلة لحساب أفضل توازي أو تشابه بين هذه التتابعات. بكل من التتابعين المختلفتين أكثر من احتمال للتوازي لذلك نحاول أن نحسب رياضياً أكثرهم تشابهاً عن طريق الأرقام المعطاه لكل العناصر السابقة، لذلك كان أفضل الأشياء هو عمل برنامج الحاسب الآلي يستطيع أن يحسب هذه التشابهات ويختار أفضلها (أعلاها قيمة).

طريقة عمل البرنامج:

نضع الخيطين في مصفوفة حتى نحاول الوصول لأفضل حساب للتشابه ما بين الخيطين كما يوضح الشكل التالي:



شكل ٧٢: صورة توضح طريقة الوصول لأفضل حساب للتشابه ما بين الخيطين.

وهناك نوعين من التوازى:

١- (التوازى الجزئى) (Local alignment) لفحص مناطق التشابه بين تتابعين مختلفين جزئيا.

٢- (التوازى الكلى) (Global alignment) لبحث عن التشابه ما بين تتابعين على الطول الكلي للتتابعين.

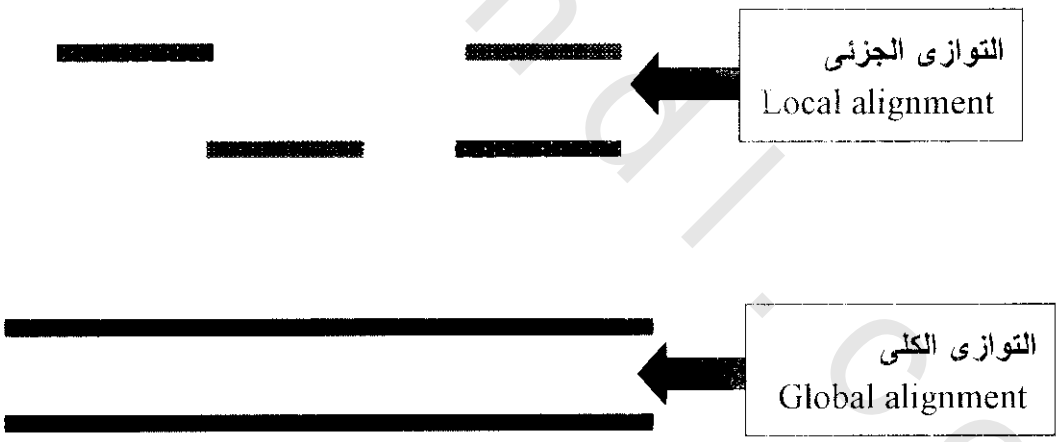
هناك تشابهه النواتج بين (التوازى الجزئى) (Local alignment) و(التوازى الكلى) (Global alignment):

١- ليس هناك أي نتيجة بالسالب.

٢- يبدأ بأعلى نتيجة بناء على المصفوفة المستخدمة ويستمر حتى صفر.

ويجدر الأشار أن التوازى الكلى (Global alignment) تم أكتشافه عن طريق العالم Needleman-Wunsch أما التوازى الجزئى (local alignment) تم أكتشافه عن طريق العالم Smith-Waterman

والشكل التالى يوضح الفرق بين الأثنين:



شكل ٧٣: صورة توضح التوازى الجزئى والتوازى الكلى.

أولاً: التوازى الجزئى (Local alignment)

ويمكن حساب التوازى الجزئى عن طريق وضع أرقام تعبرعن العناصرالوراثية المختلفة كما يلى:

فمثلا يمكن تبسيط المفاهيم كما يلى:

١. طفرة الأضافة اوالحذف = -٢ (Indel= - 2)

٢. طفرة التشابه= +١ (Match +1)

٣. طفرة الأختلاف = -١ (Mismatch -1)

وعلى ذلك يمكن حساب المثال التالى كما يلى:

المجموع = (+1x13) + (-1x2) + (-2x4) = 3

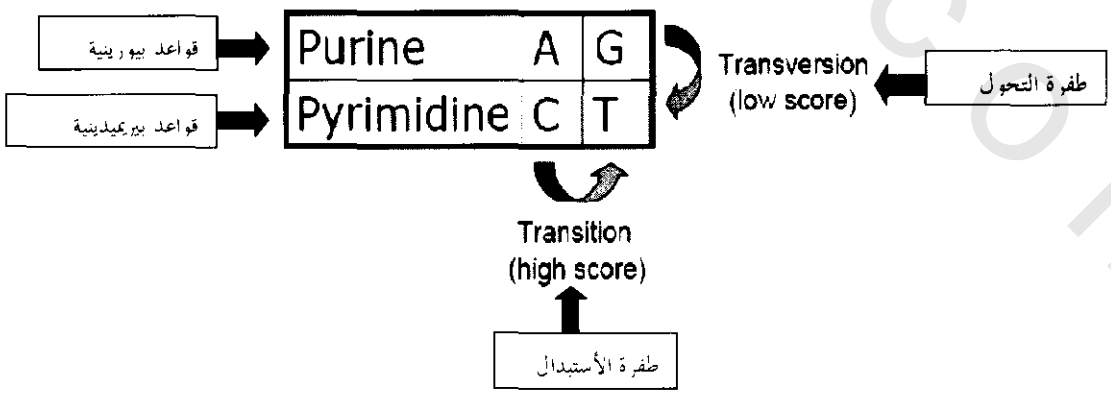
مثال أخر:

المجموع = (+1x5) + (-1x6) + (-2x11) = -23

ملحوظة: اختيار نظام درجات التقييم الطفور السابق ليست دقيقة من الناحية البيولوجية حيث أن هناك بعض الاختلافات (طفرات) تحتاج درجات تقييم أكثر عن غيرها.

فمثلا إذا تم تغيير قاعدة أو حمض أميني بحمض أميني آخر مشابه له فى الخواص (الحجم، الشحنة، الخ) لابد أن يأخذ درجة تقييم مختلفة عن تغيير حمض أميني بحمض أميني مختلف أخر مختلف فى الخاص.

المثال على ذلك:



شكل٧٤: صورة توضح أنواع طفرات الأستبدال المختلفة.

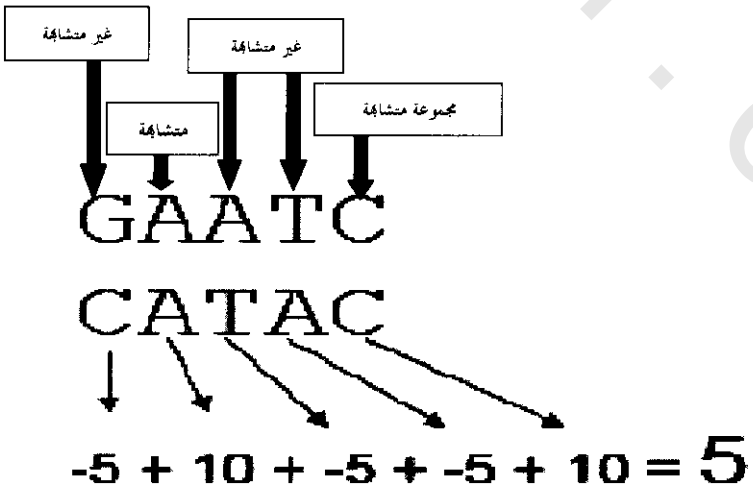
إذا حدث تبادل بين A, G (قواعد بيورينية) Purine أو بين C, T (قواعد بيريميدينية) يكون التقيم أعلى وذلك على العكس من حدوث تبادل بين T, G فإن هذا يأخذ تقييم أقل ويكون التقييم على حسب المصفوفة الجديدة التي وضعت في اعتبارها الأثر البيولوجى للطفرة والذى يزداد بتباعد خواص المجوعات الكيميائية المستبدلة (مثلا بيورين مع بيريميدين).

فمثلا يظهر الجدول التالى أن تقييم التشابة يقيم (١٠درجات) فى حين أن طفرة الأستبدال داخل نفس المجموعة الكيميائية يساوى صفر . أما طفرة الأستبدال من خارج المجموعة الكيميائية تقيم بـ (-٥) .

جدول يوضح أحد المصفوفات المقترحة لطفرات الـ DNA والتى أخذت فى الأعتبار الأثر البيولوجى للطفرة

	A	C	G	T
A	10	-5	0	-5
C	-5	10	-5	0
G	0	-5	10	-5
T	-5	0	-5	10

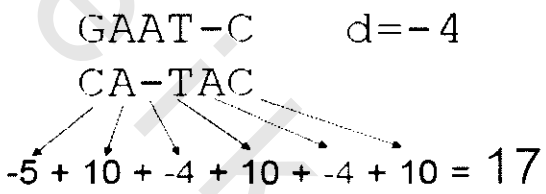
ويمكن تطبيق هذا الجدول على المثال التالى:



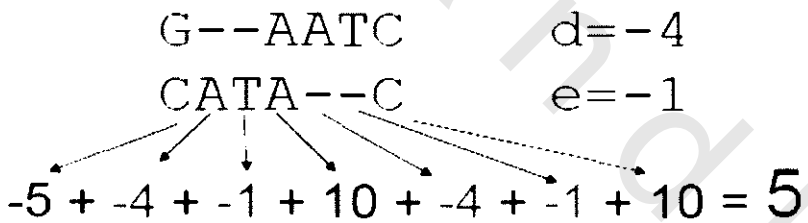
شكل ٧٥: صورة توضح أستخدام أحد المصفوفات المقترحة لتقييم التوازى والطفور.

تقييم الفراغات Scoring gaps (طفرات الحذف أو الأضافة):

يجب تحتوى المصفوفة على قيم تمثل طفرات الحذف أو الأضافة والتى تظهر فى التتابعات على شكل مسافة فارغة تمثل بخط قصير (-). وقد لوحظ أن الفراغ الأول (Gap opining) عند موقع محدد والذى يرمز له بالرمز (d) يأخذ تقييم أكبر من الفراغات التى تليه (Gap extensions) عند نفس الموقع والذى يرمز له بالرمز (e). فمثلا نجد أن الفراغ الأول (Gap opining) تأخذ تقييم(4-) فى حين أن الفراغات التى تليه (Gap extensions)) تأخذ تقييم(1-). وذلك يتضح بالمثال التالى:



مثال أخر:



تقييم عدد الأحتمالات الممكنة:

هناك أكثر من حل ممكن يمكن أن ينتج عن البرنامج وتكون نتيجة النهائية تعبر عن أفضل هذه الحلول وهى الأعلى قيمة حسابيا. حيث يمكن أن يكون هناك أكثر من حل آخر.

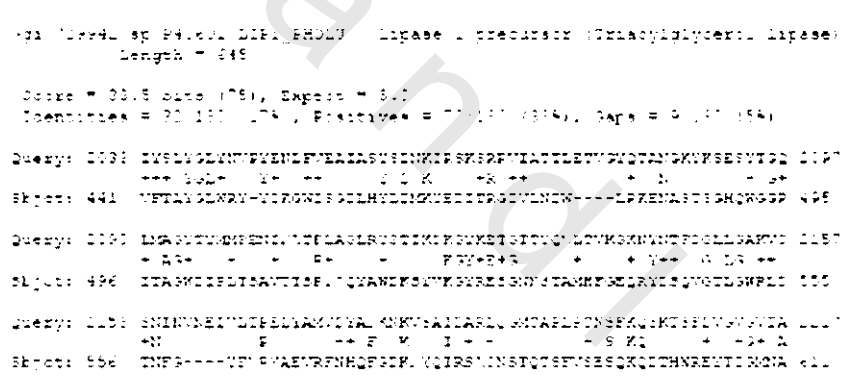
-GAATC	GAAT-C	GAATC
C-ATAC	C-ATAC	CATAC
GA-ATC	GAAT-C	GAATC-
CATA-C	CA-TAC	CA-TAC

```
GA-ATC
CATA-C

GAAT-C
CA-TAC

GAAT-C
C-ATAC

GAAT-C
-CATAC
```

أحتمالات كثيرة

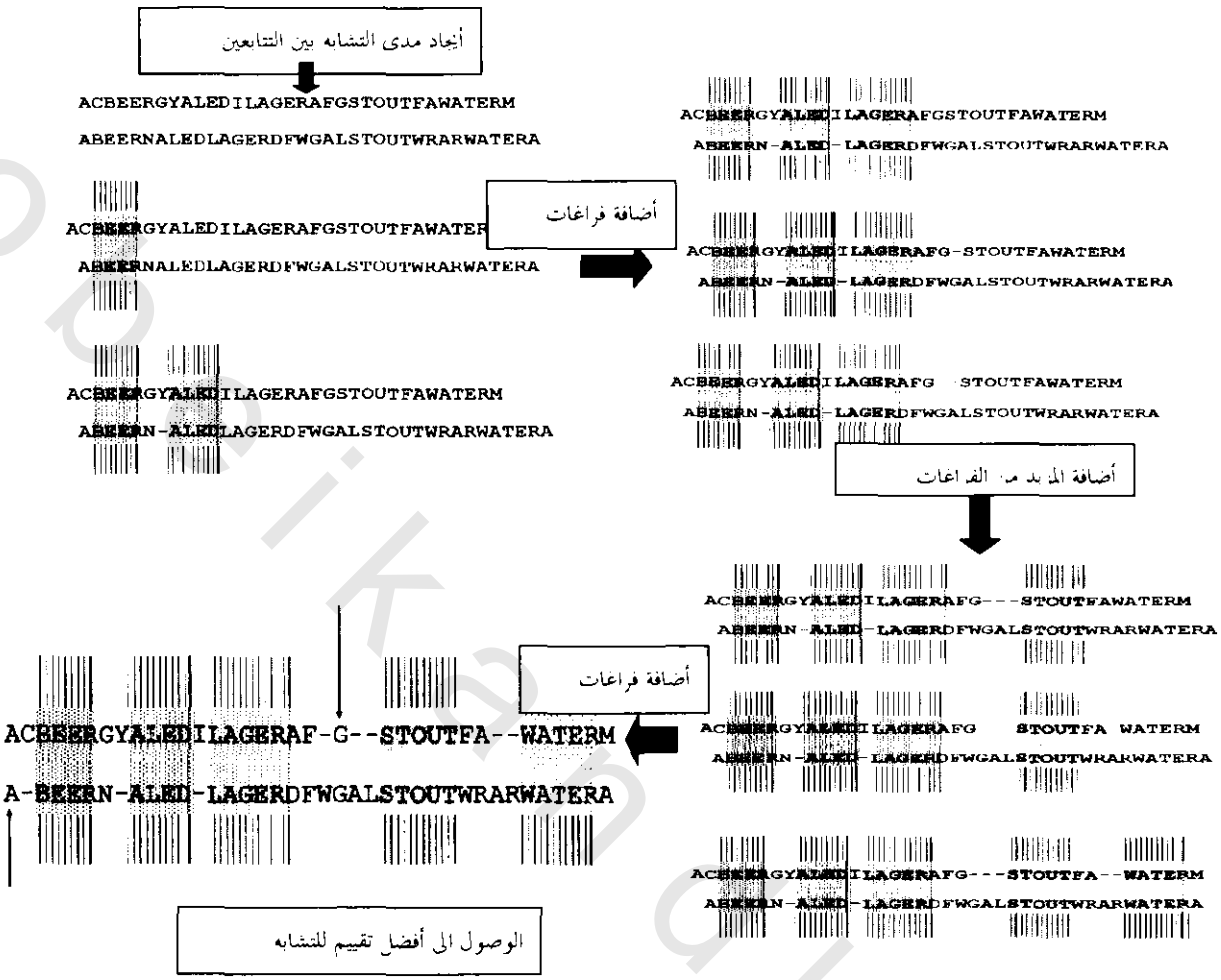
$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \approx \frac{2^{2n}}{\sqrt{\pi n}}$$

وبحساب هذه الأمثلة يمكننا معرفة أيها أفضل.



وغالبا ماتكون الفراغات الحقيقية أكثر من واحد وتتواجد بشكل متتالى

ويمكنا أن نرى من المثال التالى كيف أثرت أضافة الفراغات فى زيادة التشابه بين التتابعات بشكل كبير ولماذا نحتاج الى هذه الفراغات

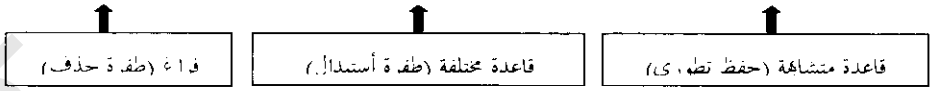


شكل ٧٦: صورة توضح كيف أثرت أضافة الفراغات فى زيادة التشابه بين التتابعات.

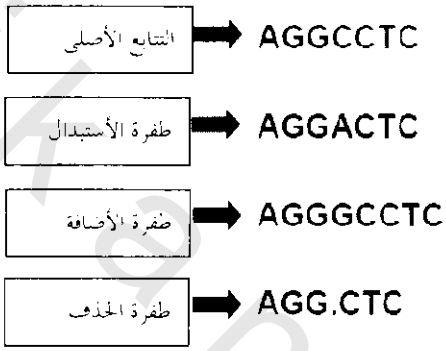
ولهذا فأنه عند حساب التوازى الجزئى بين تتابعين يجب الأخذ فى الأعتبار حساب المحصلة الكلية للفراغات ومواضع القواعد. وذلك يتم عن طريق توازى كل قاعدة فى التتابع الأول مع قاعدة أخرى أو فراغ فى الخيط الثانى كما يظهر فى الشكل التالى:

AGGCTATCACCTGACCTCCAGGCCGATGCCC
TAGCTATCACGACCGCGGTCGATTTGCCCGAC

-AGGCTATCACCTGACCTCCAGGCCGA--TGCCC---
TAG-CTATCAC--GACCGC--GGTCGATTTGCCCGAC



شكل ٧٧: شكل يوضح كيف أحتمالات التشابه والأختلاف بين القواعد فى تتابعين.



شكل ٧٨: شكل يوضح أنواع الطفرات المختلفة الممكن حدوثها على تتابع من الـ DN .

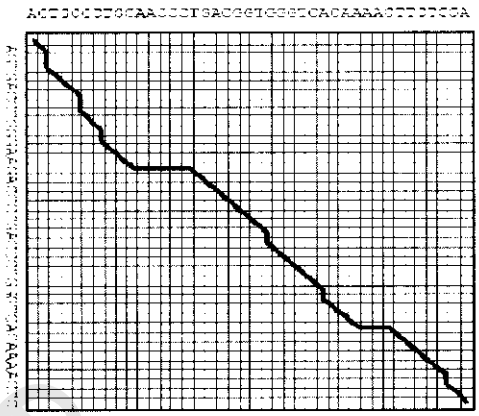
فأذا أفترضنا بأن:

التشابه (Match) = M + -عدم التشابه (Mismatch) = S- -الفراغ (Gap) = d-

لذلك فأن المجموع (F) = (عدد القواعد المتشابهة M X) – (عدد القواعد غير المتشابهة S X) – (عدد الفراغات GAP X d).

وكما يوضح (شكل ٧٩) فأن هناك عدد كبير من الأحتمالات التى يمكن توقعها عند عمل توازى بين تتابعين. وبناءعلى ذلك يتم حساب أعلى الأحتمالات ووضعها على رأس القائمة.

obeikandl.com



هناك عدد كبير جدا من الأحتمالات

$$O(2^{M+N})$$

شكل ٧٩: شكل يوضح شكل المصفوفة التى يمكن عن طريقها حساب أحتمالات التشابه بين تتابعين.

ونظرا لوجود أكثر من أحتمال ممكن عند عمل توازى بين تتابعيين فأنه قد لاتكون النتيجة التى قد يظهرها البرنامج فى البداية هى أفضل النتائج الممكنة (الشكل ٨٠) .

```
GAAT-C  GA-ATC
C-ATAC  CATA-C

GAAT-C  GAAT-C
-CATAC  CA-TAC
```

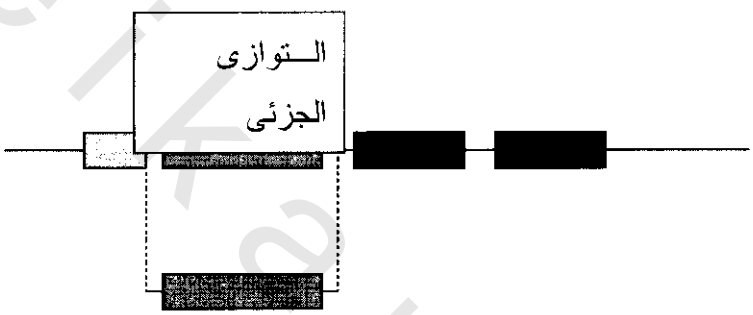
شكل ٨٠: شكل ستة أحتمالات مختلفة للتشابه بين تتابعين.

ونظرا لآن التطور الجزيئى فى الجينات المتناظره فى الكائنات المختلفة قد أشتمل على تغير فى أطوال هذه الجينات مع الأحتفاظ بالوظيفة لذلك لايتم غالبا حساب أى فراغات متواجدة عند الأطراف عند تصميم برامج التوازى الجزئى وذلك كما يظهر فى الشكل التالى . ويمكن وضح غير محدود من الفراغات عند الأطراف بدون التأثير على الناتج النهائى لحساب التوازى أو التشابه.

```
----------CTATCACCTGACCTCCAGGCCGATGCCCCTTCCGGC
GCGAGTTCATCTATCAC--GACCGC--GGTCG--------------
```

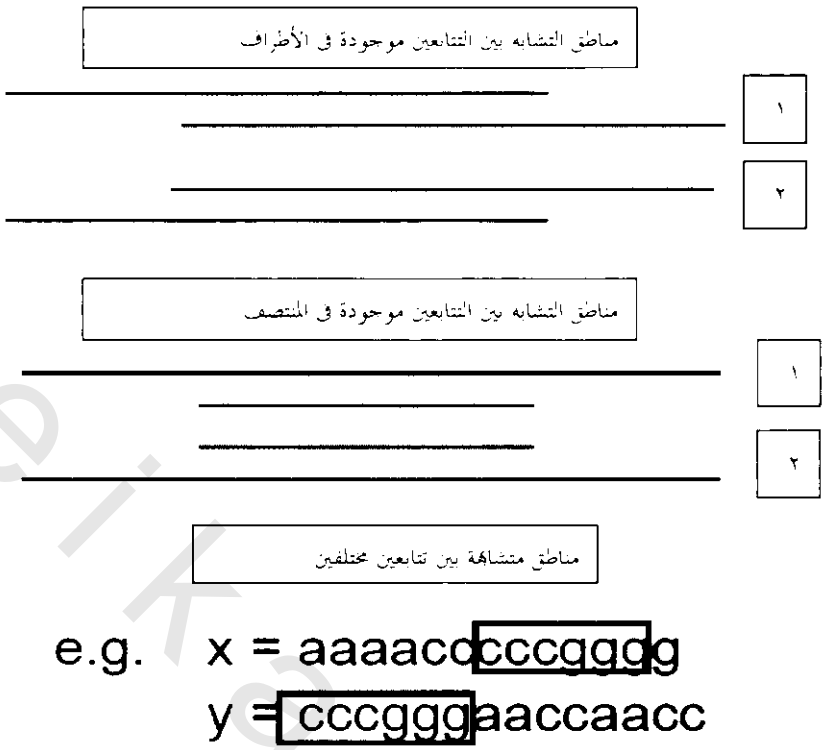
الشكل ٨١: شكل يوضح شكل الفراغات عند الأطراف التى يمكن تجاهلها عن حساب التشابه.

ونلاحظ أن الـتوازى الجزئى يمكنه أن يحدد بدقة مقدار التشابه والأختلاف بين تتابعات مختلفة فى الأطوال ولايحتاج أن تكون الأطوال متشابهة كما يوضح الشكل التالى:



شكل ٨٢: صورة توضح التوازى الجزئى بين تتابعين مختلفين فى الطول ولكن يحتويان على منطقة واحدة متشابهة.

ولذلك يمكن أستخدام التوازى الجزئى فى أكتشاف مناطق التشابه الوظيفى بين تتابعات بروتينية مختلفة الأطوال والتى تحتوى على مناطق فعالة ومتشابه وظيفيا. وقد يكون التشابهة الجزئى فى مناطق مختلفة فى التتابعين المدروسين ولكن ينجح التوازى الكلى فى أظهارها بنجاح كما يظهر فى الشكل (٨٣).

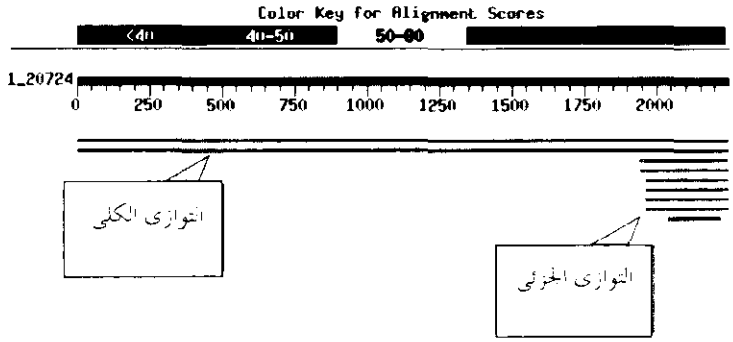


شكل (٨٣): يظهر الأماكن المختلفة التى قد يوجد بها التشابه الجزئى بين التتابعين سواء فى المنتصف أو الأطراف.

ويعمل التوازى الجزئى على معادلات بسيطة تعتمد فكرة حسابها على إيجاد مناطق متشابه بين تتابعين وهذه المناطق موجودة فى أماكن مختلفة فى كل منهما. ويفيد ذلك جدا فى أظهار التغيرات التى توجد بين نفس المجموعة من الجينات فى الكائنات المختلفة تطوريا. كما يفيد التشابه الجزئى فى الكشف عن المناطق الوظيفية المتشابهة فى البروتينات ذات الصلة.

وتجدر الأشارة أن بمقارنة جينومات الثدييات وجدأن هناك نسبة تشابه حوالى ٩٨٪ بين اى حيوانيين ثديين. وقد ظهرت مؤخرا مجموعة من المواقع والبرامج التى تقارن المحتوى الكلى لجينوم أحد الكائنات بجينوم أخر.

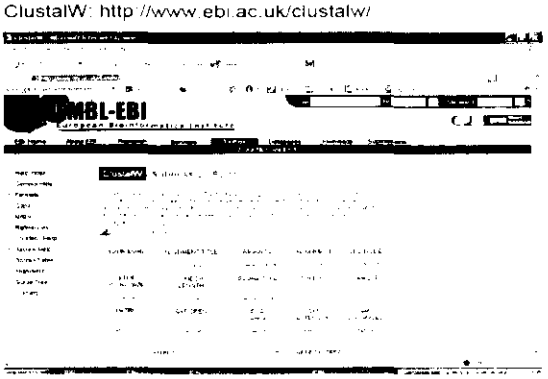
أما التوازى الكلى فيظهر الأختلاف الدقيق بين التتابعات متقاربة فى الأطوال والتركيب ولكن حدث فيها تغيير تطورى صغير يصعب ملاحظته.



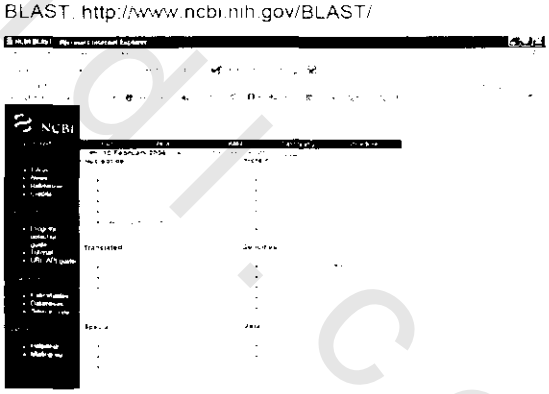
شكل ٨٤: صورة توضح الفرق التوازى الجزئى والتوازى الكلى

أفضل البرامج المستخدمة فى أظهار التوازى الجزئى هو برنامج BLAST وهو متواجد بشكل مجانى على موقع المركز الوطنى للتقنية الحيوية NCBI الموجود في الولايات كما يوجد أيضا أيضا برنامج Clustal Wعلى موقع المعمل الأوروبى للبيولوجيا الجزيئية(EMBL) والذى من خلاله يمكن أختيار عمل توازى جزئى أو كلى (شكل ٨٥).

ClustalW: http://www.ebi.ac.uk/clustalw/



BLAST: http://www.ncbi.nlm.gov/BLAST/



شكل ٨٥: أشهر البرامج وأكثرها أنتشارا لعمل التوازى الجزئى و التوازى الكلى.

ثانيًا: التوازى الكلى (Global Aligment)

يعتمد التوازى الكلى على فكرة التوازى بين كل حمض أمينى أو قاعدة نتروجينية فى التتابعات المدروسة على مستوى الأطوال الكلية لهذه التتابعات. التوازى الكلى هام جدا لعمل التوازى المتعدد للتتابعات (Multiple Sequence Alignments) ولكنه لايعطى كثيرا من المعلومات عند مقارنة تتابعيين.

التوازى الكلى أيضا غير مفيد على الأطلاق للكشف عن التشابه بين تتابعين لآن العمليات الأحصائية المستخدمة لحساب معامل الصدفة (E-Value) لاتنطبق عليه. فى التوازى الكلى أيضا لايحدث أختفاء مفاجى لآى حمض أميني أو قاعدة نتروجينية ولكن تكون التتابعات فى الناتج النهائى مشابهة تماما للتتابعات عند بداية التحليل مع أضافة بعض الفراغات عند الأختلاف فى الأطوال.

وهناك ثلاث أسباب رئيسية لعمل التوازى الكلى :

أولا: الكشف عن الاختلافات الصغيرة بين تتابعيين

وقد يحدث هذا مثلا مع تتابعات قد حدث لها تغيير عن طريق الخطأ. ويعتبر التوازى الكلى أفضل طريقة للتعرف أماكن الأختلافات الدقيقة بين تتابعين.

ثانيا: التعرف على تعدد المظاهر الوراثى (Genetic Polymorphism)

مثل الأختلافات فى قاعدة واحدة (SNPs) بين تتابعات وراثية قريبة الصلة.

ثالثا: مقارنة تتابعيين مختلفين و متداخلين جزئيا.

فى هذه الحالة يجب عمل التوازى الكلى للتتابعين مع عدم حساب قيم عدم التوازى عند الأطراف.

بعض المصطلحات الهامة فى التطور

Phylogenetics القرابة التطورية ما بين الأنواع المختلفة ويأخذ في الاعتبار أن هناك شيئان:

١ - هناك أصل لكل نوع.

٢ - هناك مسافة تطورية.

Cladogram: العلاقة بين الجينات المتشابهة التي لا يوجد بينها قرابة.

Orthologous: وجود نفس الجين بنفس الوظيفة في كائنات مختلفة تطورية مثل HSF.

Paralogous: موجود نسخة آخري من الجين في نفس الكائن نتيجة حدوث طفرات أو تضاعف وقد تقوم بوظيفة مختلفة.

Xanologous: حيث يتم إدخاله في كائن ولم يورث تطورياً مثل إدخال PT إلى داخل النباتات.

obeikandl.com