الفصل الثالث

السجلات واسترجاع المعلومات

يتناول هذا الفصل من الكتاب مهارات استرجاع المعلومات والتى تتيح الاستخدام الفعال لبنوك البيانات.

فهرسة قواعد البيانات وخصائص مصطلحات البحث:

الفهرس عبارة عن مجموعة من المؤشرات للمعلومات فى قواعد البيانات. عند البحث على شبكة المعلومات الدولية أو من خلال قاعدة بيانات للبيولوجيا الجزيئية يتم ادخال واحد أو أكثر من مصطلحات البحث ويقوم برنامج بالبحث عنها فى جداول الفهارس. وهنا يتعرف برنامج الاسترجاع على الموضوعات ذات المحتويات المتعلقة بمجال الاهتمام. على سبيل المثال عند ادخال كلمة حصان horse سوف تحصل على معلومات تختص بموضوعات عديدة تتعلق بالحصان مثل البيولوجيا الجزيئية والتربية والسلالات والقصائد الشعرية حول الحصان وغيرها. وللتركيز على موضوع محدد يسمح نظام استرجاع المعلومات باستخدام كلمات استرشادية. فمثلا بالبحث عن انزيم ديهيدروجينيز للكحول فى كبد الحصان horse liver alcohol dehydrogenase سوف يتم التعرف فقط على الموضوعات المحتوية على الكلمات الاسترشادية الأربع التى تم ادخالها وهى horse و liver و alcohol و dehydrogenase.

لذلك من المهم استخدام قواعد بيانات متخصصة بما فى ذلك قواعد بيانات البيولوجيا الجزيئية والتى تفرض تركيب معين على المعلومات لفصل فئات مختلفة من المعلومات.

تحليل البيانات المسترجعة: أحيانا يكون هناك حاجة الى الحصول على برنامج يسمح باستخدام النتائج – المسترجعة من عمليات بحث فى قواعد البيانات – كمدخلات. على سبيل المثال، تستخدم نتائج تتابع بروتين ما كمدخلات للتنقيب بواسطة برنامج PSI- BLAST. وفى هذه الحالة يجب تغذية البرنامج بنتيجة التتابع يدويا. الا أنه كما هو الحال فى التنقيب فى قواعد البيانات المتعددة، تتيح نظم المعلومات المسترجعة – فى البيولوجيا الجزيئية – تسهيلات لبدء تلك العمليات.

السجلات Archives: بالرغم من أن معرفتنا ببيانات التتابع والتركيب البيولوجى لم تكتمل بعد، الا أن حجم تلك البيانات ينمو سريعا. ويعمل العديد من العلماء على توفير البيانات أو تنفيذ مشروعات بحثية لتحليل النتائج. كما تقوم هيئات معنية بقواعد البيانات بأرشفة وتوزيع البيانات.

وتجرى عمليات أرشفة البينات المتعلقة بالمعلوماتية الحيوية بواسطة مجاميع بحثية مهتمة بالعلوم ذات الصلة.

وتتضمن عملية الجمع الأولى للبيانات المتعلقة بالجزيئات البيولوجية الكبيرة مايلى:

- تتابعات الحمض النووى بما فى ذلك مشروعات الجينوم.
- تتابعات الحمض الأمينى للبروتينات.
- تراكيب البروتين والحمض النووى.
- التراكيب البللورية للجزيئات الصغيرة.
- وظائف البروتين.
- طرز تعبير الجينات.
- المراجع والمؤلفات.

قواعد بيانات تتابع الحمض النووى:

يتكون الأرشيف العالمى لتتابع الحمض النووى من شراكة ثلاثية وهى: المركز القومى لمعلومات التكنولوجيا الحيوية بالولايات المتحدة الأمريكية، ومكتبة البيانات لمعهد البيومعلوماتية الأوروبى بالمملكة المتحدة (EMBL)، وبنك اليابان لمعلومات الدنا (المعهد القومى للوراثة باليابان). ويتم تبادل البيانات بين المجموعة يوميا. ولذلك فان البيانات الأولية تكون متطابقة بالرغم من اختلاف نماذج تخزين البيانات وطبيعة تفسيرها. وتقوم قواعد البيانات هذه بمعالجة وأرشفة وتوزيع تتابعات الدنا والرنا المجمعة من مشروعات الجينوم والمنشورات العلمية وتطبيقات براءات الاختراع. وتطلب المجلات العلمية ايداعات جديدة لتتابعات جديدة للنيكليوتيدات فى قاعدة البيانات كشرط لنشر المقالة وذلك للتأكد من حرية توفير تلك البيانات الأساسية. وكذلك الحال بالنسبة لتتابعات الحمض الأمينى وتراكيب الحمض النووى والبروتين.

تتكون قواعد بيانات تتابع الحمض النووى من مجموعات من المدخلات. لكل مدخل ملف نص يحتوى على البيانات والتفسيرات لكل تتابع مجلور. العديد من المدخلات تكون مجمعة من عدة أوراق علمية منشورة تقرر أجزاء متداخلة لتتابع كامل.

فيمايلى مثال لمدخلات تتابع دنا من مكتبة بيانات EMBL متضمنة تفسيرات وبيانات تتابع جين مثبط التربسين البنكرياسى:

The EMBL Data Library entry for the Bovine pancreatic trypsin inhibitor gene.

```
ID   BTBPTIG    standard; DNA; MAM; 3998 BP.
XX
AC   X03365; K00966;
XX
SV   X03365.1
XX
DT   18-NOV-1986 (Rel. 10, Created)
DT   20-MAY-1992 (Rel. 31, Last updated, Version 3)
XX
DE   Bovine pancreatic trypsin inhibitor (BPTI) gene
XX
KW      Alu-like  repetitive   sequence; protease inhibitor; trypsin
inhibitor.
XX
OS   Bos taurus (cow)
```

```
OC         Eukaryota;    Metazoa;  Chordata;  Craniata;  Vertebrata;
Euteleostomi; Mammalia;
OC       Eutheria;  Cetartiodactyla;  Ruminantia;   Pecora; Bovoidea;
Bovidae; Bovinae;
OC   Bos.
XX
RN   [1]
RP   1-3998
RX   MEDLINE; 86158754.
RA   Kingston I.B., Anderson S.;
RT    "Sequences encoding two trypsin inhibitors occur in strikingly
similar
RT   genomic environments";
RL   Biochem. J. 233:443-450(1986).
XX
RN   [2]
RX   MEDLINE; 84070725.
RA   Anderson S., Kingston I.B.;
RT       "Isolation of a genomic clone for bovine pancreatic trypsin
inhibitor by
RT   using a unique-sequence synthetic dna probe";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:6838-6842(1983).
XX
DR   SWISS-PROT; P00974; BPT1_BOVIN.
XX
CC   Data kindly reviewed (08-DEC-1987) by Kingston I.B.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..3998
FT                   /db_xref="taxon:9913"
FT                   /organism="Bos taurus"
FT   misc_feature    795..800
FT                   /note="pot. polyA signal"
FT   misc_feature    835..839
FT                   /note="pot. polyA signal"
FT   repeat_region   837..847
FT                   /note="direct repeat"
FT   misc_feature    930..945
FT                           /note="sequence homologous to Alu-like
consensus seq."
FT   repeat_region   1035..1045
FT                   /note="direct repeat"
FT   misc_feature    2456..2461
FT                   /note="pot. splice signal"
FT   CDS             2470..2736
FT                   /db_xref="SWISS-PROT:P00974"
FT                   /note="put. precursor"
FT                   /protein_id="CAA27062.1"
FT
/translation="PSLFNRDPPIPAAQRPDFCLEPPYTGPCKARIIRYFYNAKAGLCQ
FT                   TFVYGGCRAKRNNFKSAEDCMRTCGGAIGPWGKTGGRAEGEGKG"
FT   misc_feature    2488..2489
FT                   /note="pot. intron/exon splice junction"
FT   misc_feature    2506..2507
FT                   /note="pot. intron/exon splice junction"
FT   CDS             2512..2685
FT                   /db_xref="SWISS-PROT:P00974"
FT                   /note="trypsin inhibitor (aa 1-58)"
FT                   /protein_id="CAA27063.1"
FT
/translation="RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNF
FT                   KSAEDCMRTCGGA"
FT   misc_feature    2698..2699
FT                   /note="pot. exon/intron splice junction"
FT   misc_feature    3690..3695
FT                   /note="pot. polyA signal"
```

```
FT   misc_feature    3729..3733
FT                   /note="pot. polyA signal"
XX
SQ   Sequence 3998 BP; 1053 A; 902 C; 892 G; 1151 T; 0 other;
     aattctgata   atgcagagaa   ctggtaagga   gttctgattg   ttctgcttga
ttaaatgggt        60
     tgtaacagga   tagtgtcttg   tcctgatcct   agcattcata   tggtgtgtgt
tc:ggggcaa       120
     gtcatctgca   gtttcttcac   ctgaacaggg   ggaccaggtt   acatgagttt
cttaaaagat       180
     taccagtcat   gagtatgaag   agtttacact   ttcctgatca   atgacgtcca
tttcccatca       240
     aaatatttta   gtccaaaaga   ctcatctatc   taatgtagat   cattttctca
ccaccctct       300
     aaaaaattta   tctttcagat   atgatcattt   ctctattatg   aaattaatca
gagagttgag       360
     tgacagctga   gtgtcttccc   tccaaaggca   actgcaggaa   gagcaagaaa
tgcaatactt       420
     ttctatgagt   ttgctcgtgg   ggccaagact   gctttttcca   ggctggtaca
atagtaatca       480
     aatctcaaag   atattcttct   ttcctcctgg   ccagactatt   attttatttt
cctatcaaga       540
     tatagaaagt   tagaagtaga   ctcataatta   tataggcagg   cctcatcatc
aaatagacta       600
     acaagaattt   tatttttatct   gccttttcaa   tgactgtgca   cttggcatga
ggatgaaatg       660
     ggagatttat   tcccttgata   aatattcatg   aaatacttat   gctttttgtc
cctaaaaagc       720
     atatttcttg   atataggaaa   acagctgtaa   acaaaaggta   gtaaaataat
atgccttcta       780
     agagggatac   agacaataaa   gacgggggcg   gattcctata   ccaggtcatg
atgagtgtca       840
     tgaggaaggt   gagttatggg   gttcaggatg   ctgtagagga   tcagggaaac
cctctgtgat       900
     gaggagacat   taagcagaag   ctgccaaaaa   ggagcctggt   gtgtttgagc
agcagccagg       960
     accagggtgg   ctggagctga   gtgggtgagg   ggaggggagt   ggaaggggat
gaagcagaga      1020
     ggccatgggg   gcaggtcatg   aggaaccttc   taggacttta   taaggataaa
aattttgactc     1080
     tgagagagct   gggaaaccac   tgagggactg   gtcgttggaa   caacgagata
ggactggagt      1140
     tttaacaggg   tccttgcgac   tgcagtgtgg   agcgtggcct   ctaggggcg
aaaagcaggg      1200
     acaggggcc    cctgggcagg   tggctgcagg   ggtccagtga   gctatgatgg
agaatatata      1260
     cctgtgttgt   tccctgggtt   gattccagtt   ctcttgaata   acctgaatga
ct tgctatat     1320
     taagatatct   ggggaggctt   catcacaaat   catgattcat   taaatcttta
gtcatttgtg      1380
     attgattcaa   cctccaatcc   ctctcccctt   cctgagatgt   ggatacaatg
aaagaagtag      1440
     gaatgaaaat   tcccacaccc   aactcaggca   gttgtttccc   ctgacaactt
atcccccattc     1500
     ttggctttgc   ttgaggcttt   acaaaactca   tctccctcac   atgataaagg
actcccctttt     1560
     gctctcatct   cttaggaaat   tccaatgttt   taggagctct   gtggcaggaa
tgggatgcag      1620
     accaagttaa   tatttctttt   ataagtcaca   gtatcaatat   ttcctcaata
ttctattatt      1680
     ccagtctcca   tgaggtaacc   aaagtaaaca   ccggtgtgtt   ttctaccatg
cctttctcca      1740
     tttatggcat   gatttcctca   cacttttgta   atagtgacgg   gtcacgcagg
cctatcaacc      1800
     attggctggc   atccaggtgg   gcaccttcat   caagggataa   ctgtaaatga
gcaaccctttg     1860
```

gtggtccagt tcagccattg ccactgtcgt agccacagtg ggctcttcgc
cctcccgttc 1920
ttttgtataa aaggaacagg aatttatact gtgggaagat ggttttcttg
agacagtagc 1980
atgctatcat ctccgtgggc ccaatttcca attaaaaatg ttattcctag
ttccagcaac 2040
tcttctccgg attattggct ggccctgagg tgagcagaat gagactgggc
tcagtgatgc 2100
tttcttaacg gtggaagttt ccaccacaca catacataga aagcatagta
ttaaaaaagc 2160
cgtggatcca ttgtccagct ccagtaattt ctatacatgg agagtatttt
tatatgtgtc 2220
cctcttttgt gttacttttg aaactcatca gtagcatcat gctaattaat
gcataaacat 2280
tcatcaatgg catgtaatta tttataatat tgccctgtca ttgtcacacc
taacaacatt 2340
aataataatg tcctggaaag cagggtgtca aaaggccttt tcacgtttca
cacttctgcc 2400
caccccccat cactctctat cacaaactgg tggttttagt tgttcatctt
gtagactgag 2460
ctgtgatgac cttccctctt taaccgagat cctcccatcc ctgcagccca
gcggcctgac 2520
ttctgcctag agcctccata tacgggtccc tgcaaggcca gaattatcag
atacttctac 2580
aacgccaagg ctgggctctg ccagaccttt gtatatggcg gctgcagagc
taaaagaaac 2640
aatttcaaga gcgcagagga ctgcatgagg acctgtggtg gtgctattgg
gccctggggt 2700
aagacagggg gcagggcaga gggagagggg aagggctagg gaaagtggtg
gcgctcagaa 2760
ggccacacac ctttccaaaa aagtgatttt tttcccttgt tgcctcccaa
gagaagttgc 2820
agaagtatcc gtggattgag catgtcctcc atggaccagc ttggtgaaag
gccacccctt 2880
agaagccctg tcatcataat ctgagcctac tcacatgctc ccatttttca
gatgggaaca 2940
ctgagtcagt cactctgcag agcaagtctg gagtgctcct cagtgcccca
cctcagcctg 3000
gaaaactccc ttgtttattg ttggttatcc tggtcctggg aggactgtgg
ttgcgcattt 3060
ctgggatggt ctaggacctg tcagggtgga cagtgtccag gtctgggcct
tcagagatgt 3120
cattcagcaa gttccttttct ttttacagag aacctgtgaa ctgtgctccc
ctgagatgct 3180
gaagtatgag gaggacccac ccaaggctgg cctctatctg cttctgaaaa
atttcagcct 3240
ccttttattt cttctcaacc ctcccctcct cagcagaaat ctgtctcttt
cctcctcca 3300
caggtccact tactttagcc ctatctcatc cagtttgctc taagcaccat
gaaagcaaat 3360
cttccctttg tccctcacac ttcccacaat ttctggcaca aaggagaagg
tccagaaata 3420
ttggaggaag gaaggaatga agttccccat gactggagca tctgtagagt
ctgagattta 3480
aatctggatt cttgtcctaa tcttcctcct cacggcatcc ttaccttcat
cctccacccc 3540
accatcactg ctctccctct actggcgaaa gtagaatttc catcatcgag
ttttcagctc 3600
agtggtggga gaggtctttt catgaacgaa acctcctcct cacattgatt
tgaaggtctg 3660
tggcttcaaa gagtctggcc ttatctttaa ataaattcat attttaatta
aactaactgg 3720
agtggattgt gttgtttgca actaagaacc ttaacccata ggttccatgg
aaacggtggt 3780
ctttctcatt ttatgcagat gggtgggcag ctctccatca cctctcctca
gactcagccc 3840

```
     taccaagtag    aaggagccaa   ccccttacac   tgacatctac   ctcttatggc
cgtgccagtg      3900
     tacatgaaaa    actggatgag   agacacctca   acaagaaaac   ttttgtcctt
cacttcttgg      3960
     gccaggtcaa actttggggt gtgttatttc cctgaatt           3998
//
```

السطور التى تبدأ بFT هى عبارة عن مكون لتفسير المدخل الذى يقـــرر صفات مناطق متخصصة (تتابعات كودية CDS) بحيث تقـــرأ بواســطة برامج كمبيوتر. على سبيل المثال: لترجمة كود منطقة الى تتابع حمـــض أمينى، توجد نماذج متحكم فيها بعناية ومفردات مقيدة. ويكون من المـــهم وجود مفردات لغوية محكمة وقواميس وكلمات استرشادية وجداول مميزة حتى يمكن انشاء روابط بين قواعد بيانات مختلفة.

قواعد بيانات الجينوم:

بالرغم من أن تتابعات الجينوم تشكل مدخلات الأرشيفات المرجعية لتتـبع الحمض النووى، فان العديد من الأنواع لها قواعد بيانات خاصة والتـــى تجمع تتابع الجينوم وتفسيره مع باقى البيانات ذات العلاقة بالنوع.

مصادر على الانترنت Web Resources لروابط قواعد بيانات للكائنات:

http://www.unl.edu/stc-95/ResTools/biotools/biotools10.html

http://www.-fp.mcs.anal.gov/~gaasterland/genomes.html

http://www.hgmp.mrc.ac.uk/GenomeWeb/genome-db.html

http://www.bioinformatik.de/cgi-bin/browse/Catalog/Database/Genome_projects/

قواعـــد بيانـــات تتـــابع البروتين:

تتأتى بيانات تتابع الحمض الأمينى من ترجمة تتابعات الحمض النـــووى. يتعاون المعهد السويسرى للبيومعلوماتية مع مكتبة بيانات EMBL لـــتزود قاعدة بيانات تفسيرية لتتابعات الحمض الأمينى تســـمى SWISS-PROT. كما توجد قاعدة بيانات لتتابع البروتين تنتجـــها The PIR International والتى تشكل مجموعات فى المؤسسة القومية للبحـــوث الطبيـــة بجامعـــة جورج تاون واشنطن – الولايات المتحدة الأمريكية، ومركـــز معلومـــات

ميونيخ لتتابعات البروتين (MPIS) بألمانيا، وقاعدة بيانات اليابان الدولية لمعلومات البروتين.

وفيمايلى نموذج لمدخلات تتابع الحمض الأمينى لبروتين لمثبط التربسين البنكرياسى فى قاعدة بيانات PIR:

PIR entry for the amino-acid sequence of Bovine pancreatic trypsin inhibitor

```
ENTRY            TIBO  #type complete    iProClass View of TIBO
TITLE            basic proteinase inhibitor precursor - bovine
ALTERNATE_NAMES aprotinin; basic pancreatic trypsin inhibitor;
BPTI;
                 cationic kallikrein inhibitor; inhibitor IV
ORGANISM         #formal_name Bos primigenius taurus #common_name
cattle
   #cross-references taxon:9913
DATE             24-Apr-1984 #sequence_revision 22-Jul-1994
#text_change
                 16-Jun-2000
ACCESSIONS       S00277; A30333; S10546; S02486; S28197; A90162;
A92023;
                 A90736; A90927; A34658; A93977; S10062; A01205
REFERENCE        S00274
   #authors      Creighton, T.E.; Charles, I.G.
   #journal      J. Mol. Biol. (1987) 194:11-22
   #title        Sequences of the genes and polypeptide precursors
for two
                 bovine protease inhibitors.
   #cross-references MUID:87283904
   #accession    S00277
      ##molecule_type DNA; mRNA
      ##residues 1-100 ##label CR2
      ##cross-references GB:M20934; GB:X05274; NID:g162767;
                 PIDN:AAD13685.1; PID:g162769
REFERENCE        A90926
   #authors      Creighton, T.E.; Charles, I.G.
   #journal      Cold Spring Harb. Symp. Quant. Biol. (1987) 52:511-
519
   #title        Biosynthesis, processing, and evolution of bovine
                 pancreatic trypsin inhibitor.
   #cross-references MUID:88295740
   #accession    A30333
      ##molecule_type DNA
      ##residues 1-100 ##label CRE
      ##cross-references GB:M20934; GB:X05274; NID:g162767;
                 PIDN:AAD13685.1; PID:g162769
REFERENCE        S10546
   #authors      Kingston, I.B.; Anderson, S.
   #journal      Biochem. J. (1986) 233:443-450
   #title        Sequences encoding two trypsin inhibitors occur in
                 strikingly similar genomic environments.
   #cross-references MUID:86158754
   #accession    S10546
      ##molecule_type DNA
      ##residues 34-97 ##label KIN
REFERENCE        S02485
   #authors      Fioretti, E.; Angeletti, M.; Fiorucci, L.; Barra,
D.;
                 Bossa, F.; Ascoli, F.
   #journal      Biol. Chem. Hoppe-Seyler (1988) 369(Suppl.):37-42
```

```
   #title       Aprotinin-like isoinhibitors in bovine organs.
   #cross-references MUID:89076531
   #accession   S02486
      ##molecule_type protein
      ##residues 36-93 ##label FIO
REFERENCE       S28197
   #authors     Ikekita, M.; Jone, C.S.; Kamo, M.; Tsugita, A.;
Kizuki,
                K.; Moriya, H.
   #journal     Protein Seq. Data Anal. (1992) 5:7-11
   #title       Purification and characterization of the major
cationic
                kallikrein inhibitor in bovine pituitary gland.
   #cross-references MUID:93150003
   #accession   S28197
      ##molecule_type protein
      ##residues 36-93 ##label IKE
REFERENCE       A90162
   #authors     Kassell, B.; Laskowski, M.
   #journal     Biochem. Biophys. Res. Commun. (1965) 20:463-468
   #title       The basic trypsin inhibitor of bovine pancreas. V.
The
                disulfide linkages.
   #cross-references MUID:66083012
   #contents    annotation; disulfide bonds
   #accession   A90162
      ##molecule_type protein
      ##residues 36-93 ##label KAS
REFERENCE       A92023
   #authors     Anderer, F.A.; Hornle, S.
   #journal     J. Biol. Chem. (1966) 241:1568-1572
   #title       The disulfide linkages in kallikrein inactivator of
bovine
                lung.
   #cross-references MUID:66171231
   #contents    annotation; disulfide bonds
   #accession   A92023
      ##molecule_type protein
      ##residues 36-93 ##label AN2
REFERENCE       A90736
   #authors     Chauvet, J.; Acher, R.
   #journal     Bull. Soc. Chim. Biol. (1967) 49:985-1000
   #title       La structure covalente d'un inhibiteur
polypeptidique de
                la trypsine (inhibiteur de Kunitz et Northrop).
   #cross-references MUID:68012003
   #contents    annotation; disulfide bonds
   #accession   A90736
      ##molecule_type protein
      ##residues 36-93 ##label CHA
REFERENCE       A90927
   #authors     Dlouha, V.; Pospisilova, D.; Meloun, B.; Sorm, F.
   #journal     Collect. Czech. Chem. Commun. (1968) 33:1363-1365
   #title       Sequence of residues 18-20 in pancreatic trypsin
                inhibitor.
   #accession   A90927
      ##molecule_type protein
      ##residues 36-93 ##label DLO
REFERENCE       A93410
   #authors     Huber, R.; Kukla, D.; Ruhlmann, A.; Epp, O.;
Formanek, H.
   #journal     Naturwissenschaften (1970) 57:389-392
   #title       The basic trypsin inhibitor of bovine pancreas. I.
                Structure analysis and conformation of the
polypeptide
                chain.
   #cross-references MUID:70255230
```

```
   #contents     annotation; X-ray crystallography of basic protease
                 inhibitor, 2.5 angstroms
REFERENCE        A34658
   #authors      Lewis, R.V.; Ray, P.; Coguill, R.; Kruggel, W.
   #journal      Biochem. Biophys. Res. Commun. (1990) 167:543-547
   #title        Presence of pancreatic trypsin inhibitor in adrenal
                 medullary chromaffin cells.
   #cross-references MUID:90211226
   #accession    A34658
      ##molecule_type protein
      ##residues 36-53,55-81.##label LEW
REFERENCE        A93977
   #authors      Anderson, S.; Kingston, I.B.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1983) 80:6838-6842
   #title        Isolation of a genomic clone for bovine pancreatic
trypsin
                 inhibitor by using a unique-sequence synthetic DNA
probe.
   #cross-references MUID:84070725
   #accession    A93977
      ##molecule_type DNA
      ##residues 'PSLFNRDPPIPA',34-97,'GKTGGRAEGEGKG' ##label AND
      ##cross-references GB:X03365; GB:K00966; NID:g142;
PIDN:CAA27062.1;
                 PID:g1364183
REFERENCE        S00371
   #authors      Siekmann, J.; Wenzel, H.R.; Schroeder, W.;
Tschesche, H.
   #journal      Biol. Chem. Hoppe-Seyler (1988) 369:157-163
   #title        Characterization and sequence determination of six
                 aprotinin homologues from bovine lungs.
   #cross-references MUID:88221840
   #accession    S10062
      ##molecule_type protein
      ##residues 36-66,'P',68-82,'S',84-93 ##label SIE
      ##experimental_source lung
      ##note the authors designated this protein as isoaprotinin 2
COMMENT          Basic proteinase inhibitor is an intracellular
polypeptide
                 found in many tissues, probably located in granules
of
                 connective tissue mast cells.
GENETICS
   #introns      34/1; 98/1
CLASSIFICATION   #superfamily basic proteinase inhibitor; animal
                 Kunitz-type proteinase inhibitor homology
KEYWORDS         serine proteinase inhibitor
FEATURE
   1-20                    #domain signal sequence #status predicted
#label
                           SIG\
   21-35                   #domain propeptide #status predicted
#label PRO\
   36-100                  #product basic proteinase inhibitor
#status
                           experimental #label MAT\
   40-90                   #domain animal Kunitz-type proteinase
inhibitor
                           homology #label BPI\
   40-90,49-73,65-86       #disulfide_bonds #status experimental\
   50                      #inhibitory_site Lys (trypsin,
chymotrypsin,
                           kallikrein, plasmin) #status experimental
SUMMARY          #length 100 #molecular_weight 10903


SEQUENCE
```

```
            5        10        15        20        25        30
  1 M K M S R L C L S V A L L V L L G T L A A S T P G C D T S N
 31 Q A K A Q R P D F C L E P P Y T G P C K A R I I R Y F Y N A
 61 K A G L C Q T F V Y G G C R A K R N N F K S A E D C M R T C
 91 G G A I G P W E N L
```

---

in the PIR1 section of the Protein Sequence Database, release 71.00,
31-Dec-2001, assembled and annotated by the PIR-International.
Copyright 2000 PIR-International.

PDB structures most related to TIBO:
1CBWD (36-93) 100.0%; 1BZ5E (36-93) 100.0%; 9PTI (36-91) 100.0%
1BZXI (36-93) 100.0%; 1B0CB (36-93) 100.0%; 1CBWI (36-93) 100.0%
1B0CD (36-93) 100.0%; 1B0CE (36-93) 100.0%; 6PTI (36-93) 100.0%
1BHCA (36-93) 100.0%; 1BHCC (36-93) 100.0%; 1MTND (36-93) 100.0%
1BHCE (36-93) 100.0%; 1MTNH (36-93) 100.0%; 1BHCG (36-93) 100.0%
1PIT (36-93) 100.0%; 1BHCI (36-93) 100.0%; 1TPAI (36-93) 100.0%
1BPI (36-93) 100.0%; 2HEXA (36-93) 100.0%; 1BTHQ (36-93) 100.0%
2HEXB (36-93) 100.0%; 1BZ5B (36-93) 100.0%; 2HEXC (36-93) 100.0%
1BZ5D (36-93) 100.0%; 2HEXD (36-93) 100.0%; 1B0CC (36-93) 100.0%
2HEXE (36-93) 100.0%; 1BHCD (36-93) 100.0%; 2KAII (36-93) 100.0%
1BHCH (36-93) 100.0%; 2PTCI (36-93) 100.0%; 1BTHP (36-93) 100.0%
2TGPI (36-93) 100.0%; 1BZ5C (36-93) 100.0%; 2TPII (36-93) 100.0%
1BHCB (36-93) 100.0%; 3TGII (36-100) 100.0%; 1BHCJ (36-93) 100.0%
3TGJI (36-100) 100.0%; 1B0CA (36-93) 100.0%; 3TPII (36-93) 100.0%
1BZ5A (36-93) 100.0%; 4PTI (36-93) 100.0%; 1BHCF (36-93) 100.0%
5PTI (36-93) 100.0%; 1FAN (36-93) 98.3%; 1BPT (36-93) 98.3%
1NAG (36-93) 98.3%; 1BTI (36-93) 98.3%; 4TPII (36-93) 98.3%
8PTI (36-93) 98.3%; 1AALA (36-93) 96.6%; 1AALB (36-93) 96.6%
7PTI (36-93) 96.6%; 1BRBI (36-93) 94.8%; 1QLQA (36-93) 93.1%

SCOP: 1CBW ; 1BZ5 ; 9PTI ; 1BZX ; 1B0C ; 6PTI ; 1BHC ; 1MTN ; 1PIT ; 1TPA
; 1BPI ; 2HEX ; 1BTH ; 2KAI ; 2PTC ; 2TGP ; 2TPI ; 3TGI ; 3TGJ ; 3TPI ; 4PTI
; 5PTI ; 1FAN ; 1BPT ; 1NAG ; 1BTI ; 4TPI ; 8PTI ; 1AAL ; 7PTI ; 1BRB ; 1QLQ

CATH: 1CBW ; 1BZ5 ; 9PTI ; 1BZX ; 1B0C ; 6PTI ; 1BHC ; 1MTN ; 1PIT ; 1TPA
; 1BPI ; 2HEX ; 1BTH ; 2KAI ; 2PTC ; 2TGP ; 2TPI ; 3TGI ; 3TGJ ; 3TPI ; 4PTI
; 5PTI ; 1FAN ; 1BPT ; 1NAG ; 1BTI ; 4TPI ; 8PTI ; 1AAL ; 7PTI ; 1BRB ; 1QLQ

```
FSSP:   1CBW ; 1BZ5 ; 9PTI ; 1BZX ; 1B0C ; 6PTI ; 1BHC ; 1MTN ;
1PIT ; 1TPA
; 1BPI ; 2HEX ; 1BTH ; 2KAI ; 2PTC ; 2TGP ; 2TPI ; 3TGI ; 3TGJ ;
3TPI ; 4PTI
; 5PTI ; 1FAN ; 1BPT ; 1NAG ; 1BTI ; 4TPI ; 8PTI ; 1AAL ; 7PTI ;
1BRB ; 1QLQ


MMDB:   1CBW ; 1BZ5 ; 9PTI ; 1BZX ; 1B0C ; 6PTI ; 1BHC ; 1MTN ;
1PIT ; 1TPA
; 1BPI ; 2HEX ; 1BTH ; 2KAI ; 2PTC ; 2TGP ; 2TPI ; 3TGI ; 3TGJ ;
3TPI ; 4PTI
; 5PTI ; 1FAN ; 1BPT ; 1NAG ; 1BTI ; 4TPI ; 8PTI ; 1AAL ; 7PTI ;
1BRB ; 1QLQ


ALIGNMENTS containing TIBO:
     FA2061 basic proteinase inhibitor - 328.8 1.0
     SA0572 basic proteinase inhibitor superfamily 328.8
     M01603 basic proteinase inhibitor - 1561.0 1.0
Associated Alignments:
     DA1053 animal Kunitz-type proteinase inhibitor homology
```

قواعد البيانات المشاركة لـ SWISS-PROT:

يوجد اثنان من قواعد البيانات المشاركة ل SWISS-PROT وهما ENZYME DB & PROSITE.

يقوم ENZYME DB بتخزين المعلومات التالية عن الانزيمات:

- EC Number وهو وسيلة تعريف رقمية تم اقرارها بواسطة مفوضية الانزيمات المنبثقة من الاتحاد الدولى للكيمياء الحيوية والبيولوجيا الجزيئية، (يمكن الرجوع الى الموقع التالى: http://www.chem.qmw.ac.uk/iubmb/enzyme/)
- الاسم الموصى به.
- الاسم البديل، ان وجد.
- النشاط المحفز.
- المرافقات.
- مؤشرات الى SWISS-PROT وبنوك المعلومات الأخرى.
- مؤشرات الى الأمراض المصاحبة للخلل فى النشاط الانزيمى.

والمثال التالى يوضح المخلات فى ENZYME DB حيث:

ID = التعريف

DE = الوصف = الاسم الرسمى

AN = الاسم البديل

CA = النشاط التحفيزى

CF = المرافقات

CC = التعليقات

DR = مرجع قاعدة البيانات (SWISS-PROT).

A Sample Entry in ENZYME DB

```
ID 1.14.17.3
DE PEPTIDYLGLYCINE MONOOXYGENASE.
AN PEPTIDYL ALPHA-AMIDATING ENZYME.
AN PEPTIDYLGLYCINE 2-HYDROXYLASE.
CA PEPTIDYLGLYCINE + ASCORBATE + O(2) = PEPTIDYL(2-HYDROXYGLYCINE)+
CA DEHYDROASCORBATE + H(2)O.
CF COPPER.
CC -!- PEPTIDYLGLYCINES WITH A NEUTRAL AMINO ACID RESIDUE IN THE
CC PENULTIMATE POSITION ARE THE BEST SUBSTRATES FOR THE ENZYME.
CC -!- THE ENZYME ALSO CATALYZES THE DISMUTATATION OF THE PRODUCT TO
CC GLYOXYLATE AND THE CORRESPONDING DESGLYCINE PEPTIDE AMIDE.
DR P10731, AMD_BOVIN ; P19021, AMD_HUMAN ; P14925, AMD_RAT ;
DR P08478, AMD1_XENLA; P12890, AMD2_XENLA;
```

PROSITE :

تحتوى على طرز لمتبقيات مجاميع البروتينات مثل الطراز (البصمة أو القالب) الذى يظهر عادة فى عائلة من البروتينات القريبة بسبب متطلبات أماكن الارتباط والتى تحد من التحول فى عائلة بروتين ما.

PIR وقواعد البيانات المشاركة:

يدير PIR عدة قواعد بيانات خاصة بالبروتينات مثل:

- PIR-PSD : وهى قاعدة البيانات الأساسية لتتابع البروتين.
- iProClass : تقسيم البروتينات طبقا للتركيب والوظيفة.

- ASDB : قاعدة بيانات التفسير والمماثلة، وكل مدخل متصل بقائمة من التتابعات المماثلة.
- P/R-NREF : تجميع لأكثر من ٨٠٠٠٠٠ تتابع بروتين من مصادر متاحة.
- NRL3D : قاعدة بيانات لتتابعات وتفسيرات لبروتينات معروفة التركيب وموجودة ببنك معلومات البروتين.
- ALN : قاعدة بيانات لمحازاة تتابع البروتين.
- RESID : قاعدة بيانات لتحورات التركيب التساهمى للبروتين.

وقد أنشأ PIR موقع IESA (Integrated Environment for Sequence Analysis) لاسترجاع المعلومات والمعاملة الحسابية لها. ويمكن الرجوع الى موقع PIR على الوب:

http://pir.georgetown.edu

قواعد بيانات التراكيب:

يقوم أرشيف قواعد بيانات التركيب بتفسير وتوزيع مجاميع من المنسقات الذرية. ويعتبر بنك بيانات البروتين Protein Data Bank (PDA) من أفضل قواعد البيانات الخاصة بتراكيب الجزيئات البيولوجية الكبيرة. ويحتوى البنك على تراكيب بروتينات وأحماض نووية و قليل من الكربوهيدرات. ويدار البنك بواسطة the Research Collaboratory for Structural Bioinformatics (RCSB).

والموقع الالكترونى الأساسى لبنك بيانات البروتين هو: http://www.rcsb.org. وهناك مواقع رسمية للبنك لأماكن عديدة حول العالم منها أوروبا، سينغافورة، اليابان، والبرازيل.

وتحتوى الصفحة الأساسية للبنك على روابط اتصال ملفات البيانات ذاتها، ومواد تعليمية، أخبار، نشرات دورية وبرامج بحث متخصصة لمعالجة التراكيب.

مثال: يمكن الحصول على بيانات بروتين ثيوردوكسين فى بكتريا *E. coli* (Protein data bank entry 2trx, *E. coli* thioredoxin) من خلال الموقع التالى:

http://www.rcsb.org/pdb/cgi/explore.cgi?job=download;pdbId=2TRX;page=0;pid=197301010382789&opt=show&format=PDB

وتضمن المدخلات البيانات التالية:

- نوع البروتين ومصدره.
- مراجع لوصف طرق تحديد التركيب.
- بيانات تفصيلية لتجارب تحديد التركيب متضمنة نتائج الأشعة السينية واحصائيات للشكل الفراغي للتركيب.
- تتابع الحمض الأمينى.
- الجزيئات الاضافية مثل المرافقات، المثبطات، جزيئات الماء.
- التركيب الثانوي: الشكل المطوى، المنفرد.
- المواقع ثنائية الكبريت.
- المنسقات الذرية.

وقد حدد بنك بيانات البروتين كود تعرف رباعى لكل تركيب موجود بالبنك، يبدأ برقم (من ١ الى ٩). وأصبح من السهل استرجاع تركيب ما وذلك باستخدام كود التعريف الخاص به. حيث يمكن ادخال الكود على الصفحة الرئيسية لموقع RCSB ثم اختيار Explorer للحصول على ملخص لبيانات التركيب من صفحة واحدة كما هو مبين:

Protein data bank entry Structure Explorer - 2trx, *E. coli* thioredoxin
http://www.rcsb.org/pdb/cgi/explore.cgi?pid=197301010382789&page=0&pdbId=2TRX

RCSB PDB
PROTEIN DATA BANK

Structure Explorer - 2TRX

Summary Information

Summary Information
View Structure
Download/Display File
Structural Neighbors
Geometry
Other Sources
Sequence Details
Crystallization Info

Top of Form
Explore
SearchLite SearchFields
Bottom of Form

Title: **Crystal structure of thioredoxin from Escherichia coli at 1.68 A resolution.**

Compound: **Thioredoxin**

Authors: **S. K. Katti, D. M. LeMaster, H. Eklund**

Exp. Method: **X-ray Diffraction**

Classification: **Electron Transport**

Source: **Escherichia coli**

Primary Citation: Katti, S. K., LeMaster, D. M., Eklund, H.: **Crystal structure of thioredoxin from Escherichia coli at 1.68 A resolution.** ***J Mol Biol*** **212** ***pp.*** **167 (1990)**

Deposition Date: **19-Mar-1990** *Release Date:* **15-Oct-1991**

Resolution [Å]: **1.68** *R-Value:* **0.165**

Space Group: **C 2**

Unit Cell: dim [Å]: *a* **89.50** *B* **51.06** *c* **60.45**
alpha **90.00** *Beta* **113.50** *gamma* **90.00**

Polymer Chains: **A, B** *Residues:* **216**

Atoms: **1842**

Chemical Components: ("HET" groups)

ID (needs Rasmol)	Name	Formula	Retrieve All PDB IDs Containing
CU	COPPER (II) ION	2(Cu)	CU
MPD	2-METHYL-2,4-PENTANEDIOL	7($C_6H_{14}O_2$)	MPD

CATH: Structural Classification

PDBSum: Summary of PDB Structure

SCOP: Structural Classification

RCSB

والصفحة السابقة بها روابط اتصال للمعلومات التالية:

- مراجع للبيانات الموجودة من خلال قاعدة بيانات المراجع PubMed.
- صور للتراكيب ويحتاج ذلك وجود برنامج لاظهار الصور.
- مداخل لملفات البيانات.
- قوائم بالتراكيب الأخرى ذات الصلة طبقا لتصنيف تراكيب البروتين.
- التحليل الكيميائى الفراغى، وتوزيع أطوال الروابط والزوايا.
- مصادر لمعلومات حول المدخلات.
- التتابع والتراكيب الثانوية.
- تفاصيل حول الشكل البللورى وطرق انتاج البللورت.

وفى حالة عدم معرفة كود التعرف يمكن استخدام أداة بسيطة من على الصفحة الرئيسية لبنك بيانات البروتين وهى Search Lite والتى تسمح باستخدام كلمات استرشادية key words.

مصادر على الانترنت Web Resources لتراكيب البروتين والحمض النووى:

الصفحة الرئيسية لبنك بيانات البروتين:

http://www.rcsb.org

الصفحة الرئيسية لقاعدة بيانات EBI التركيب الجزيئى الماكرو:

http://msd.ebi.ac.uk/

الصفحة الرئيسية ل BioMagResBank:

http://www.bmrb.wisc.edu/

للبحث فى بنك بيانات البروتين:

الصفحة الرئيسية ل SCOP (Structural Classification of Protein):

http://scop.mrc-lmb.cam.ac.uk/scop/

قائمة بأماكن التصفح:

http://pdp-browsers.ebi.ac.uk/browse_it.shtml

أداة البحث OCA:

http://oca.ebi.ac.uk/oca-bin/ocamain

قاعدة بيانات التركيب الرباعى للبروتين:

http://pqs.ebi.ac.uk/

تقارير جودة التركيب:

http://www.cmbi.kun.nl/gv/pdbreport

مصادر على الانترنت Web Resources لقواعد بيانات لعائلات معينـــة من بروتين:

Protein kinase :

http://www.sdsc.edu/kinases/

HIV proteases

http://www-fbsc.ncifcrf.gov/HIVdb/

Icosahedral viruses :

http://mmtsb.scripps.edu/viper/main.html

قاعدة بيانات علم المناعة Immunology :

IGMT : (International ImMunoGene Tics database):

http://imgt.cines.fr

وهى قاعدة بيانات متكاملة عالية الجودة، ومتخصصة فى:

Immunoglobulins (Ig)، و T-Cell receptors (TcR) ، و

Histocompatibility Complex (MHC) molecules of all vertebrate species.

KABAT : http://immuno.bme.nwu.edu./

وهى قاعدة بيانات لتتابعات البروتينات ذات الأهمية المناعية.

MHCPEP : http://wehih.wehi.edu.au/mhcpep/

وهى قاعدة بيانات لمعقد البيبتيدات الرئيسية المرتبطة والمتوافقة هيستولوجيا.

مجموعات من روابط الاتصال لقواعد بيانات لعائلات معينة من البروتين:

http://www2.ebi.ac.uk/msd/Links?family.shtml

قواعد بيانات ناتج تعبير الجين والبروتيومكس Expression and proteomics databases:

من المعرف أن الدنا يصنع الرنا ويقوم الرنا بصنع البروتين. وتحتوى قواعد بيانات الجينوم على تتابعات الدنا. كما تسجل قواعد بيانات التعبير الجينى مقاييس مستويات الرنا الرسول mRNA وذلك عبر (short terminal sequences of cDNA synthesized from mRNA) ESTs مع وصف طرز نسخ الجين. بينما تسجل قواعد بيانات البروتيومات proteomics مقاييس للبروتينات مع وصف طرز ترجمة الجين.

تقدم مقارنات طرز التعبير الجينى معلومات موثقة عن:

(١) وظيفة وآلية فعل نواتج الجين.

(٢) كيف تنسق الكائنات عملية التحكم فى العمليات الأيضية تحت ظروف مختلفة. على سبيل المثال: الخميرة تحت ظروف هوائية ولاهوائية.

(٣) الاختلافات فى تعبئة الجين أثناء مراحل مختلفة من دورة الخلية وتطور الكائن.

(٤) آليات المقاومة للمضادات الحيوية فى البكتريا وبالتالى اقتراح أماكن مستهدفة لاستحداث أدوية.

(٥) الاستجابة لتحدى طفيل ما.

(٦) الاستجابة لأنواع وجرعات مختلفة من الأدوية للوصول الى علاج فعال.

وهناك عدة قواعد بيانات ل ESTs، تحتوى مدخلات معظمها على مجالات تشير الى موقع نسيج و/ أو مكون تحت خلوى، حالة التطور، ظروف النمو، والمستوى الكمى للتعبير. حاليا تحتوى dbEST من خلال بنك الجين على ما يقرب من تسعة ملايين مدخل من ٣٤٨ نوع.

بعض مجموعات EST متخصصة فى أنسجة معينة (عضلات، أسنان) أو أنواع. كما توجد جهود لربط طرز التعبير بمعلومات أخرى حول الكائن. فعلى سبيل المثال ينسق مشروع تطور الجرذان بين بيانات تعبير الجين والتشريح التطورى.

تتيح العديد من قواعد البيانات وسائل للربط بين ESTs فى أنواع مختلفة مثل ربط التماثل فى الانسان والجرذان، أو العلاقات بين جينات أمراض الانسان وبروتينات الخميرة. وهناك مجموعات أخرى متخصصة فى نوع من البروتين مثل سيتوكينات. كما يوجد جهدا هائلا للتركيز على السرطان من حيث تكامل المعلومات عن الطفرات واعادة التنظيم للكروموسومات والتغيرات فى طرز التعبير وذلك للتعرف على التغيرات الوراثية أثناء تكوين الأورام ونموها.

بالرغم من العلاقة الشديدة بين طرز النسخ وطرز الترجمة، الا أن القياسات المباشرة للمحتوى البروتينى للخلايا والأنسجة – البروتيومكس proteomics تتيح معلومات أضافية قيمة. وبسبب عدلات النسخ المختلفة لمختلف الرنا الرسول فان قياسات البروتينات تعطى مباشرة وصفا أكثر دقة لطرز تعبير الجين بالمقارنة بقياسات النسخ. ويمكن اكتشاف تحورات ما بعد الترجمة فقط عن طريق فحص البروتينات.

يتضمن تحليل البروتيوم الفصل والتعريف والتقدير الكمى لبروتينات العينة. ويعتمد ذلك على فرد البروتين بواسطة الجيل ثنائى الأبعاد

وتعريف كل مكون بواسطة مطياف الكتلة. وتقوم قواعد بيانات البروتيوم بتخزين صور الجيل وتفسيراتها على هيئة طرز بروتين. كما تظهر بعض قواعد البيانات صورا وتسمح باختيار تفاعلى للنقط. وباختيار نقطة معينة تفتح نافذة لمدخلات البيانات المتطابقة. ويوجد مدخل بيانات لكل بروتين يسجل المعلومات التالية:

- تعريف البروتين.
- الكمية.
- الوظيفة.
- آلية الفعل.
- طراز التعبير.
- مكان التواجد تحت الخلوى.
- البروتينات ذات الصلة.
- تحورات ما بعد الترجمة.
- التفاعلات مع البروتينات الأخرى.
- روابط اتصال بقواعد بيانات أخرى.

وقد ساهمت البيومعلوماتية فى انشاء وتطوير قواعد البيانات هذه، وكذلك ايجاد نظم خوارزمية لمقارنة وتحليل طرز البروتينات المحتوية عليها تلك القواعد.

قواعد بيانات المسارات الأيضية Databases of metabolic pathways:

تقوم قاعدة KEGG (Kyoto Encyclopedia of Gens and Genomes) بجمع الجينوم، ونواتج الجين ووظائفها، وعمل تكامل بين المعلومات البيوكيميائية والوراثية. وتركز KEGG على تفاعلات شبكات التجميع الجزيئى والشبكات الأيضية والمنظمة.

١) وتنظم KEGG خمسة أنواع من البيانات من خلال نظام شامل وهى:

٢) كتالوج للجينات يحتوى على جزيئات وتتابعات معينة.

٣) خرئط الجينوم. حيث يتم التكامل بين الجينات طبقا لظهورهم على الكروموسومات.

٤) خرائط المسارات والتى تصف شبكات الأنشطة الأيضية والمنظمة للجزيئات. وتساهم فى امكانية ايجاد مسار ايضى حقيقى فى كائن معين عن طريق مضاهاة بروتينات هذا الكائن مع انزيمات مسارات مرجعية.

٥) جداول أرثولوج. تستخدم بيانات تلك الجداول لربط انزيم فىكائن ما بالانزيمات ذات الصلة فى كائنات أخرى. ويتيح ذلك تحليل العلاقات بين المسارات الأيضية فى كائنات مختلفة.

تتيح KEGG امكانية أخذ مجموعة من الانزيمات من كائن واختبار مدى تكاملها مع مسارات أيضية معروفة. وفى حالة ظهور فراغ فى مسارما فان ذلك قد يرجع الى غياب انزيم أو مسار أيضى غير متوقع.

مداخل الى السجلات :

تمتلك قواعد بيانات الحمض النووى وتتابعات البروتين امكانات هائلة لاسترجاع المعلومات وتحليلها من خلال عديد من العمليات والتى تتضمن:

١. استرجاع التتابعات من قاعدة البيانات.

٢. مقارنة التتابع.

٣. ترجمة تتابعات الدنا الى تتابعات بروتين.

٤. تحليل التركيب والتنبؤ.

٥. طرز التعرف.

٦. الأشكال الجزيئية. وتستخدم تلك الأشكال فى:

- خرطنة الأجزاء التى يعتقد أن لها وظيفة ما الى اطار ثلاثى الأبعاد لبروتين.
- تقسيم ومقارنة طرز الطى فى البروتينات.
- تحليل الاختلافات بين التراكيب شديدة القرابة أو بين الأوضاع الفراغية لجزئ ما.
- دراسة تفاعل جزئ بسيط مع بروتين كمحاولة لتحديد وظيفة أو لاكتشاف دواء.
- الحصول على نموذج تفاعلى لتحسين جودة صور الأشعة السينية لمقاييس تراكيب البروتين.
- تصميم ونمذجة تراكيب جديدة.

طرق الدخول إلى قواعد بيانات البيولوجيا الجزيئية:

ENTREZ: هى نقطة البداية لاسترجاع تتابعات وتراكيب من على الموقع التالى:

http://www.ncbi.n1m.nih.gov/ENTREZ/

وتقدم ENTREZ طرق دخول عبر الأقسام الآتية لقواعد البيانات:

- Protein
- Peptide
- Nucleotide
- Structure
- Genome
- Popset- information about populations
- OMIM- Online Mendelian Inheritance in Man

فمثلا للبحث فى قاعدة البروتين: ادخل على موقع ENTREZ، ثم اختر بروتين، ثم ادخل نوع البروتين المطلوب البحث عنه، ثم اضغط على GO وسيقوم البرنامج باظهار الاجابات.

وللبحث فى قاعدة بيانات النيكليوتيدات: اختر نيكليوتيد من على موقع ENTREZ واضغط على INDEX ثم اختر الكائن من القائمة وادخل اسم الموضوع فى صندوق البحث، ثم اضغط على AND ثم اتبع التعليمات من القائمة.

OMIM:

وهى قاعدة بيانات للجينات البشرية والعيوب الوراثية. وتعمل بتكامل تام مع ENTREZ.

نظام استرجاع التتابع The Sequence Retrieval System (SRS):

يستطيع نظام SRS البحث فى ١٤١ من قواعد بيانات البروتين، وتتابعات النيكليوتيد، والمسارات الأيضية، والتراكيب ثلاثية الأبعاد، والوظائف، والجينوم، والأمراض.

مصدر تعريف البروتين The Protein Identification Resource (PIR):

يمكن البحث باستخدام SRS من على المواقع التالية :

فى الولايات المتحدة الأمريكية :

http://www.-nbrf.georgetown.edu/pirwww/search/textpsd.html

فى أوروبا: http://www.mips.gsf.de

ExPASY-Expert Protein Analysis System :

وهو نظام استرجاع وتحليل المعلومات لمعهد البيومعلوماتية السويسرى بالتعاون مع معهد البيومعلوماتية الأوروبى والذى ينتج قواعد بيانات تتابع البروتين SWISS-PROT وقاعدة TrEMBL التى تحتوى على تراجم تتابعات النيكلوتيد. وفتح الصفحة الرئيسية ل ExPASY

(http://www.expasy.ch) واختيار SWISS-PROT و TrEMBL يعطى مداخل توصل الى أدوات استرجاع المعلومات.

:ENSEMBL

وهو عبارة عن مصدر معلومات شامل للجينوم البشرى (http://www.ensemble.org). وتهدف الى جمع وتفسير كل المعلومات المتاحة عن تتابعات الدنا البشرى، وروابط اتصال الى تتابع جينوم قياسى، واتاحة هذه المعلومات لعديد من العلماء. ويمكن للمستخدم تعريف مناطق فى التتابع عبر عدة أدوات بحث:

- BLAST ويستخدم للبحث عن تتابع أو جزء منه.
- التصفح، ويبدأ على مستوى الكروموسوم ثم يتعمق داخله.
- العلاقة بالأمراض بواسطة OMIM.
- ENSEMBL ID.
- أداة بحث عامة للنصوص.