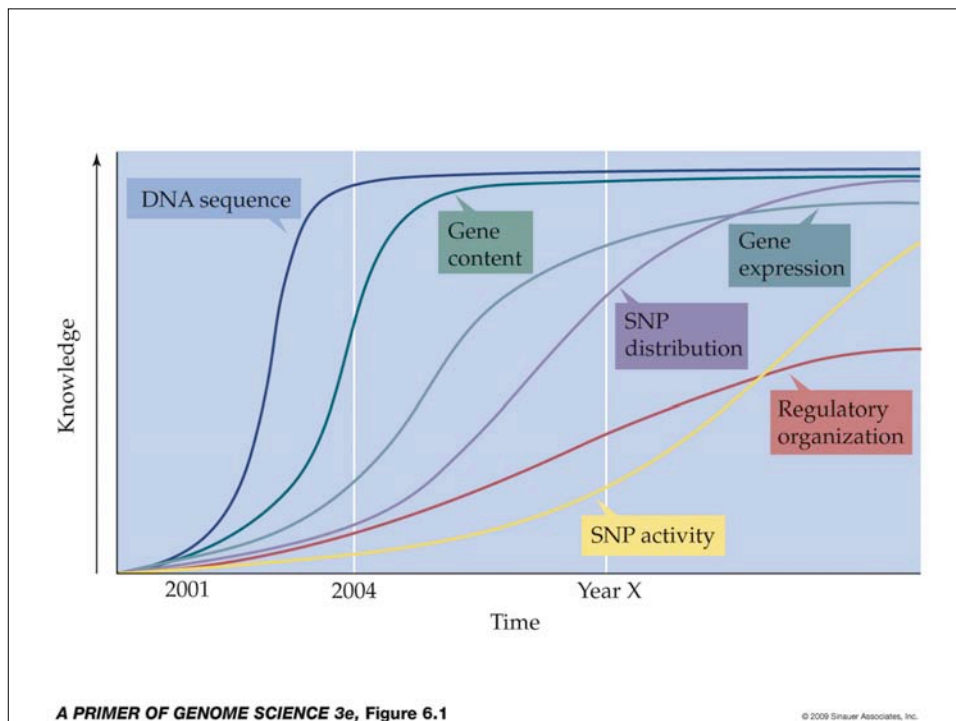


Proteomics and functional genomics

Biosciences 741: Genomics
Fall, 2011
Week 10



Functional Genomics

- The field of “functional genomics” represents the next stage in genomics, in which we will move beyond the (relatively) simple clerical exercise of “sequence annotation” to experimental determination of biological function(s) at the whole genome level.
- In most cases, functional genomics is based on the methods of molecular genetics, but adapted to high throughput, in one of the following ways:
- “forward genetics” uses random mutagenesis to identify the set of genes that affect any given trait.
- “reverse genetics” uses molecular sequence information to construct loss-of-function mutations without needing to know the phenotype.
- “fine structure genetics” is used to systematically manipulate the structure and function of genes, in order to obtain a more detailed picture of the functions of each part of the molecule, or the functions in each part of the body, etc.
- Chip-on-chip and DNA microarrays are also part of functional genomics.

Forward genetics

PROTEOMICS AND FUNCTIONAL GENOMICS 303

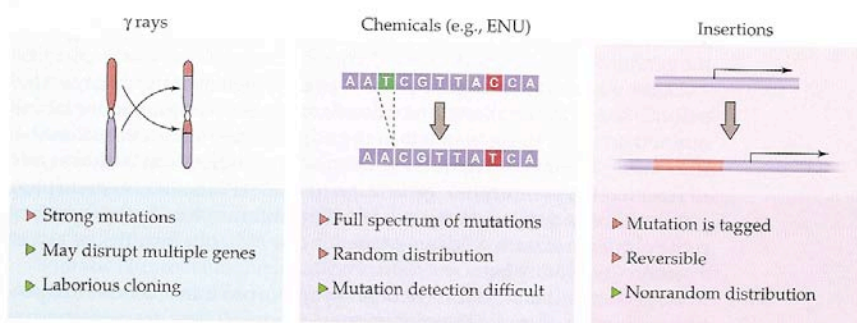
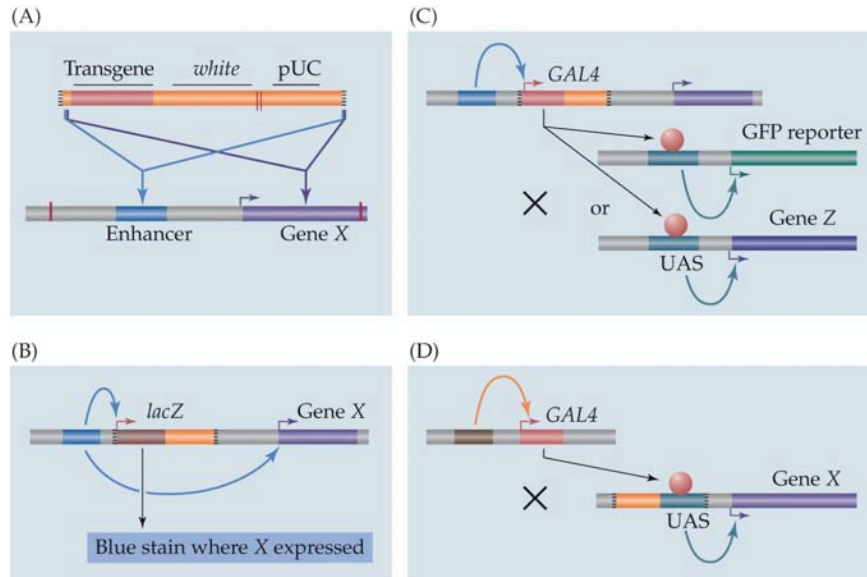


Figure 5.18 The three major types of mutagen. Gamma irradiation induces chromosome inversions, translocations, and large deletions. Chemicals such as ENU induce point mutations and deletions of a few bases. Transposable element insertions can occur anywhere in a gene, but are often found in upstream regulatory regions. Advantages (red arrowheads) and disadvantages (green arrowheads) of each mutagen are indicated.

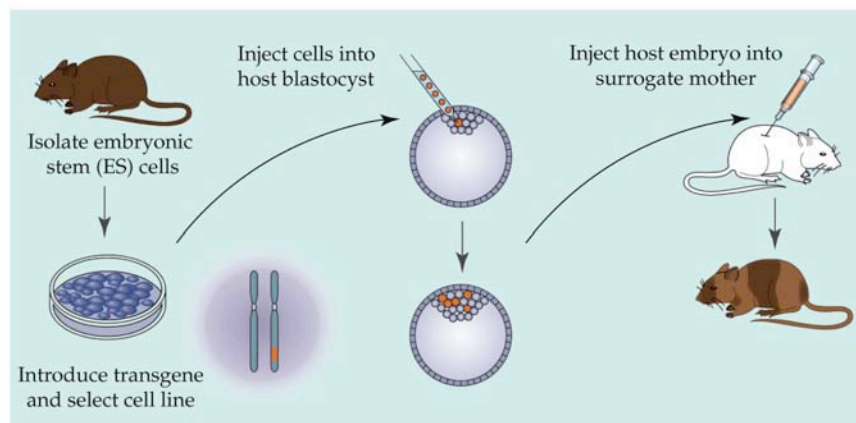
Transposon mutagenesis, enhancer trap screens, the GAL4-UAS expression system



A PRIMER OF GENOME SCIENCE 3e, Figure 5.24

© 2009 Sinauer Associates, Inc.

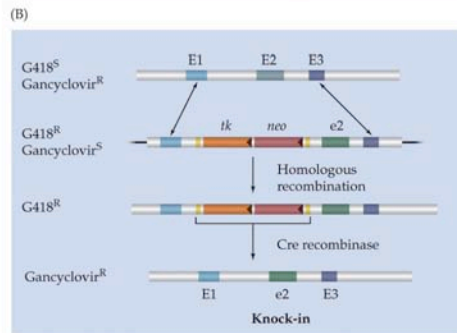
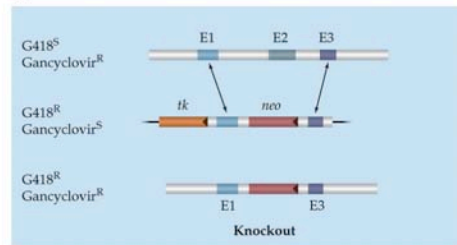
Box 5.3 - production of transgenic mice from ES cells



A PRIMER OF GENOME SCIENCE 3e, Box 5.3, Figure B

© 2009 Sinauer Associates, Inc.

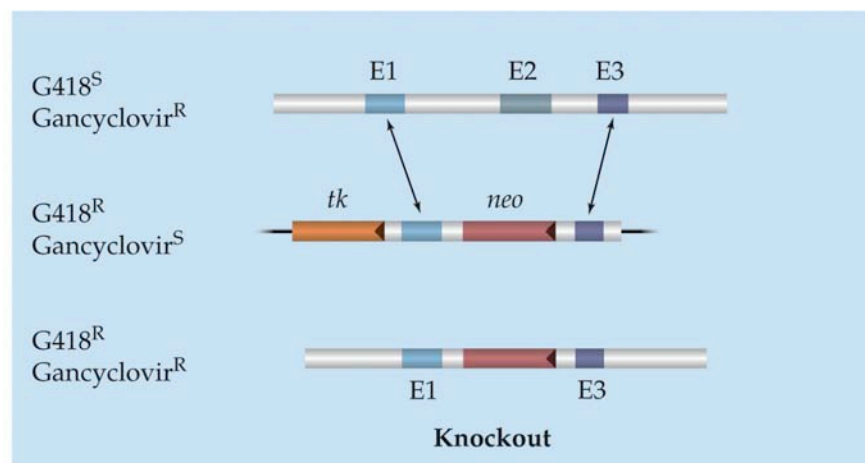
Figure 5.20 - Analyzing gene function with targeted mutations (transgenic knockout and knock-in mice)



A PRIMER OF GENOME SCIENCE 3e, Figure 5.20

© 2009 Sinauer Associates, Inc.

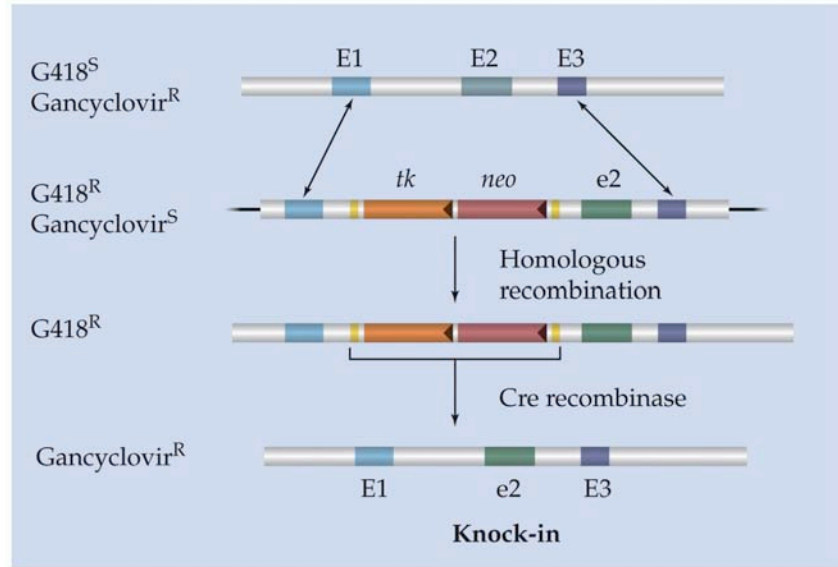
(A)



A PRIMER OF GENOME SCIENCE 3e, Figure 5.20 (Part 1)

© 2009 Sinauer Associates, Inc.

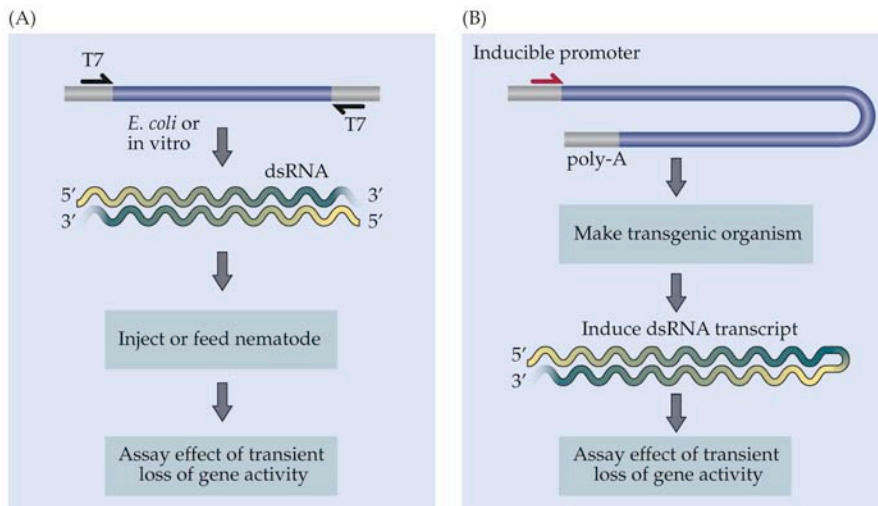
(B)



A PRIMER OF GENOME SCIENCE 3e, Figure 5.20 (Part 2)

© 2009 Sinauer Associates, Inc.

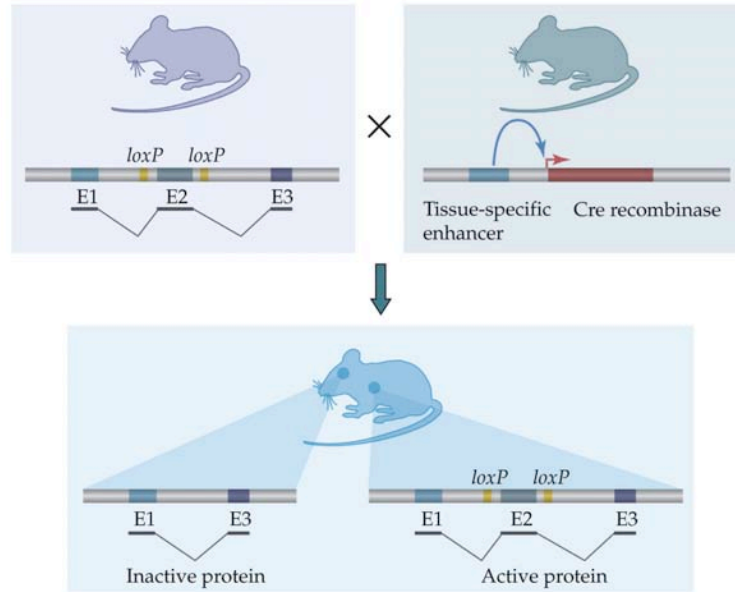
Figure 5.21 - RNA interference



A PRIMER OF GENOME SCIENCE 3e, Figure 5.21

© 2009 Sinauer Associates, Inc.

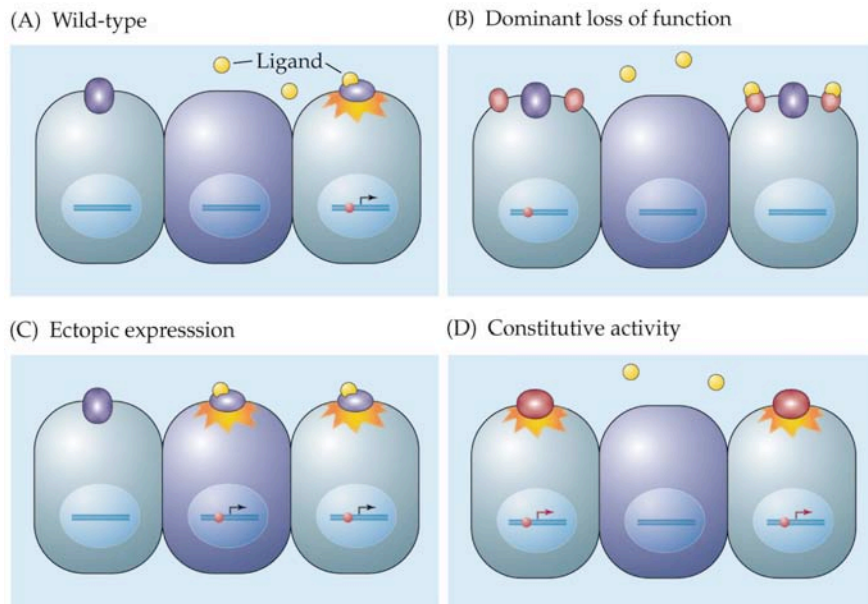
Figure 5.25 - Floxing mice to knock out gene function in a specific tissue or developmental stage



A PRIMER OF GENOME SCIENCE 3e, Figure 5.25

© 2009 Sinauer Associates, Inc.

Dominant gain- and loss-of-function in transgenic mice



A PRIMER OF GENOME SCIENCE 3e, Figure 5.22

© 2009 Sinauer Associates, Inc.

Proteomics

- Proteomics as a complement to gene expression analysis
- Basic methods - 2d gels, affinity chromatography, immunohistochemistry
- Mass spectrometry
- Two-hybrid screens of protein binding
- Protein microarrays
- Computational prediction of protein structure & function
- Drug development as an example of an integrated genomics approach to problem solving.

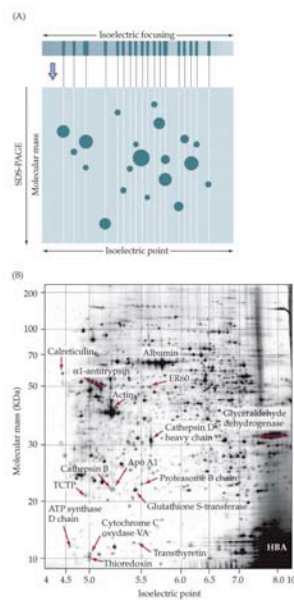
Proteomics vs. DNA microarrays

- It has been estimated that translational + post-translational regulation accounts for about 1/3 of all gene regulation in yeast (L. Hood). These are often quantitative (1.5 fold) rather than qualitative (on or off).
- In principle, protein measurements could be more accurate because they are more directly related to the gene's biological function.
- In practice, practical issues can out-weigh the above advantages. For example, protein chips often have far fewer genes than DNA chips.
- Denaturation of proteins is irreversible.
- Post-translational modifications are complex, and may require specialized reagents adapted to specific gene products.
- The technology for working with integral membrane proteins is rudimentary at best.

Two-dimensional gel electrophoresis

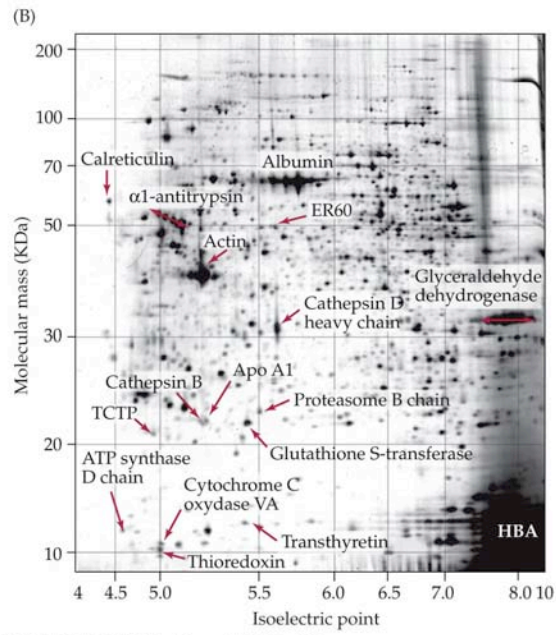
- First dimension is isoelectric focusing (separation by charge in a capillary).
- Second dimension is SDS-PAGE (separation by molecular weight in an acrylamide slab gel).
- The technique is capable of resolving several thousand proteins (this often requires multiple isoelectric focusing columns).
- Post-translational modifications are only partially resolved.
- Quantitation and sensitivity are limited.
- In principle, proteins can be identified by using specific antibodies, or specific mutants, or by obtaining N-terminal sequences from individual spots.
- In practice, amino acid sequencing from 2-d spots has given mixed results. Proteins are often blocked, impure, or not present in sufficient amounts for amino acid sequencing.

Figure 5.4 Two-dimensional polyacrylamide gel electrophoresis



A PRIMER OF GENOME SCIENCE 3e, Figure 5.4

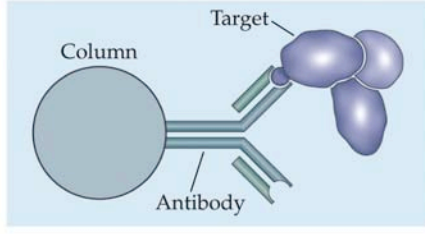
© 2009 Sinauer Associates, Inc.



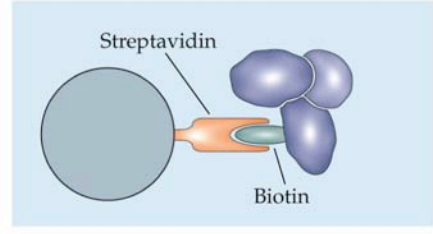
A PRIMER OF GENOME SCIENCE 3e, Figure 5.4 (Part 2)

© 2009 Sinauer Associates, Inc.

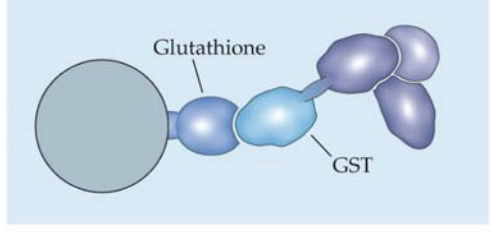
(A) Co-immunoprecipitation



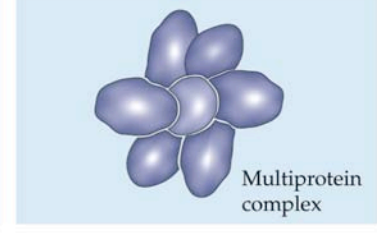
(B) Biotin-affinity chromatography



(C) GST-fusion chromatography

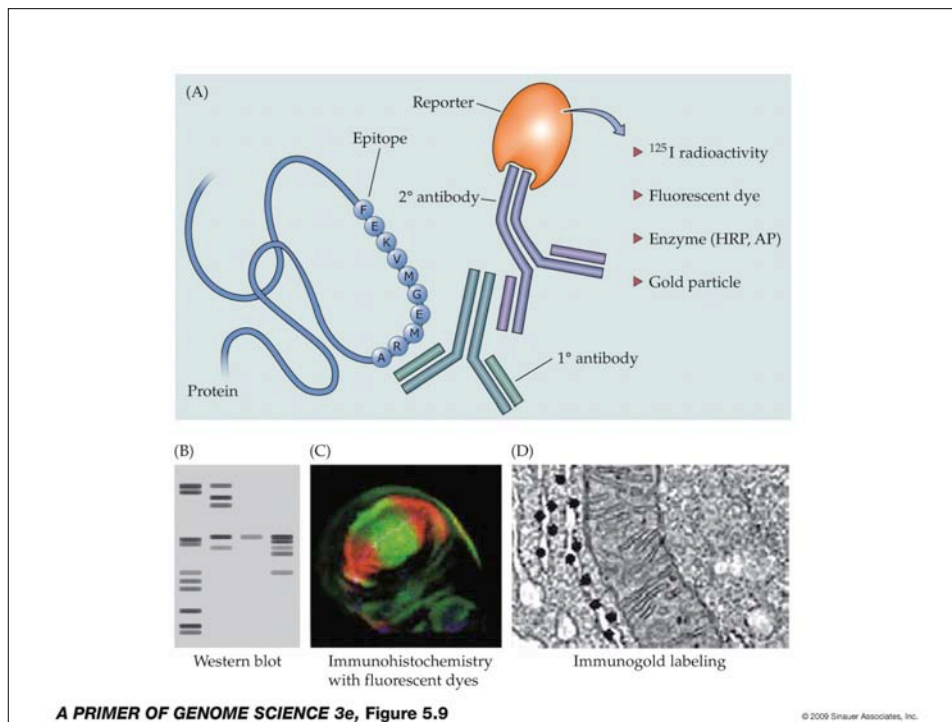


(D) Direct purification



A PRIMER OF GENOME SCIENCE 3e, Figure 5.5

© 2009 Sinauer Associates, Inc.



Mass spectrometry

- In principle, ionization time-of-flight mass spectrometry provides an extremely accurate measurement of molecular weight (time of flight to detector) and abundance (charge transferred to the detector) of macromolecules, including small to medium-sized proteins.
- In practice, resolution is roughly comparable to that of 2-d gels.
- Quantitation is good but is limited by vagaries of surface chemistry and the requirements for extensive sample preparation.
- Speed and sensitivity are excellent.
- Identification of proteins are possible from the molecular weight, but post-translational modifications complicate this issue.
- Mass spec is not a preparative technique; other methods must be used to isolate specific proteins for sequencing.
- Membrane proteins are somewhat difficult to analyze.

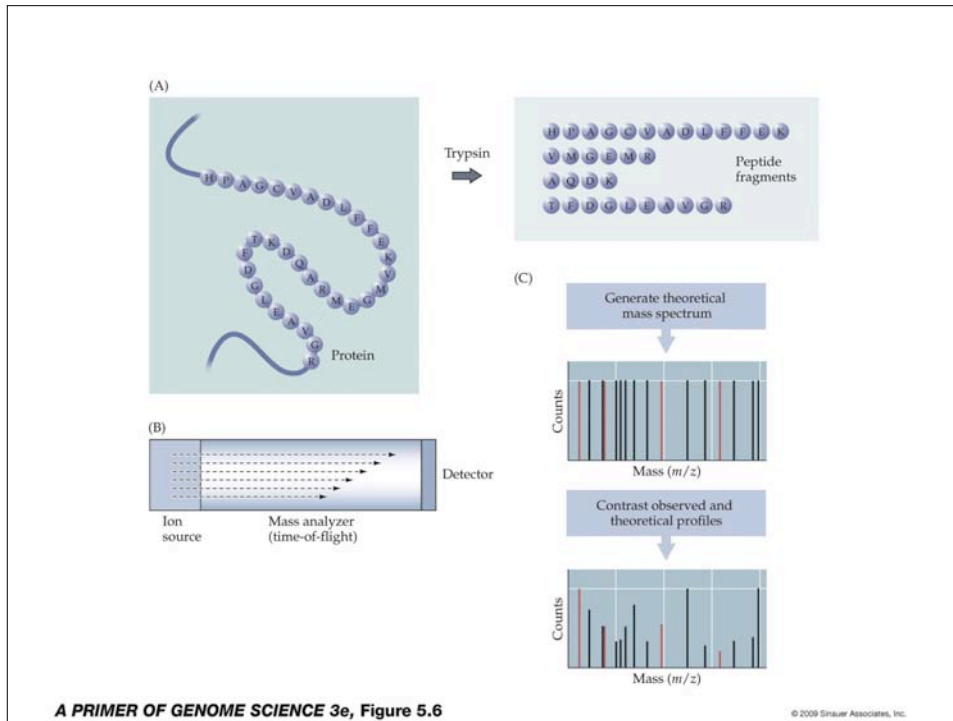
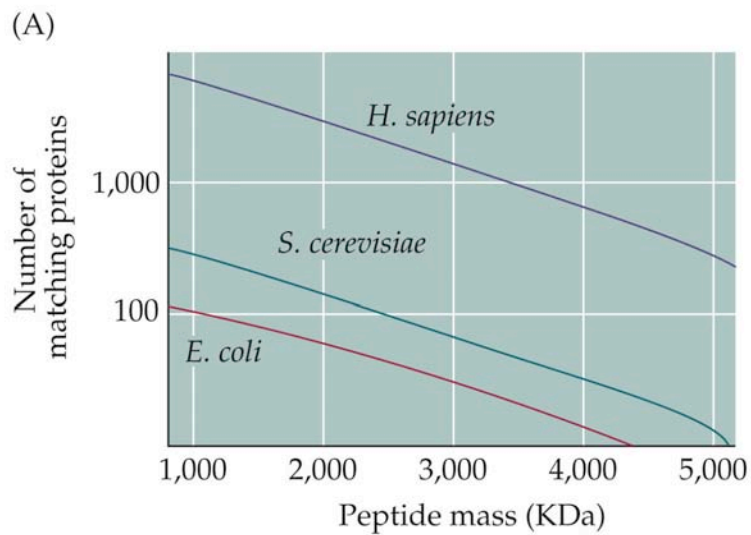


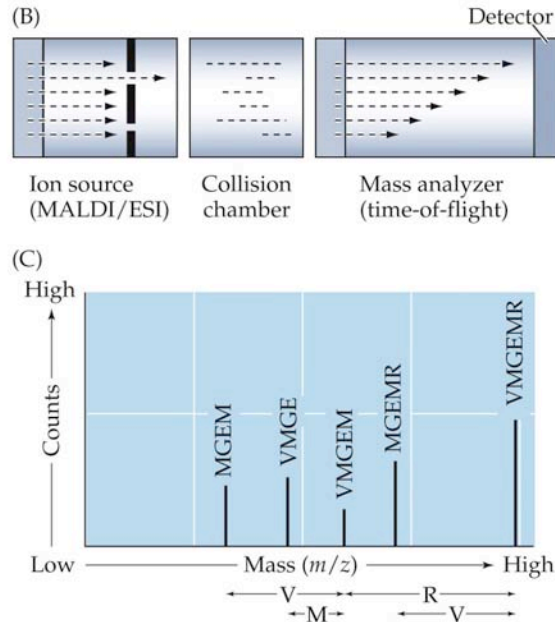
Figure 5.7 Peptide sequencing by tandem mass spectrometry



A PRIMER OF GENOME SCIENCE 3e, Figure 5.7 (Part 1)

© 2009 Sinauer Associates, Inc.

Figure 5.7 Peptide sequencing by tandem mass spectrometry



A PRIMER OF GENOME SCIENCE 3e, Figure 5.7 (Part 2)

© 2009 Sinauer Associates, Inc.

Two-hybrid screens of protein binding

- Two-hybrid screens are based on cDNA expression libraries, in which the insert is expressed (usually in yeast) as a gene fusion to a DNA binding domain or a transcriptional activator domain.
- A second such library can be transformed into the same yeast cells (usually one clone at a time) by selecting for two antibiotics (one on each plasmid).
- Positive interactions are detected by expression of the reporter gene (such as lacZ).
- Positive colonies are selected visually and sequenced to determine the identity of the two interacting genes.
- Two-hybrid screens have been completed on a genomic scale in *E. coli*, yeast, and *Drosophila*. Similar results will also be forthcoming for vertebrate proteomes.
- Two-hybrid screens do not work well for membrane proteins, or proteins regulated by post-translational modifications.

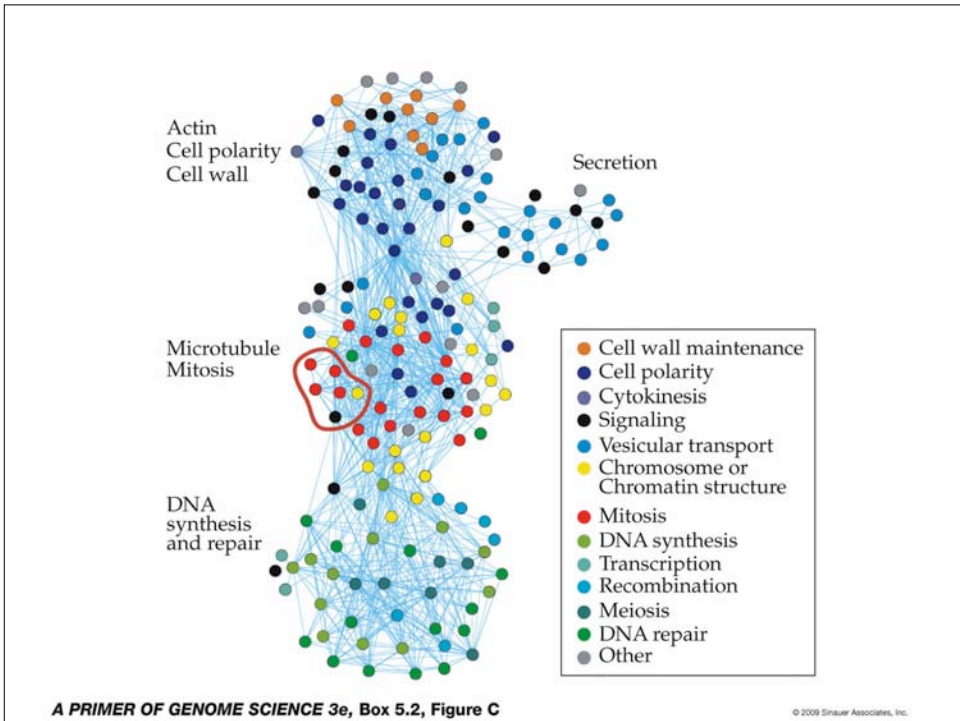
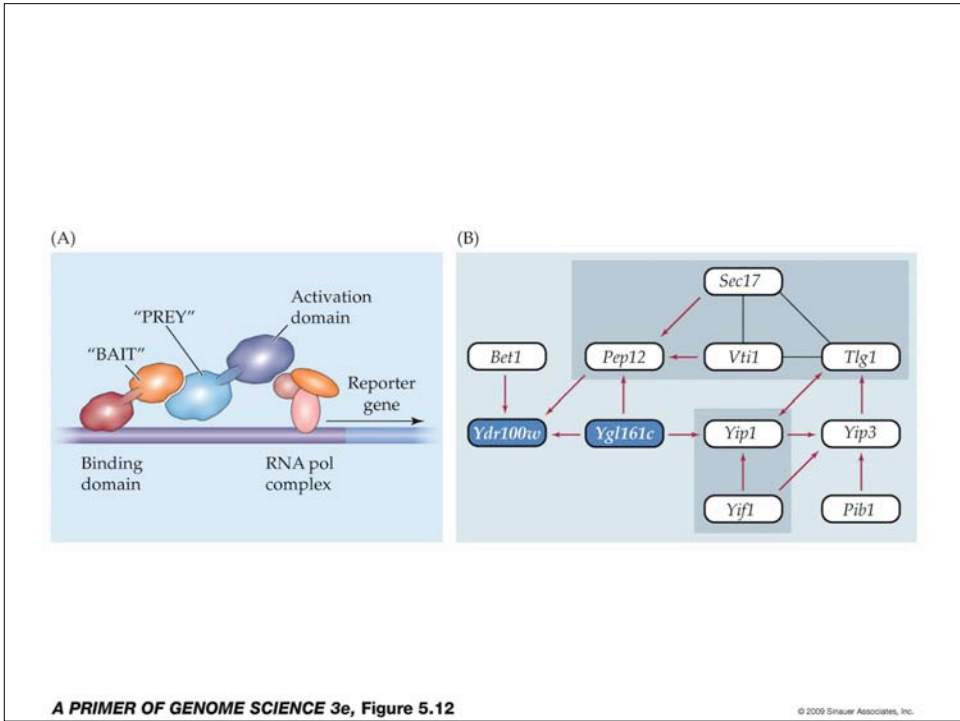
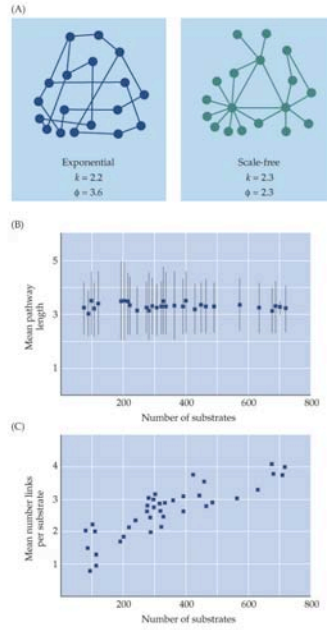


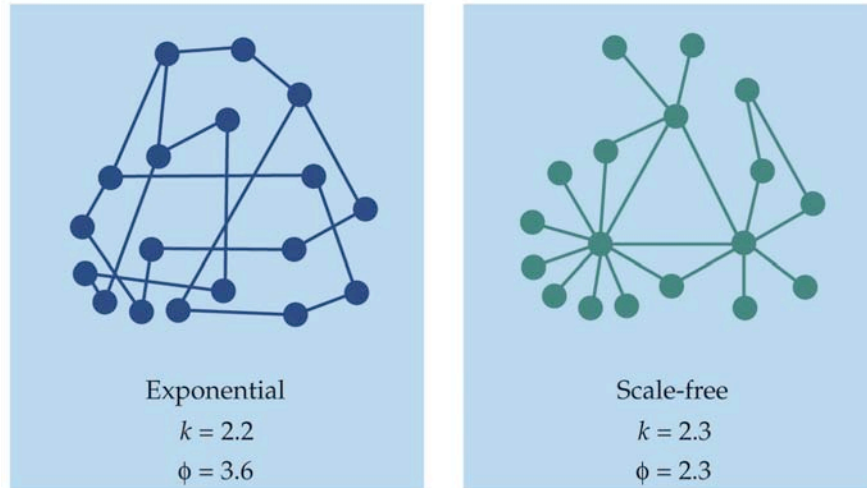
Figure 6.8 Large-scale organization of metabolic networks



A PRIMER OF GENOME SCIENCE 3e, Figure 6.8

© 2009 Sinauer Associates, Inc.

(A)

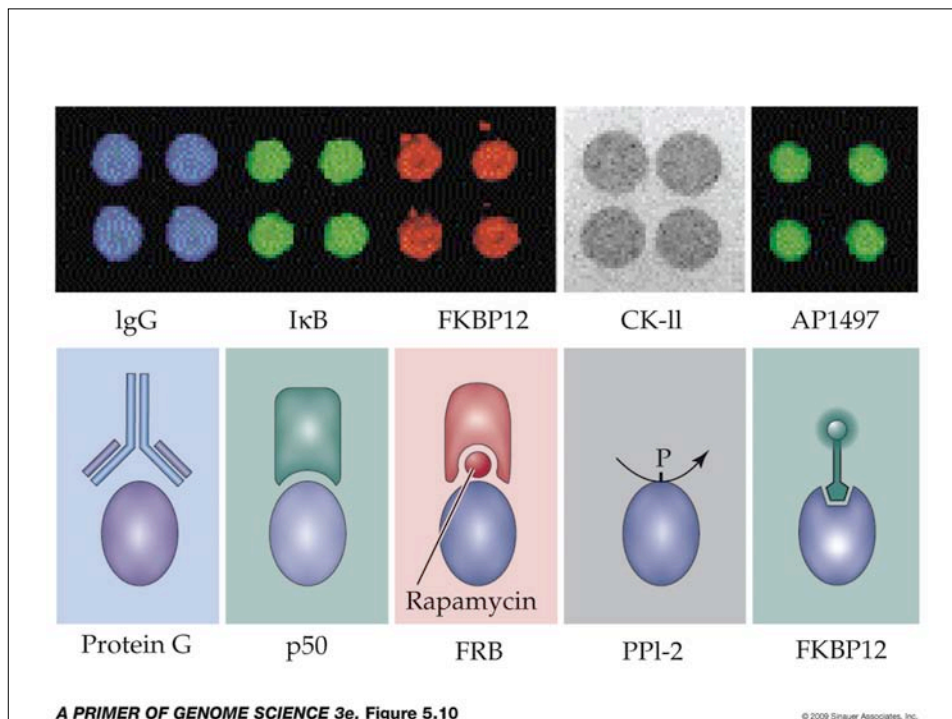


A PRIMER OF GENOME SCIENCE 3e, Figure 6.8 (Part 1)

© 2009 Sinauer Associates, Inc.

Protein microarrays

- Proteins can be printed on microarrays, and cross-linked to the substrate using techniques similar to those used to print DNA microarrays.
- The problem of protein denaturation can be minimized by selecting antibodies that are capable of binding to the denatured protein.
- Antibodies can be used to detect critical post-translational modifications (such as phosphorylation).
- The development and testing of specific monoclonal antibodies to each protein is expensive and time-consuming, but once developed they can be mass-produced.
- So far, attempts to do this on a large scale have been limited to about 2,000 proteins per slide in yeast (L. Hood), but further attempts can be expected.

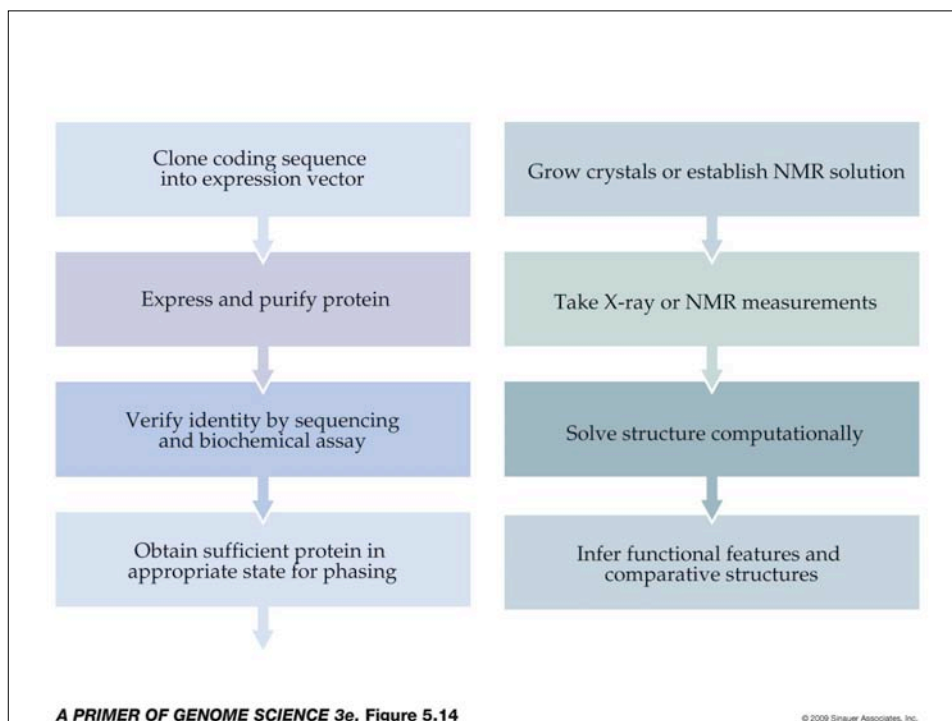
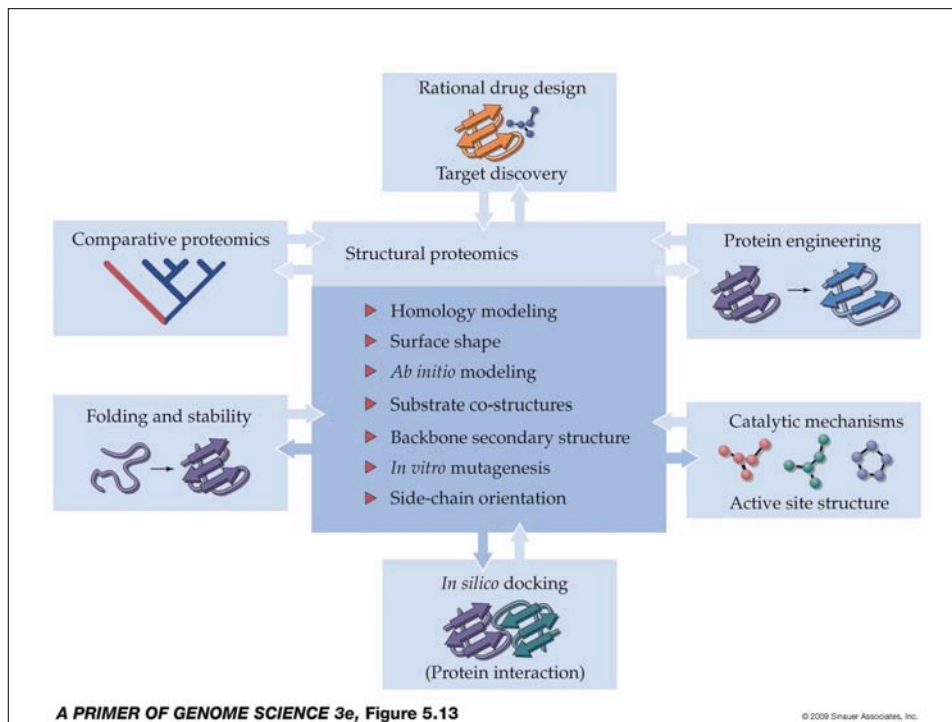


Structure & Function Prediction

- Functions can be predicted from amino acid sequences of proteins, based on the known functions of gene family members.
- Functions can also be predicted, in a more general way, based on the presence of particular peptide motifs that occur in multiple gene families.
- Functions can be tested in transgenic mutants (mouse or fly) or by RNA interference (usually in cell culture).
- Stereotyped secondary structures (alpha helix, transmembrane segments, beta-pleated sheet) depend on local amino acid composition in a fairly well-understood way and can be predicted with about 80% confidence from amino acid sequences.
- Tertiary structures can be experimentally determined by X-ray crystallography.
- Predicting tertiary structures is still difficult, but a combination of energy minimization and phylogenetic comparison to proteins of known structure works better than either method alone.

Drug development

- The vast majority of all pharmacological drugs work by binding to a specific protein (often a G-protein coupled receptor).
- Similarly, most side effects of drugs are attributable to binding to other proteins, in addition to the desired target.
- Thus, a considerable amount of drug design is really a proteomics problem, and can (should) be assayed by proteome-level binding studies with trial drugs.
- Many of the recent generation of drugs (such as Ventolin and Viagra) were designed with the benefit of sequence data that allowed investigators to test previously unknown gene family members that would be the most likely to cause side effects.
- One recent refinement of this technique is the use of X-ray crystallography to determine the molecular structure of the candidate drug while it is bound to its correct (or incorrect) ligand (Don Abraham).
- Another refinement is the design of ligands for “orphan receptors” of unknown function (but promising parentage, such as leptin).



(A) Primary

```
MKVLRLICFIALLISSLEADKCKEREKIILVSSANEIDVRPCPLNPNEHKGITITWKD
DSKTPVSTEQASRIHQHKEKLWFVPAKVEDSGHYCVVRNSSYCLR IKISAKFVNEPNL
CYNAQAIFKQKLPVAGDGLVCPYMEFFKNENNELPKLQWYKCKPLLDNIHFGVKDR
LIVMVAEKHRGNYTCHASYTYLGKQYPITRVI EFITLEENKPTRPVI VSPANETMEVDL
GSQILICNVTGQLSDIAYKWNGSVI DEDDPVLGEDYYSVENPANKRRSTLITVLNISE
IESRFYKHPFTCFAKNTHGIDAAYIQLIYPVTNFKHMIGICVTLTVIIVCSVFIYKIFK
IDIVLWYRDS CYDFLPIKASDGKTYDAYILYPKTVGEGSTSDCDIFVFKVLPVLEKQCG
YKLFYGRDDYVGEDI VEVINENVKSRRLIIILVRETSGF SWLGGSSSEEQIAMYNALVQ
DGIKVVLELEKI QDYKMPESI KPIKQKHGAIRWSGDFTQGPQSAKTRFWKNVRYHMPV
QRRSPSSKHQLLSPATKEKLQREAHVPLG
```

(B) Secondary



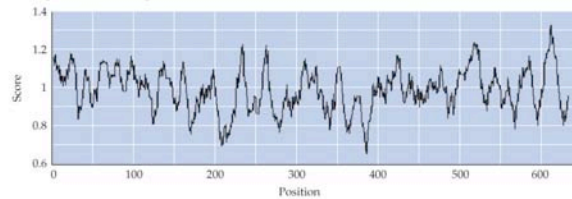
(C) Tertiary



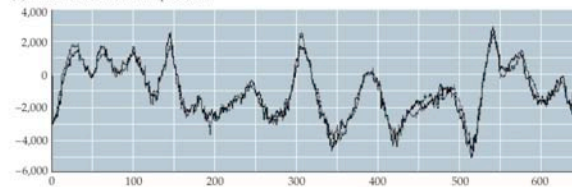
A PRIMER OF GENOME SCIENCE 3e, Figure 5.2

© 2009 Sinauer Associates, Inc.

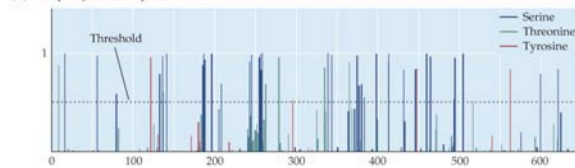
(A) Alpha-helical content prediction



(B) Transmembrane domain prediction



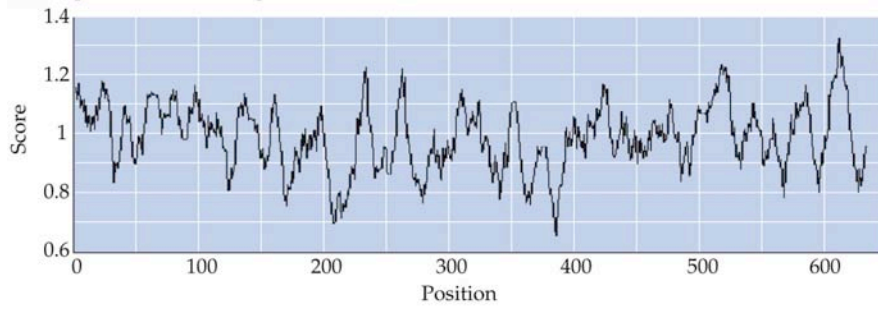
(C) Phosphorylation site prediction



A PRIMER OF GENOME SCIENCE 3e, Figure 5.3

© 2009 Sinauer Associates, Inc.

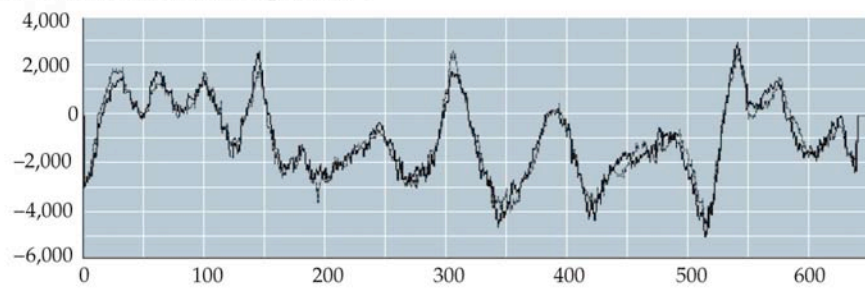
(A) Alpha-helical content prediction



A PRIMER OF GENOME SCIENCE 3e, Figure 5.3 (Part 1)

© 2009 Sinauer Associates, Inc.

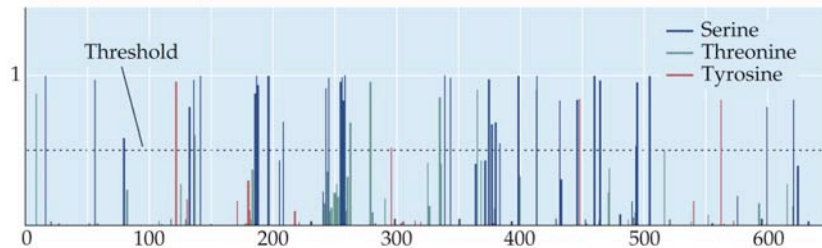
(B) Transmembrane domain prediction



A PRIMER OF GENOME SCIENCE 3e, Figure 5.3 (Part 2)

© 2009 Sinauer Associates, Inc.

(C) Phosphorylation site prediction



A PRIMER OF GENOME SCIENCE 3e, Figure 5.3 (Part 3)

© 2009 Sinauer Associates, Inc.

Discussion questions (week 10)

- Discuss the advantages and disadvantages of (1) Chip-on-chip vs. (2) two-hybrid screens of protein binding, for the purpose of building models of gene interaction networks and genetic pathways.
- Discuss the advantages and disadvantages of forward, reverse, and fine-structure genetics. Under what circumstances would each be expected to be more likely to find the key result(s)?
- Why aren't transcript and protein expression profiles always in agreement? Which is more closely related to biological function?
- Discuss some of the advantages and disadvantages of RNA interference as a way of manipulating gene expression.
- Discuss the use of comparative genomics, structure prediction, X-ray crystallography, and expression cloning in drug discovery and design.

Discussion Questions (continued)

- Discuss the findings of Kondrashov et al. (2011), concerning the role of ribosomal proteins in the regulation of Hox genes, in the context of the widespread use of “housekeeping genes” for the normalization of protein or DNA microarray data.
- What is “polysome profiling”. How can it be used to separate mRNAs that are being translated from those that are not? How did Kondrashov et al. (2011) use this technique to identify which specific Hox genes were being regulated by RPL38?