Moc/Bio and Nano/Micro

Lee and Stowell

Moc/Bio-Lecture 4

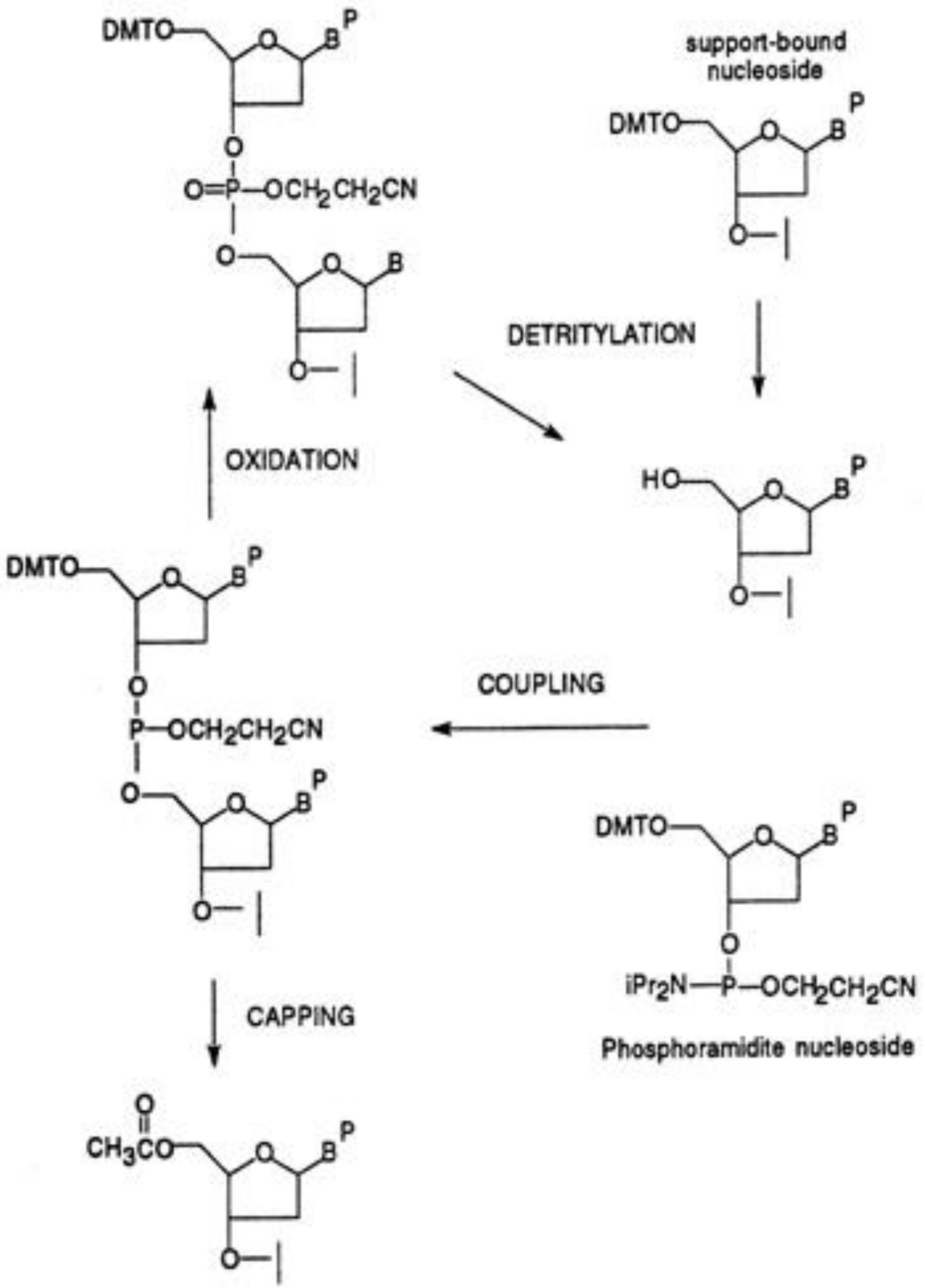
Production, Purification and Characterization of Biomolecules

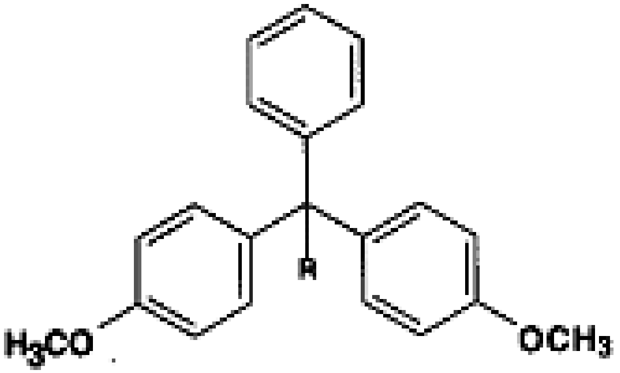
- **DNA**
- **RNA**
- **Proteins**

Reading material

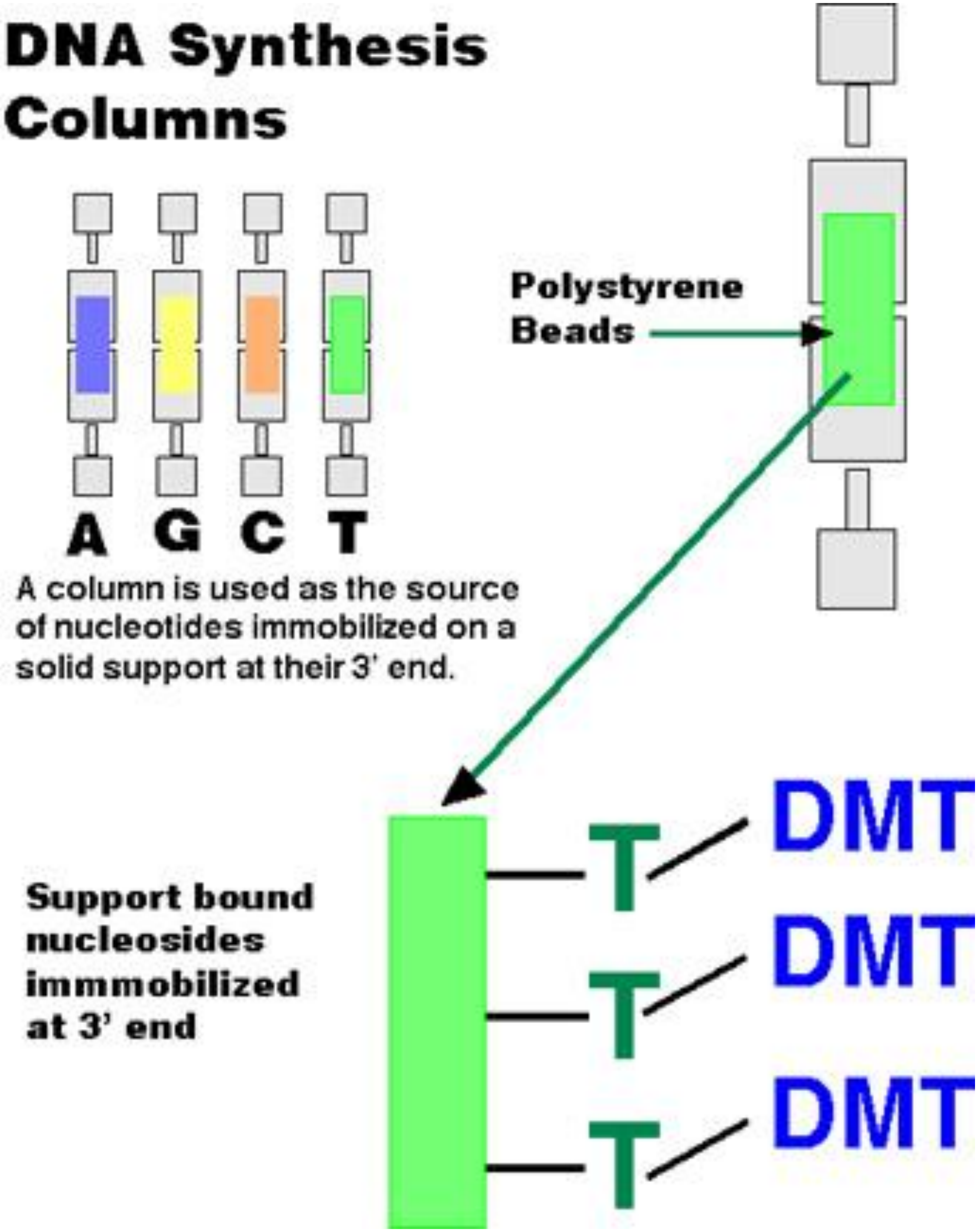
- http://www.ncbi.nlm.nih.gov/books/NBK21589/
- http://www.ncbi.nlm.nih.gov/books/NBK21654/
- http://www.ncbi.nlm.nih.gov/books/NBK21505/

DNA/RNA synthesis


support-bound nucleoside
DMTO
HO
O
B
P
O=P—OCH₂CH₂CN
P—OCH₂CH₂CN
iPr₂N—P—OCH₂CH₂CN
CH₃CO
DETRITYLATION
COUPLING
OXIDATION
CAPPING
Phosphoramidite nucleoside


R
H₃CO
OCH₃

DNA Synthesis Columns

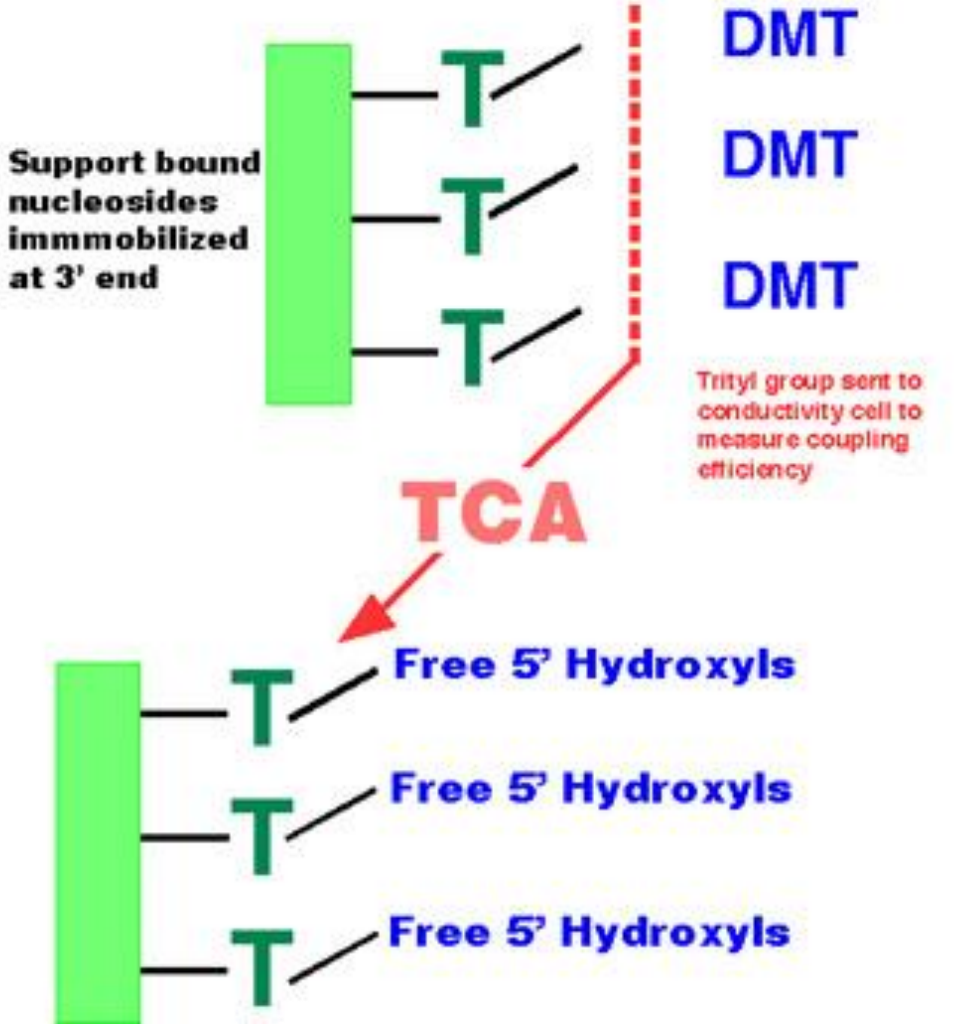


A column is used as the source of nucleotides immobilized on a solid support at their 3' end.

Support bound nucleosides immmobilized at 3' end

The first step in oligo synthesis is to remove the acid-labile, dimethoxytrityl (DMT) protecting group on the 5'-hydroxyl of the support bound nucleoside

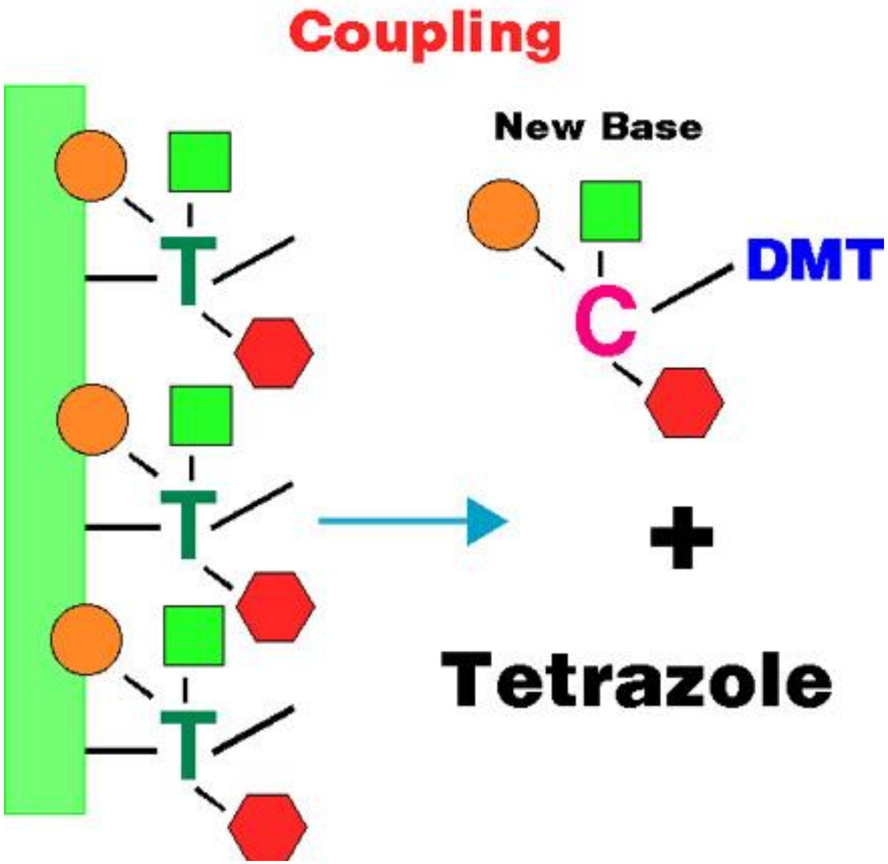
Detritylation

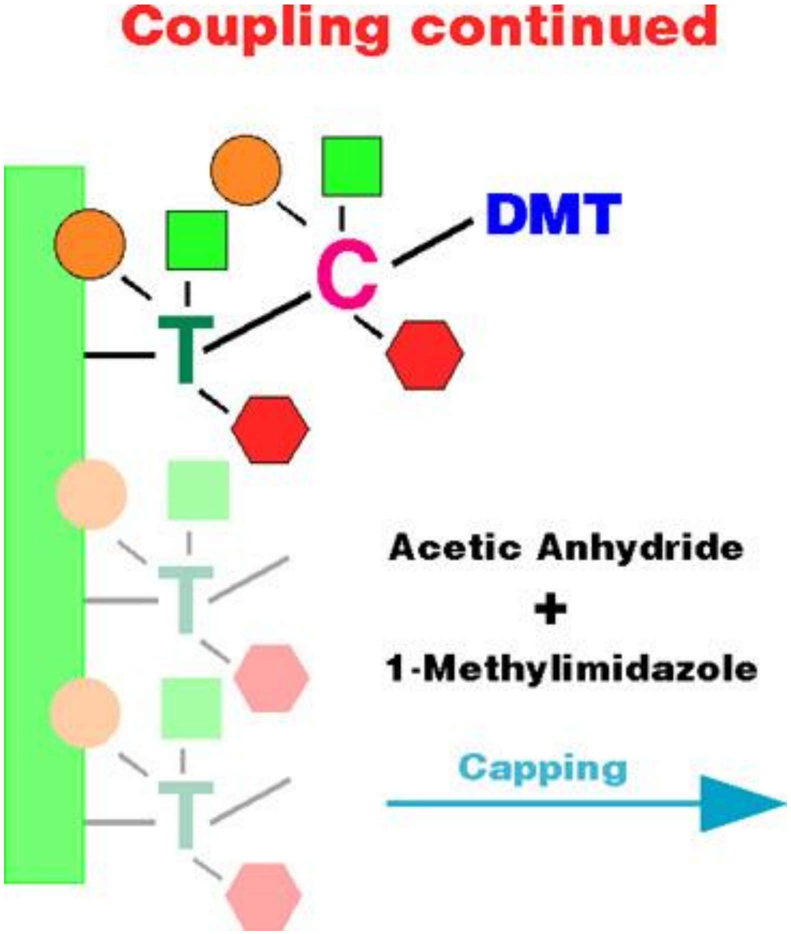


Support bound nucleosides immmobilized at 3' end

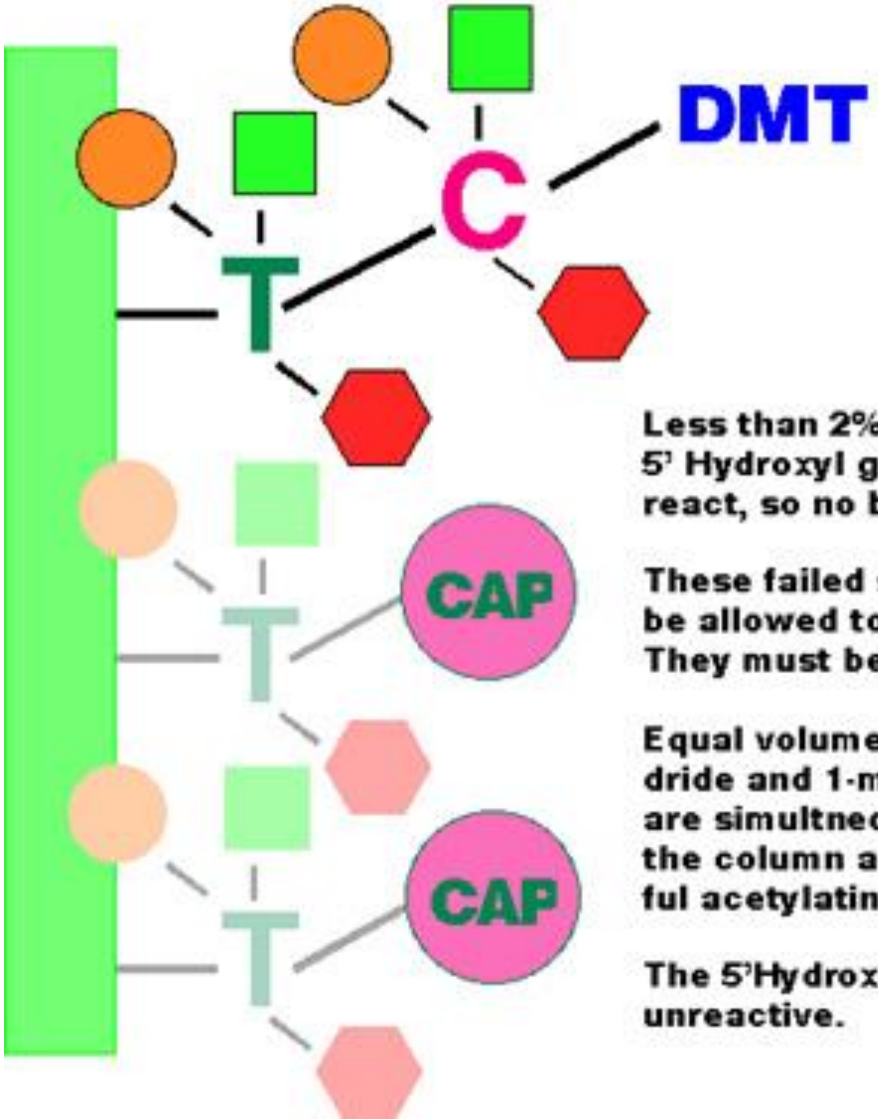
Trityl group sent to conductivity cell to measure coupling efficiency

Coupling


Coupling
New Base
DMT
C
T
T
T
+
Tetrazole


Coupling continued
DMT
C
T
T
T
Acetic Anhydride
+
1-Methylimidazole
Capping

Capping



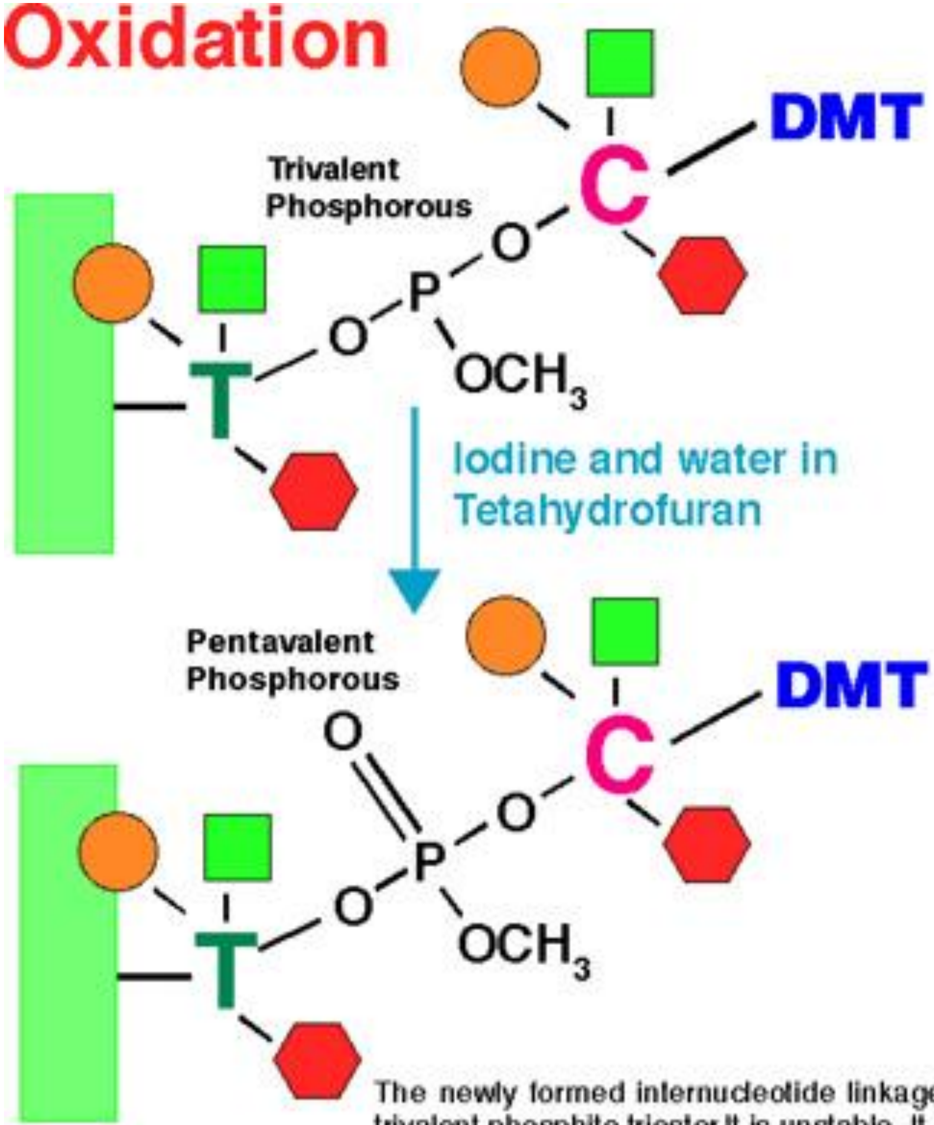
Less than 2% of the free 5' Hydroxyl groups do not react, so no base is added.

These failed sequences cannot be allowed to grow.
They must be capped.

Equal volumes of acetic anhydride and 1-methylimidazole are simultneously delivered to the column acting as a powerful acetylating agent.

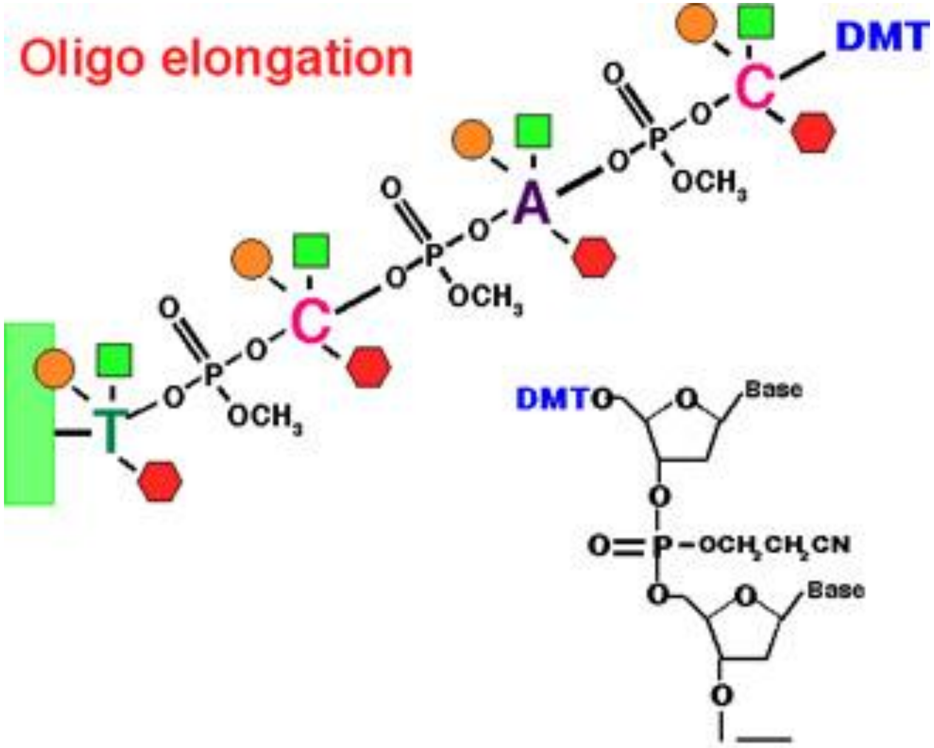
The 5'Hydroxyls are rendered unreactive.

Oxidation



The newly formed internucleotide linkage is a trivalent phosphite triester It is unstable. It must be oxidized to a stable pentavalent phosphate triester. Iodine is used as a oxidant with water as the oxygen doner

Oligo elongation

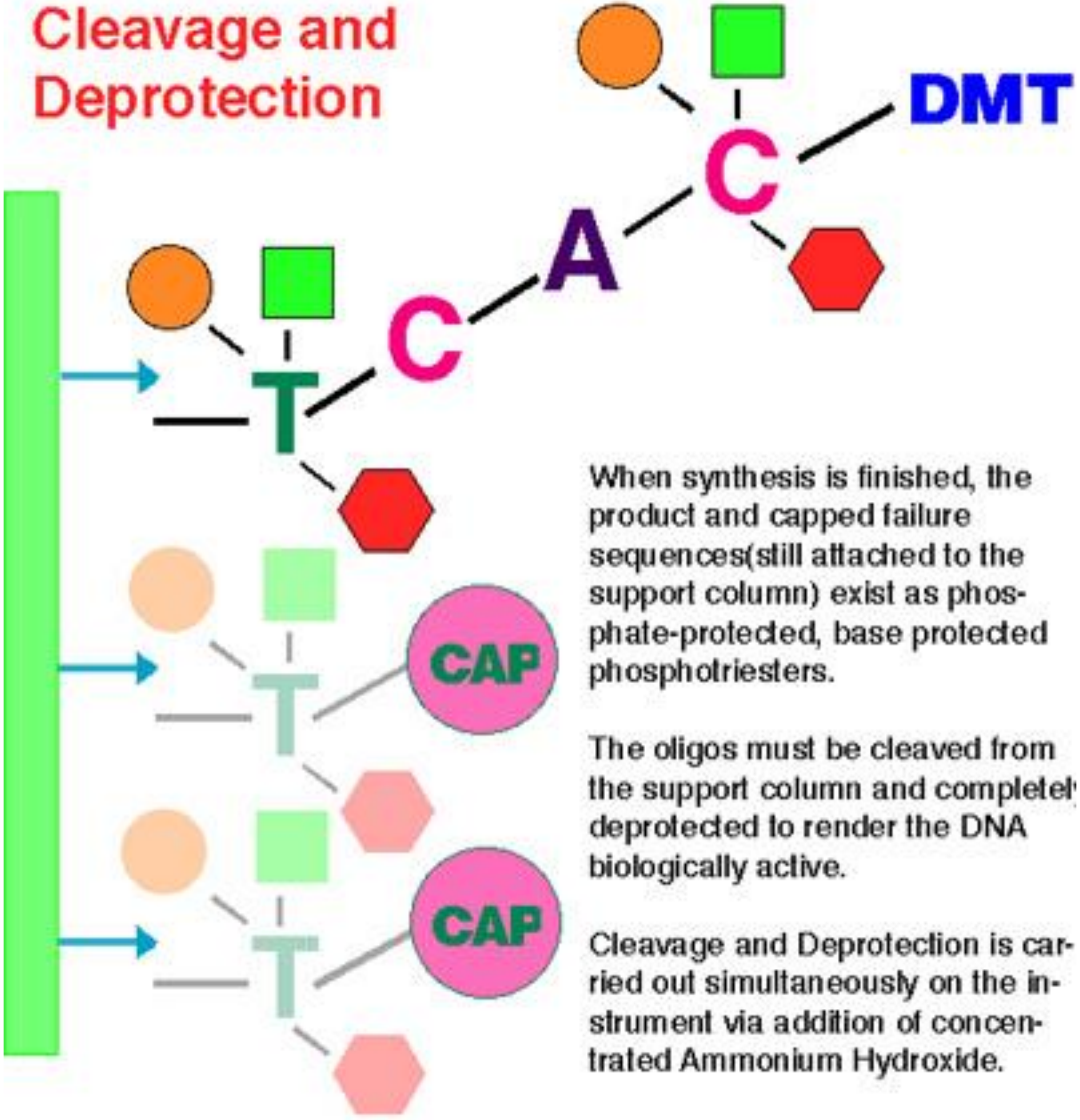


Following oxidation, a cycle of nucleotide addition is complete. The 5' terminus of the oligomer is protected by the DMT group. DNA synthesis continues by removing the DMT group and repeating another cycle of base addition. This is done until DNA of the specified length has been fully synthesized.

The oligos are usually synthesized with "Trityl off"
when purifying by gel electrophoresis or
ion exchange HPLC.

Syntheses are left "Trityl on"
when purifying by OPC or trityl-specific,
reverse phase HPLC.

Cleavage and Deprotection

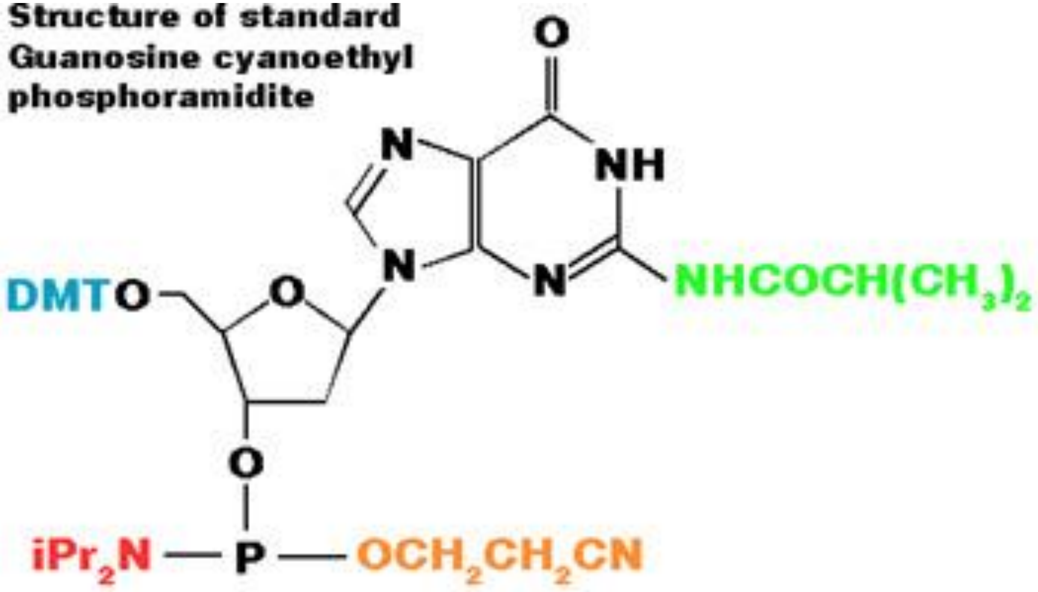


When synthesis is finished, the product and capped failure sequences(still attached to the support column) exist as phosphate-protected, base protected phosphotriesters.

The oligos must be cleaved from the support column and completel deprotected to render the DNA biologically active.

Cleavage and Deprotection is carried out simultaneously on the instrument via addition of concentrated Ammonium Hydroxide.

Structure of standard Guanosine cyanoethyl phosphoramidite



Phosphoramidites are chemically modified nucleosides. There are four groups added.

- A diiopropylamino phosphoramidite on a 3' trivalent moiety, which is stable until tetrazole is added
- A 3' B-cyanoethyl protecting group which prevents side reactions and aids solubility and is removed in the ammonia during deprotection
- A dimethoxytrityl protecting group on the 5' hydroxyl
- A benzoyl protecting group on the exocyclic amine of A and C and an isobutyryl protecting group on the exo cyclic amine of G. There is no exocyclic amine on T, so there is no protecting group present

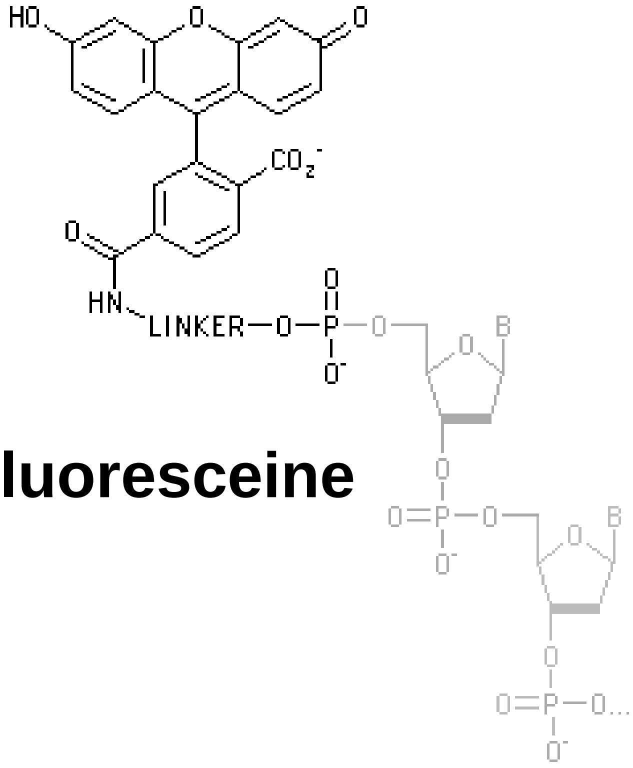
Summary

- Rapid 2-3 days for short oligos
- Cheap, ~$.25 per/base/mg
- Efficiency about 98-99% per cycle
- Length limited by efficiency (~130)
    - HW Problem

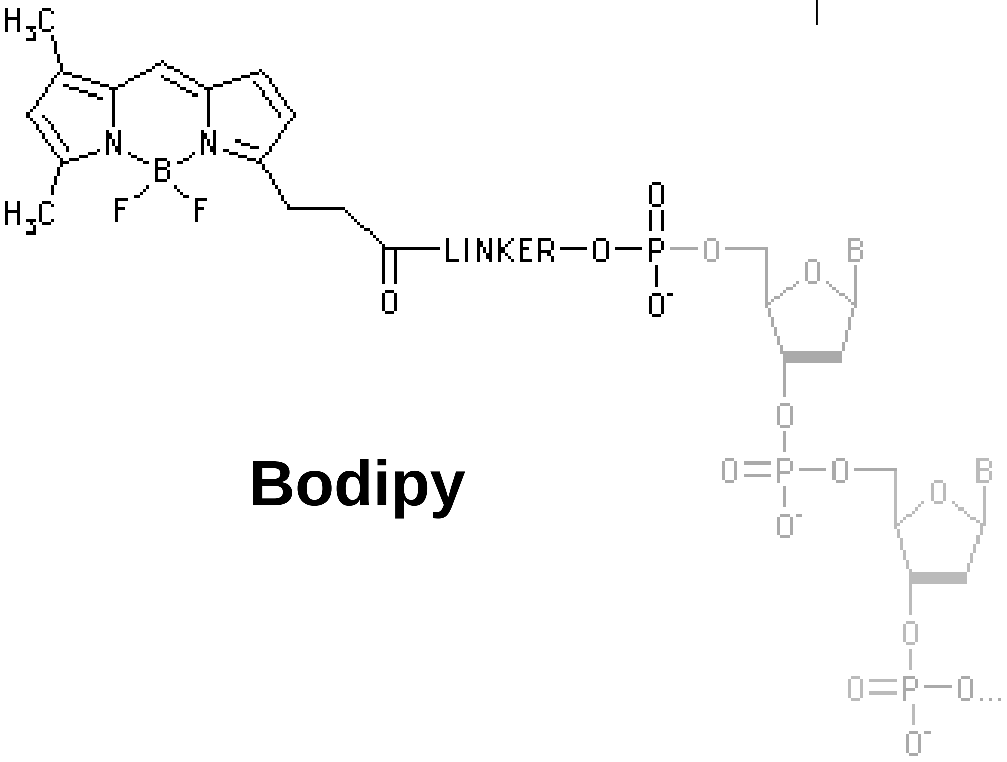
Types of labels for DNA/RNA

- Fluorescent labels
  - Abs 336-800nm
- Functional labels
  - Amine, thiol, carboxyl, aldehyde
- Affinity Tags
  - Biotin
  - Digoxigenen
  - TEG

Fluorescent examples



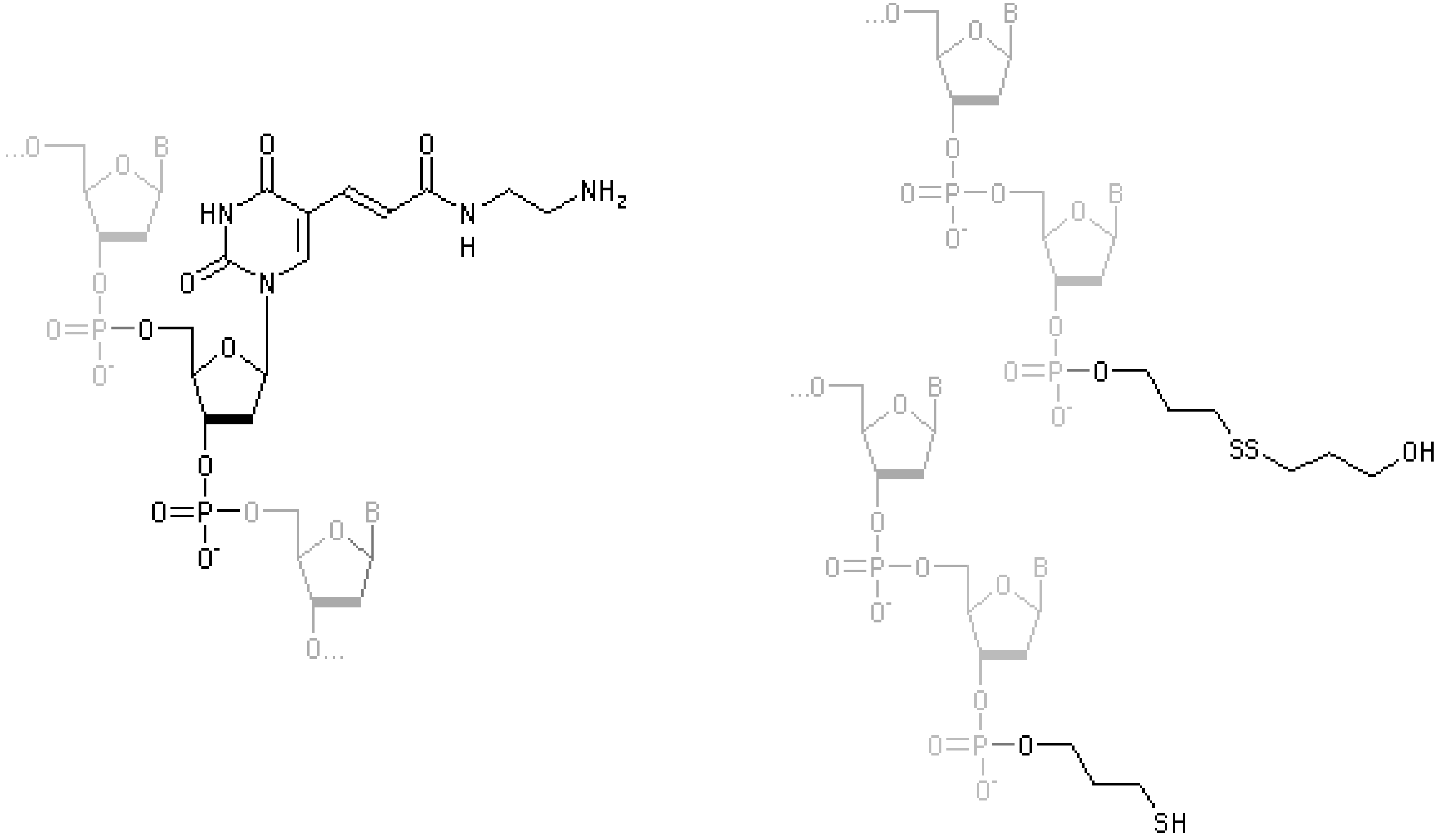
Fluoresceine



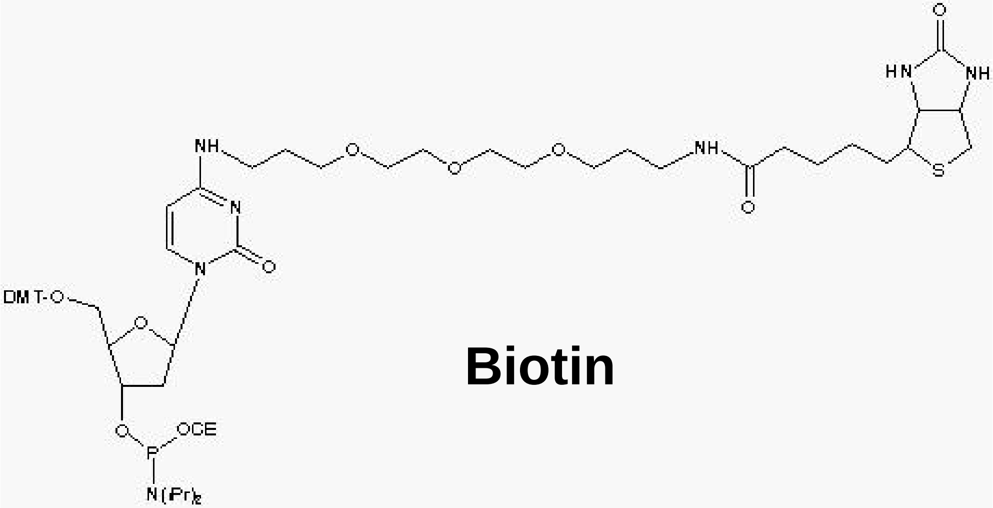
Bodipy

Absorbance from 336-800nm Emission from 377-820nm

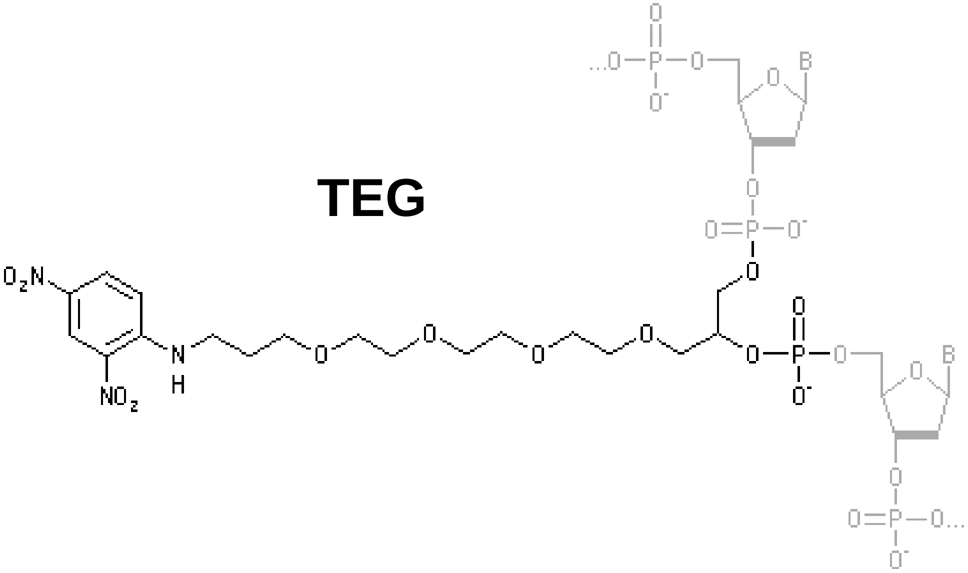
Functional examples



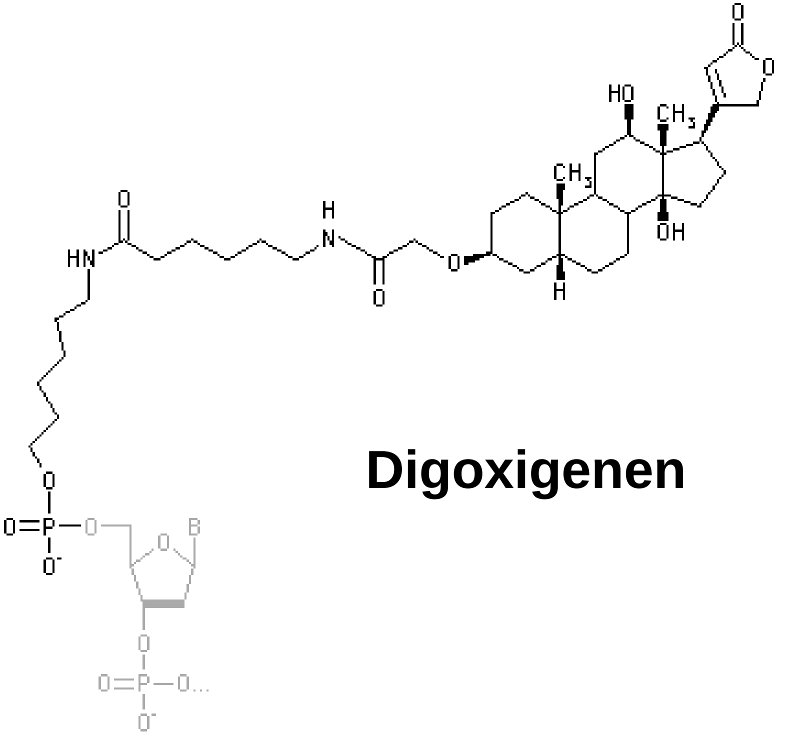
Affinity examples



Biotin



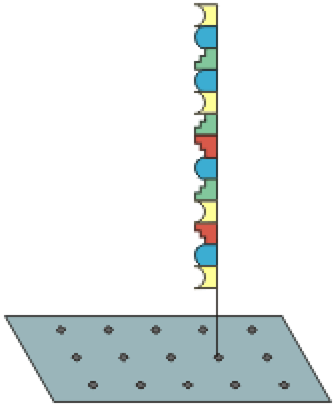
TEG



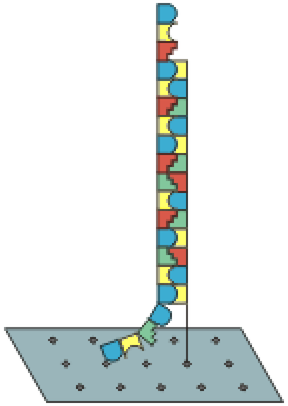
Digoxigenen

Example for single mol SNP

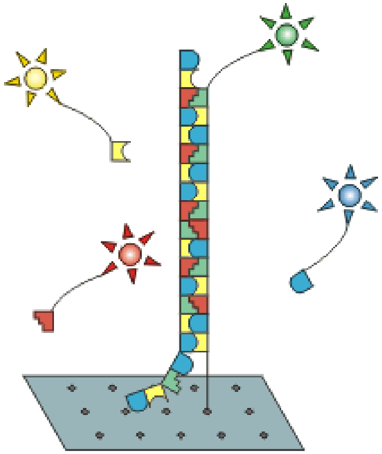
Chip array with probe oligos



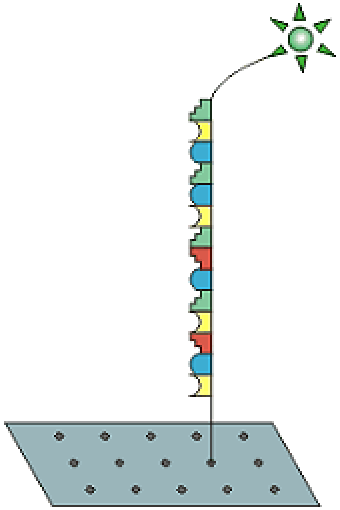
Add PCR amplified target DNA

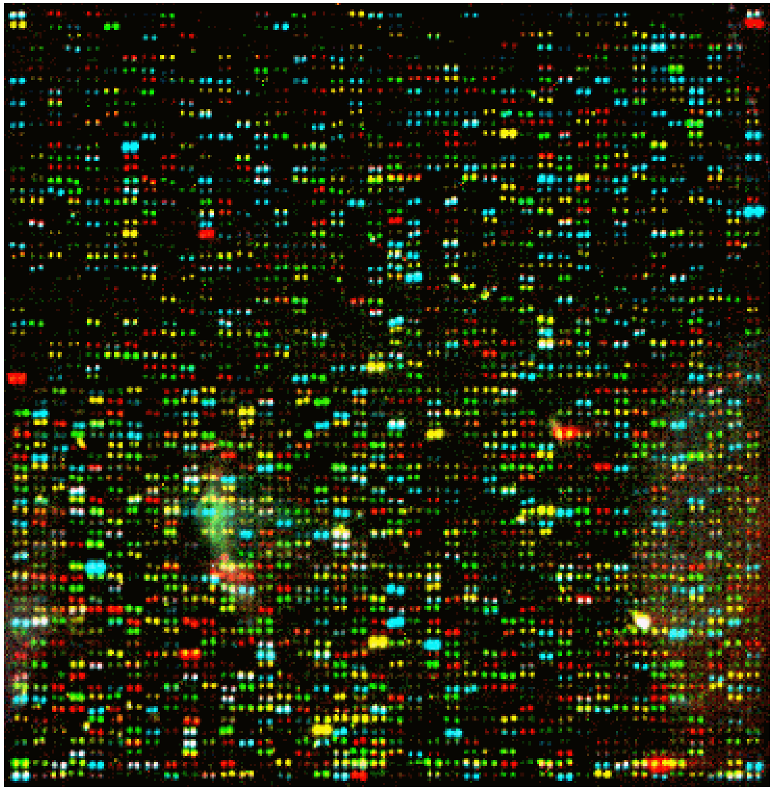


Add fluorescent bases



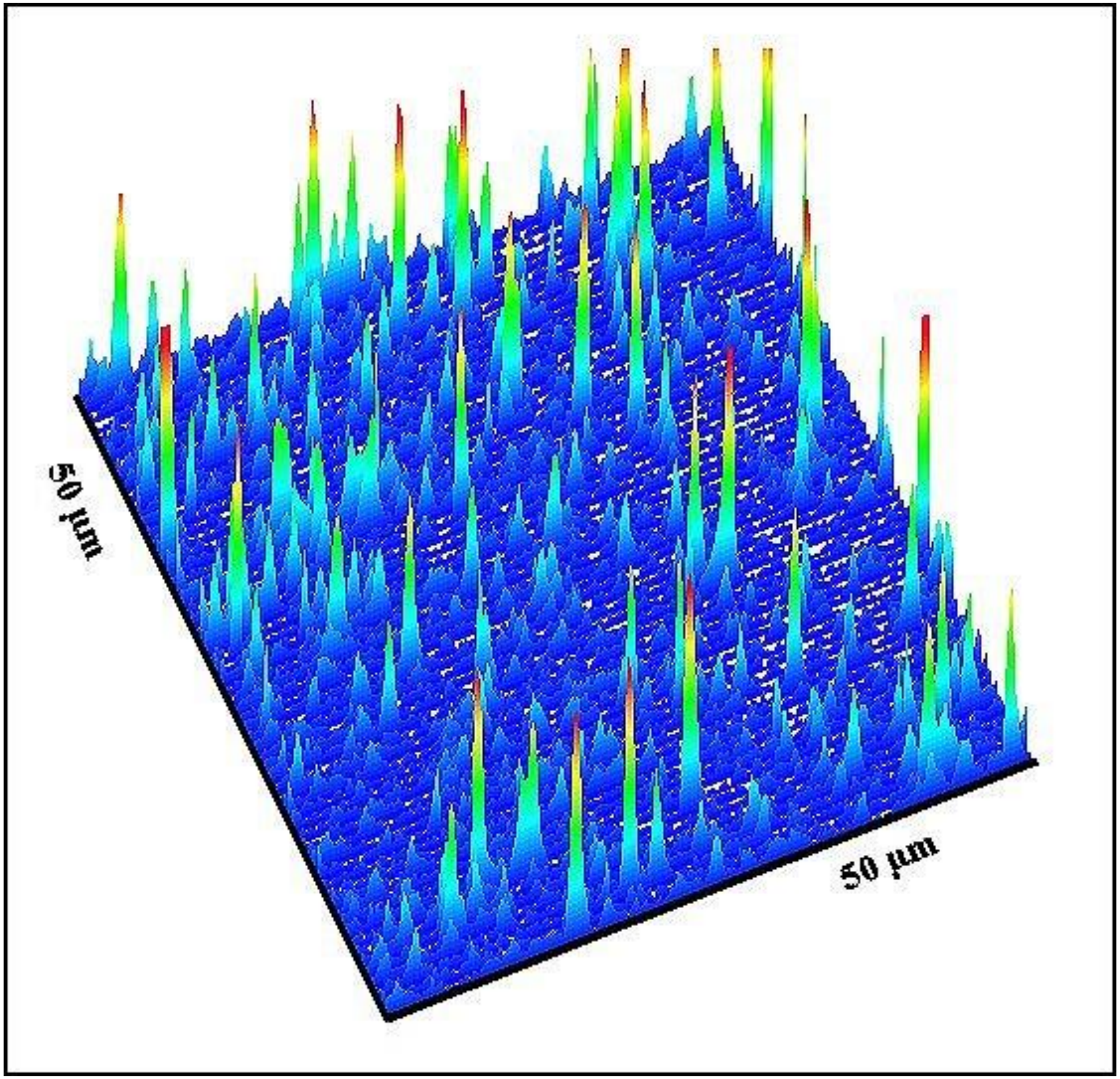
Link wash and read





Spot size 100 microns

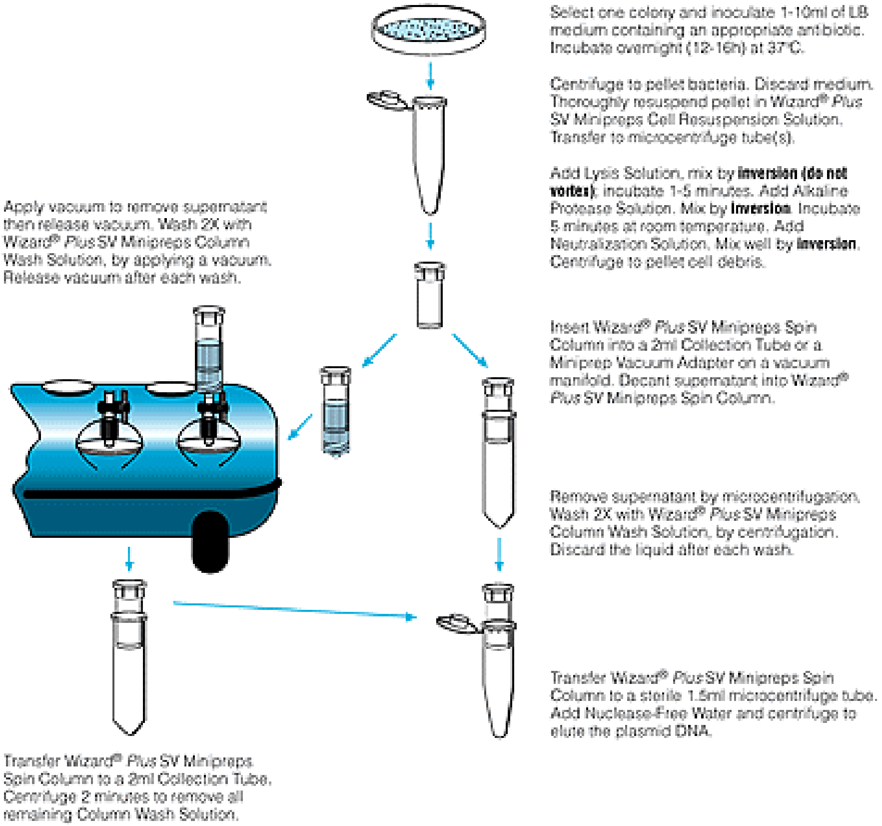
With single molecule imaging



Spot size < 1 micron

So we can probe 10^4 times more samples

Isolation of Genomic/Plasmid



Select one colony and inoculate 1-10ml of LB medium containing an appropriate antibiotic. Incubate overnight (12-16h) at 37°C.

Centrifuge to pellet bacteria. Discard medium. Thoroughly resuspend pellet in Wizard® *Plus* SV Minipreps Cell Resuspension Solution. Transfer to microcentrifuge tube(s).

Add Lysis Solution, mix by **inversion (do not vortex)**; incubate 1-5 minutes. Add Alkaline Protease Solution. Mix by **inversion**. Incubate 5 minutes at room temperature. Add Neutralization Solution. Mix well by **inversion**. Centrifuge to pellet cell debris.

Insert Wizard® *Plus* SV Minipreps Spin Column into a 2ml Collection Tube or a Miniprep Vacuum Adapter on a vacuum manifold. Decant supernatant into Wizard® *Plus* SV Minipreps Spin Column.

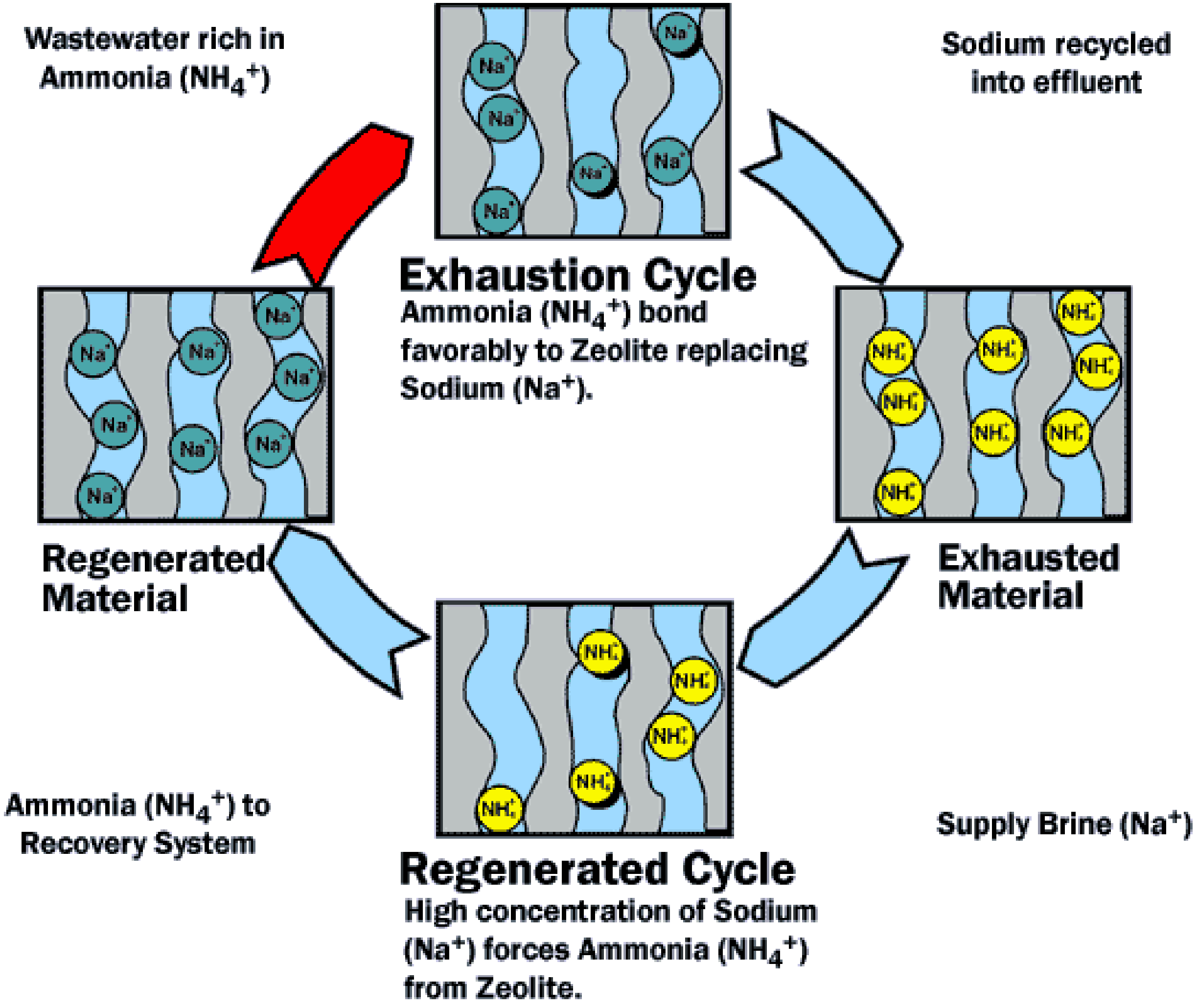
Remove supernatant by microcentrifugation. Wash 2X with Wizard® *Plus* SV Minipreps Column Wash Solution, by centrifugation. Discard the liquid after each wash.

Transfer Wizard® *Plus* SV Minipreps Spin Column to a sterile 1.5ml microcentrifuge tube. Add Nuclease-Free Water and centrifuge to elute the plasmid DNA.

Apply vacuum to remove supernatant then release vacuum. Wash 2X with Wizard® *Plus* SV Minipreps Column Wash Solution, by applying a vacuum. Release vacuum after each wash.

Transfer Wizard® *Plus* SV Minipreps Spin Column to a 2ml Collection Tube. Centrifuge 2 minutes to remove all remaining Column Wash Solution.

Ion exchange



Wastewater rich in Ammonia (NH_4^+)

Sodium recycled into effluent

Exhaustion Cycle
Ammonia (NH_4^+) bond favorably to Zeolite replacing Sodium (Na^+).

Regenerated Material

Exhausted Material

Ammonia (NH_4^+) to Recovery System

Supply Brine (Na^+)

Regenerated Cycle
High concentration of Sodium (Na^+) forces Ammonia (NH_4^+) from Zeolite.

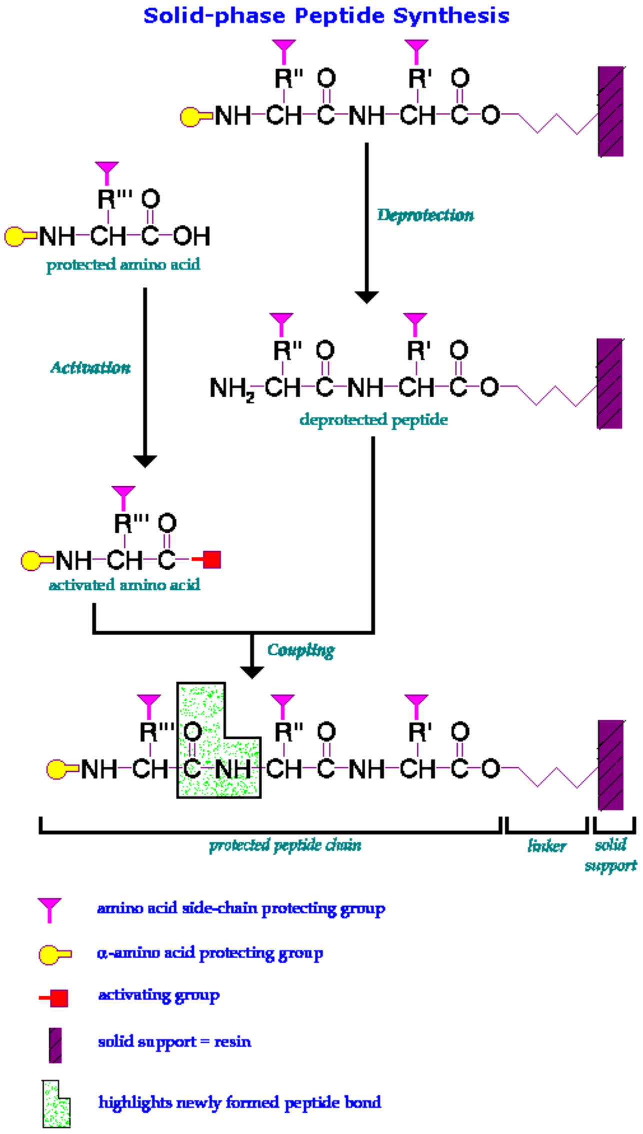
Proteins

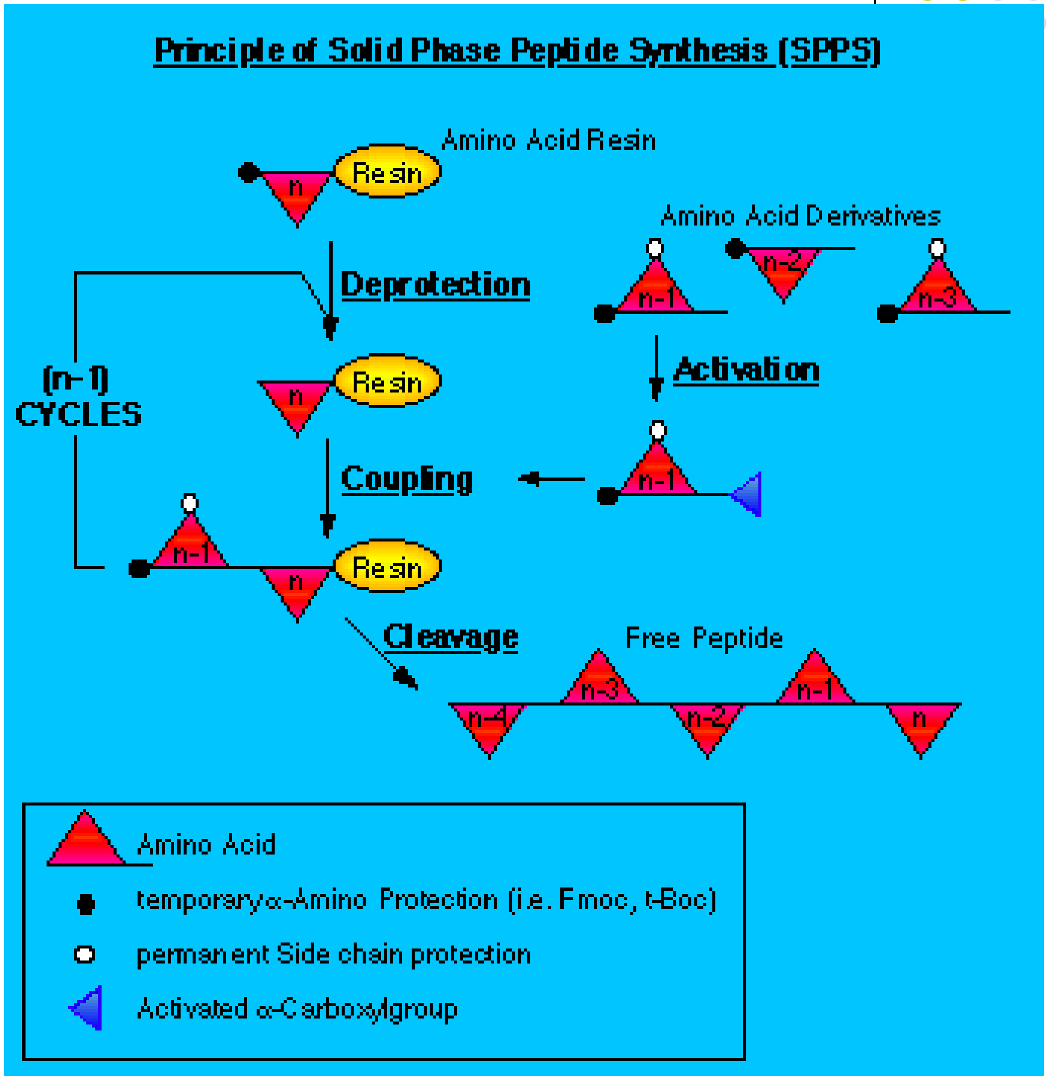
- Chemical synthesis
  - Small peptides and proteins
  - Small quantities for large proteins
- Biochemical synthesis
  - Small peptides to large proteins
  - Relatively small scale at present
- Biological synthesis
  - Small to large proteins
  - Small to Industrial scale

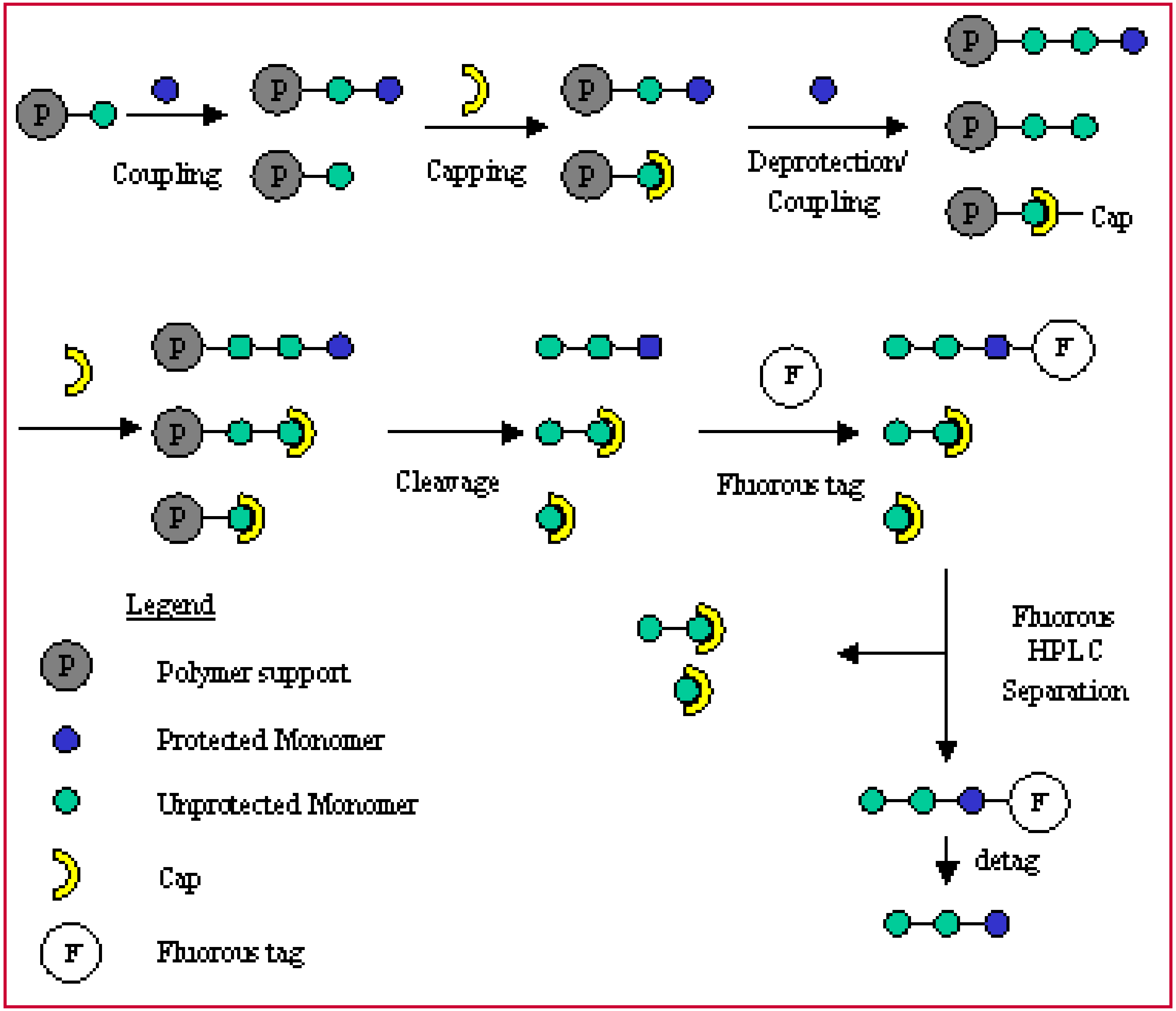
Chemical synthesis

- Synthesis
  - 1-50 amino acids
  - Cost ~ $10/residue/mg
- Synthesis + ligation
  - ~150 amino acids
  - Cost ….well more.

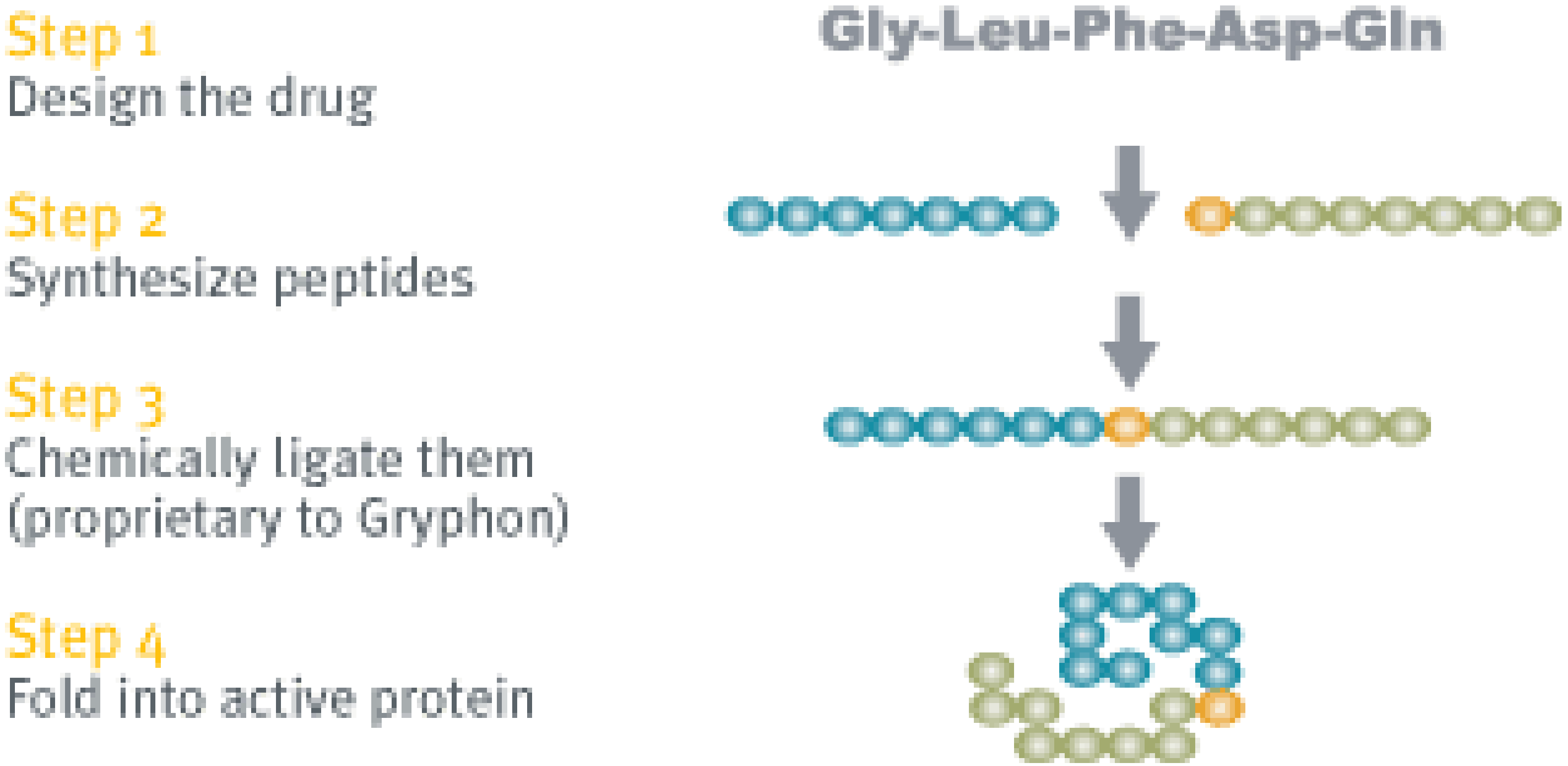
Chemical synthesis






P
Coupling
Capping
Deprotection/
Coupling
Cap
Cleavage
F
Fluorous tag
Fluorous
HPLC
Separation
detag
Legend
Polymer support
Protected Monomer
Unprotected Monomer
Cap
Fluorous tag

Synthesis + Ligation



Step 1
Design the drug

Step 2
Synthesize peptides

Step 3
Chemically ligate them
(proprietary to Gryphon)

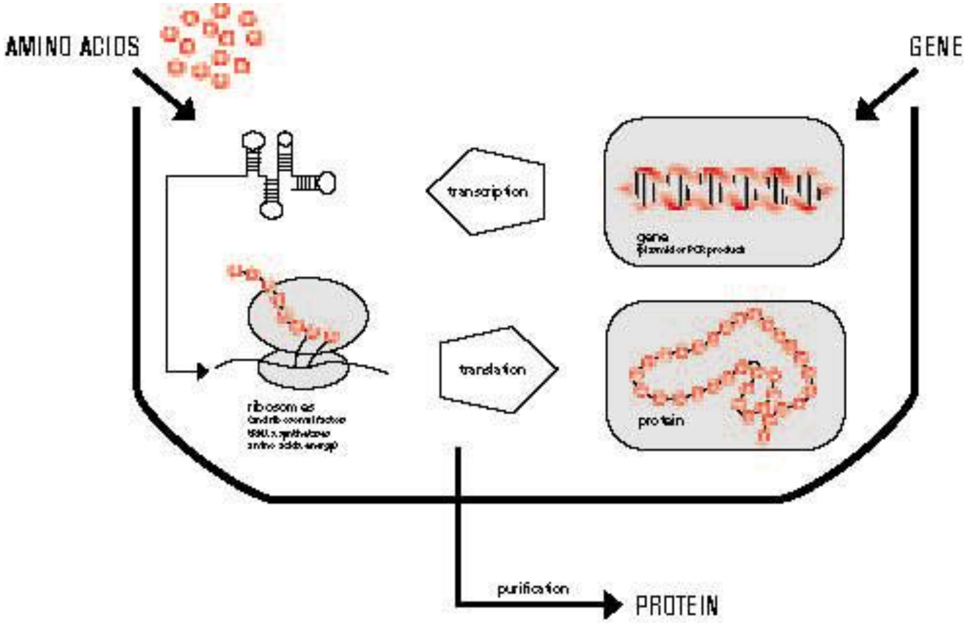
Step 4
Fold into active protein

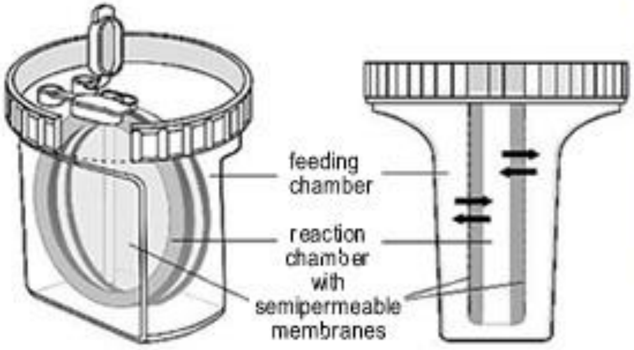
S.B.H. Kent, P. Dawson, Synthesis of Native Proteins by Chemical Ligation, Annual Review of Biochemistry 69, 925-962 (2000).

Biochemical synthesis

- In vitro translation
  - Uses all the machinery of the cell for protein synthesis
  - 0.1 to 1mg of protein
  - ~$500/per run

Over 30 different proteins produced this way





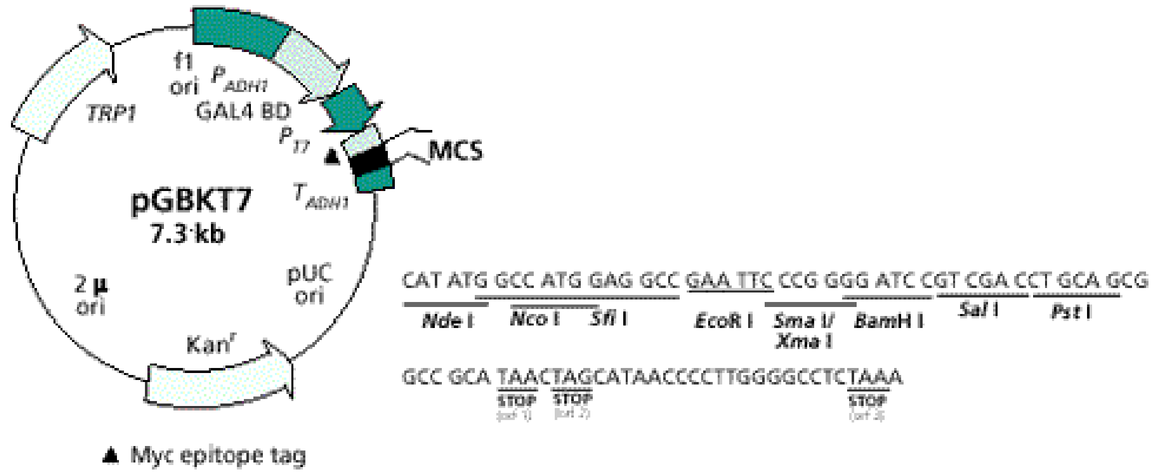
Expression

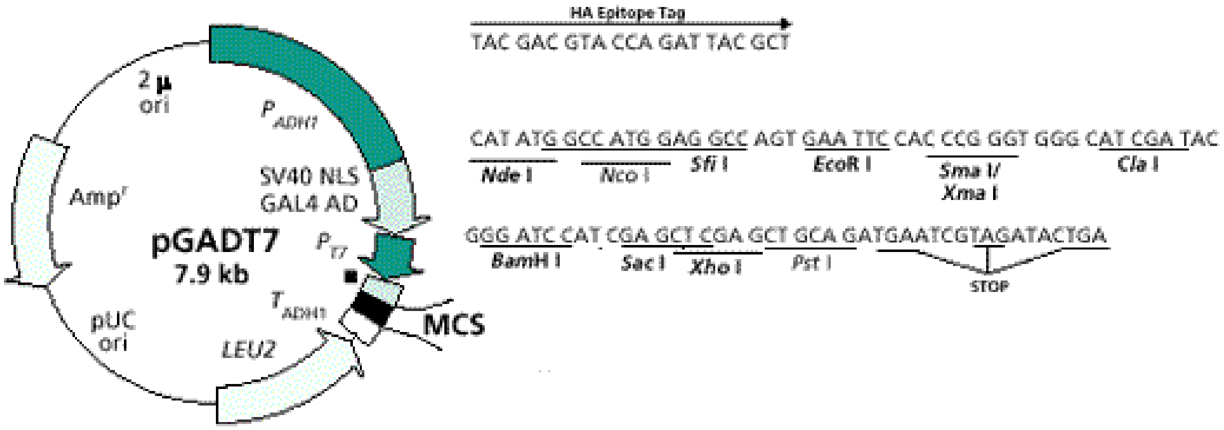
- Use of cellular machinery to make your proteins
  - Uses DNA transfection/transformation to modify a host for protein expression.
  - From micrograms to kilograms of protein
  - Cheapest route to proteins, 0.1$/mg or less

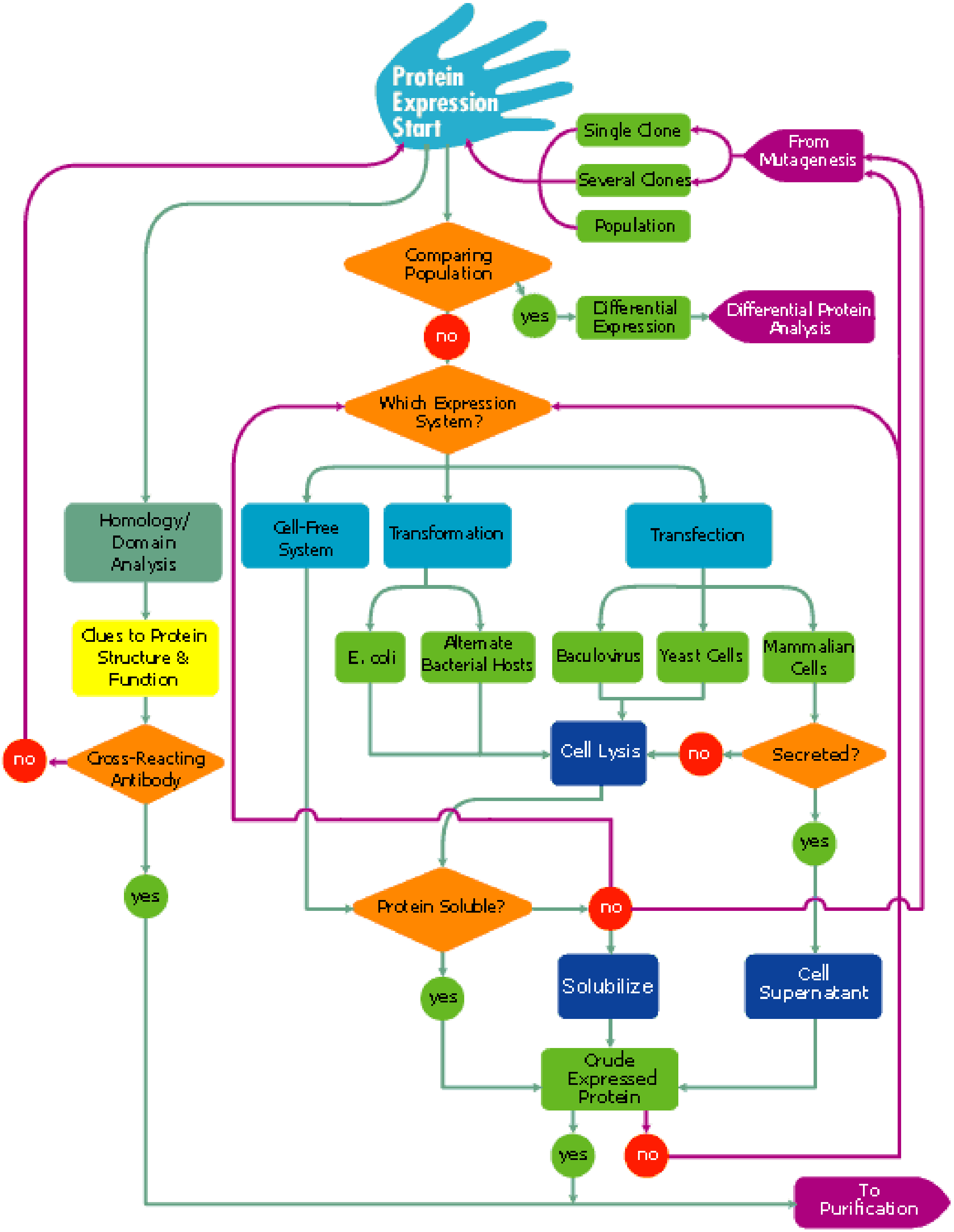
The needed components

- DNA of target protein
  - Via PCR
- Expression vector
  - Many, many commercially available
- Expression host
  - Bacterial, yeast, insect, mammalian

Expression vector sample






Protein Expression Start
Single Clone
Several Clones
Population
From Mutagenesis
Comparing Population
yes
Differential Expression
Differential Protein Analysis
no
Which Expression System?
Homology/ Domain Analysis
Cell-Free System
Transformation
Transfection
Clues to Protein Structure & Function
E. coli
Alternate Bacterial Hosts
Baculovirus
Yeast Cells
Mammalian Cells
no
Cross-Reacting Antibody
Cell Lysis
no
Secreted?
yes
yes
Protein Soluble?
no
yes
Solubilize
Cell Supernatant
Crude Expressed Protein
yes
no
To Purification

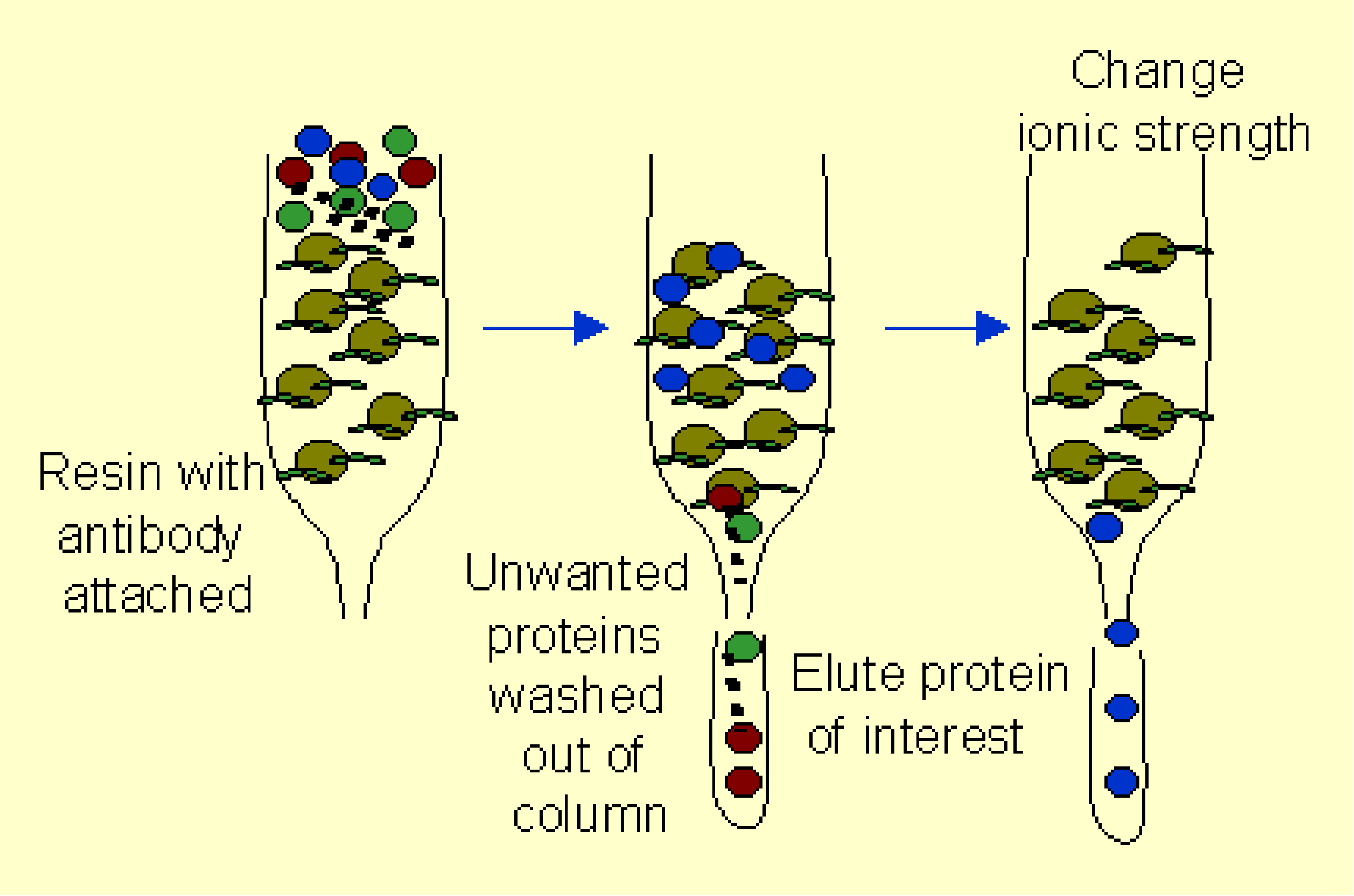
Which expression system

- Bacterial
  - Easiest and cheapest
- Yeast
  - Next easiest cost about the same
- Insect
  - Expensive, time consuming
- Mamallian
  - Most expensive and time consuming

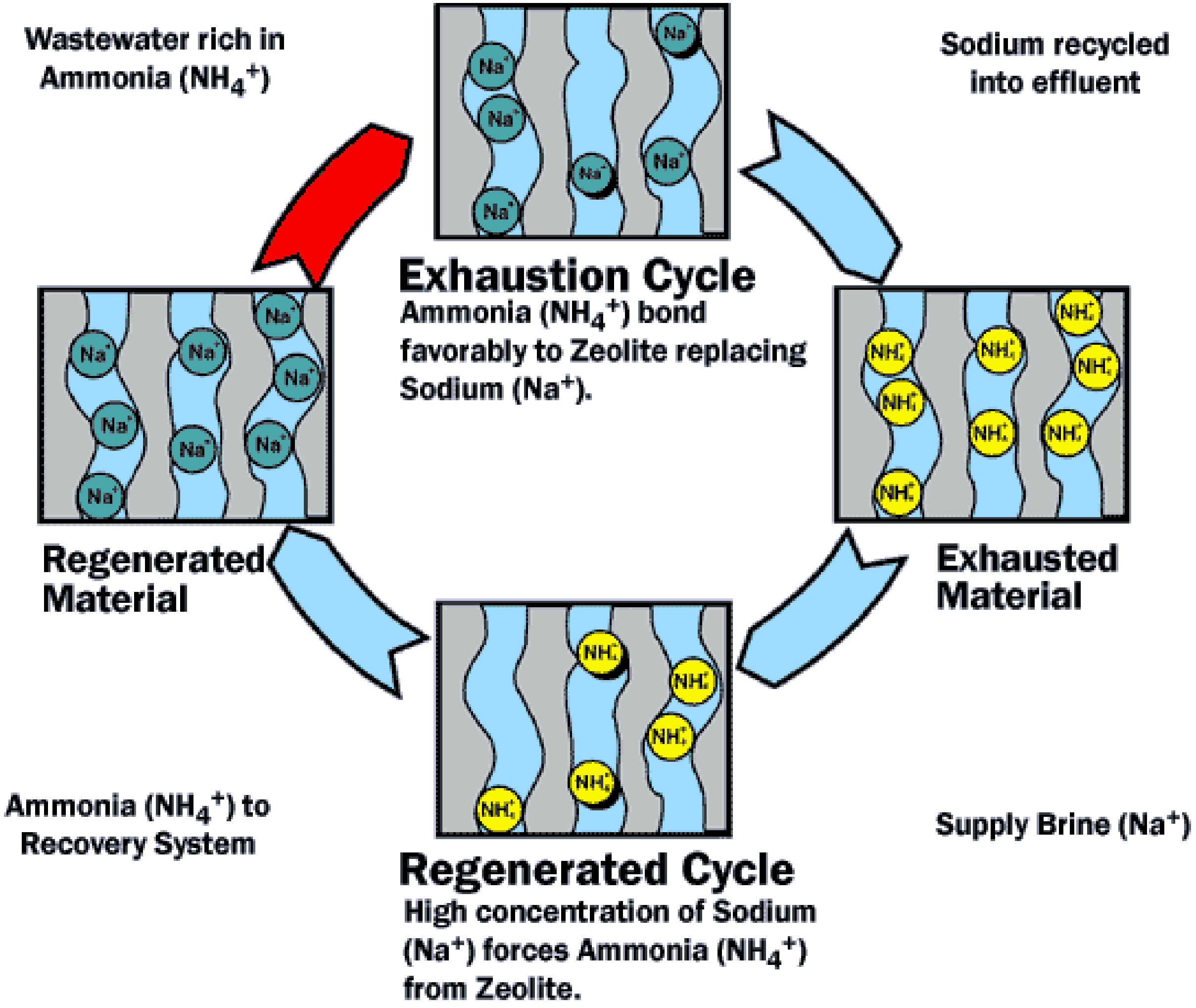
Purification

- Affinity
  - Best but can be expensive and difficult
  - Easy when a “tagged” protein is being purified
- Ion exchange
  - Simple and universal
- Size exclusion

Affinity


Change
ionic strength
Resin with
antibody
attached
Unwanted
proteins
washed
out of
column
Elute protein
of interest

Ion exchange



Wastewater rich in Ammonia (NH_4^+)

Sodium recycled into effluent

Exhaustion Cycle

Ammonia (NH_4^+) bond favorably to Zeolite replacing Sodium (Na^+).

Regenerated Material

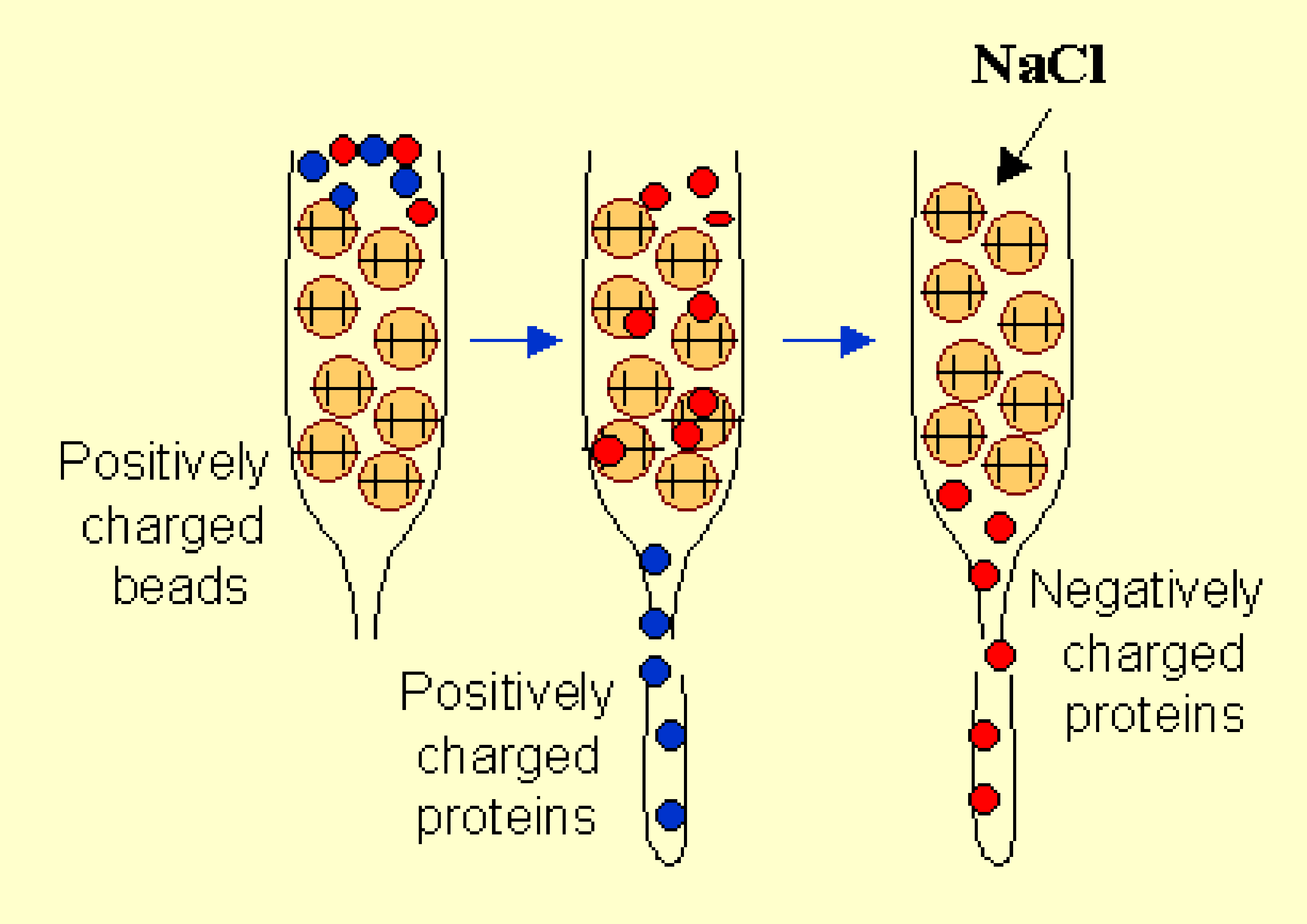
Exhausted Material

Ammonia (NH_4^+) to Recovery System

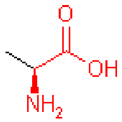
Supply Brine (Na^+)

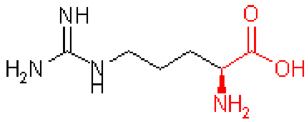
Regenerated Cycle

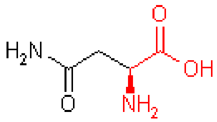
High concentration of Sodium (Na^+) forces Ammonia (NH_4^+) from Zeolite.

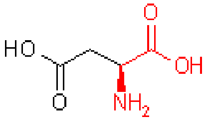

NaCl
Positively charged beads
Positively charged proteins
Negatively charged proteins

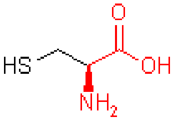
Different side chains different pKa

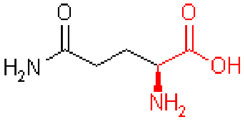

Ala

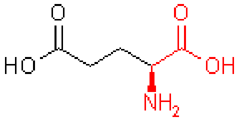

Arg

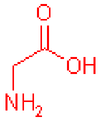

Asn

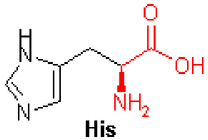

Asp

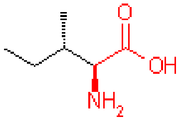

Cys

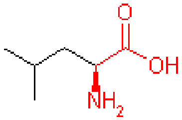

Gln

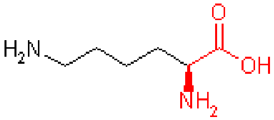

Glu

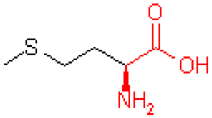

Gly

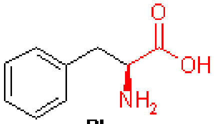

His

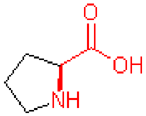

Ile

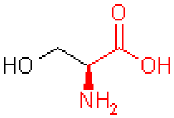

Leu

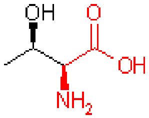

Lys

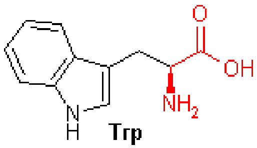

Met

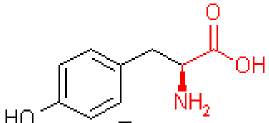

Phe

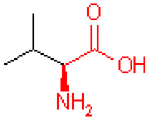

Pro

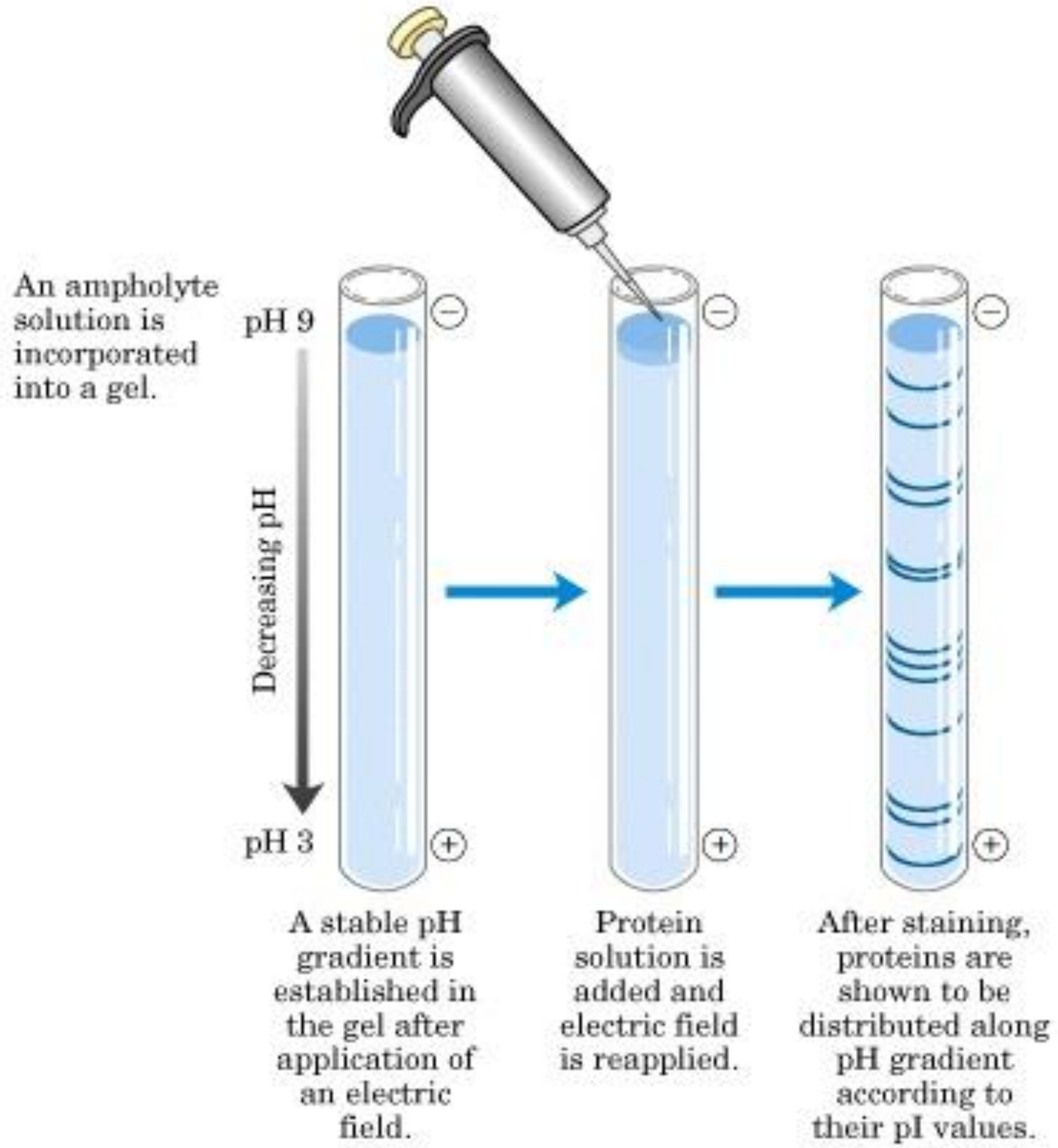

Ser


Thr


Trp


Tyr


Val



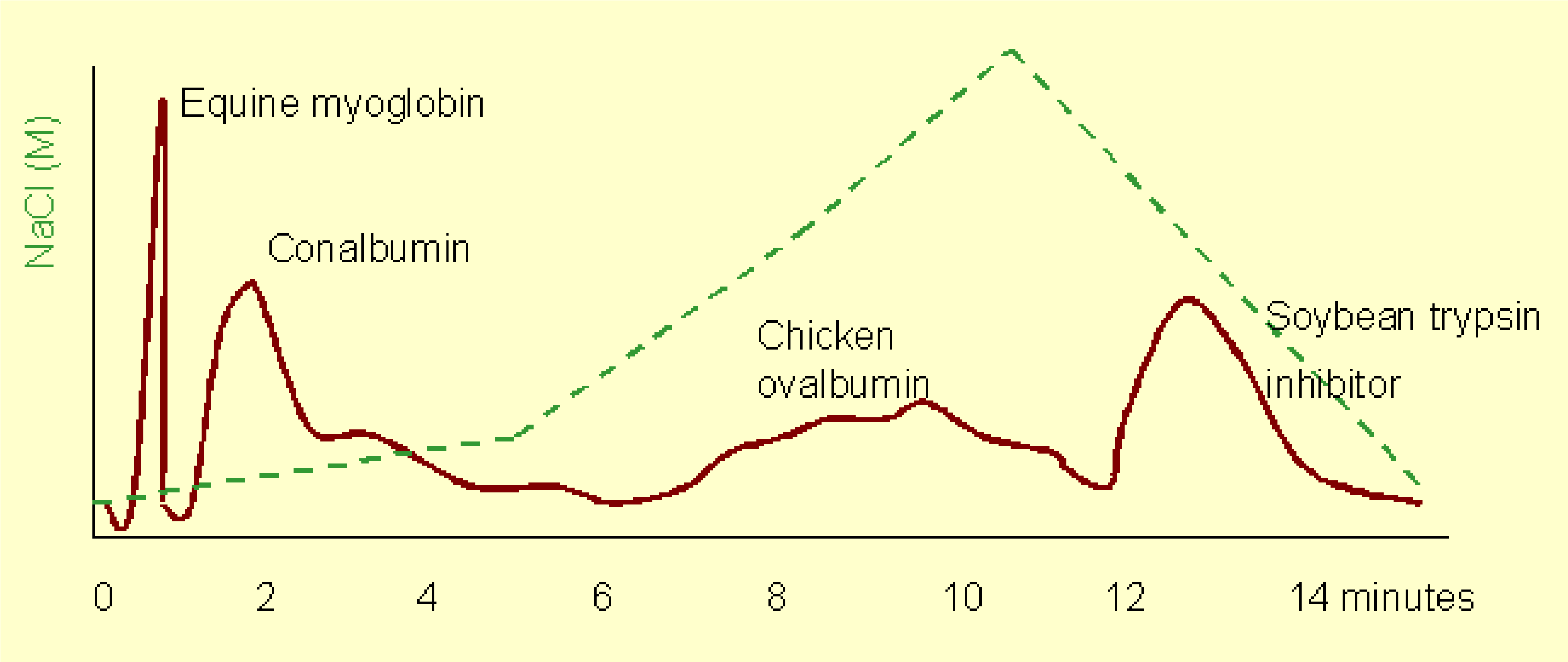
An ampholyte solution is incorporated into a gel.

A stable pH gradient is established in the gel after application of an electric field.

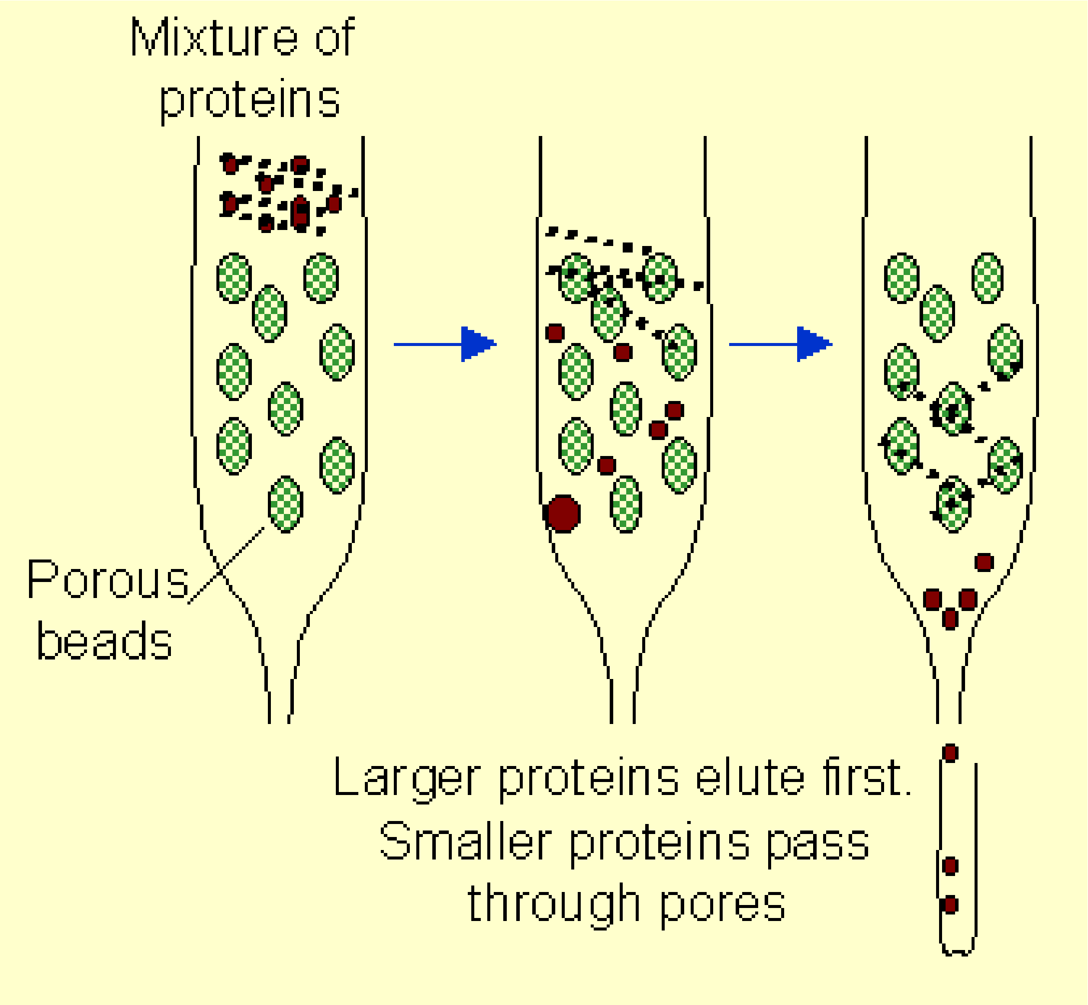
Protein solution is added and electric field is reapplied.

After staining, proteins are shown to be distributed along pH gradient according to their pI values.

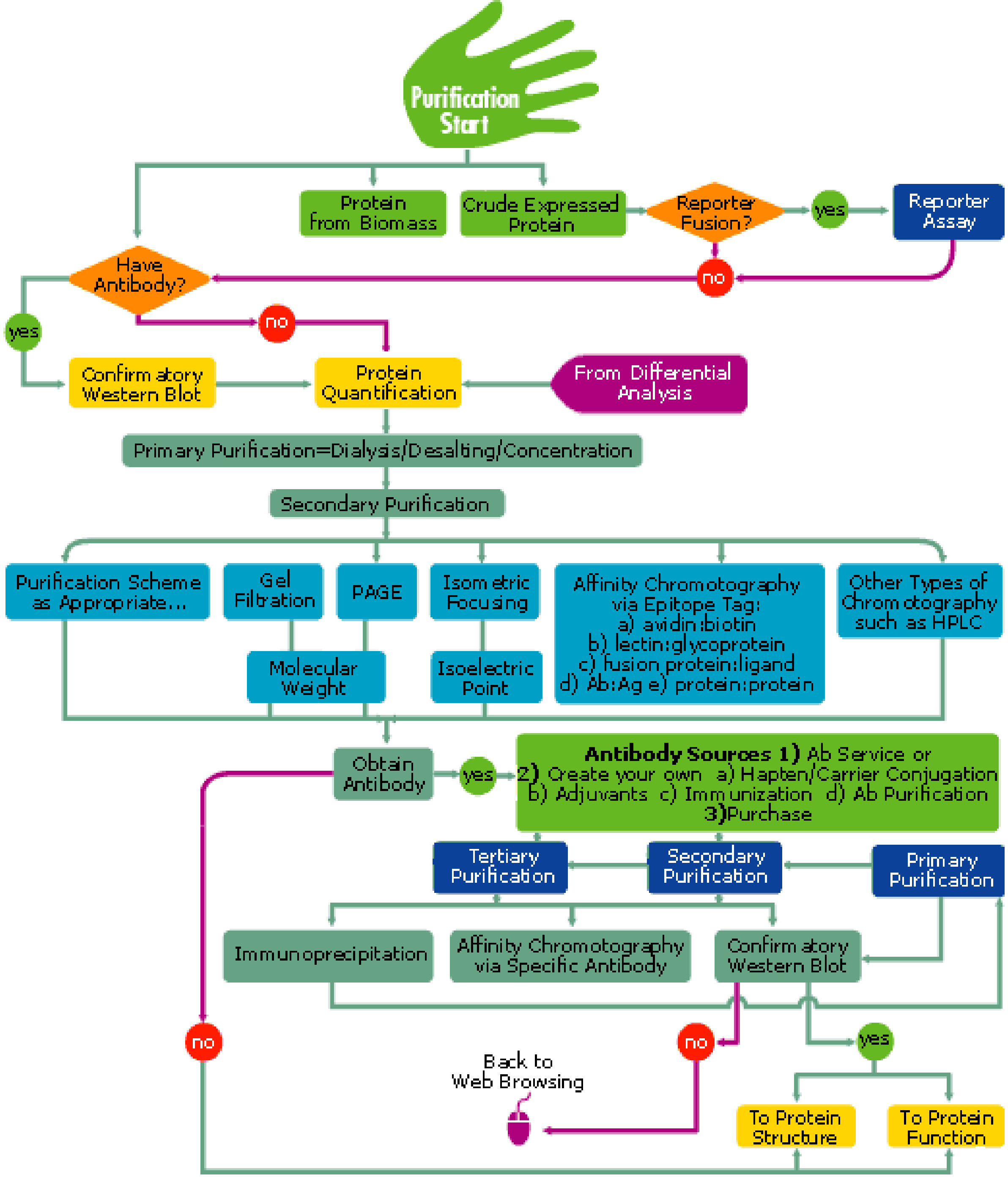
http://scansite.mit.edu/calc_mw_pi.html


NaCl (M)
Equine myoglobin
Conalbumin
Chicken ovalbumin
Soybean trypsin inhibitor
0
2
4
6
8
10
12
14 minutes

Size exclusion



Larger proteins elute first. Smaller proteins pass through pores



Purification Start

Protein from Biomass

Crude Expressed Protein

Reporter Fusion?

yes

Reporter Assay

no

Have Antibody?

yes

no

Confirmatory Western Blot

Protein Quantification

From Differential Analysis

Primary Purification=Dialysis/Desalting/Concentration

Secondary Purification

Purification Scheme as Appropriate...

Gel Filtration

PAGE

Isometric Focusing

Affinity Chromatography via Epitope Tag:
a) avidin:biotin
b) lectin:glycoprotein
c) fusion protein:ligand
d) Ab:Ag e) protein:protein

Other Types of Chromotography such as HPLC

Molecular Weight

Isoelectric Point

Obtain Antibody

yes

Antibody Sources 1) Ab Service or **2)** Create your own a) Hapten/Carrier Conjugation b) Adjuvants c) Immunization d) Ab Purification **3)**Purchase

Tertiary Purification

Secondary Purification

Primary Purification

Immunoprecipitation

Affinity Chromotography via Specific Antibody

Confirmatory Western Blot

no

no

yes

Back to Web Browsing

To Protein Structure

To Protein Function

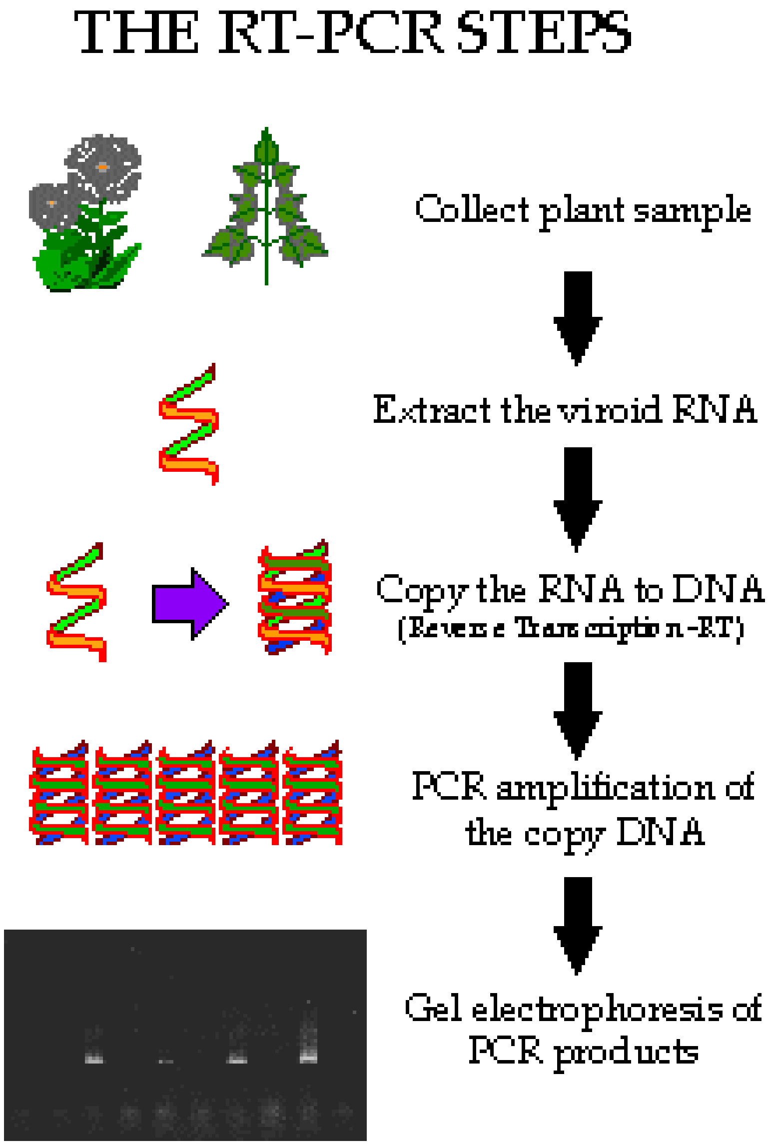
Workshop

- Protein synthesis and protein purification are much more expensive and complicated relative to DNA or RNA synthesis and purification. Why is this?

Characterization of biomolecules

- DNA/RNA
  - RT-PCR
  - Sanger sequencing method
- Proteins
  - Electrophoresis
  - Edman degradation sequencing
  - Active site mapping
  - Structural techniques

RNA characterization



THE RT-PCR STEPS

Collect plant sample

Extract the viroid RNA

Copy the RNA to DNA
(Reverse Transcription-RT)

PCR amplification of
the copy DNA

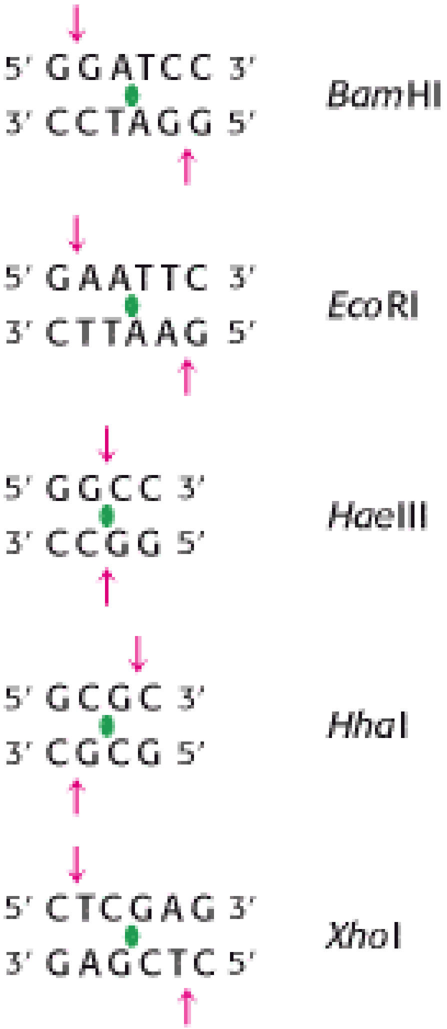
Gel electrophoresis of
PCR products

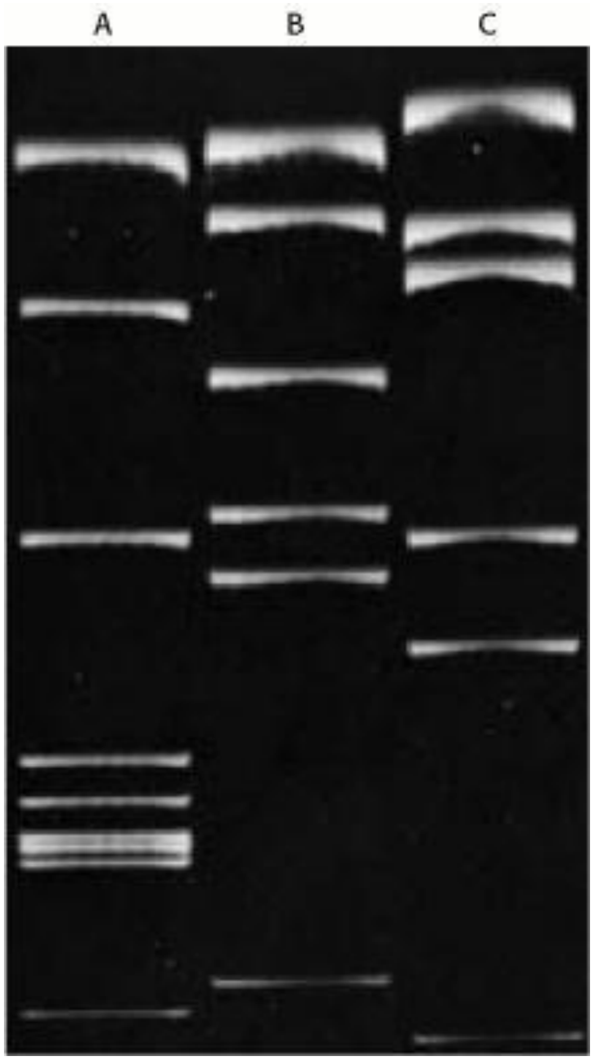
We can follow up with DNA characterization methods such as sequencing

RT-PCR key issues

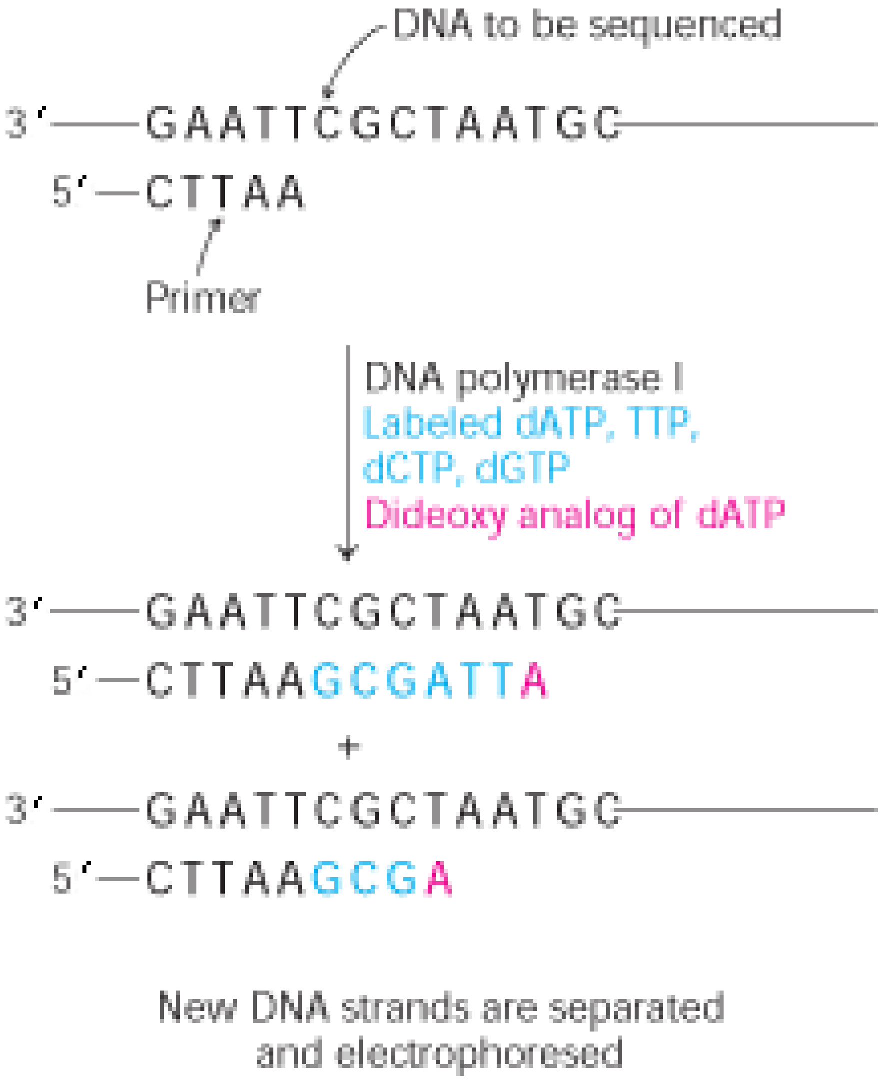
- We can use a poly T primer for RT but we need a forward primer for the PCR
- Forward primers can be
  - Gene specific if you know your target
  - Random primers for general analysis

DNA characterization
Restriction digest





DNA sequencing



DNA to be sequenced

3′—GAATTCGCTAATGC—

5′—CTTAA

Primer

DNA polymerase I

Labeled dATP, TTP, dCTP, dGTP

Dideoxy analog of dATP

3′—GAATTCGCTAATGC—

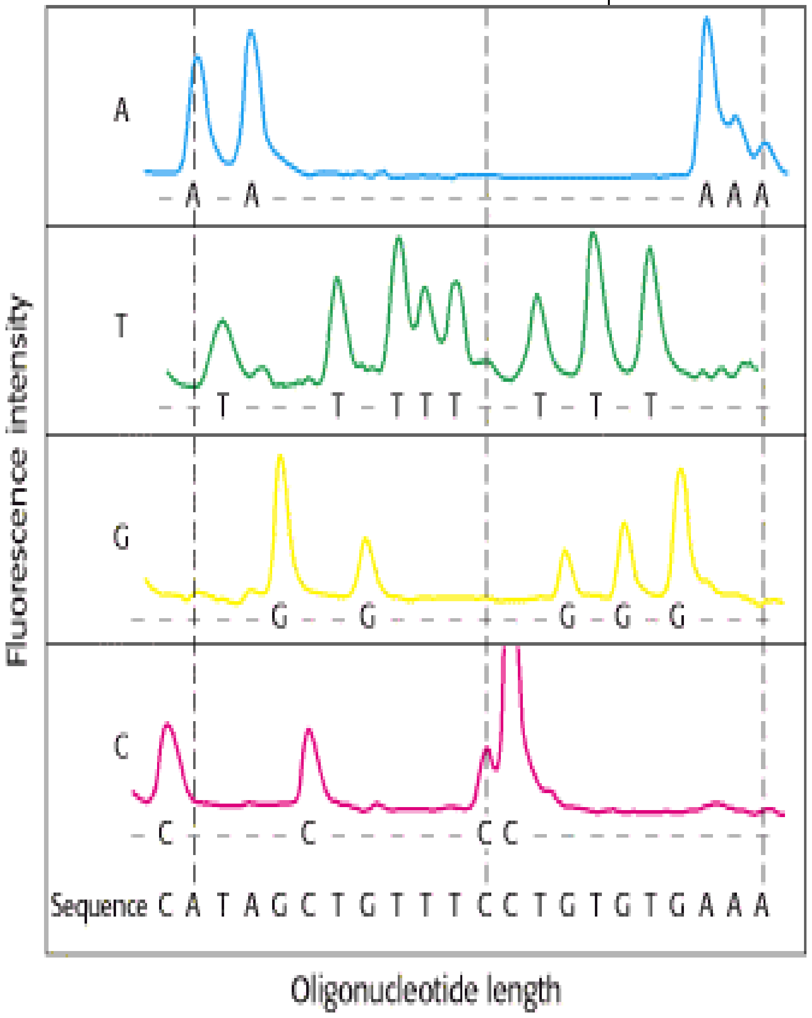
5′—CTTAAGCGATTA

+

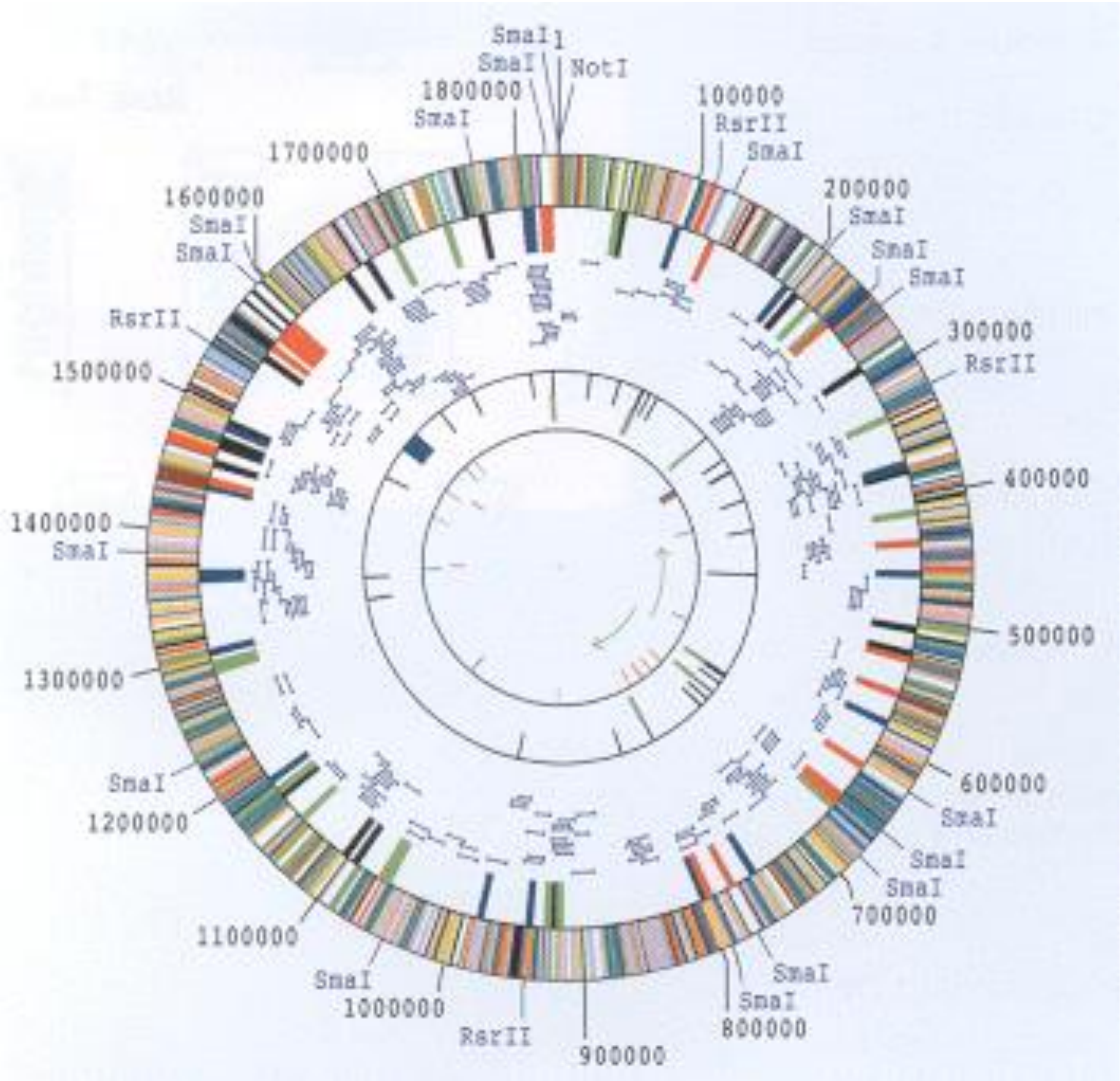
3′—GAATTCGCTAATGC—

5′—CTTAAGCGA

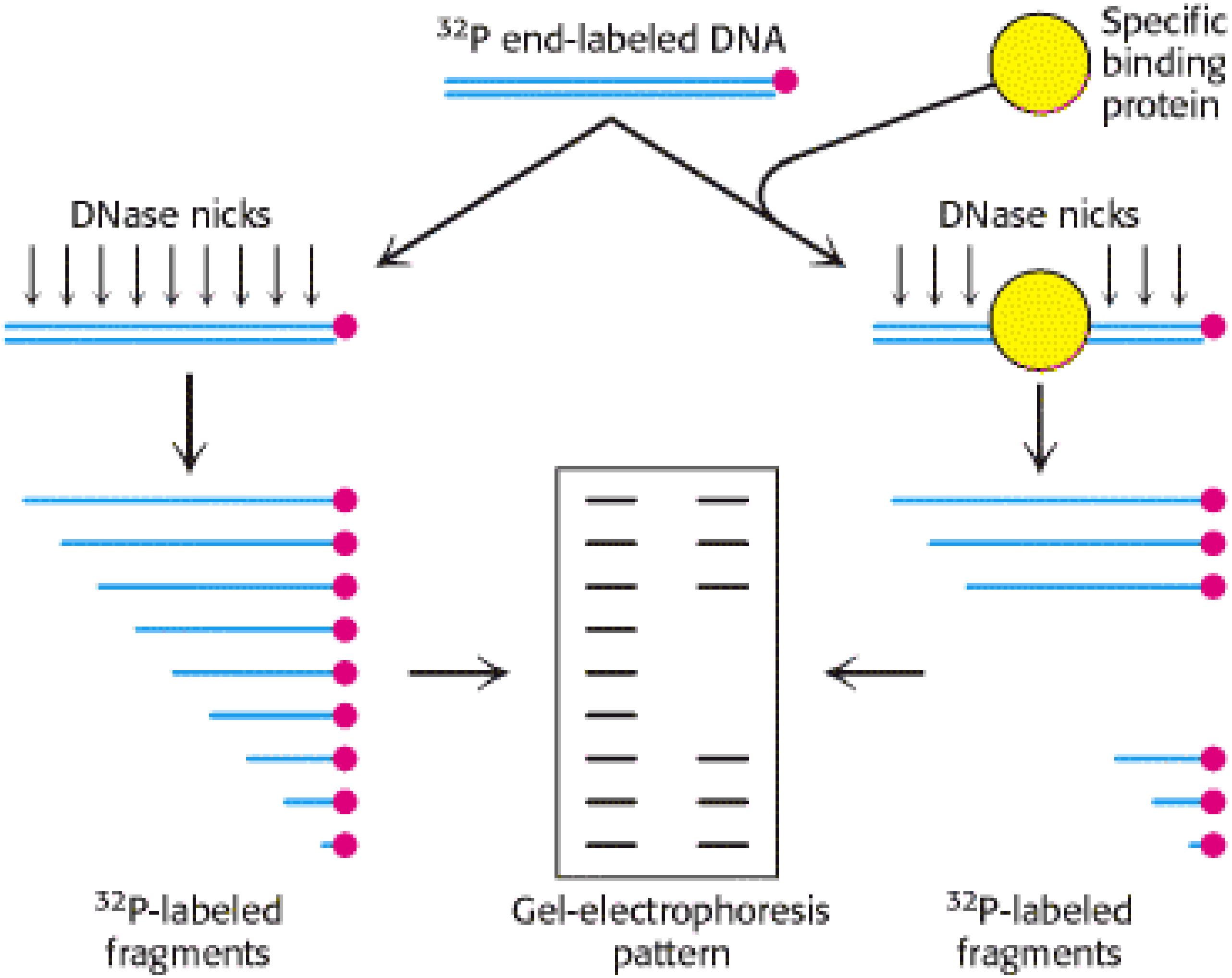
New DNA strands are separated and electrophoresed



Genome sequencing



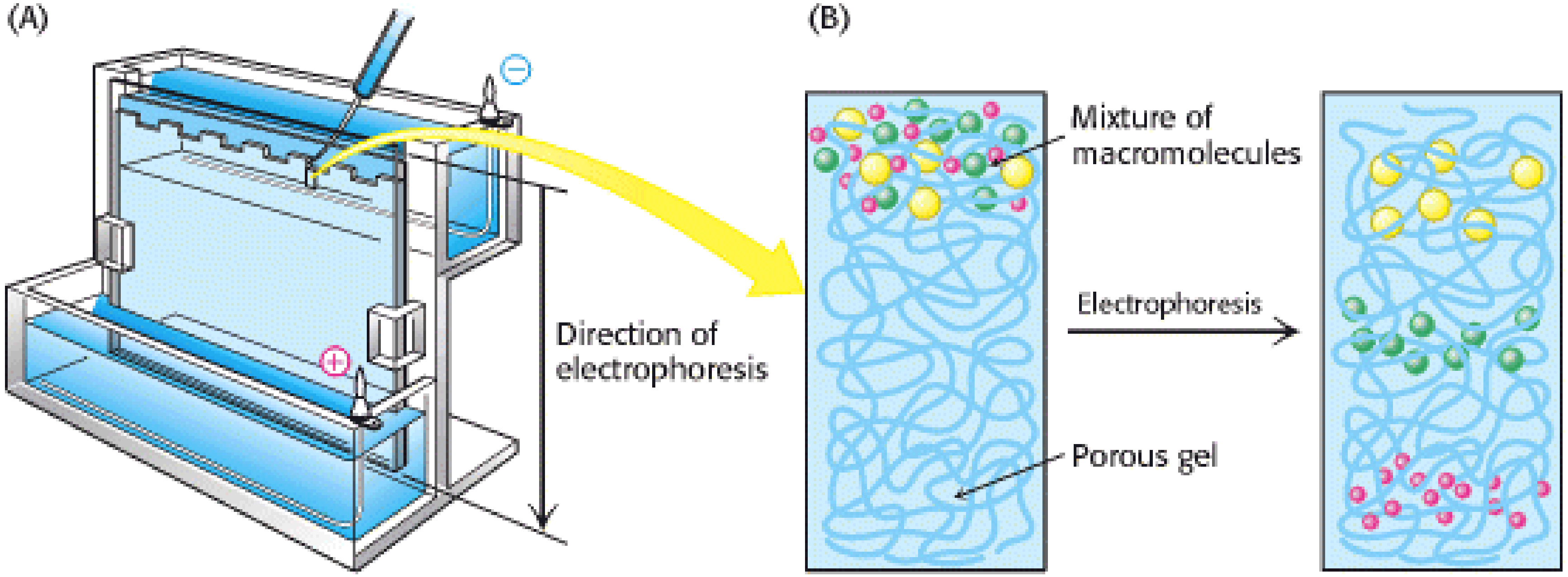
DNA footprinting



DNA/RNA characterization

- RT-PCR for RNA
- Restriction digest and sequencing of DNA
  - Sequence up 500 bp
  - Combine sequence from digested fragments to achieve complete genomic sequence

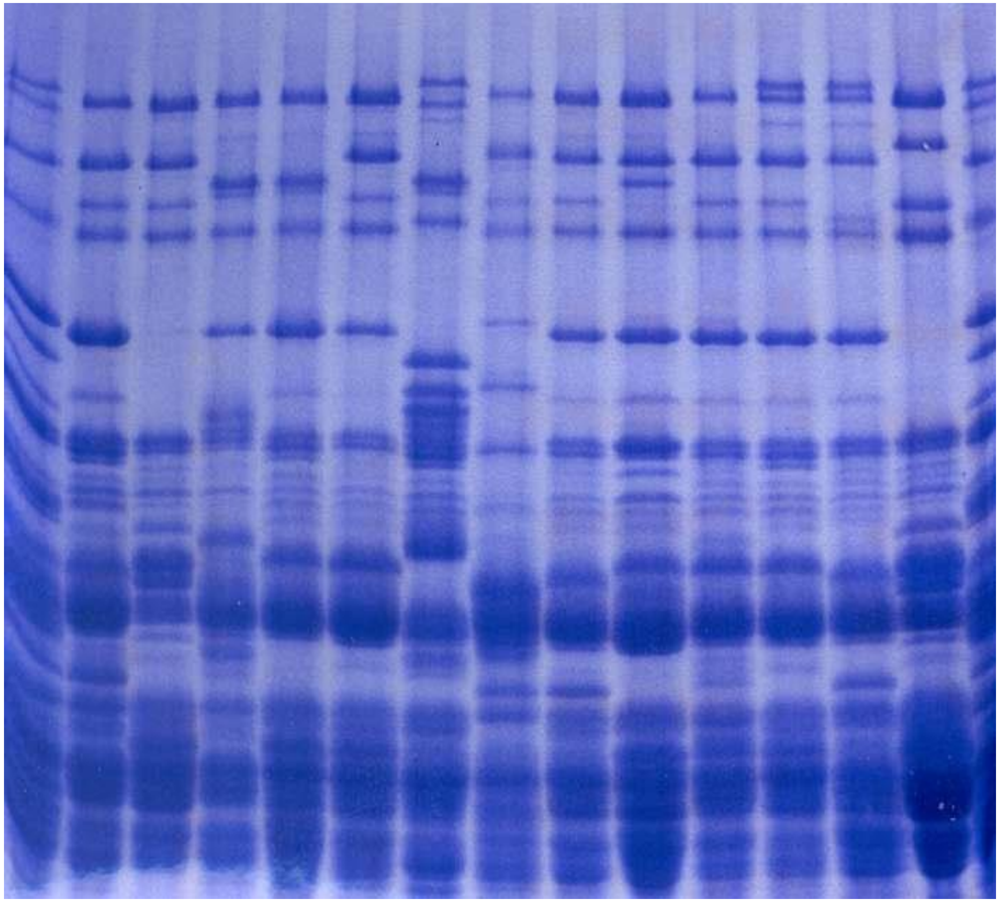
Proteins Electrophoresis

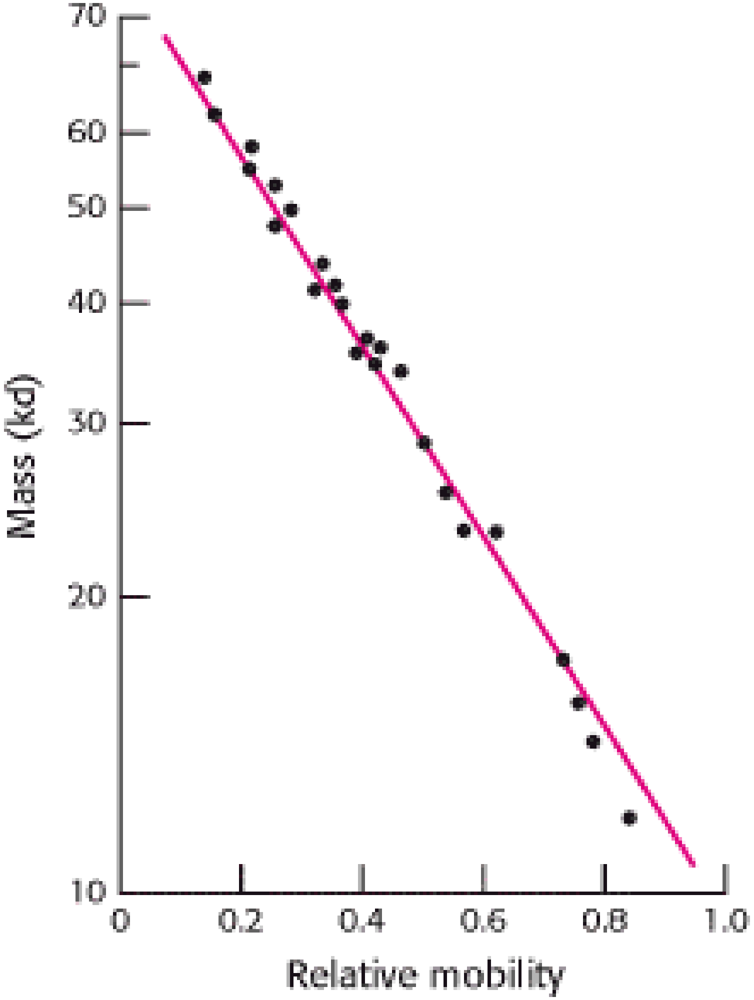


Velocity = Ez/f

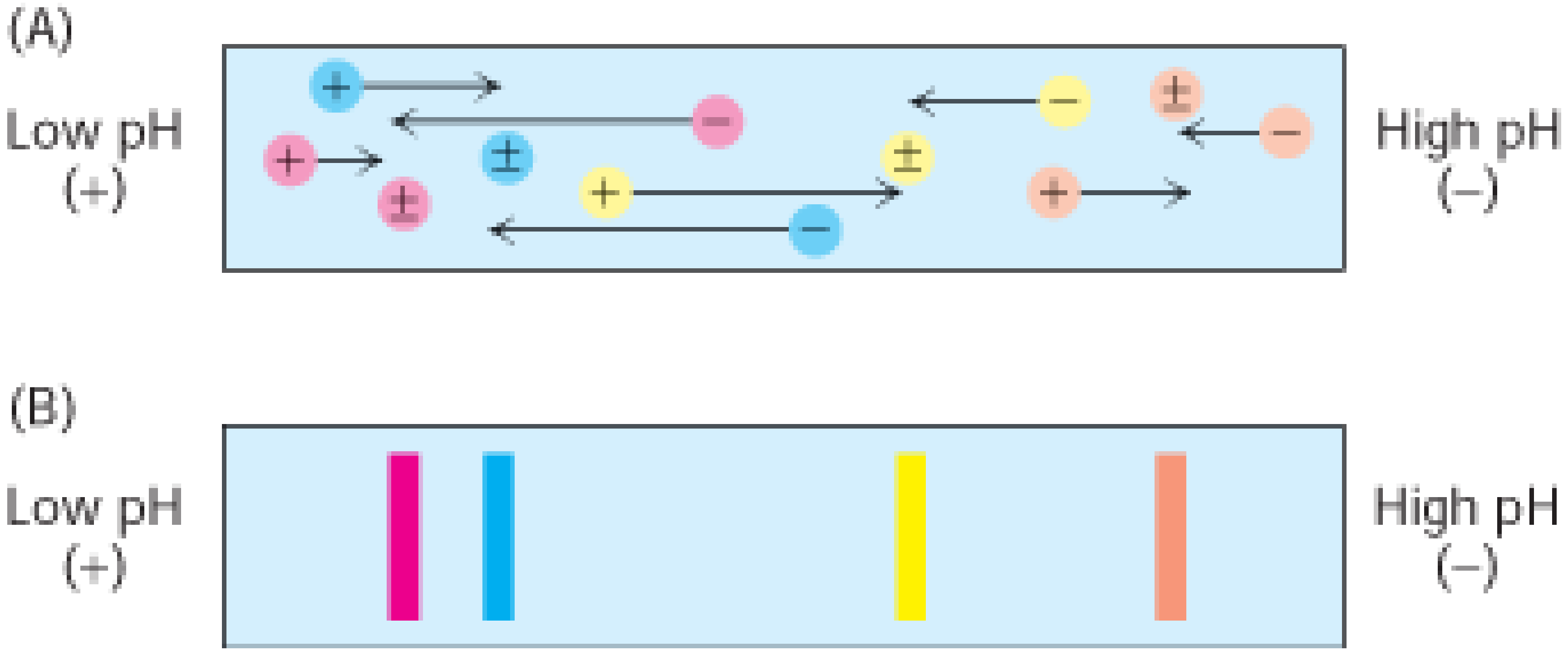
f=6πnr

Mobility proportional to mass





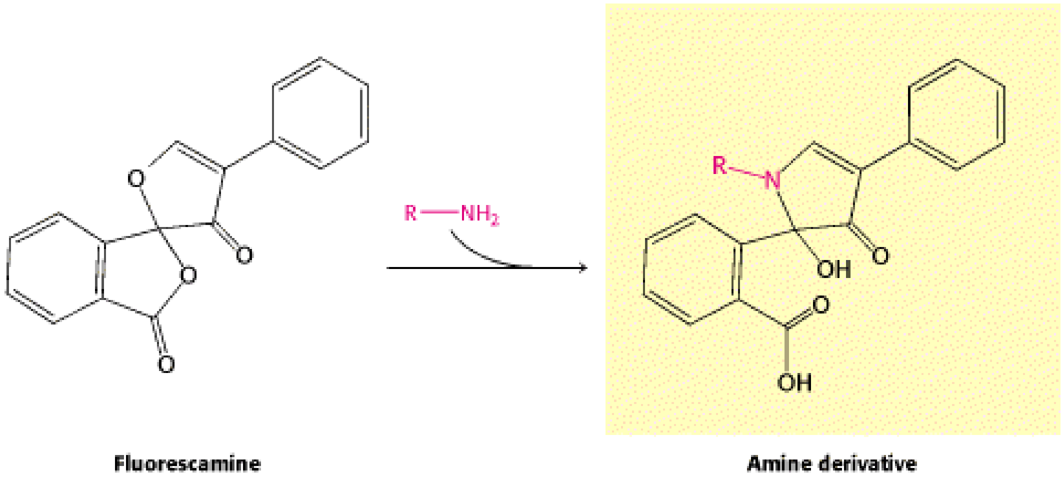
Isoelectric focusing



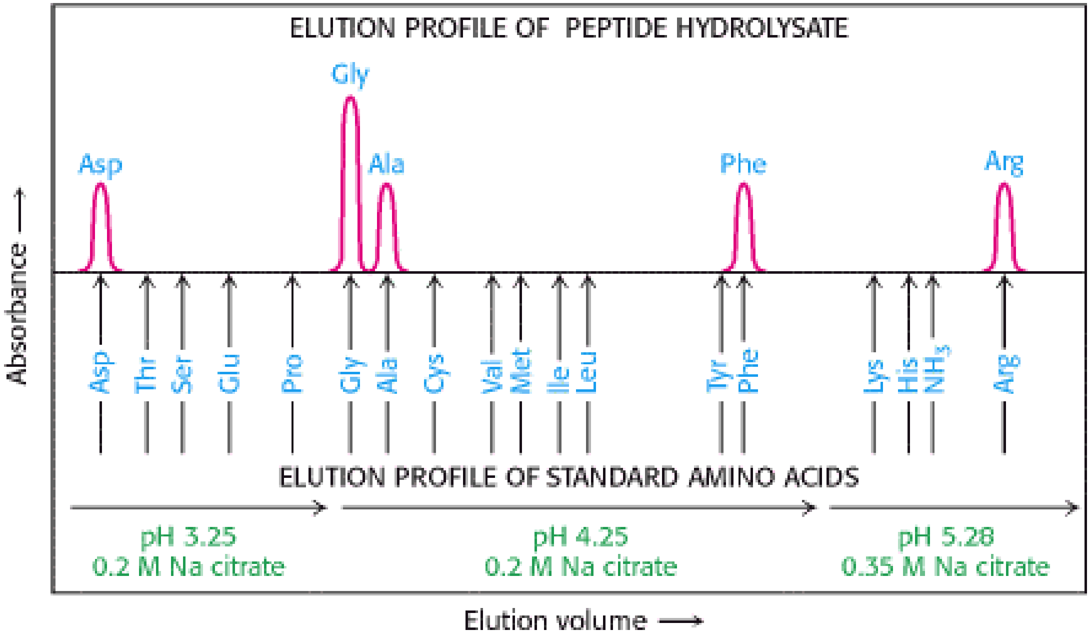
$V = Ez/f$

$f=6\pi nr$

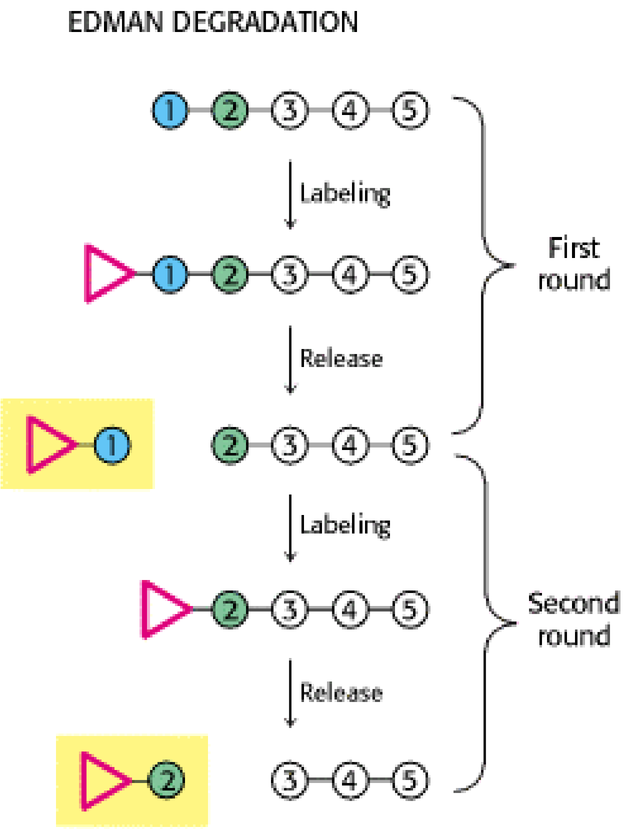
Amino acid composition

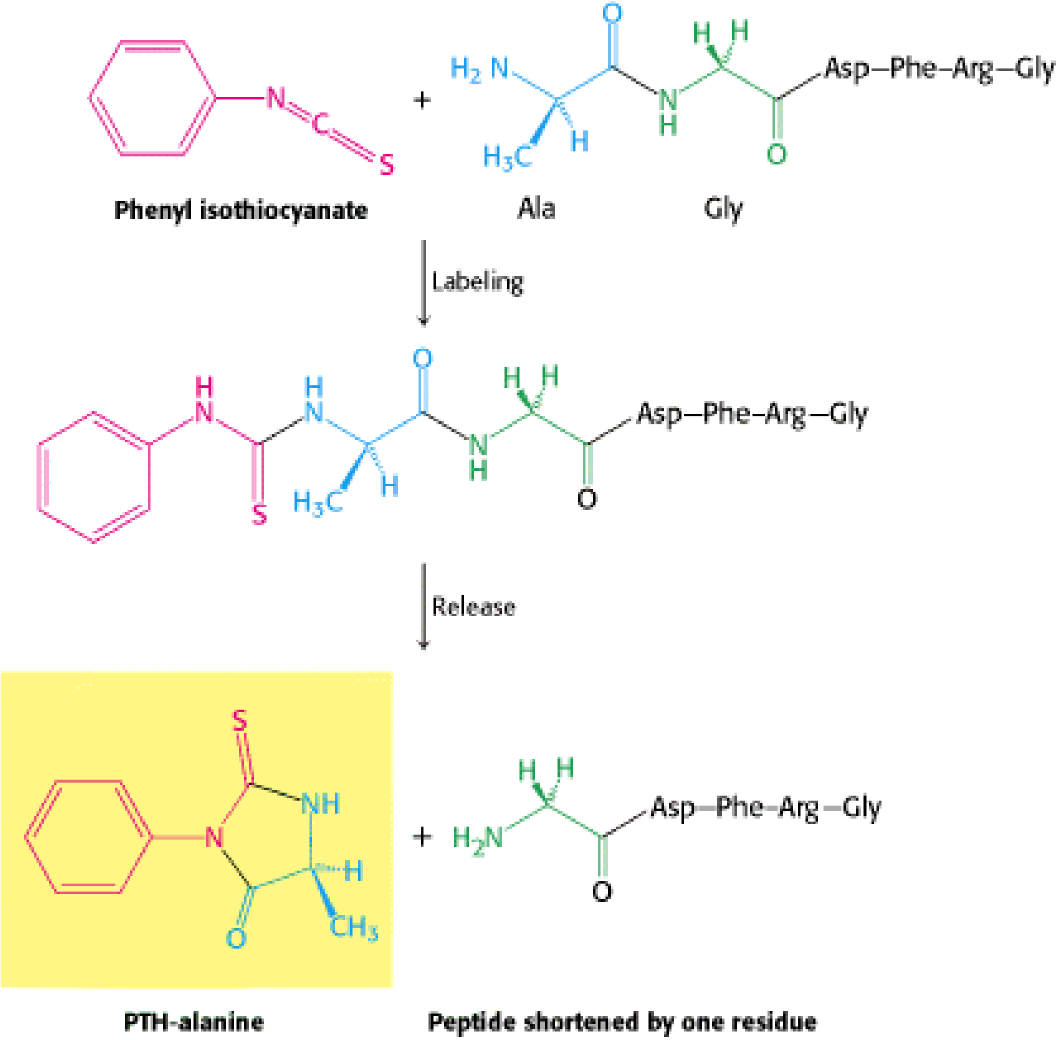


10 pico moles of amino acid

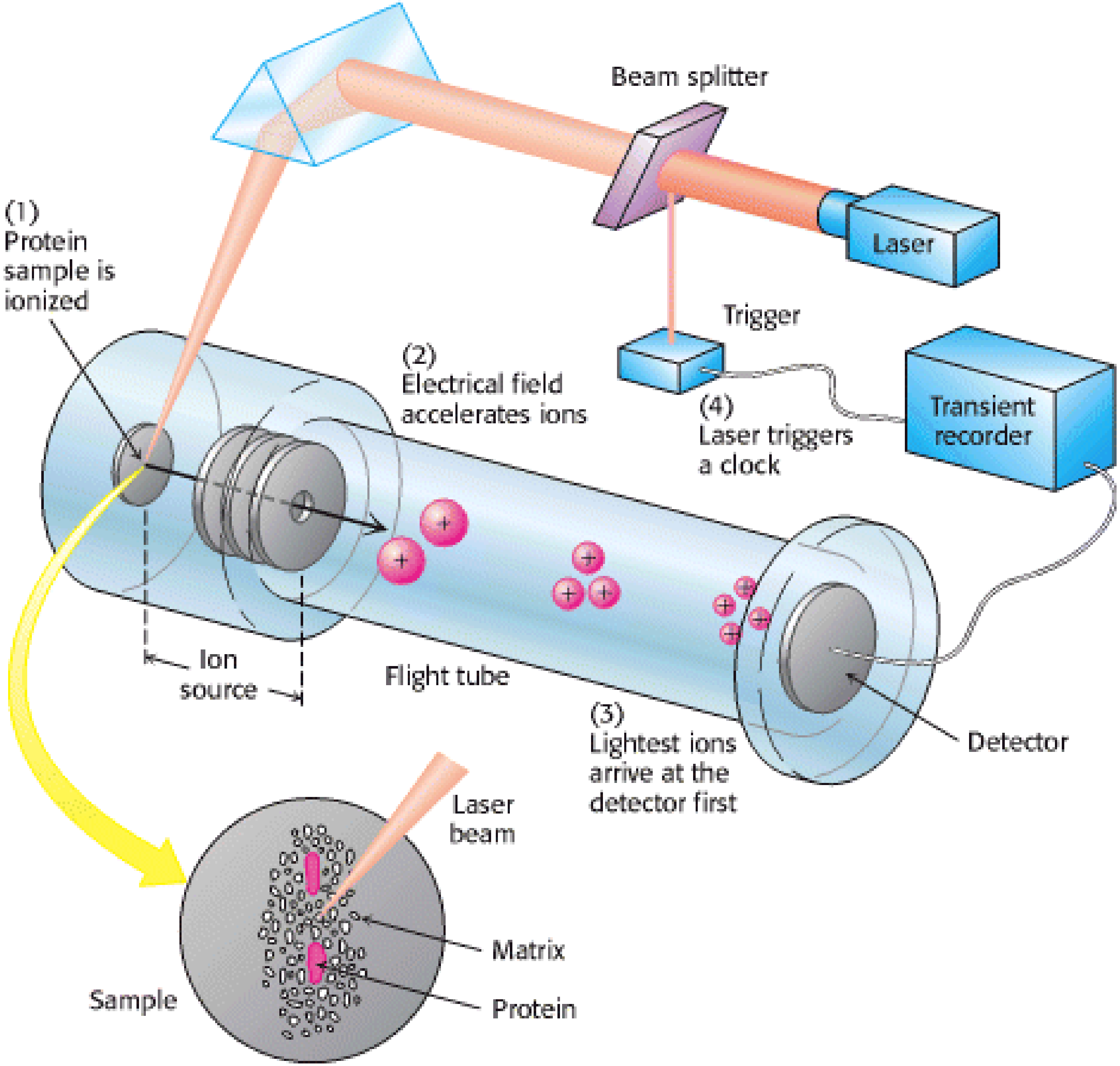


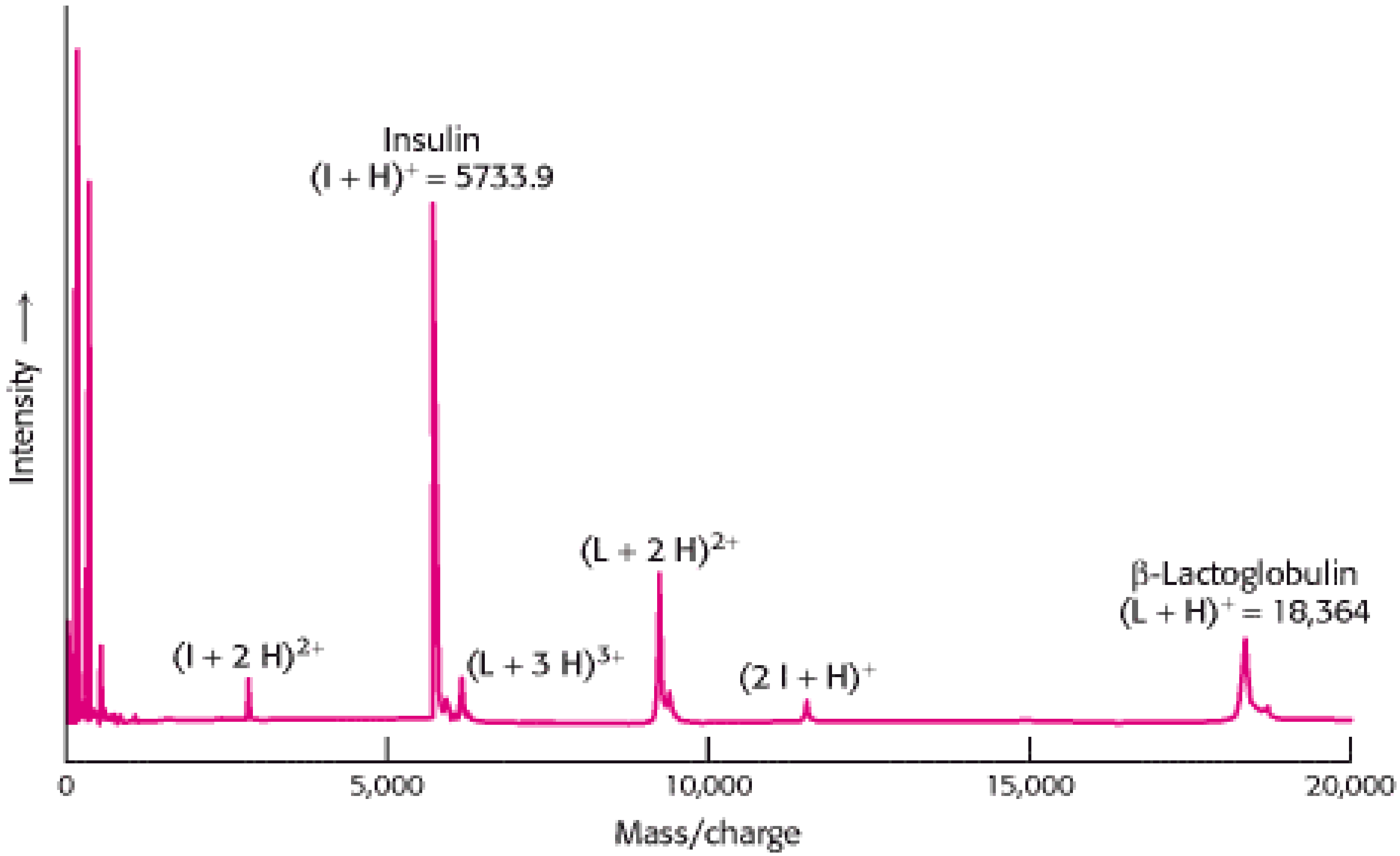
Protein sequence





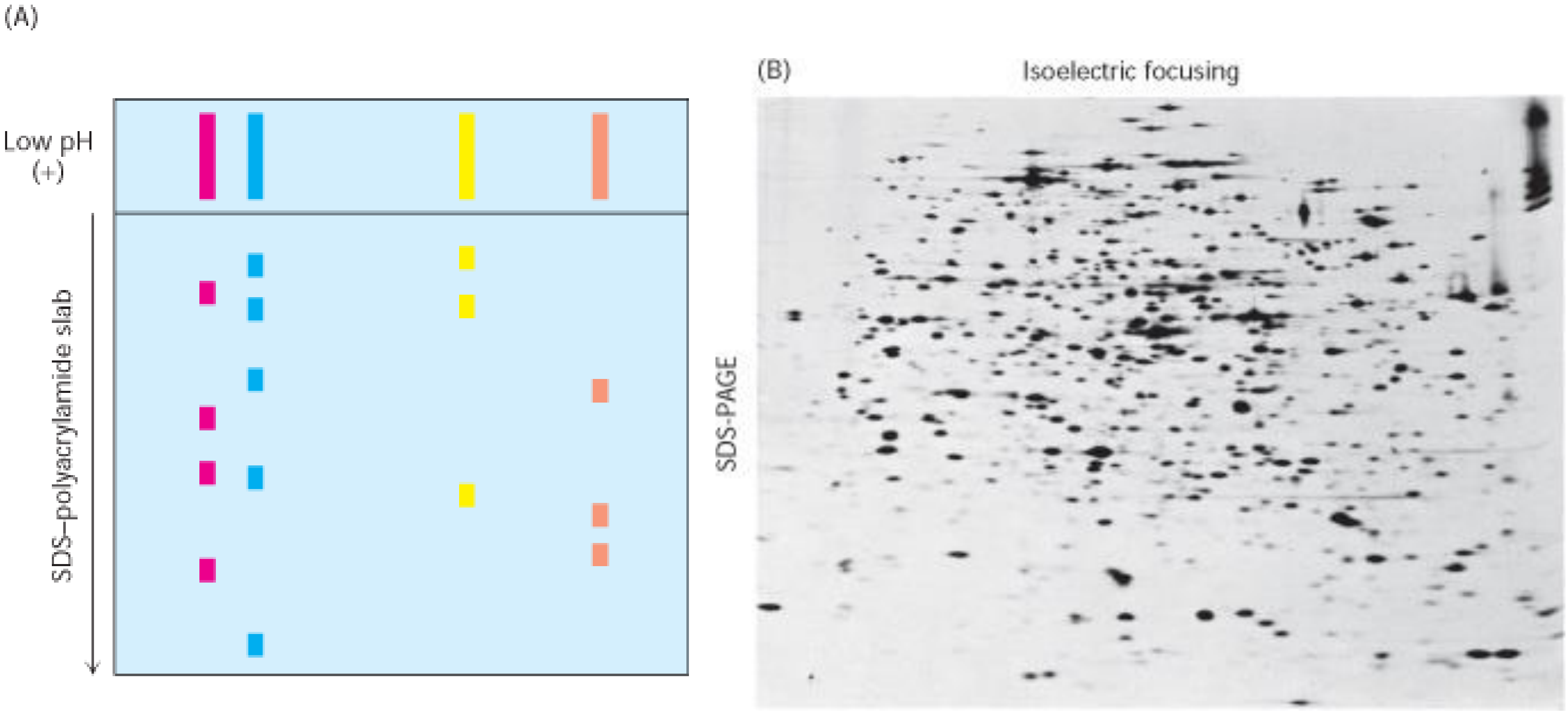
Mass spec analysis



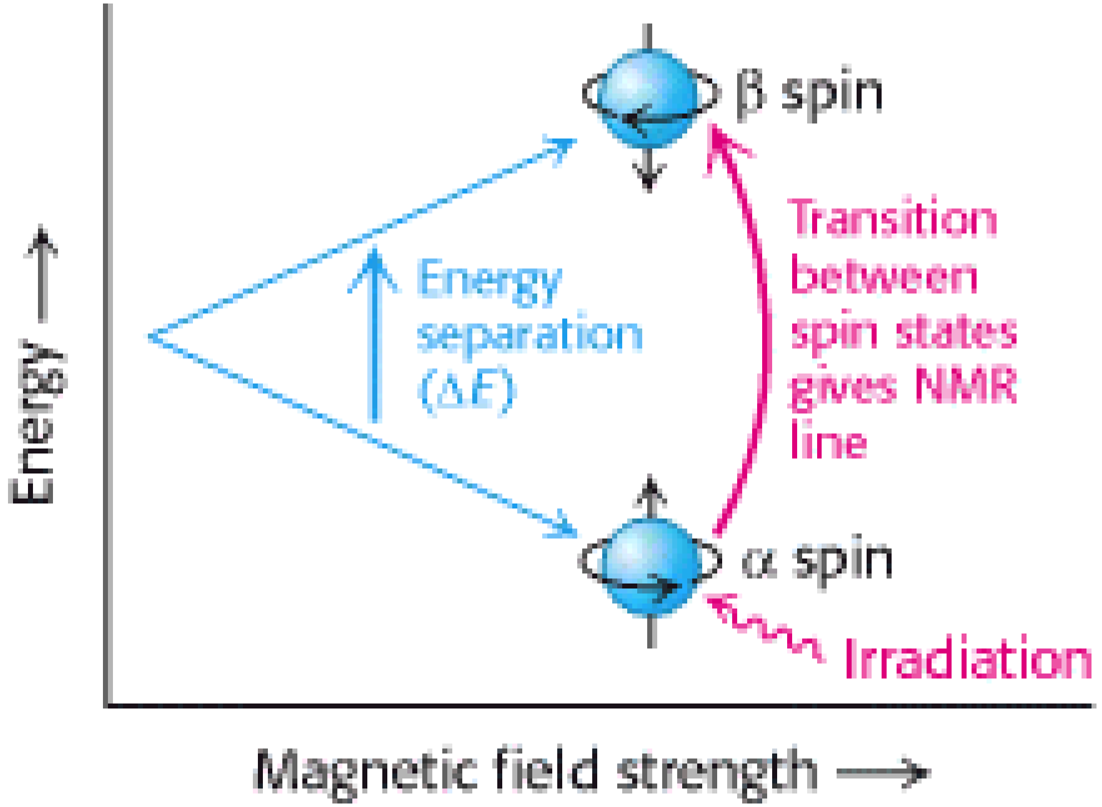

Insulin
(I + H)+ = 5733.9
(I + 2 H)2+
(L + 3 H)3+
(L + 2 H)2+
(2 I + H)+
β-Lactoglobulin
(L + H)+ = 18,364
Intensity ⟶
0
5,000
10,000
15,000
20,000
Mass/charge

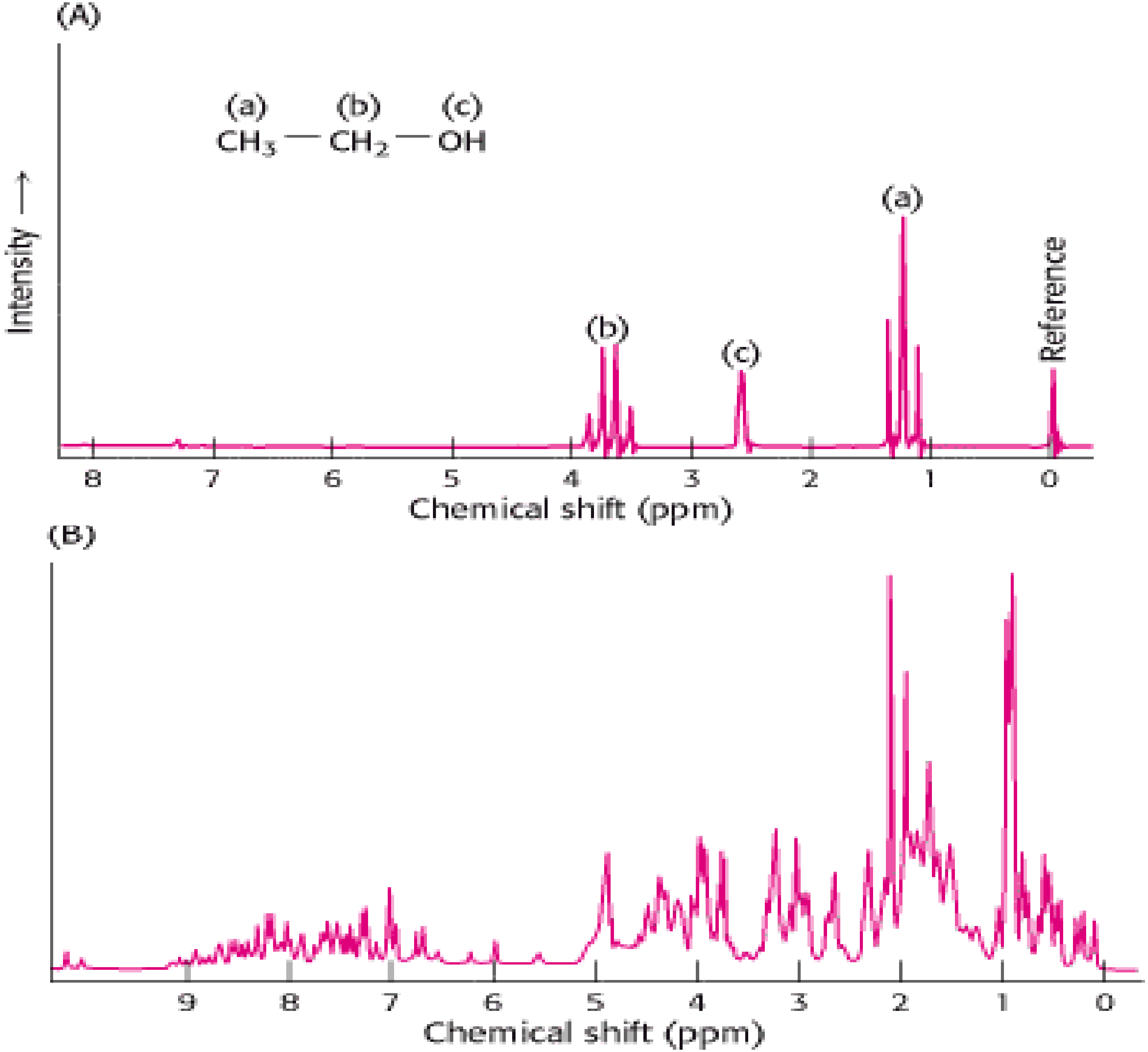
2D gels and mass spec

80% of proteins identified for yeast

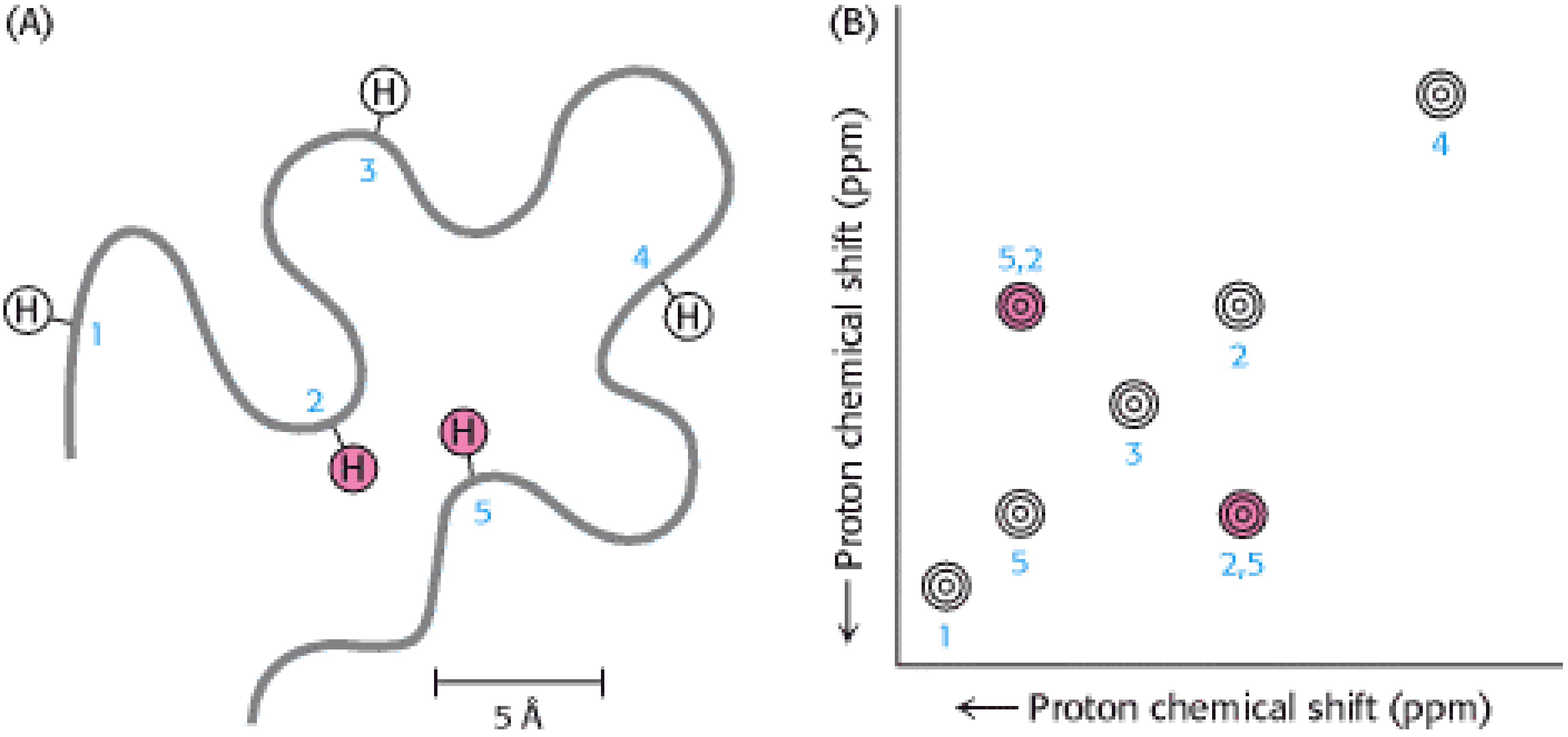


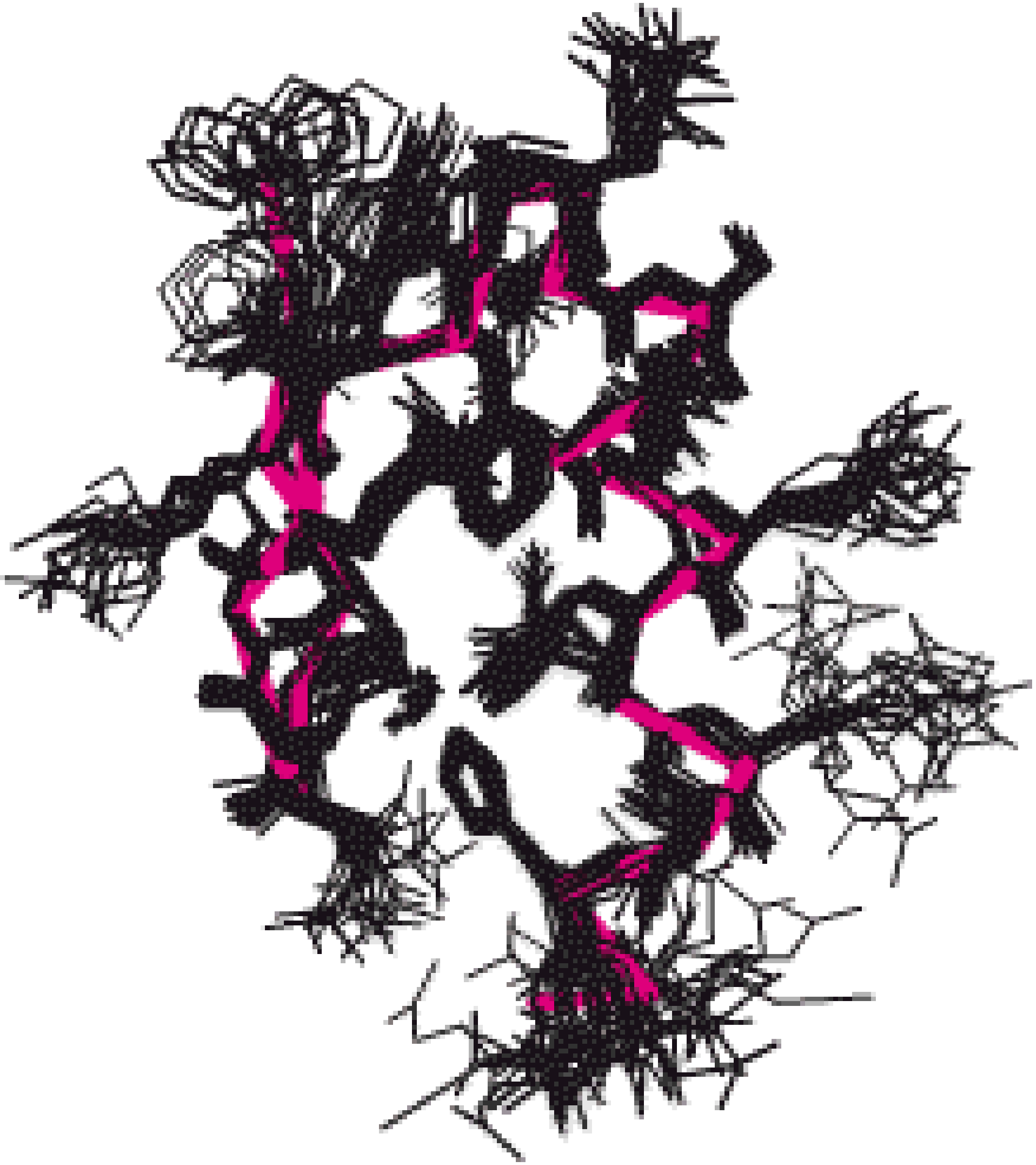
Proteins
NMR




(A)
(a) (b) (c)
CH3 — CH2 — OH
Intensity →
(a)
(b)
(c)
Reference
8 7 6 5 4 3 2 1 0
Chemical shift (ppm)
(B)
9 8 7 6 5 4 3 2 1 0
Chemical shift (ppm)

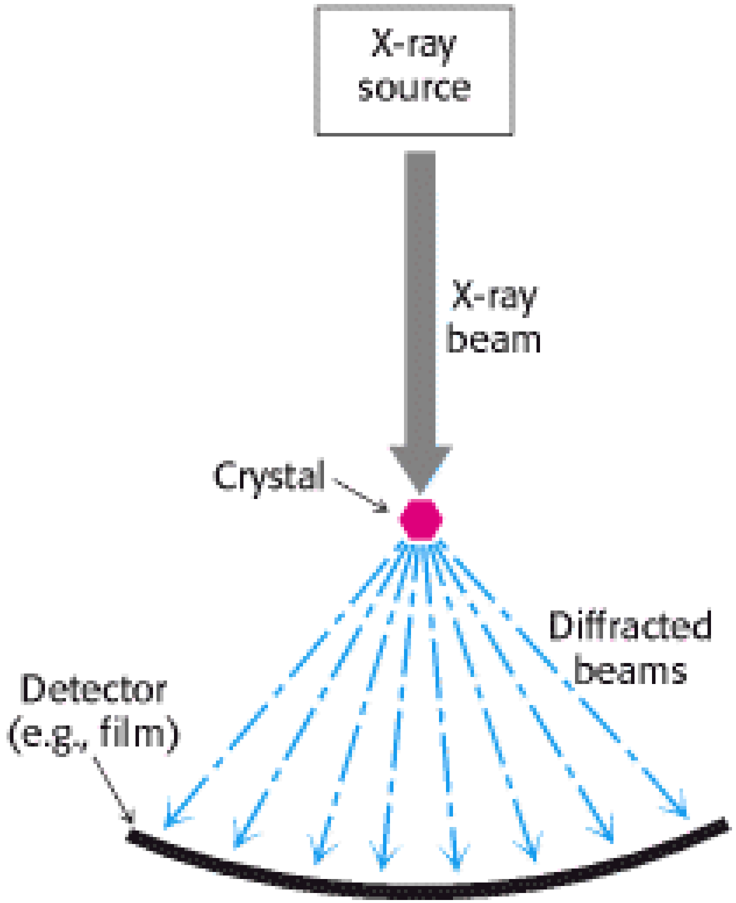
2D NMR

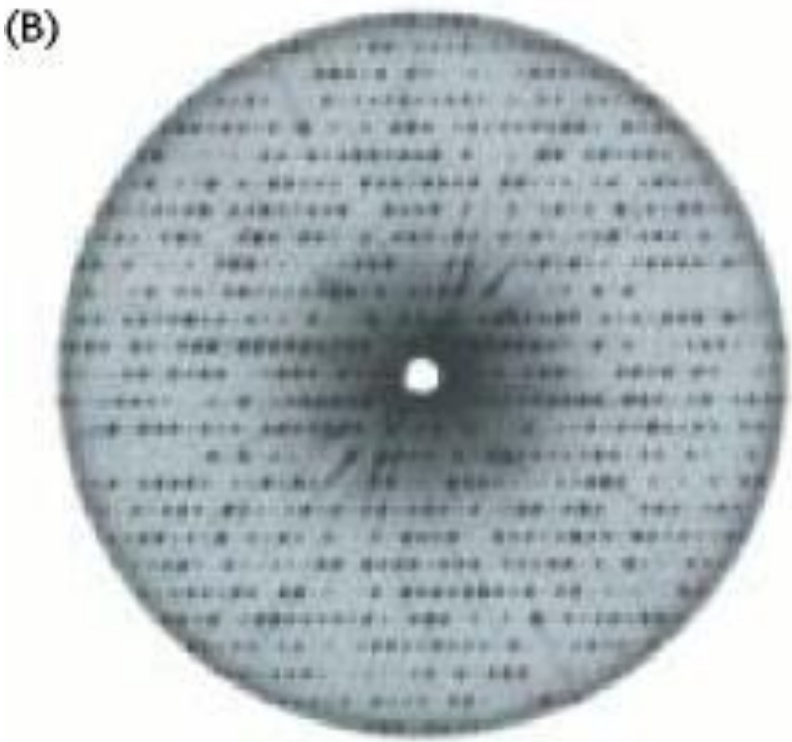

(A)
H
3
4
H
1
H
2
H
H
5
5 Å
(B)
Proton chemical shift (ppm)
4
5,2
2
3
5
2,5
1
Proton chemical shift (ppm)



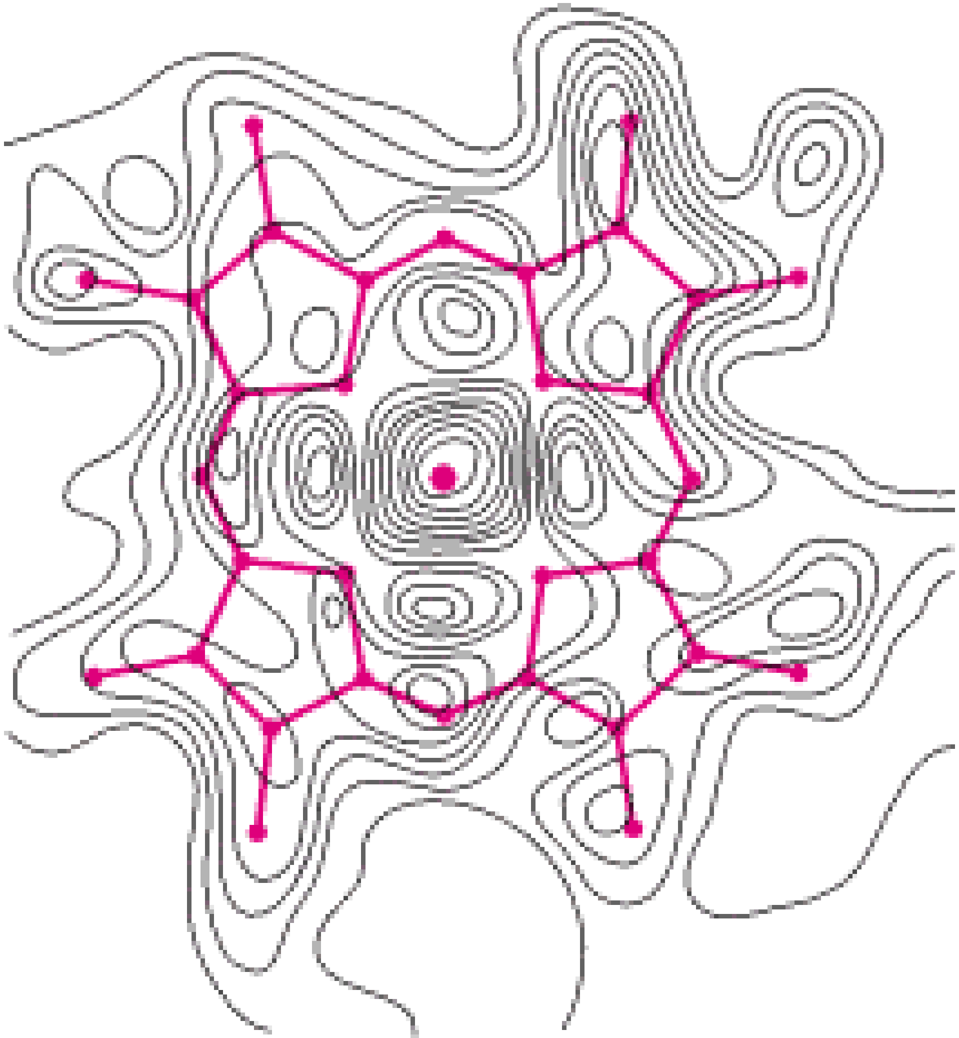
Proteins

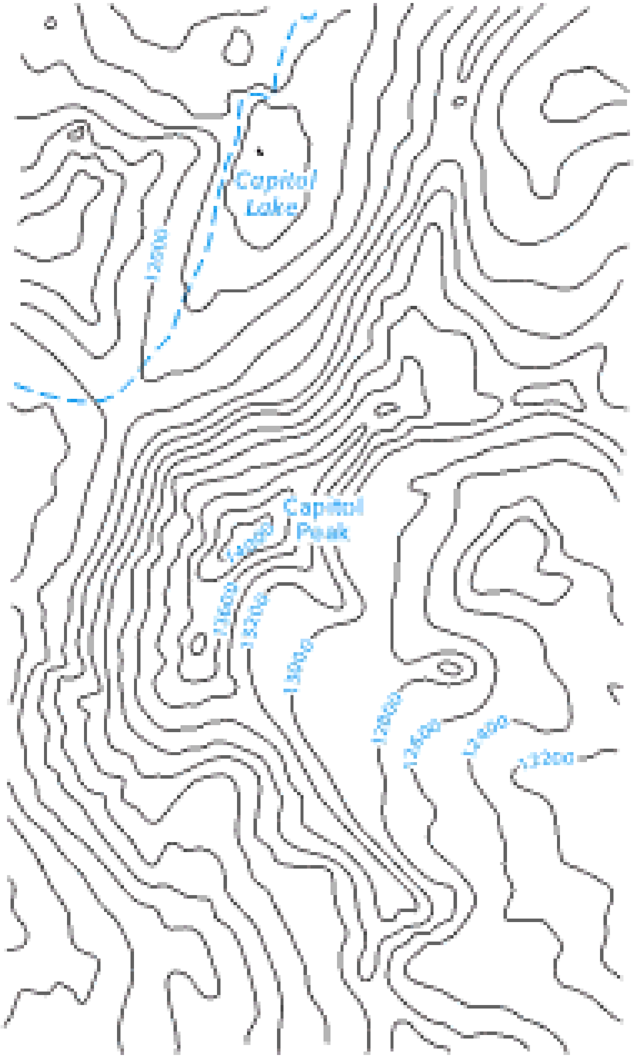
X-ray



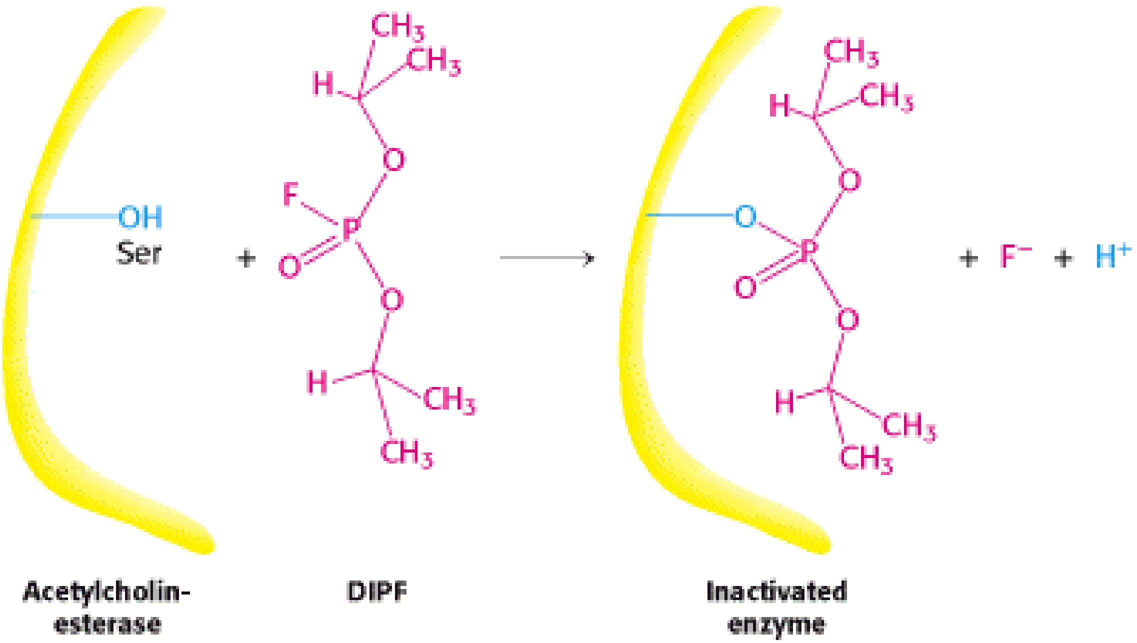


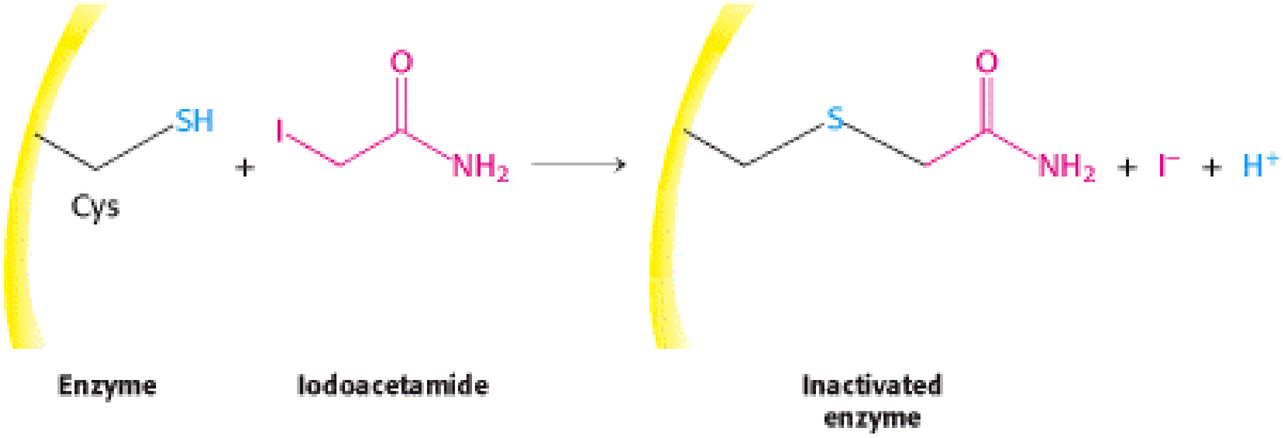
Electron density





Active site mapping





Computational methods

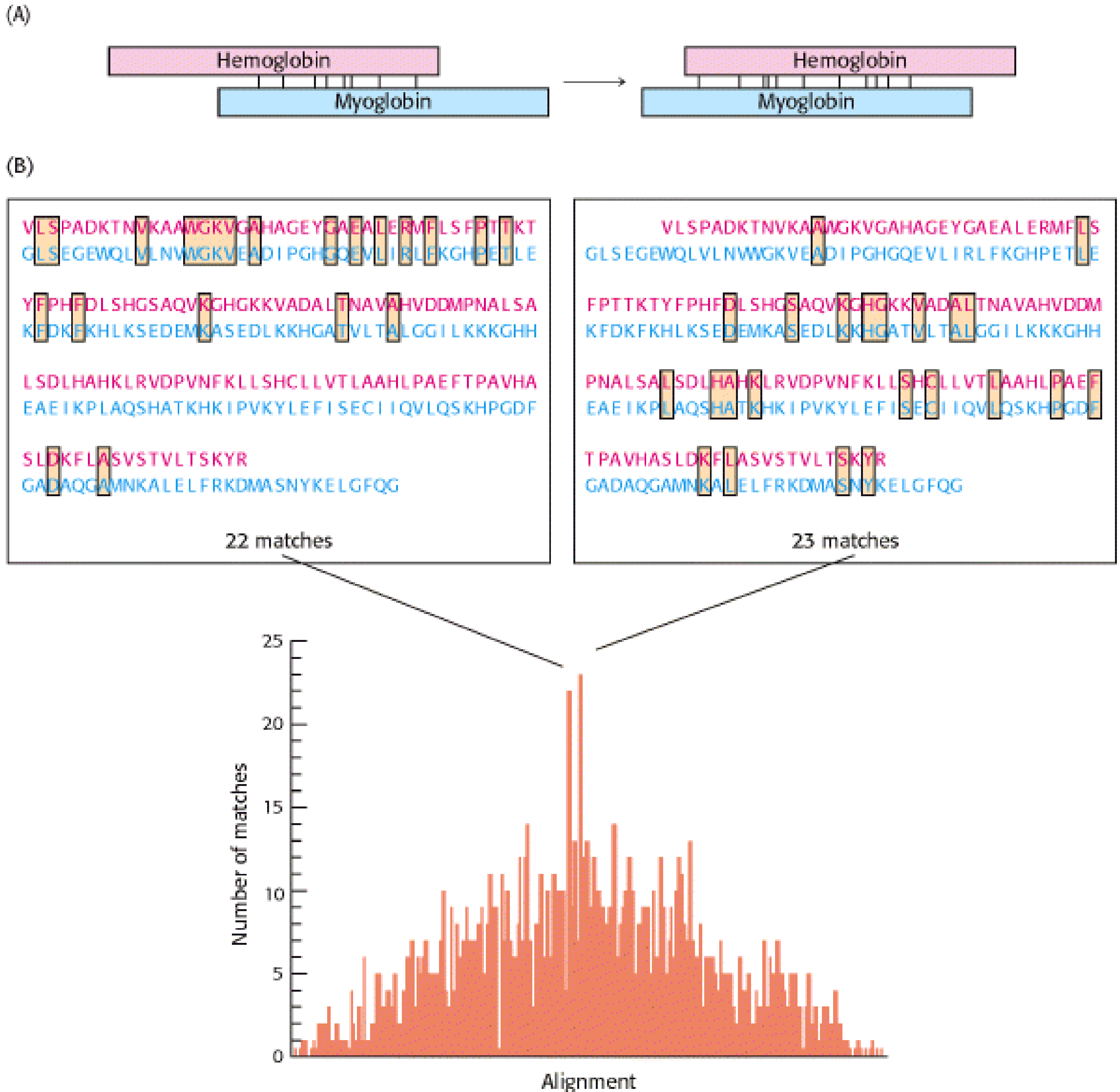
- Sequence comparison

Human hemoglobin (α chain)

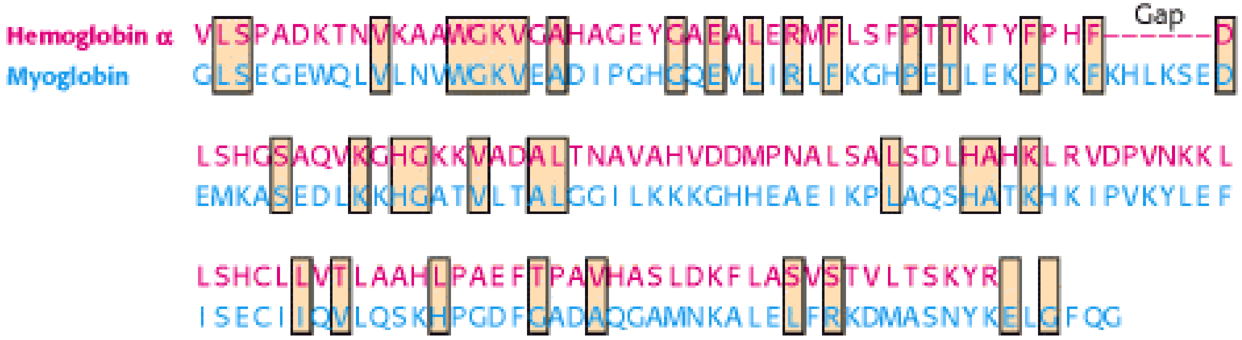
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHG
SAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLS
HCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR

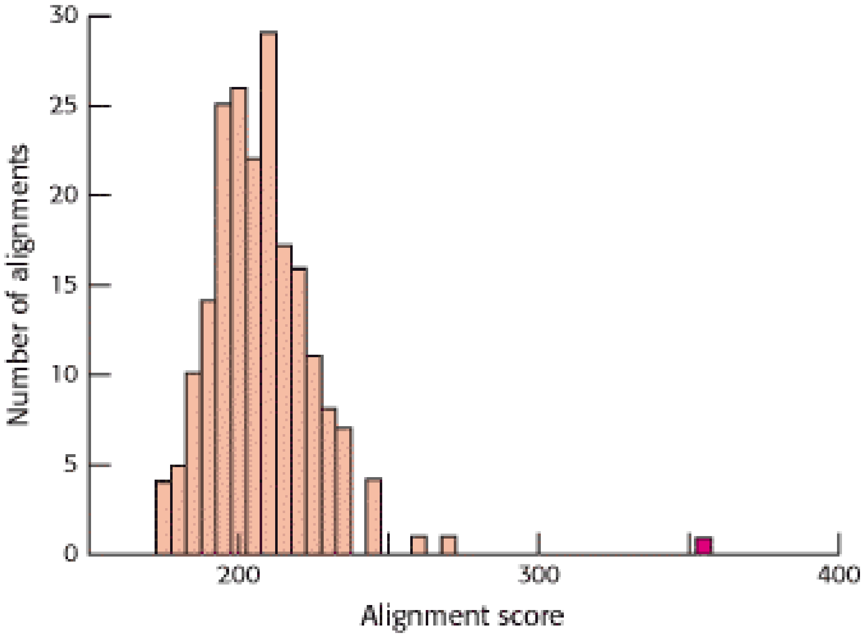
Human myoglobin

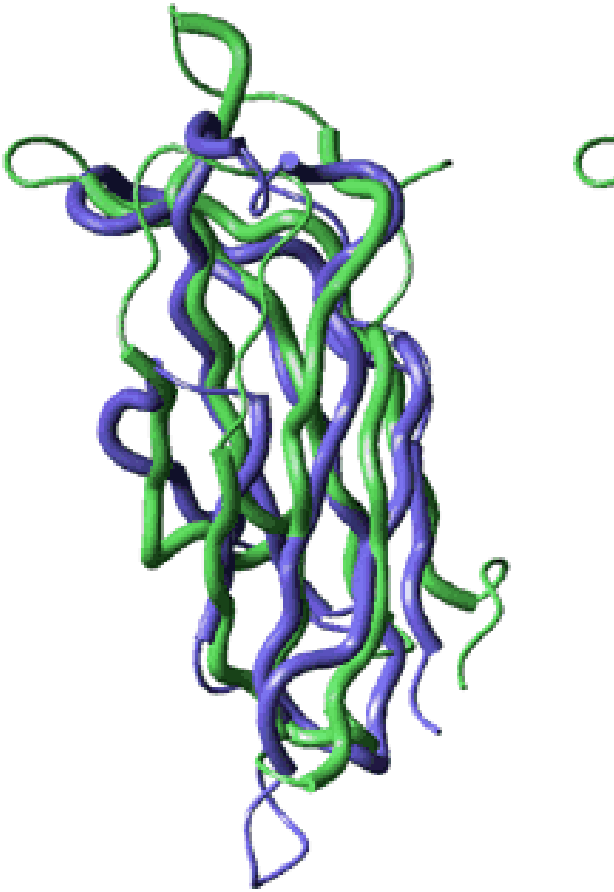
GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKS
EDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVK
YLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG


(A)
Hemoglobin
Myoglobin
Hemoglobin
Myoglobin
(B)
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKT
GLSEGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLE
YFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSA
KFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILKKKGHH
LSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHA
EAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDF
SLDKFLASVSTVLTSKYR
GADAQGAMNKALELFRKDMASNYKELGFQG
22 matches
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLS
GLSEGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLE
FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDM
KFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILKKKGHH
PNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEF
EAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDF
TPAVHASLDKFLASVSTVLTSKYR
GADAQGAMNKALELFRKDMASNYKELGFQG
23 matches
25
20
15
10
5
0
Number of matches
Alignment

Optimizing alignments

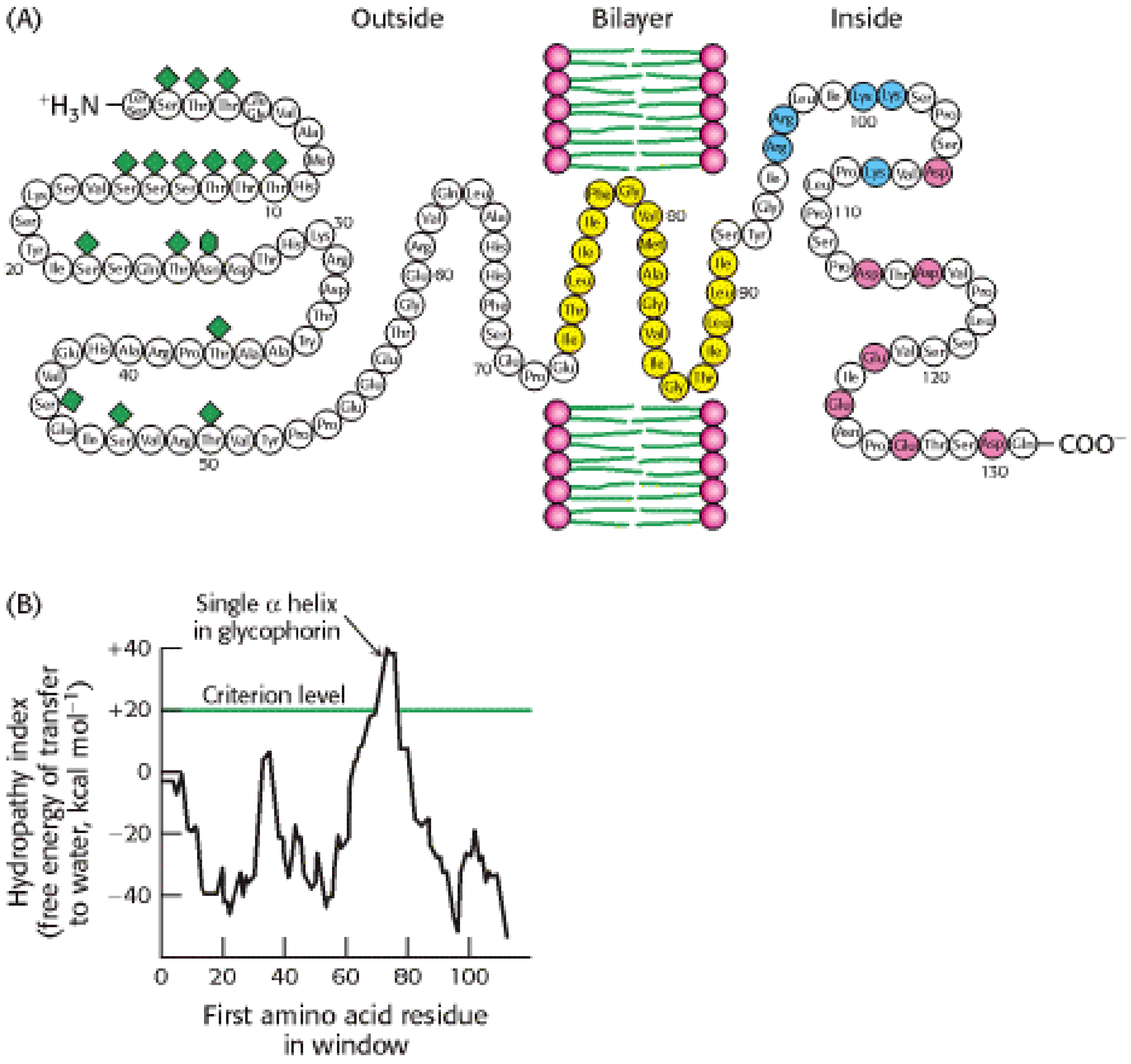






http://www.ebi.ac.uk/services

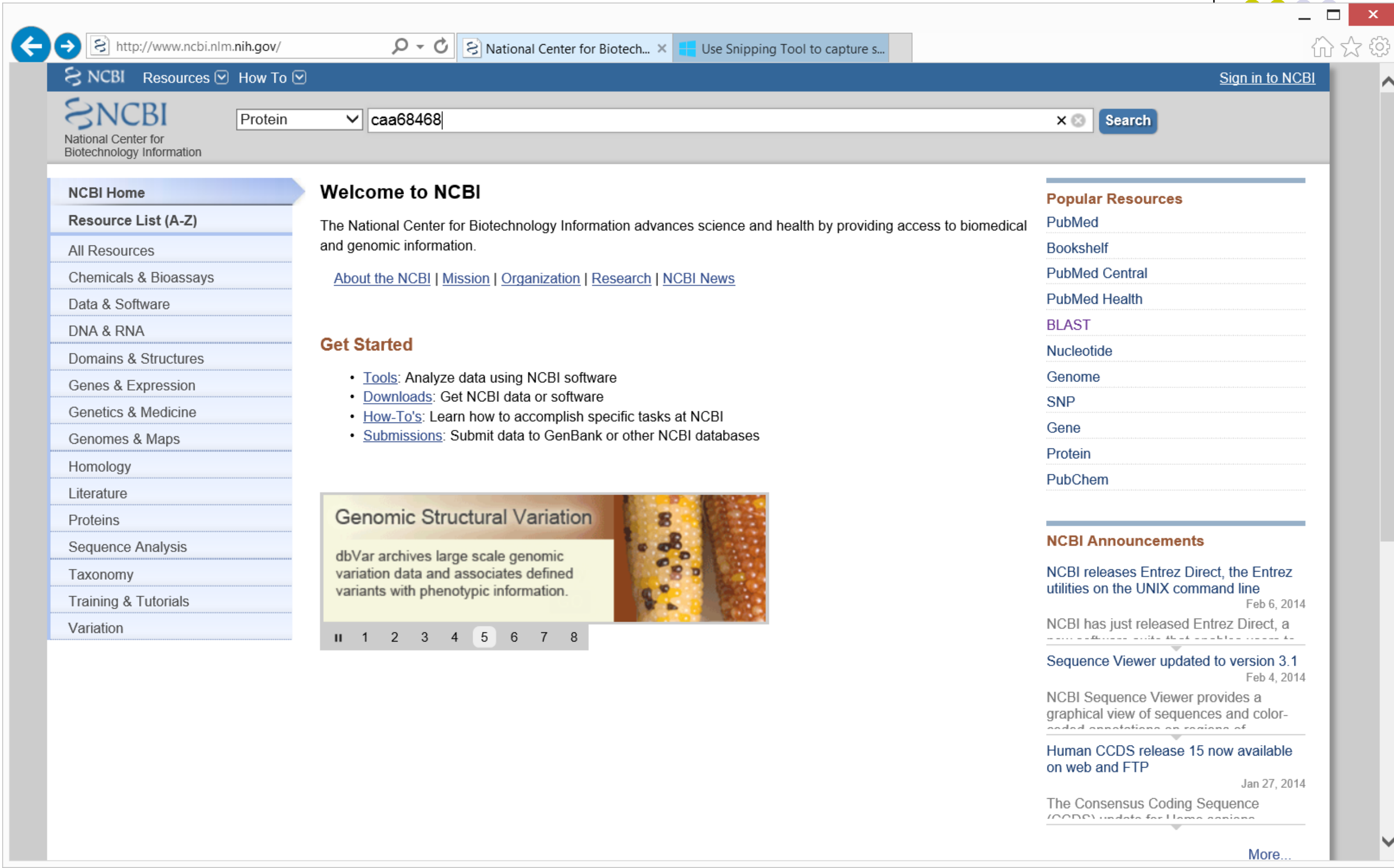
Hydropathy analysis



Tools

- http://www.ebi.ac.uk/services/
- http://us.expasy.org/
- http://blast.ncbi.nlm.nih.gov/Blast.cgi

Getting a sequence



Protein: caa68468

NCBI Home

Resource List (A-Z)

- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | NCBI News

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Genomic Structural Variation

dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

NCBI releases Entrez Direct, the Entrez utilities on the UNIX command line
Feb 6, 2014

NCBI has just released Entrez Direct, a

Sequence Viewer updated to version 3.1
Feb 4, 2014

NCBI Sequence Viewer provides a graphical view of sequences and color-

Human CCDS release 15 now available on web and FTP
Jan 27, 2014

The Consensus Coding Sequence

More...

conalbumin [Gallus gallus]

GenBank: CAA68468.1

```
LOCUS       CAA68468                 705 aa            linear   VRT 23-OCT-2008
DEFINITION  conalbumin [Gallus gallus].
ACCESSION   CAA68468
VERSION     CAA68468.1  GI:295721
DBSOURCE    embl accession Y00407.1
KEYWORDS    .
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Testudines + Archosauria group; Archosauria; Dinosauria;
            Saurischia; Theropoda; Coelurosauria; Aves; Neognathae;
            Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (residues 1 to 705)
  AUTHORS   Jeltsch,J.M., Hen,R., Maroteaux,L., Garnier,J.M. and Chambon,P.
  TITLE     Sequence of the chicken ovotransferrin gene
  JOURNAL   Nucleic Acids Res. 15 (18), 7643-7645 (1987)
   PUBMED   3658709
REFERENCE   2  (residues 1 to 705)
  AUTHORS   Jeltsch,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-1987) Jeltsch J.-M., Inserm U.184 and L.G.M.E. du
            C.N.R.S., Faculte de Medecine, 11, rue Humann, F-67085 Strasbourg
COMMENT     On Jun 16, 1993 this sequence version replaced gi:63132.
            Data kindly reviewed (30-OCT-1987) by Jeltsch J.-M.
FEATURES             Location/Qualifiers
     source          1..705
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
```

Analyze this sequence

Protein 3D Structure

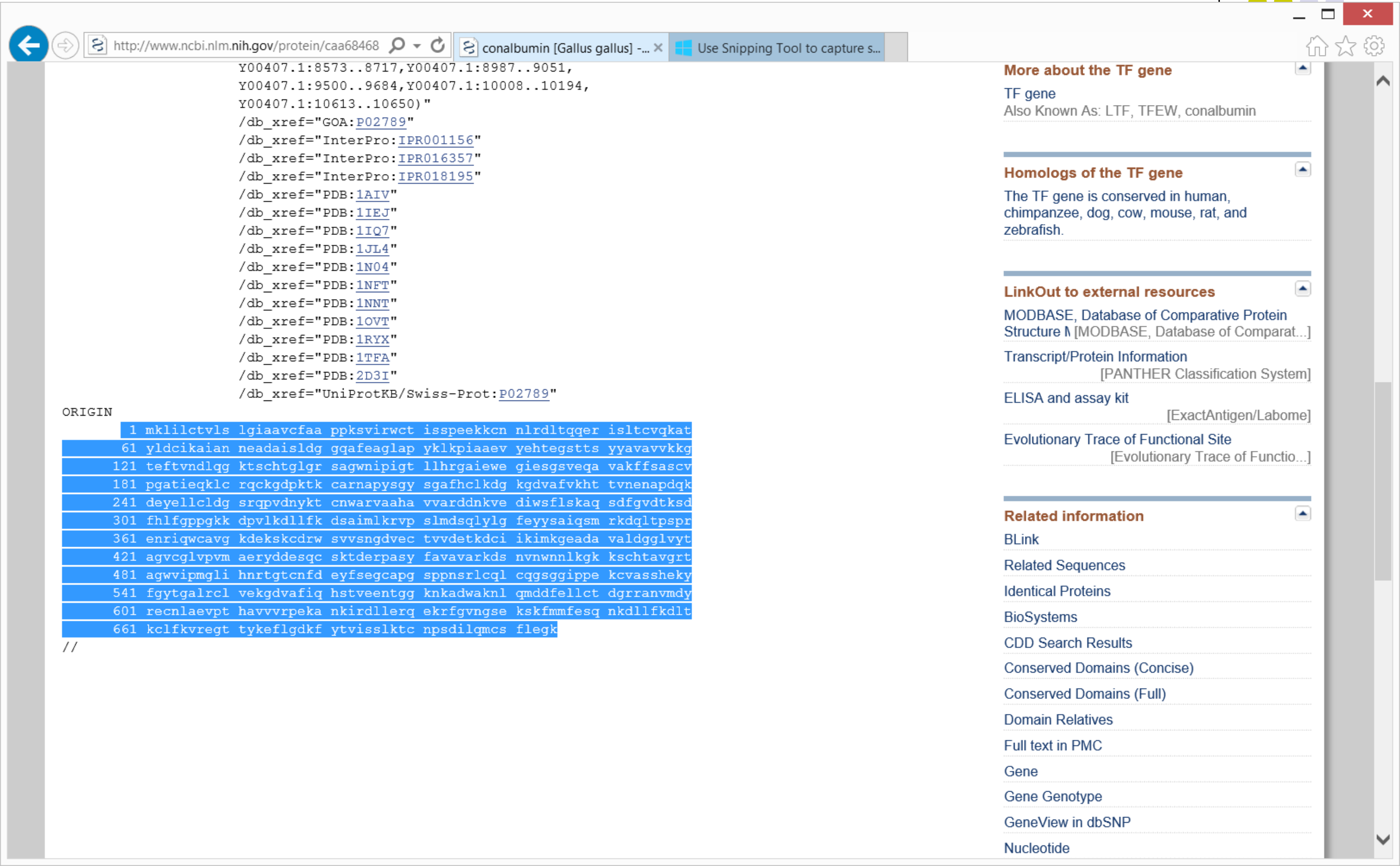
Refined Crystallographic Structure Of Hen Ovotransferrin At 2.4
PDB: 1OVT
Source: Gallus gallus
Method: X-Ray Diffraction
Resolution: 2.4 Å

Articles about the TF gene

pH-dependent conformational transitions in conalbumin (ovotrai [Cell Biochem Biophys. 2011]

Unexpected differences in the behavior of ovotransferrin at the [J Colloid Interface Sci. 2011]

Identification of novel antioxidative peptides derived from a thermo [J Agric Food Chem. 2010]



```
                     Y00407.1:8573..8717,Y00407.1:8987..9051,
                     Y00407.1:9500..9684,Y00407.1:10008..10194,
                     Y00407.1:10613..10650)"
                     /db_xref="GOA:P02789"
                     /db_xref="InterPro:IPR001156"
                     /db_xref="InterPro:IPR016357"
                     /db_xref="InterPro:IPR018195"
                     /db_xref="PDB:1AIV"
                     /db_xref="PDB:1IEJ"
                     /db_xref="PDB:1IQ7"
                     /db_xref="PDB:1JL4"
                     /db_xref="PDB:1N04"
                     /db_xref="PDB:1NFT"
                     /db_xref="PDB:1NNT"
                     /db_xref="PDB:1OVT"
                     /db_xref="PDB:1RYX"
                     /db_xref="PDB:1TFA"
                     /db_xref="PDB:2D3I"
                     /db_xref="UniProtKB/Swiss-Prot:P02789"
ORIGIN
        1 mklilctvls lgiaavcfaa ppksvirwct isspeekkcn nlrdltqqer isltcvqkat
       61 yldcikaian neadaisldg gqafeaglap yklkpiaaev yehtegstts yyavavvkkg
      121 teftvndlqg ktschtglgr sagwnipigt llhrgaiewe giesgsveqa vakffsascv
      181 pgatieqklc rqckgdpktk carnapysgy sgafhclkdg kgdvafvkht tvnenapdqk
      241 deyellcldg srqpvdnykt cnwarvaaha vvarddnkve diwsflskaq sdfgvdtksd
      301 fhlfgppgkk dpvlkdllfk dsaimlkrvp slmdsqlylg feyysaiqsm rkdqltpspr
      361 enriqwcavg kdekskcdrw svvsngdvec tvvdetkdci ikimkgeada valdgglvyt
      421 agvcglvpvm aeryddesqc sktderpasy favavarkds nvnwnnlkgk kschtavgrt
      481 agwvipmgli hnrtgtcnfd eyfsegcapg sppnsrlcql cqgsggippe kcvassheky
      541 fgytgalrcl vekgdvafiq hstveentgg knkadwaknl qmddfellct dgrranvmdy
      601 recnlaevpt havvvrpeka nkirdllerq ekrfgvngse kskfmmfesq nkdllfkdlt
      661 kclfkvregt tykeflgdkf ytvisslktc npsdilqmcs flegk
//
```

More about the TF gene

TF gene
Also Known As: LTF, TFEW, conalbumin

Homologs of the TF gene

The TF gene is conserved in human, chimpanzee, dog, cow, mouse, rat, and zebrafish.

LinkOut to external resources

MODBASE, Database of Comparative Protein Structure M [MODBASE, Database of Comparat...]

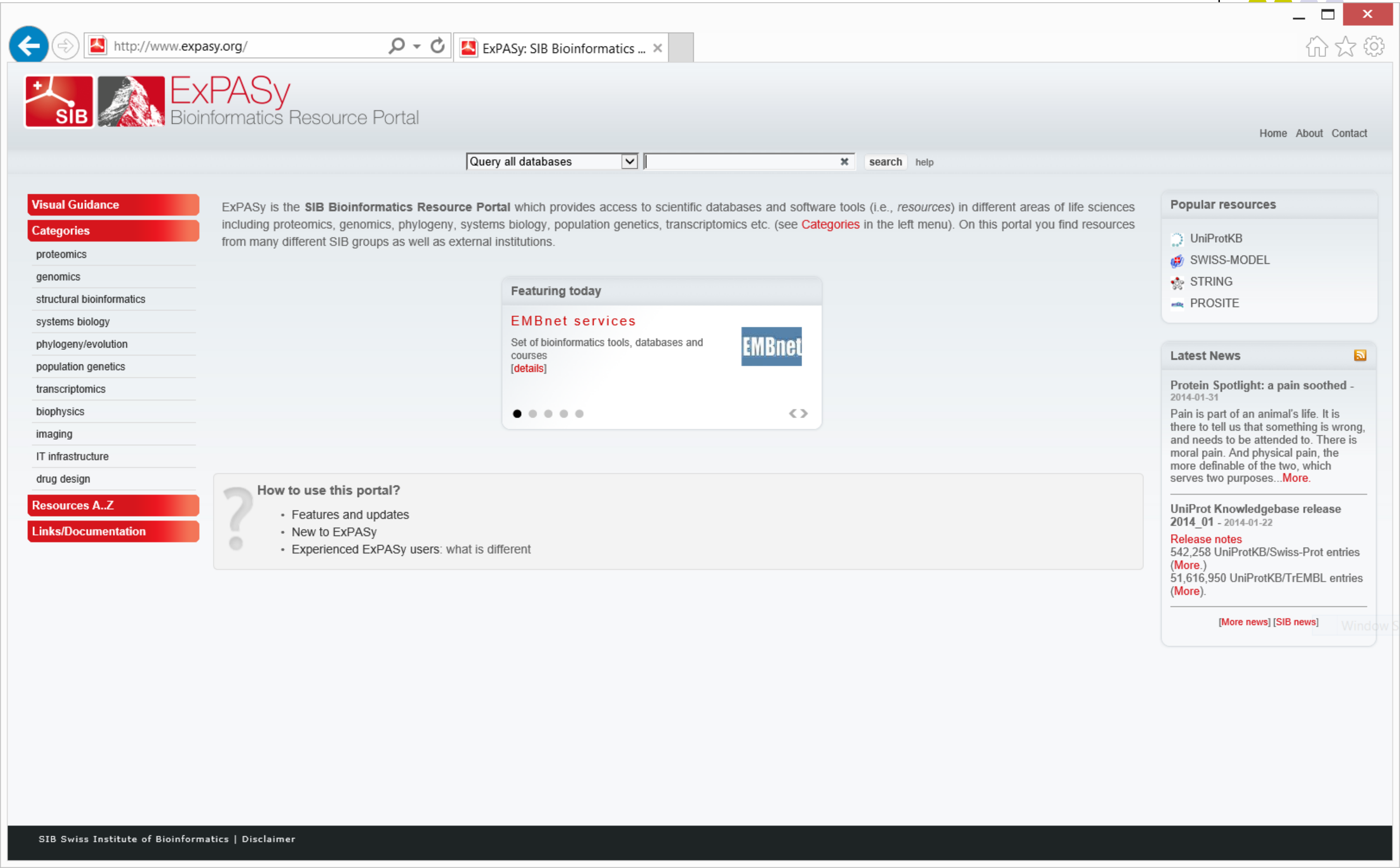
Transcript/Protein Information [PANTHER Classification System]

ELISA and assay kit [ExactAntigen/Labome]

Evolutionary Trace of Functional Site [Evolutionary Trace of Functio...]

Related information

BLink
Related Sequences
Identical Proteins
BioSystems
CDD Search Results
Conserved Domains (Concise)
Conserved Domains (Full)
Domain Relatives
Full text in PMC
Gene
Gene Genotype
GeneView in dbSNP
Nucleotide



ExPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resources*) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see Categories in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.

Featuring today

EMBnet services

Set of bioinformatics tools, databases and courses
[details]

How to use this portal?

- Features and updates
- New to ExPASy
- Experienced ExPASy users: what is different

Latest News

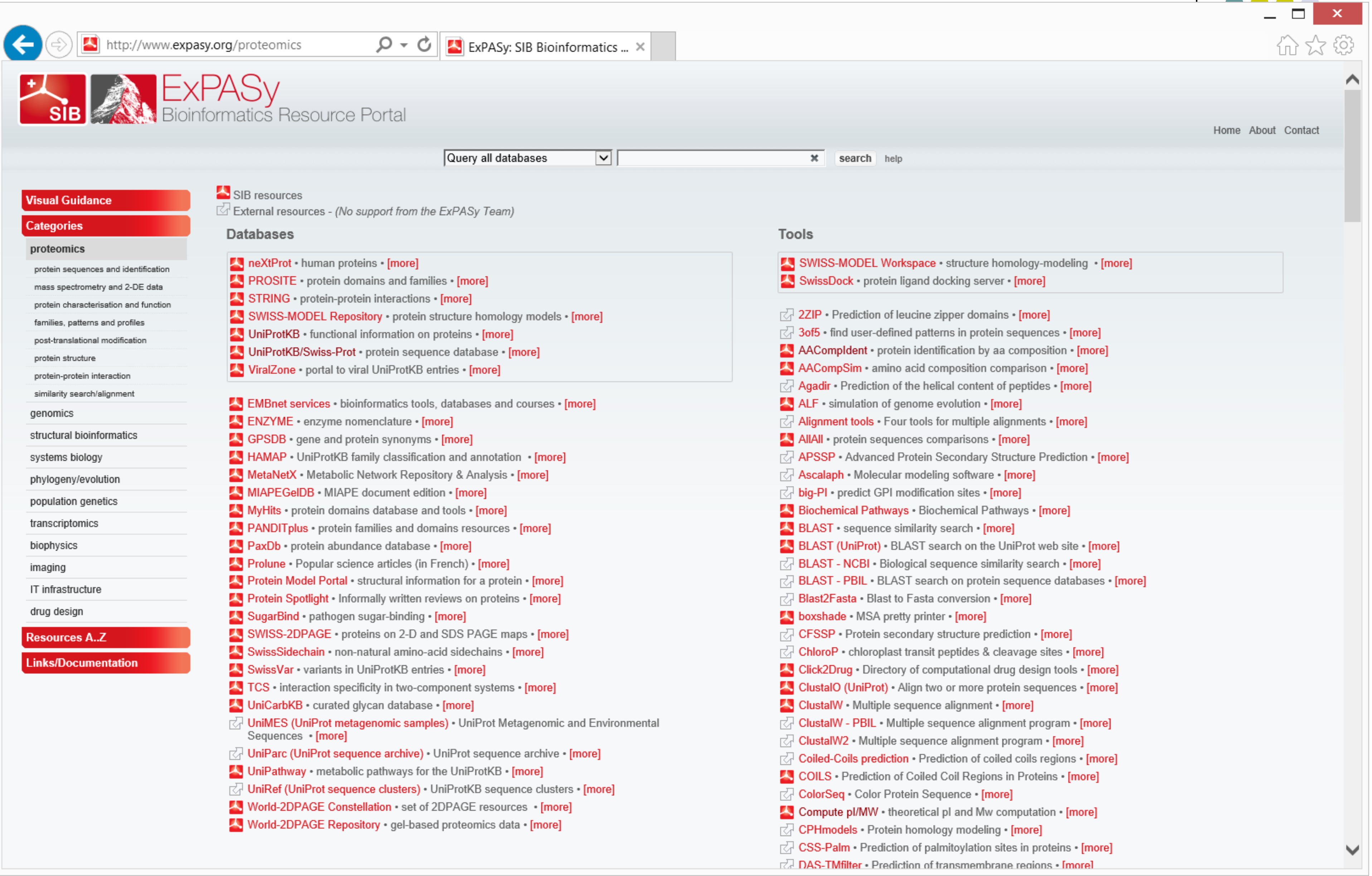
Protein Spotlight: a pain soothed - 2014-01-31

Pain is part of an animal's life. It is there to tell us that something is wrong, and needs to be attended to. There is moral pain. And physical pain, the more definable of the two, which serves two purposes...More.

UniProt Knowledgebase release 2014_01 - 2014-01-22

Release notes
542,258 UniProtKB/Swiss-Prot entries (More.)
51,616,950 UniProtKB/TrEMBL entries (More).

[More news] [SIB news]



http://www.expasy.org/proteomics

ExPASy: SIB Bioinformatics ...

SIB

ExPASy
Bioinformatics Resource Portal

Home About Contact

Query all databases | search | help

Visual Guidance

Categories

- **proteomics**
  - protein sequences and identification
  - mass spectrometry and 2-DE data
  - protein characterisation and function
  - families, patterns and profiles
  - post-translational modification
  - protein structure
  - protein-protein interaction
  - similarity search/alignment
- genomics
- structural bioinformatics
- systems biology
- phylogeny/evolution
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- drug design

Resources A..Z

Links/Documentation

SIB resources
External resources - *(No support from the ExPASy Team)*

Databases

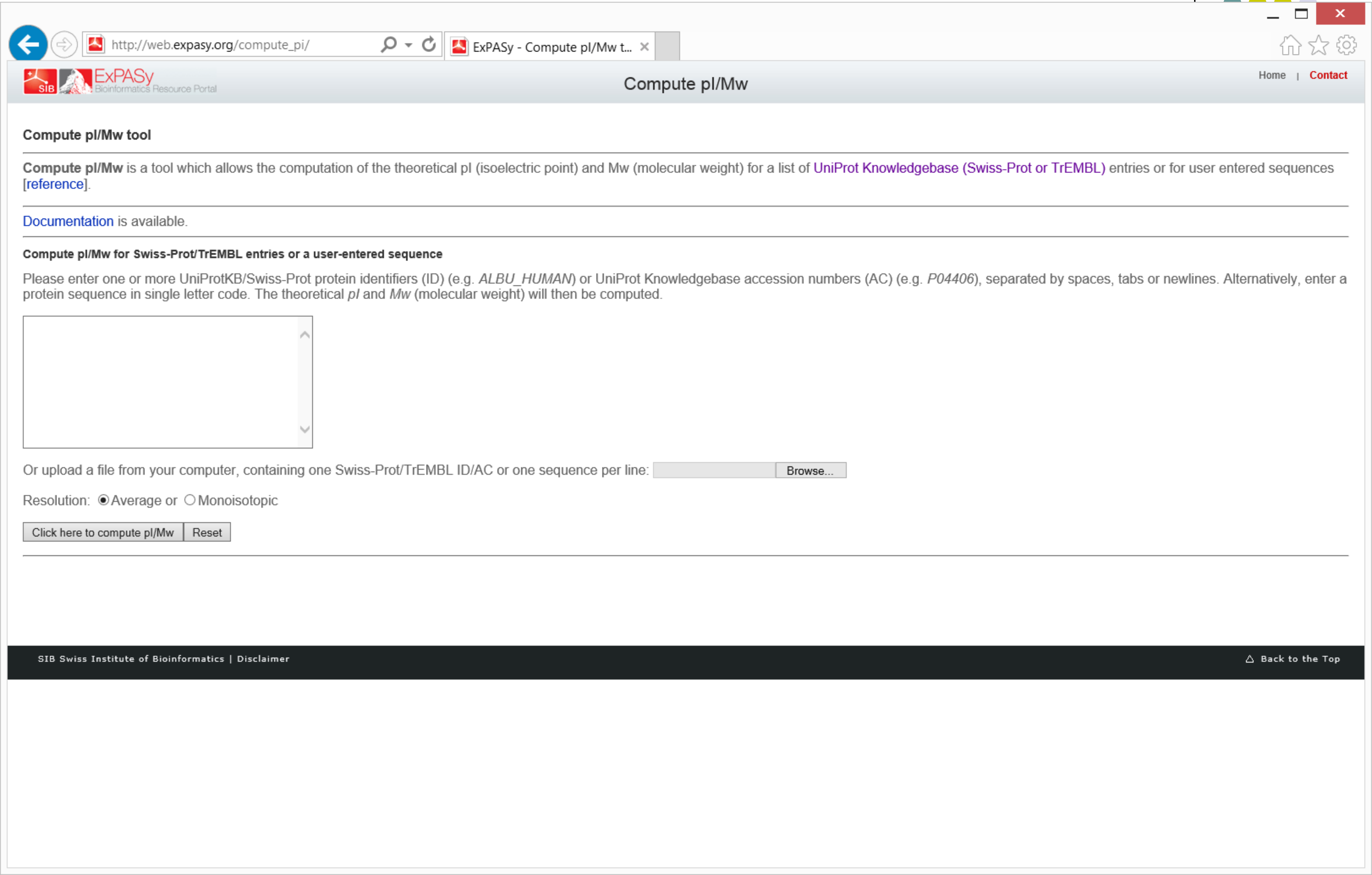
- neXtProt • human proteins • [more]
- PROSITE • protein domains and families • [more]
- STRING • protein-protein interactions • [more]
- SWISS-MODEL Repository • protein structure homology models • [more]
- UniProtKB • functional information on proteins • [more]
- UniProtKB/Swiss-Prot • protein sequence database • [more]
- ViralZone • portal to viral UniProtKB entries • [more]

- EMBnet services • bioinformatics tools, databases and courses • [more]
- ENZYME • enzyme nomenclature • [more]
- GPSDB • gene and protein synonyms • [more]
- HAMAP • UniProtKB family classification and annotation • [more]
- MetaNetX • Metabolic Network Repository & Analysis • [more]
- MIAPEGelDB • MIAPE document edition • [more]
- MyHits • protein domains database and tools • [more]
- PANDITplus • protein families and domains resources • [more]
- PaxDb • protein abundance database • [more]
- Prolune • Popular science articles (in French) • [more]
- Protein Model Portal • structural information for a protein • [more]
- Protein Spotlight • Informally written reviews on proteins • [more]
- SugarBind • pathogen sugar-binding • [more]
- SWISS-2DPAGE • proteins on 2-D and SDS PAGE maps • [more]
- SwissSidechain • non-natural amino-acid sidechains • [more]
- SwissVar • variants in UniProtKB entries • [more]
- TCS • interaction specificity in two-component systems • [more]
- UniCarbKB • curated glycan database • [more]
- UniMES (UniProt metagenomic samples) • UniProt Metagenomic and Environmental Sequences • [more]
- UniParc (UniProt sequence archive) • UniProt sequence archive • [more]
- UniPathway • metabolic pathways for the UniProtKB • [more]
- UniRef (UniProt sequence clusters) • UniProtKB sequence clusters • [more]
- World-2DPAGE Constellation • set of 2DPAGE resources • [more]
- World-2DPAGE Repository • gel-based proteomics data • [more]

Tools

- SWISS-MODEL Workspace • structure homology-modeling • [more]
- SwissDock • protein ligand docking server • [more]

- 2ZIP • Prediction of leucine zipper domains • [more]
- 3of5 • find user-defined patterns in protein sequences • [more]
- AACompIdent • protein identification by aa composition • [more]
- AACompSim • amino acid composition comparison • [more]
- Agadir • Prediction of the helical content of peptides • [more]
- ALF • simulation of genome evolution • [more]
- Alignment tools • Four tools for multiple alignments • [more]
- AllAll • protein sequences comparisons • [more]
- APSSP • Advanced Protein Secondary Structure Prediction • [more]
- Ascalaph • Molecular modeling software • [more]
- big-PI • predict GPI modification sites • [more]
- Biochemical Pathways • Biochemical Pathways • [more]
- BLAST • sequence similarity search • [more]
- BLAST (UniProt) • BLAST search on the UniProt web site • [more]
- BLAST - NCBI • Biological sequence similarity search • [more]
- BLAST - PBIL • BLAST search on protein sequence databases • [more]
- Blast2Fasta • Blast to Fasta conversion • [more]
- boxshade • MSA pretty printer • [more]
- CFSSP • Protein secondary structure prediction • [more]
- ChloroP • chloroplast transit peptides & cleavage sites • [more]
- Click2Drug • Directory of computational drug design tools • [more]
- ClustalO (UniProt) • Align two or more protein sequences • [more]
- ClustalW • Multiple sequence alignment • [more]
- ClustalW - PBIL • Multiple sequence alignment program • [more]
- ClustalW2 • Multiple sequence alignment program • [more]
- Coiled-Coils prediction • Prediction of coiled coils regions • [more]
- COILS • Prediction of Coiled Coil Regions in Proteins • [more]
- ColorSeq • Color Protein Sequence • [more]
- Compute pI/MW • theoretical pI and Mw computation • [more]
- CPHmodels • Protein homology modeling • [more]
- CSS-Palm • Prediction of palmitoylation sites in proteins • [more]
- DAS-TMfilter • Prediction of transmembrane regions • [more]



Compute pI/Mw

Compute pI/Mw tool

Compute pI/Mw is a tool which allows the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for a list of UniProt Knowledgebase (Swiss-Prot or TrEMBL) entries or for user entered sequences [reference].

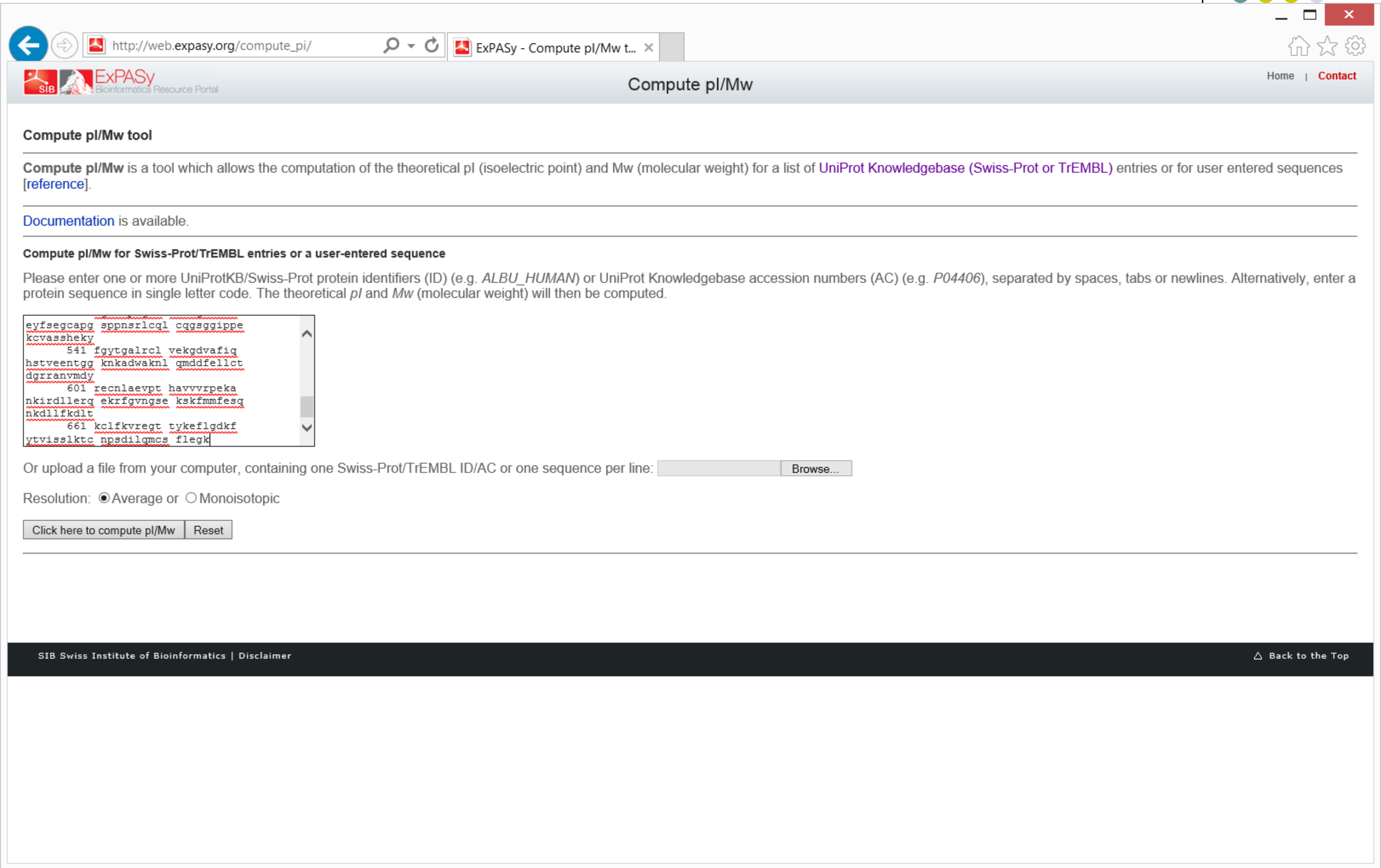
Documentation is available.

Compute pI/Mw for Swiss-Prot/TrEMBL entries or a user-entered sequence

Please enter one or more UniProtKB/Swiss-Prot protein identifiers (ID) (e.g. *ALBU_HUMAN*) or UniProt Knowledgebase accession numbers (AC) (e.g. *P04406*), separated by spaces, tabs or newlines. Alternatively, enter a protein sequence in single letter code. The theoretical *pI* and *Mw* (molecular weight) will then be computed.

Or upload a file from your computer, containing one Swiss-Prot/TrEMBL ID/AC or one sequence per line:

Resolution: ● Average or ○ Monoisotopic



Compute pI/Mw

Compute pI/Mw tool

Compute pI/Mw is a tool which allows the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for a list of UniProt Knowledgebase (Swiss-Prot or TrEMBL) entries or for user entered sequences [reference].

Documentation is available.

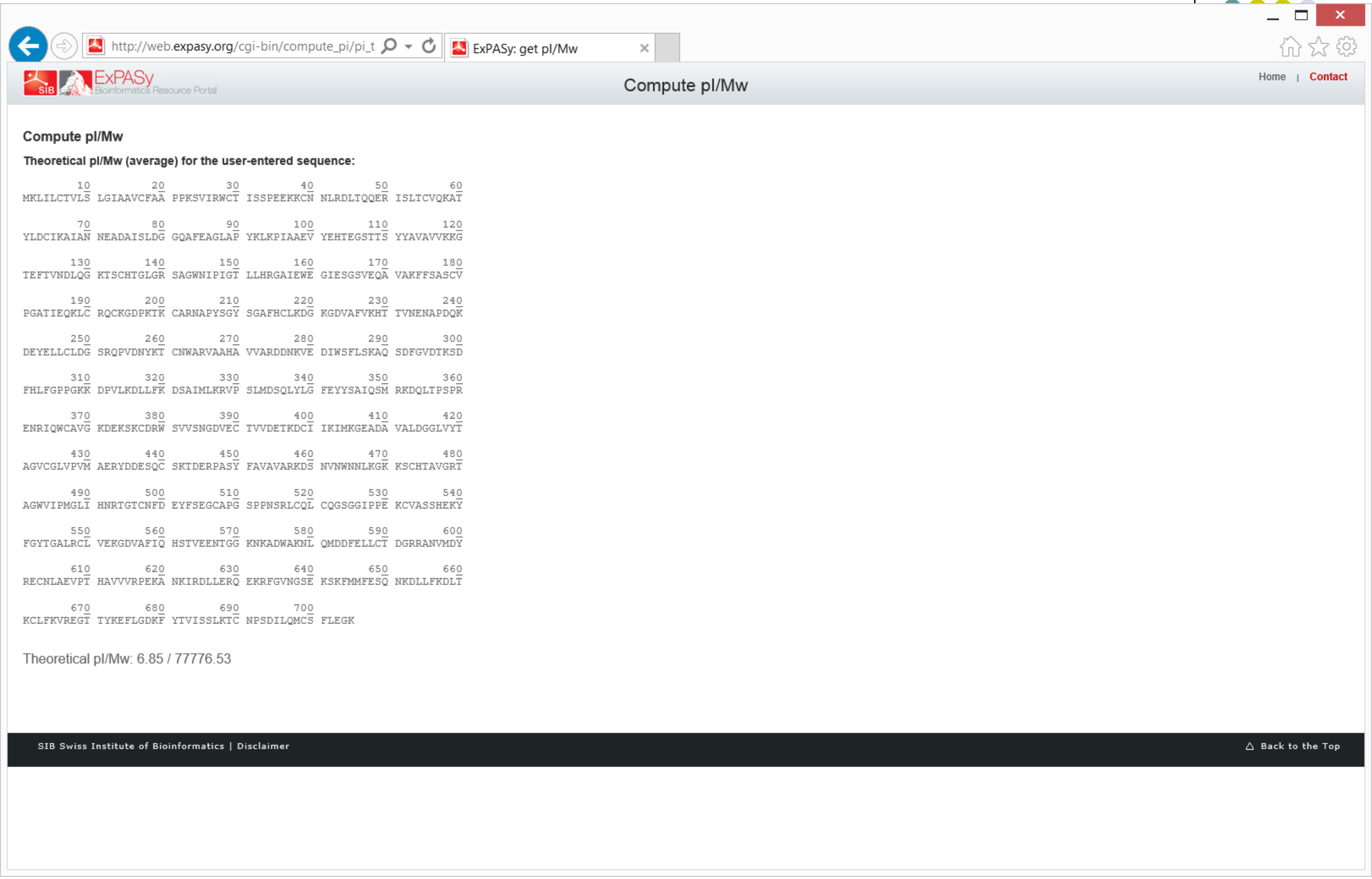
Compute pI/Mw for Swiss-Prot/TrEMBL entries or a user-entered sequence

Please enter one or more UniProtKB/Swiss-Prot protein identifiers (ID) (e.g. *ALBU_HUMAN*) or UniProt Knowledgebase accession numbers (AC) (e.g. *P04406*), separated by spaces, tabs or newlines. Alternatively, enter a protein sequence in single letter code. The theoretical *pI* and *Mw* (molecular weight) will then be computed.

```
eyfsegcapg sppnsrlcql cqgsggippe
kcvassheky
       541 fgytgalrcl vekgdvafiq
hstveentgg knkadwaknl qmddfellct
dgrranvmdy
       601 recnlaevpt havvvrpeka
nkirdllerq ekrfgvngse kskfmmfesq
nkdllfkdlt
       661 kclfkvregt tykeflgdkf
ytvisslktc npsdilqmcs flegk
```

Or upload a file from your computer, containing one Swiss-Prot/TrEMBL ID/AC or one sequence per line: Browse...

Resolution: ● Average or ○ Monoisotopic



Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MKLILCTVLS LGIAAVCFAA PPKSVIRWCT ISSPEEKKCN NLRDLTQQER ISLTCVQKAT

        70         80         90        100        110        120
YLDCIKAIAN NEADAISLDG GQAFEAGLAP YKLKPIAAEV YEHTEGSTTS YYAVAVVKKG

       130        140        150        160        170        180
TEFTVNDLQG KTSCHTGLGR SAGWNIPIGT LLHRGAIEWE GIESGSVEQA VAKFFSASCV

       190        200        210        220        230        240
PGATIEQKLC RQCKGDPKTK CARNAPYSGY SGAFHCLKDG KGDVAFVKHT TVNENAPDQK

       250        260        270        280        290        300
DEYELLCLDG SRQPVDNYKT CNWARVAAHA VVARDDNKVE DIWSFLSKAQ SDFGVDTKSD

       310        320        330        340        350        360
FHLFGPPGKK DPVLKDLLFK DSAIMLKRVP SLMDSQLYLG FEYYSAIQSM RKDQLTPSPR

       370        380        390        400        410        420
ENRIQWCAVG KDEKSKCDRW SVVSNGDVEC TVVDETKDCI IKIMKGEADA VALDGGLVYT

       430        440        450        460        470        480
AGVCGLVPVM AERYDDESQC SKTDERPASY FAVAVARKDS NVNWNNLKGK KSCHTAVGRT

       490        500        510        520        530        540
AGWVIPMGLI HNRTGTCNFD EYFSEGCAPG SPPNSRLCQL CQGSGGIPPE KCVASSHEKY

       550        560        570        580        590        600
FGYTGALRCL VEKGDVAFIQ HSTVEENTGG KNKADWAKNL QMDDFELLCT DGRRANVMDY

       610        620        630        640        650        660
RECNLAEVPT HAVVVRPEKA NKIRDLLERQ EKRFGVNGSE KSKFMMFESQ NKDLLFKDLT

       670        680        690        700
KCLFKVREGT TYKEFLGDKF YTVISSLKTC NPSDILQMCS FLEGK
```

Theoretical pI/Mw: 6.85 / 77776.53