Bioanalytical Chemistry Text

www.pharmed.uz/_books/4_bioanalyt_chem2004.pdf

Biopolymers are key biomolecules

Proteins
Nucleic acids
Lipids
Polysacharides

The aim of bioanalytical chemistry is to:

1. Identify and determine quantities of specific biopolymers (qualitative and quantitative analyses)
2. Find what other molecules they interact with (screening for affinity)
3. Determine quantitative parameters of interaction (kinetic methods)

Analysis of Biopolymers: Identification and quantitation

Properties of Biopolymers that can be Used for their Analysis

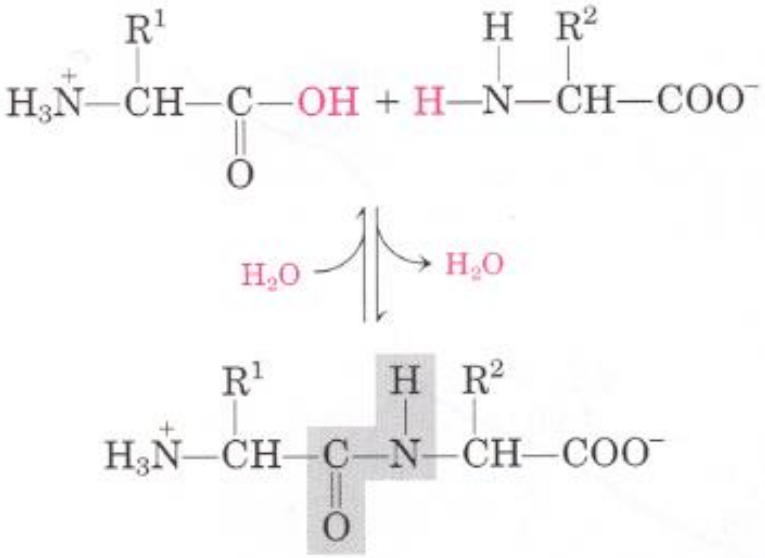
- Size
- Mass
- Shape
- Charge
- Isoelectric point
- Electron structure
- Affinity interactions
- Chemical reactivity

Structure of Proteins

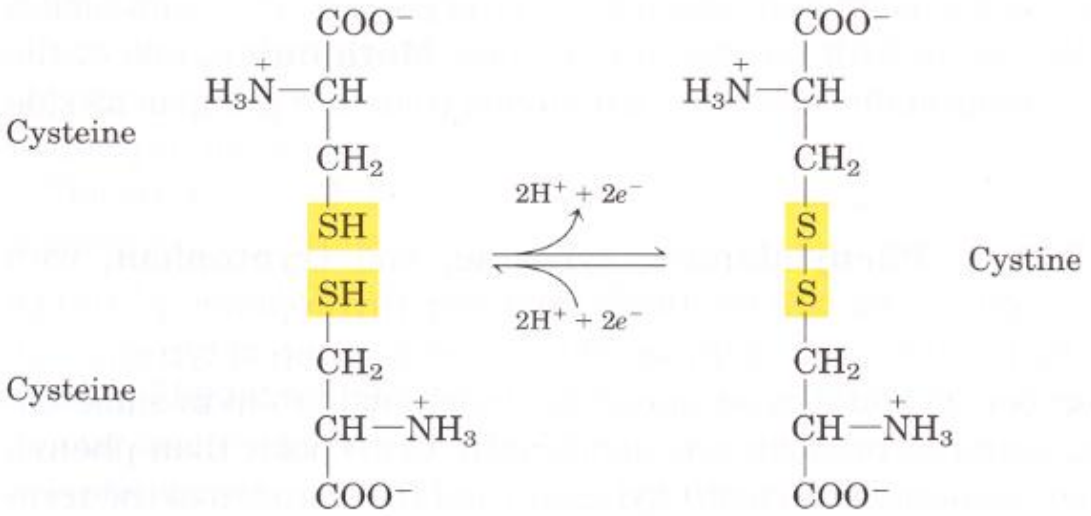
Structure units - 20 amino acids
Primary structure – amino acid sequence (-C-N- bonds)
Secondary structure - α-helix, β-sheet (H-bonds)
Tertiary structure – 3D structure (S-S bonds, H-bonds)
Quaternary structure – 3D structure of >1 subunits (weak interactions)

Covalent Bonds in Proteins

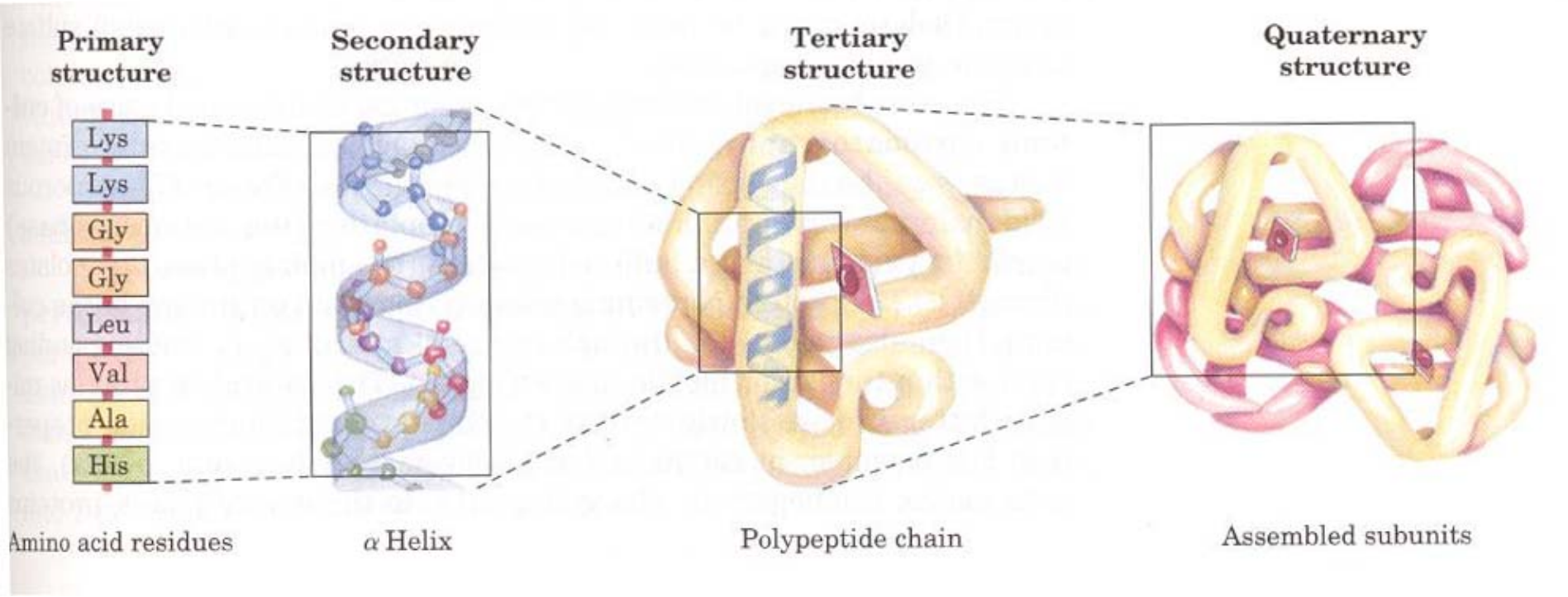
Peptide bond



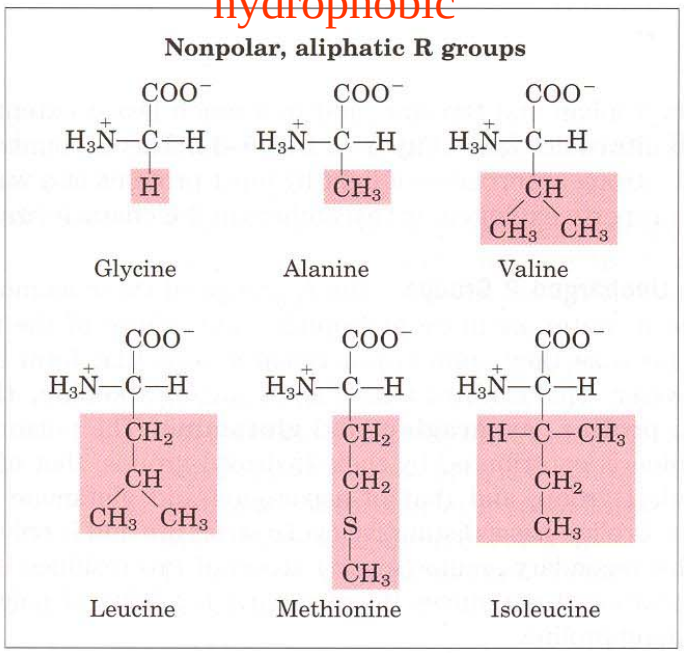
Disulfide bond

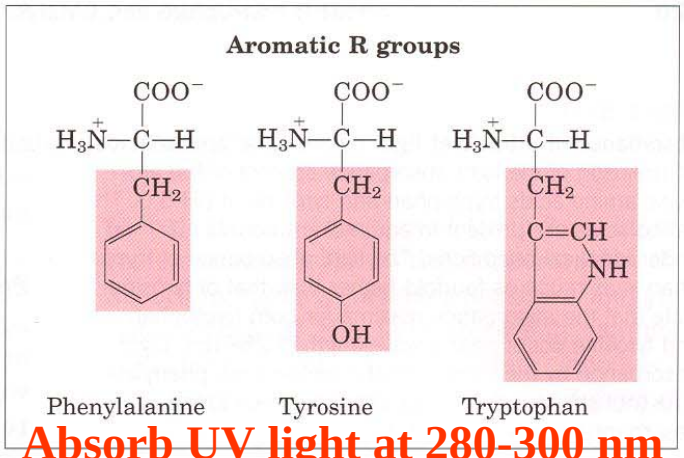


Four Levels of Protein Structuring

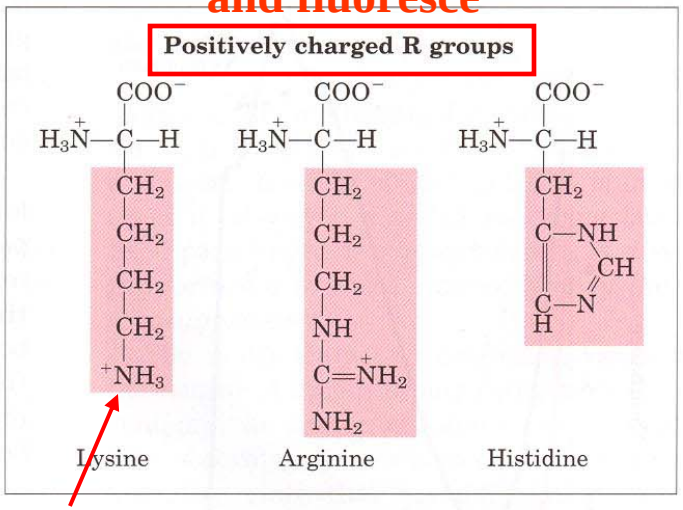


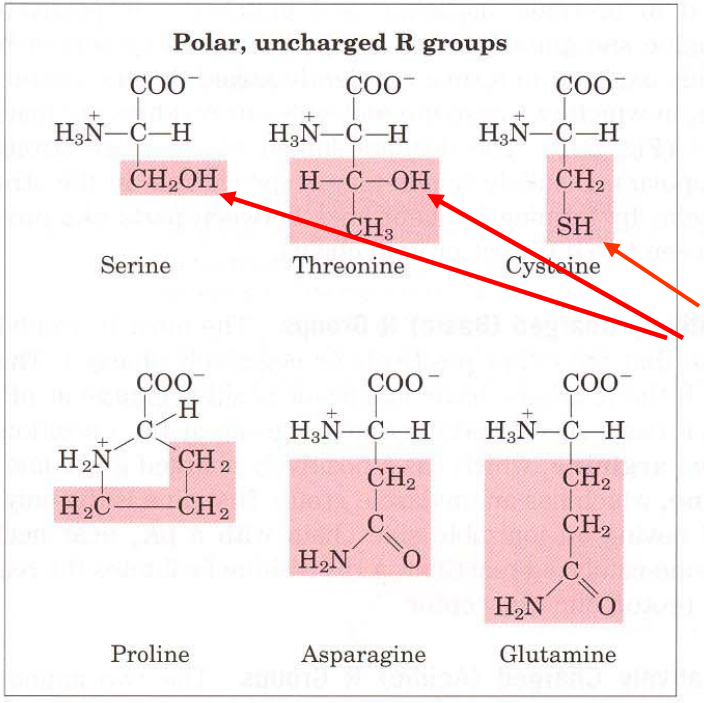
hydrophobic

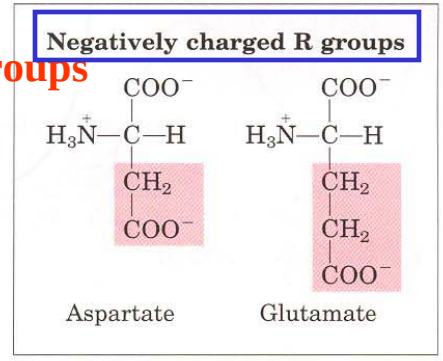




Absorb UV light at 280-300 nm and fluoresce







Reactive groups in R groups

20 Amino Acids

pKa Values of Amino Acids Define the Charge of Protein

Properties and Conventions Associated with the Standard Amino Acids

Amino acid	Abbreviated names		M_r	pK_a values: pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)
Nonpolar, aliphatic R groups									
Glycine	Gly	G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala	A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val	V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu	L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile	I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met	M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups									
Phenylalanine	Phe	F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp	W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups									
Serine	Ser	S	105	2.21	9.15		5.68	−0.8	6.8
Proline	Pro	P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr	T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys	C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn	N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln	Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups									
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups									
Aspartate	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	6.3
Average			120						

Important for terminal amino acids only

Structure of DNA

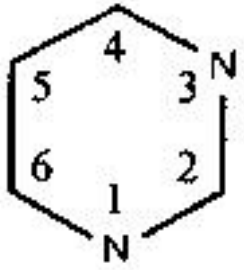
Structure units – 4 nucleotides

“Primary structure” – ssDNA sequence (-O-C- bonds)

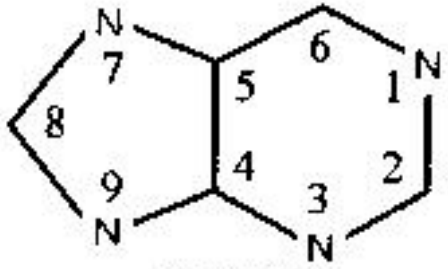
“Secondary structure” – ddDNA, double helix (H bonds: A-T, C-G)

“Tertiary structure” – Nucleosomes (weak interactions)

“Quaternary structure” – Chromatin (weak interactions)

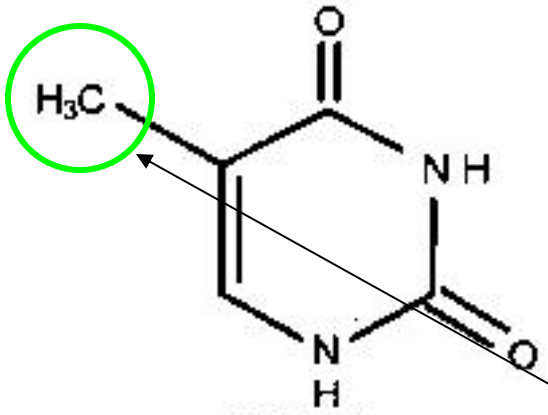


PYRIMIDINE

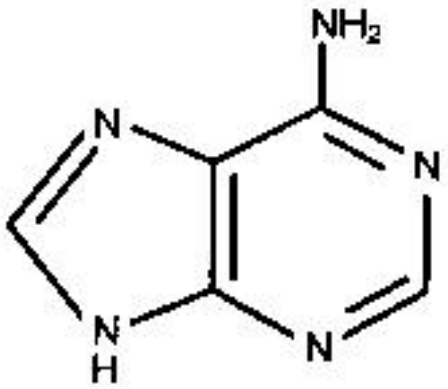


PURINE

4 Bases Found in DNA

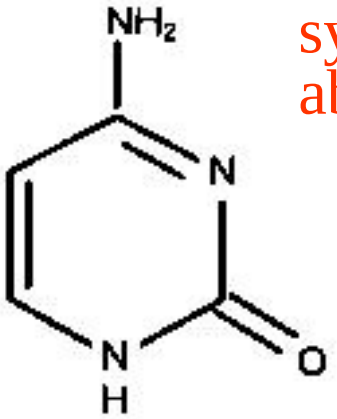


THYMINE

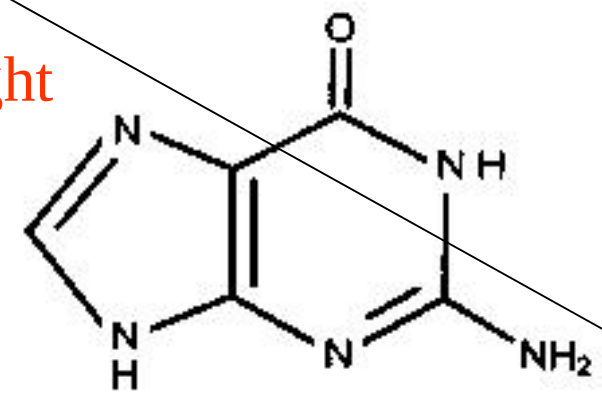


ADENINE

Conjugate systems absorb light



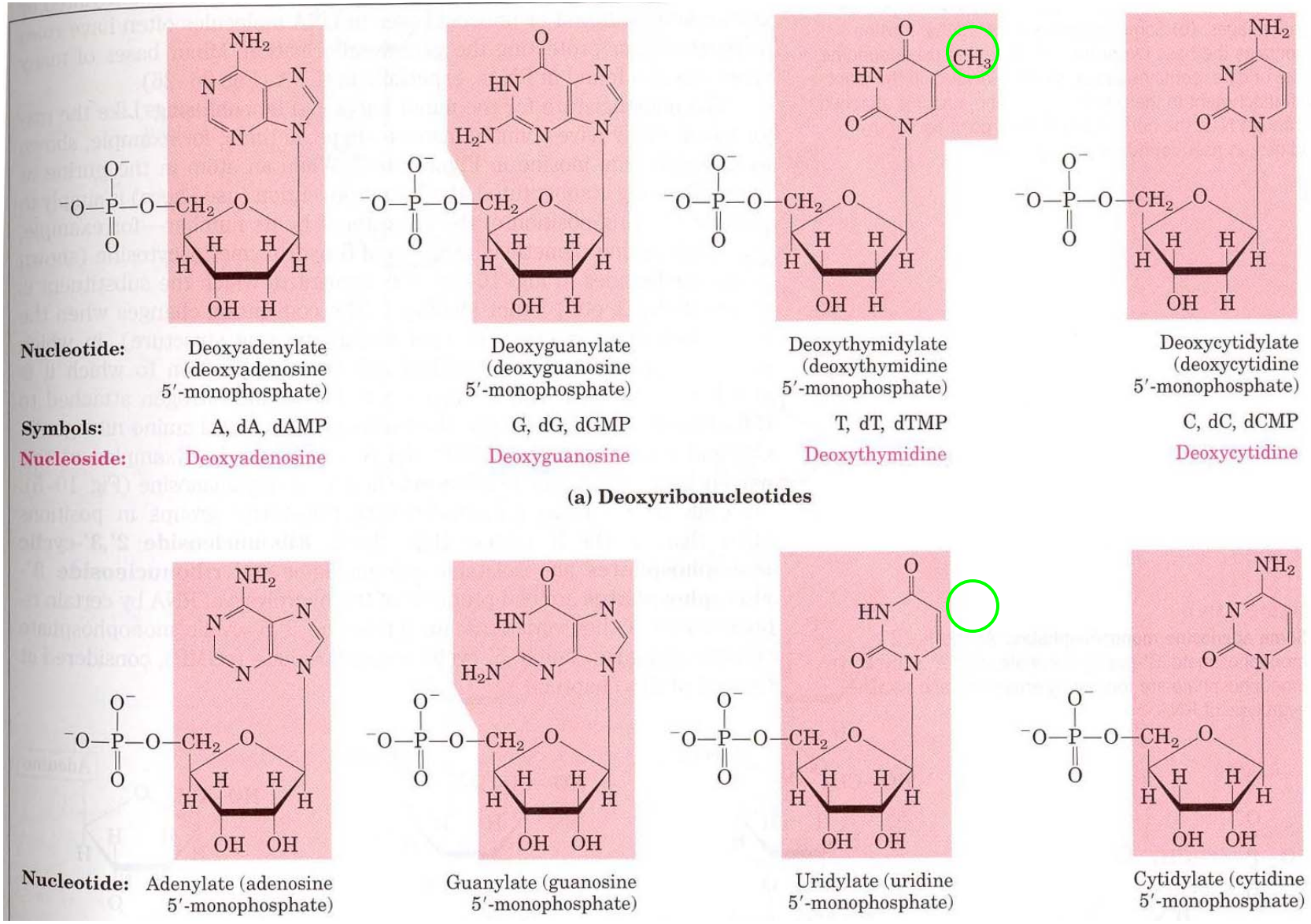
CYTOSINE



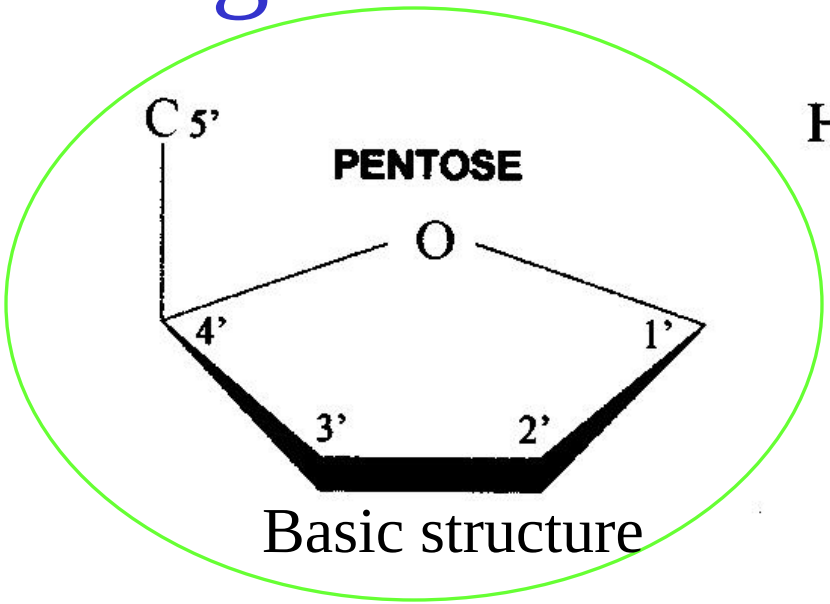
GUANINE

Note: In RNA Uridine replaces Thymine. Uridine does not have the CH_3 group

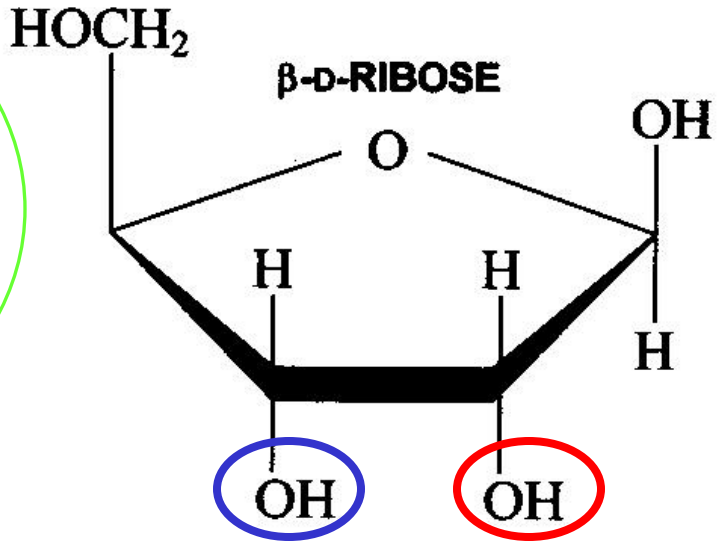
Nucleotides



Sugars found in DNA and RNA

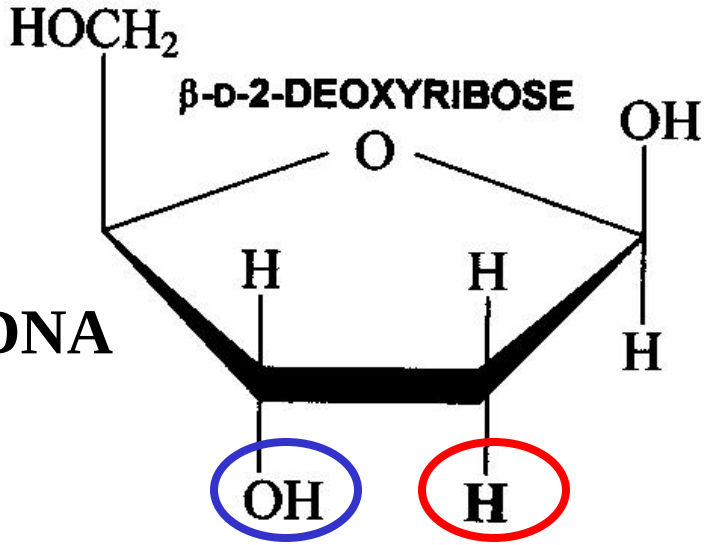


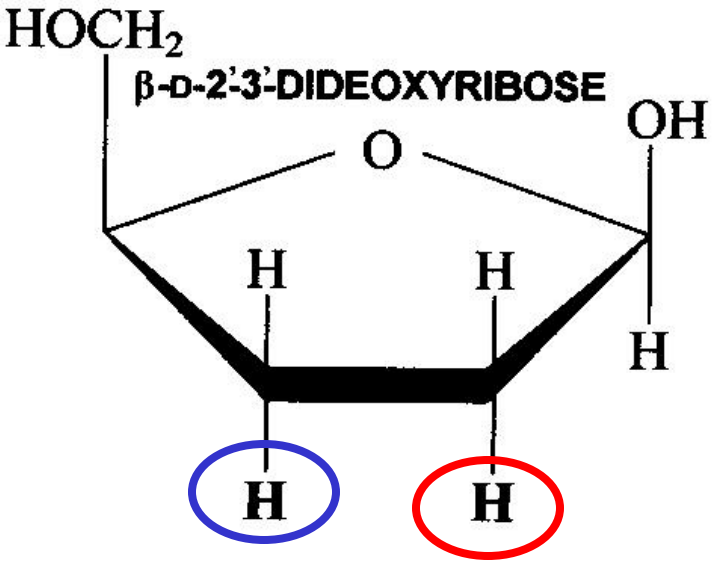
Basic structure



In RNA

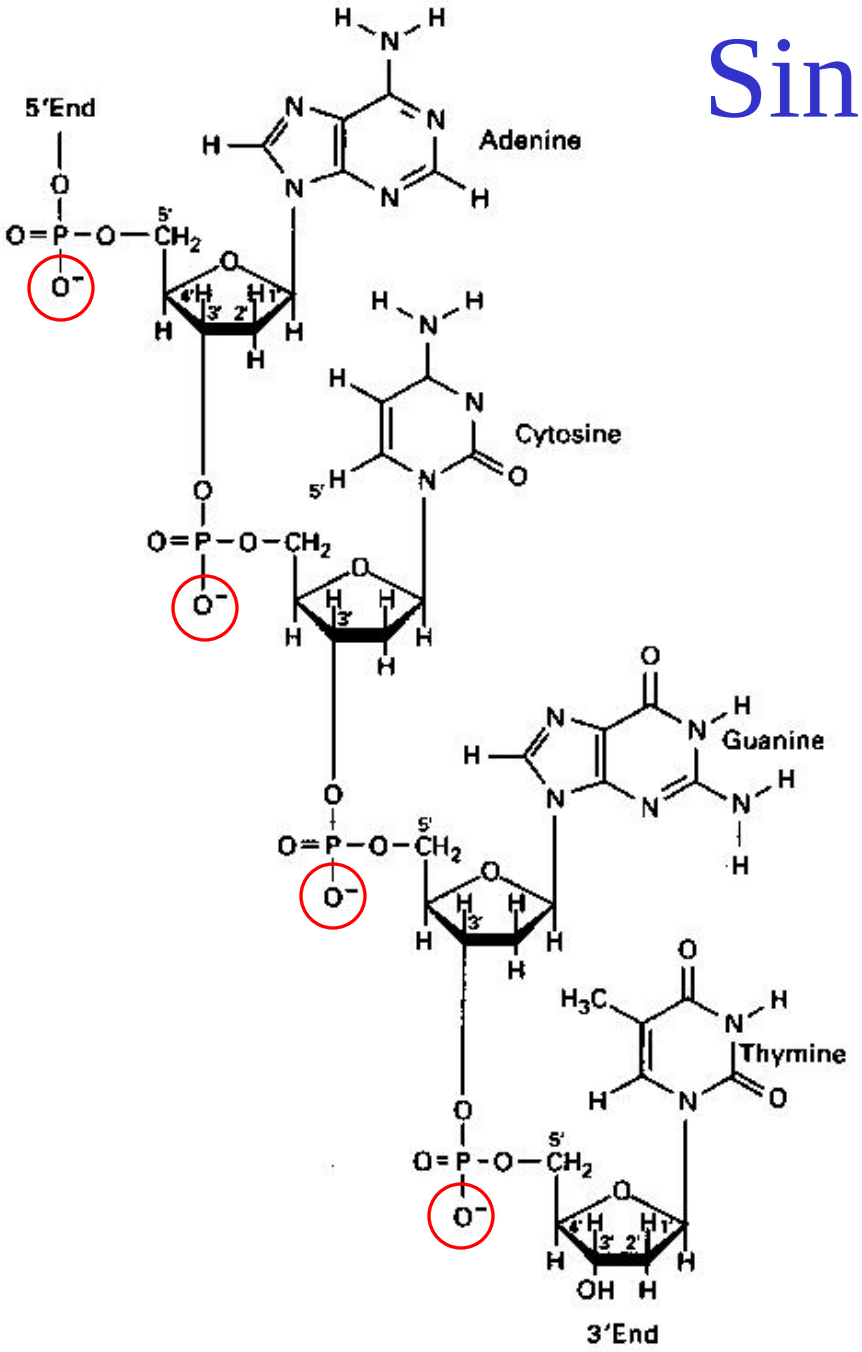
In DNA





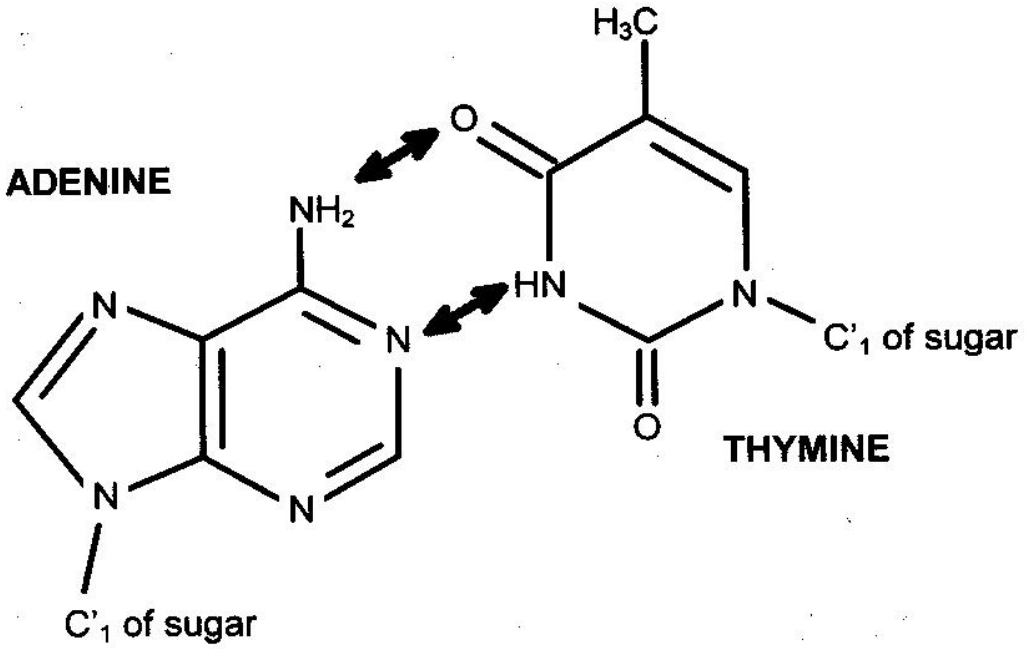
Stopper
Used in DNA
sequencing

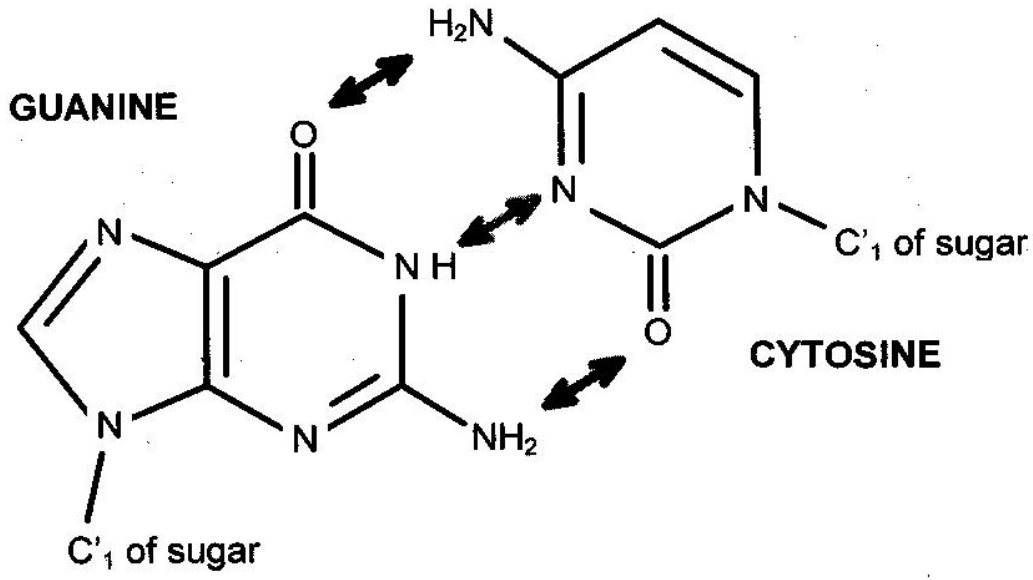
Single-stranded DNA (ssDNA)



Sugars are involved in forming covalent bonds in ssDNA (primary structure)

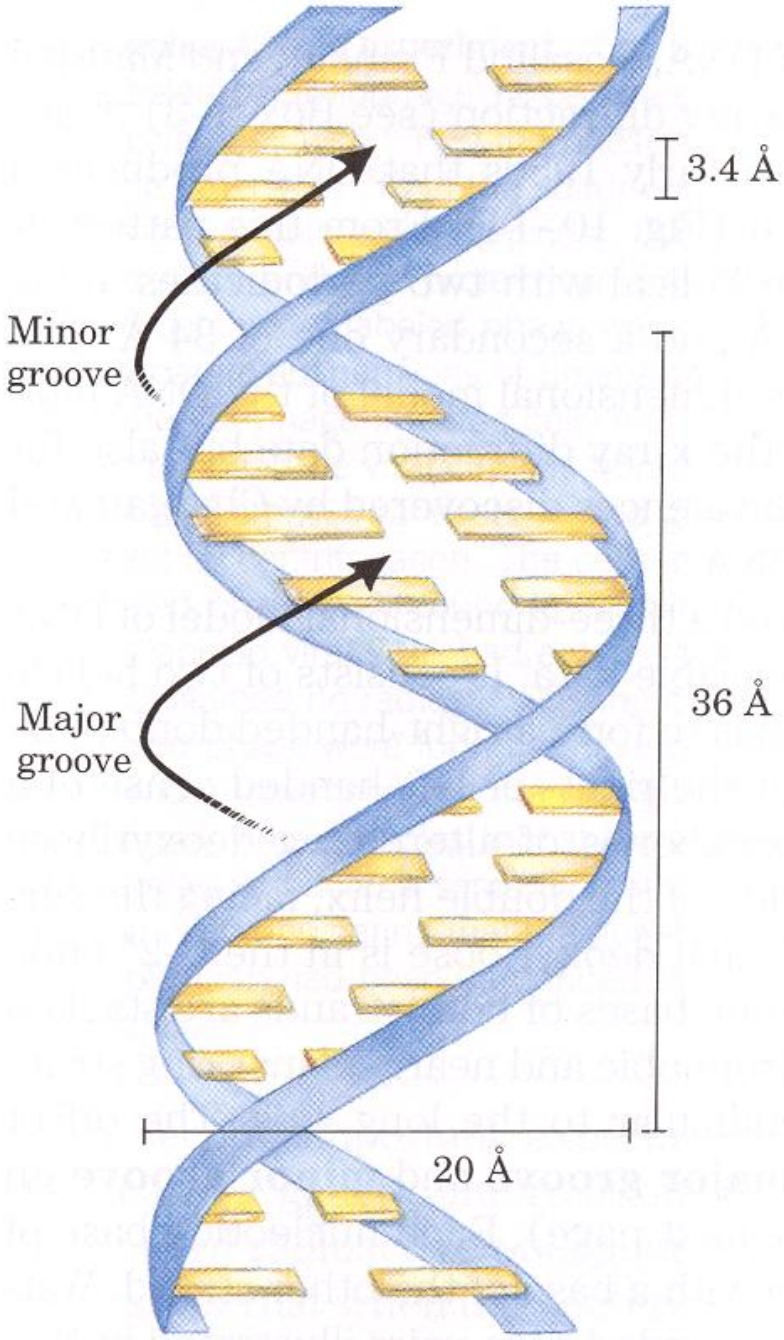
The charge of DNA molecule is negative due to phosphate groups. The charge to mass ratio is approximately constant.





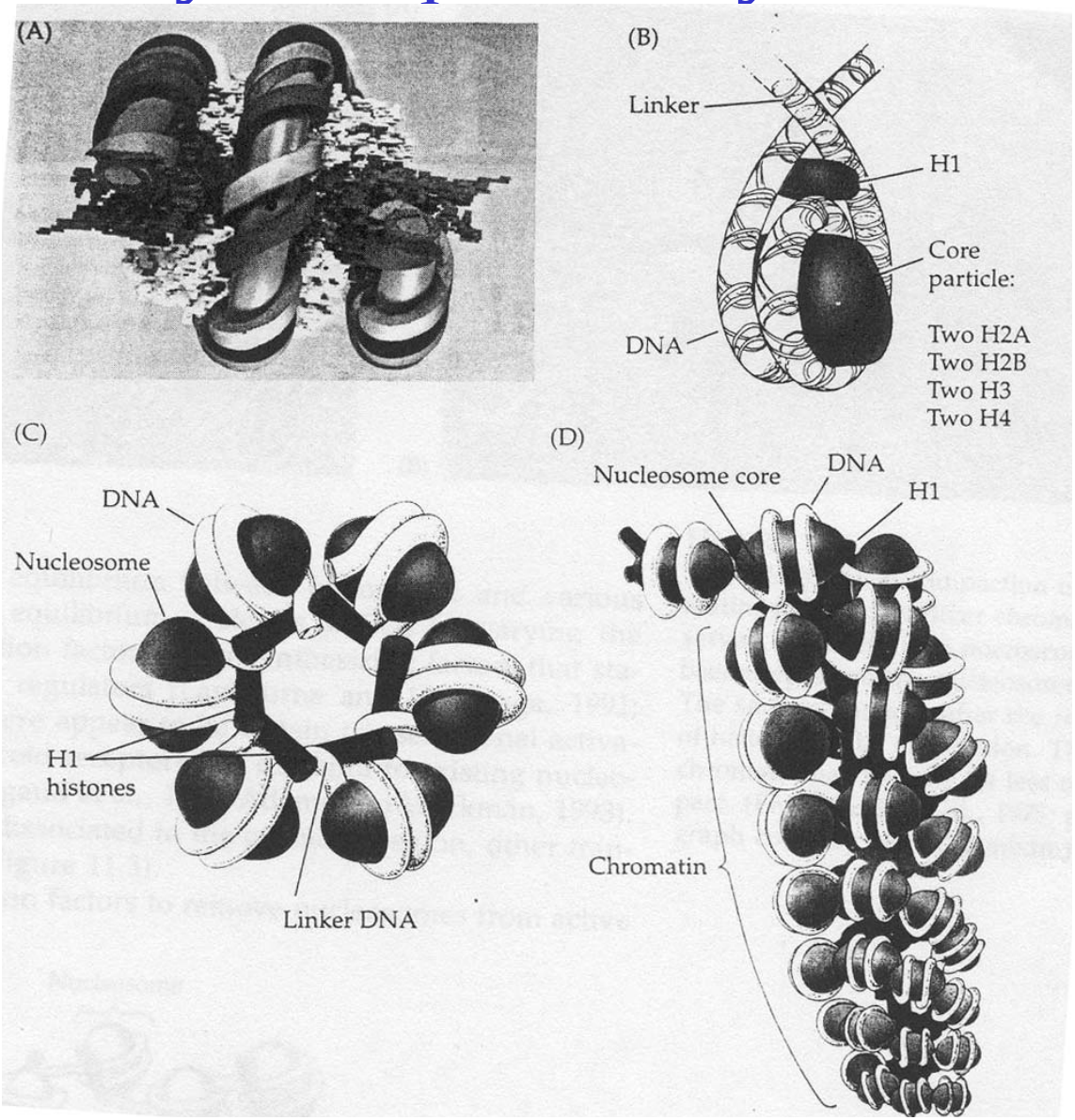
H-bonds in A-T and G-C pairs form dsDNA (secondary structure)

The G-C bond is 3/2 stronger than the A-T one



dsDNA

Chromatin Structure
(tertiary and quaternary structures)



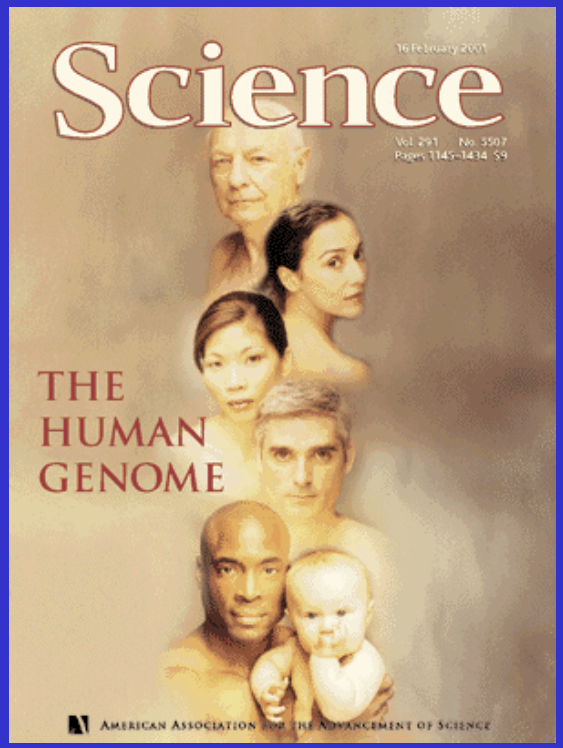
Parameters that can be used to distinguish proteins from proteins and DNA from DNA

	Proteins	DNA
1. Size	Yes	Yes
2. Mass	Yes	Yes
3. Shape	Yes	Yes (e.g. ds vs. ss)
4. Charge	Yes	Yes
5. Isoelectric point	Yes	No
6. Charge to mass ratio	Yes	No
7. Electron structure	Yes (heme)	No
8. Affinity interactions	Yes	Yes

Quantitation of Biopolymers in Complex Mixtures

- Purification (centrifugation – chapter 13, precipitation, dialysis, chromatography - chapter 14)
- Separation (chromatography – chapter 14, electrophoresis – chapters 10-12, MS – chapter 15)
- Detection (absorbance – chapter 1, fluorescence, affinity – chapters 5 - 7, MS - chapters 15)

Separation is necessary when detection is not specific or not specific enough



Capillary Electrophoresis (CE)

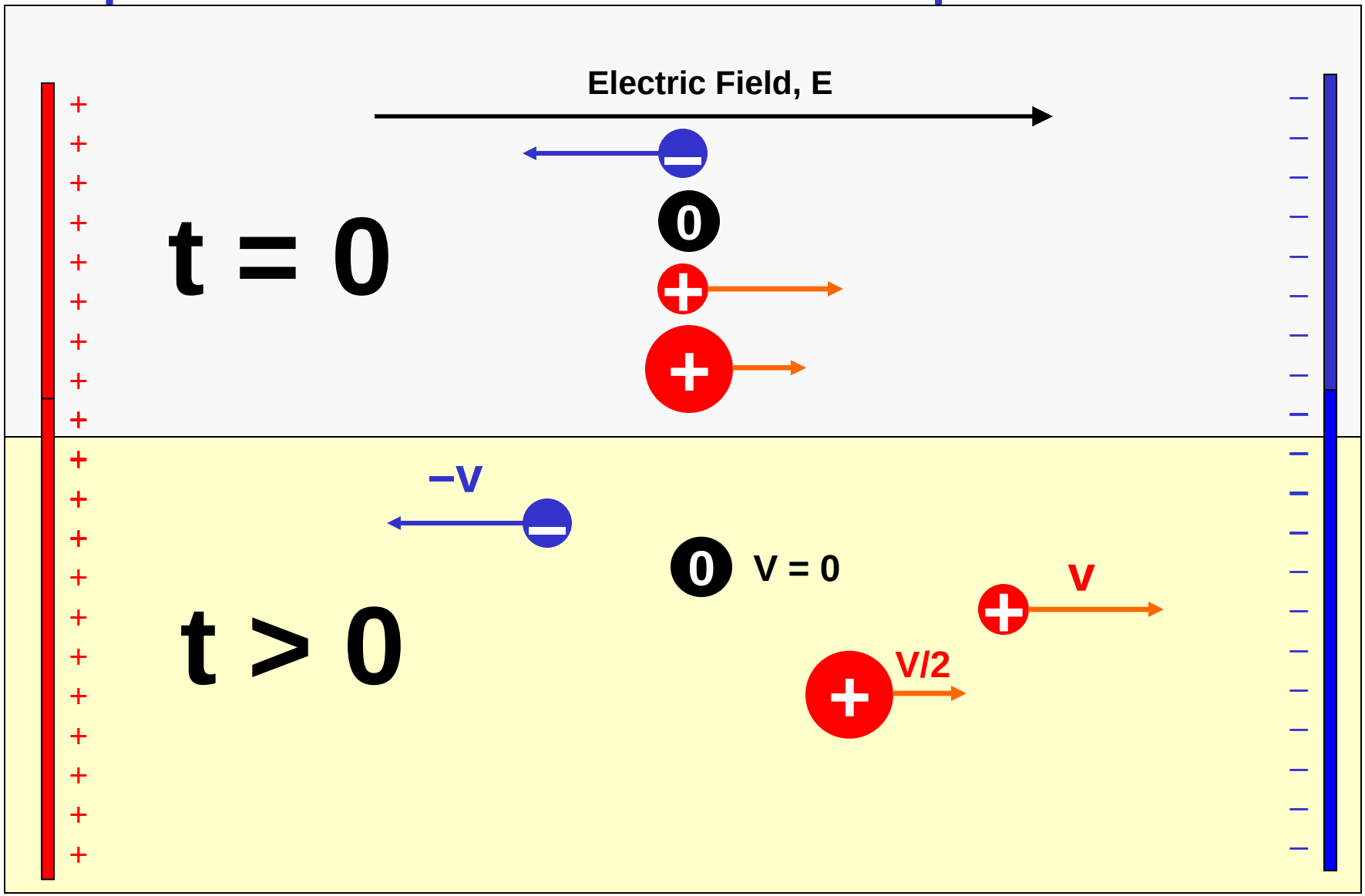
Unsung Heroes

"An ocean apart, Dovichi at the University of Alberta in Canada and Kambara at the Hitachi Co. in Tokyo independently hit upon a sequencing technology that greatly advanced the human genome project. The method, used in today's high-speed machines, uses laser beams to scan DNA being pumped through numerous capillary tubes, simultaneously identifying the bases by color-coded chemical tags."

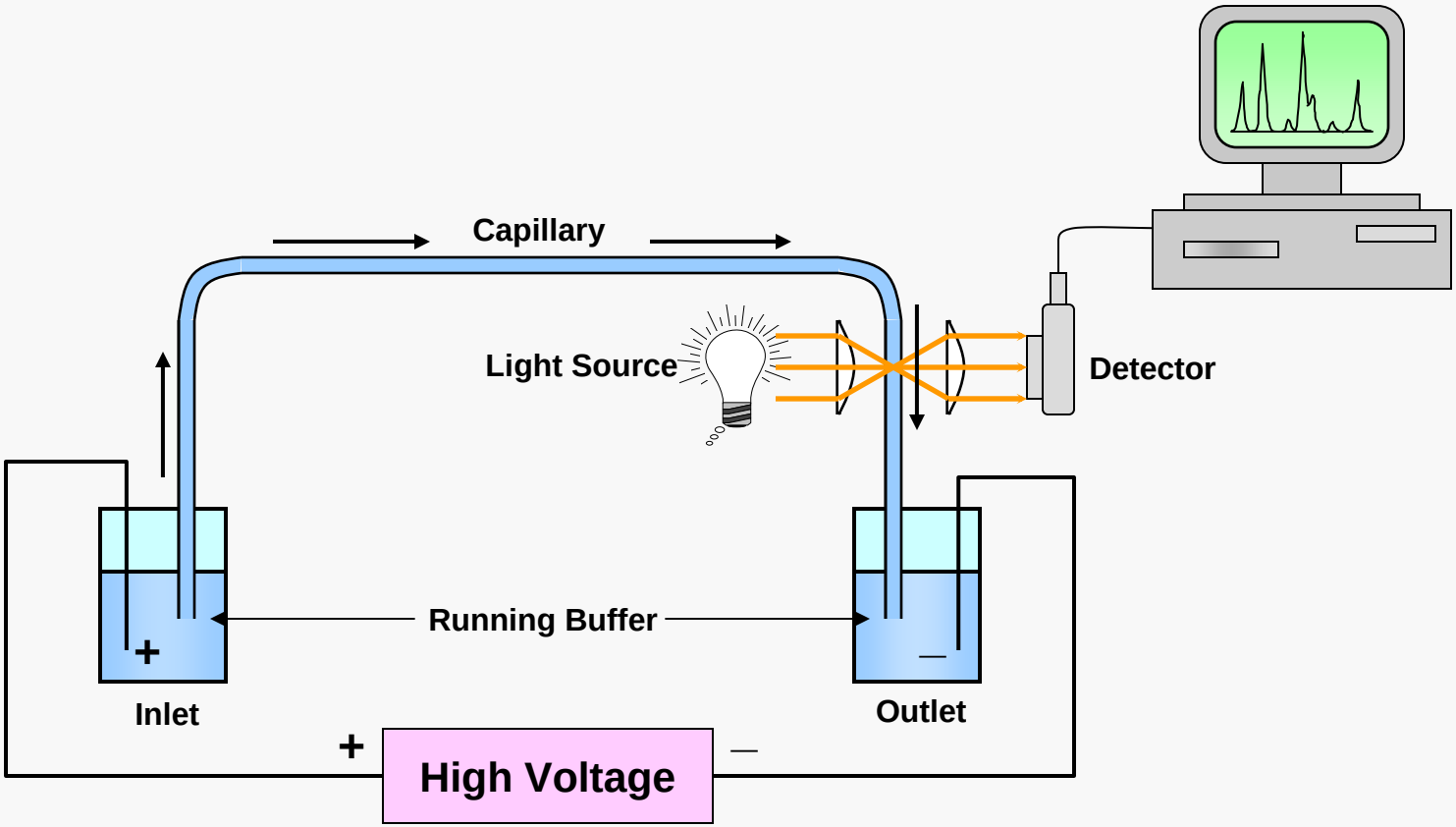
Science 2001, Feb 16 2001: 1207

Electrophoresis

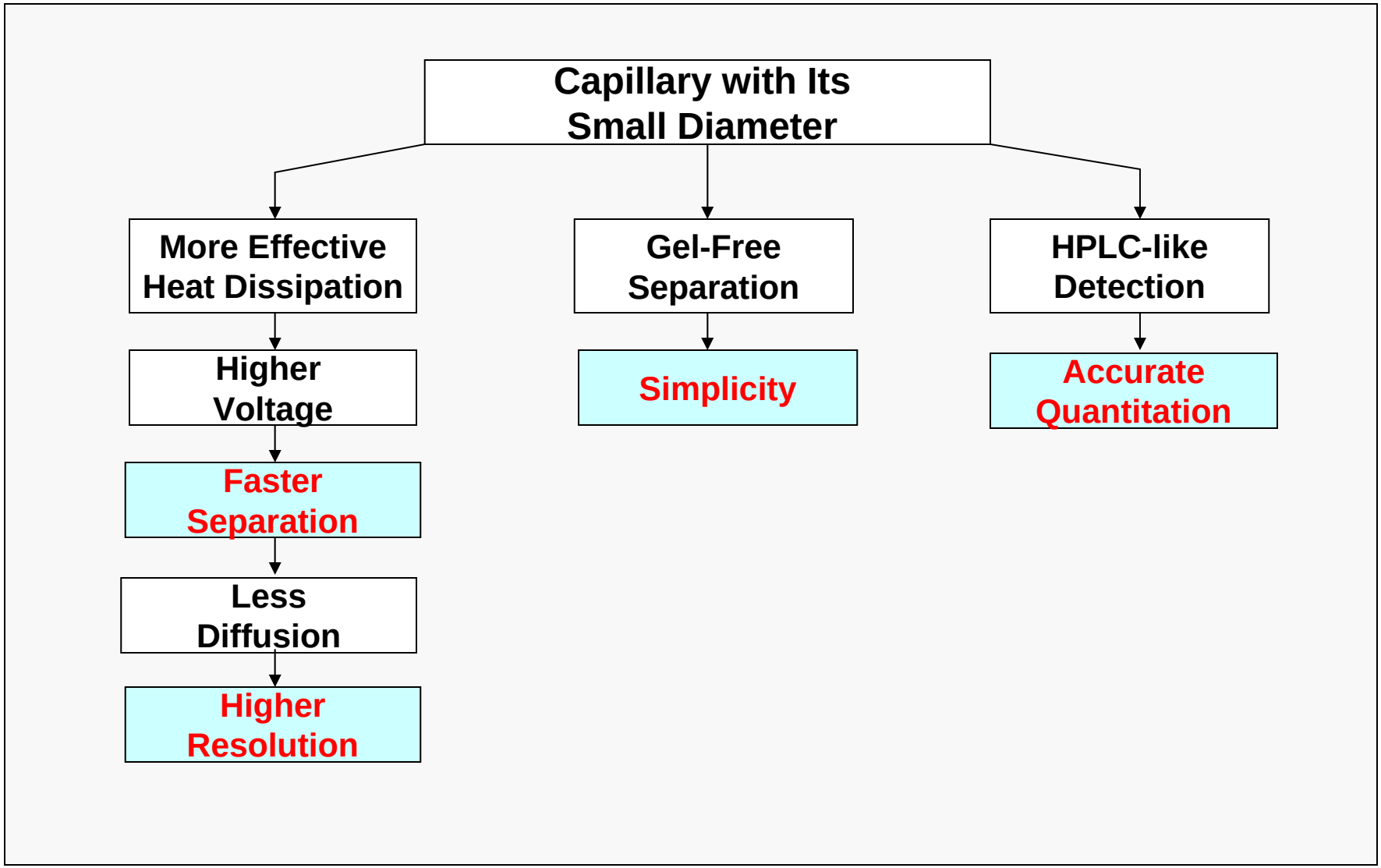
Separation based on differences in electrophoretic mobilities



Capillary Electrophoresis Instrument

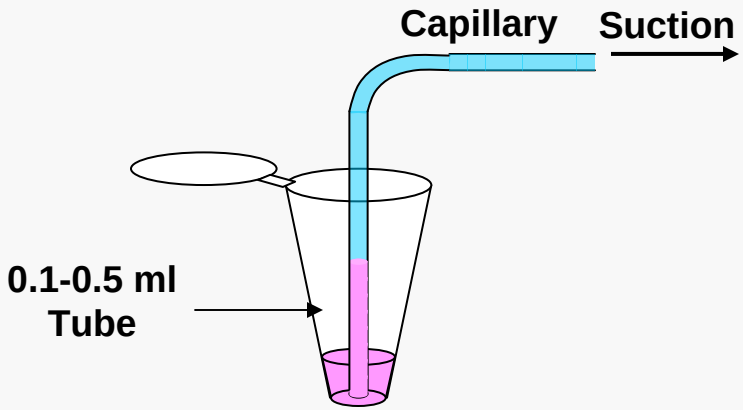


Why Electrophoresis in a Capillary?

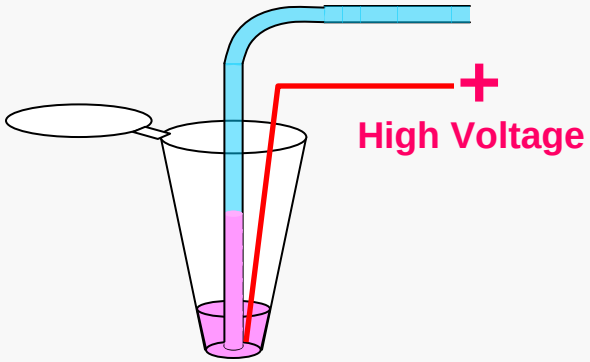


Sample Injection

Injection by Suction (pressure)

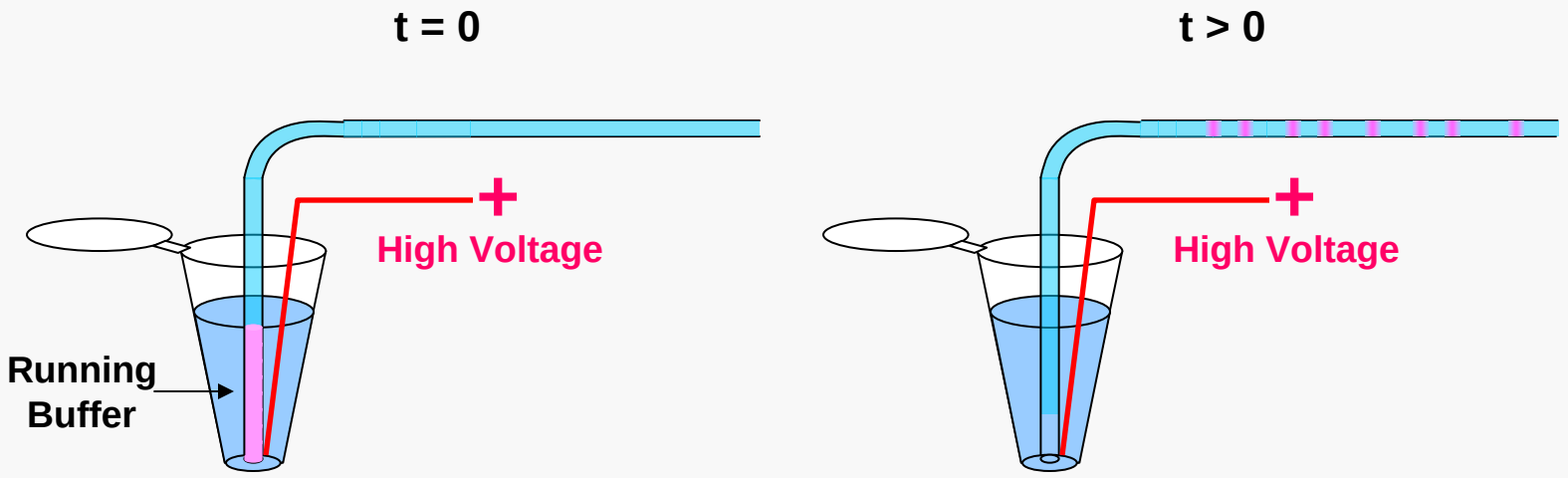


Electrokinetic Injection



Minimal Sample Volume ~1 μL
Typical Injection Time 1-5 s
Typical Injected Volume ~10^{-4} μL

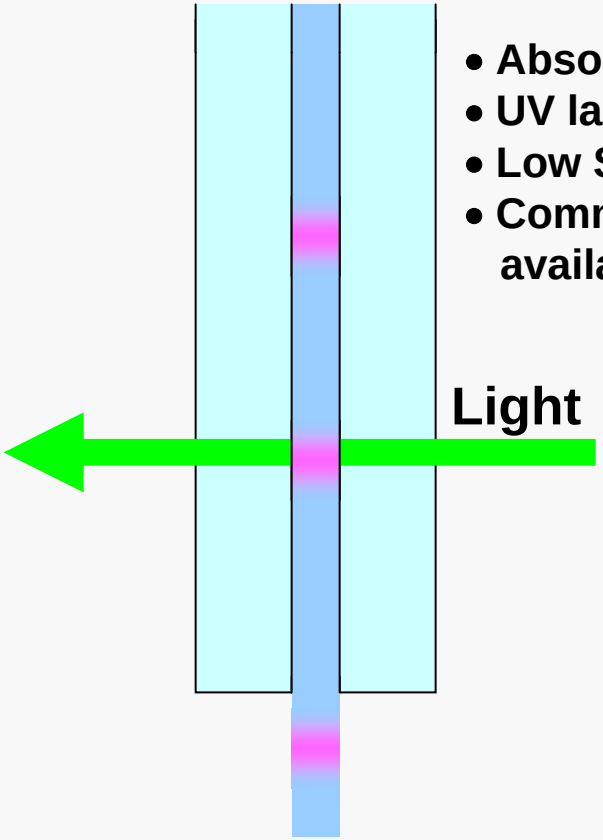
Separation



Typical Electric Field 400 V/cm
Typical Separation Time 10 min

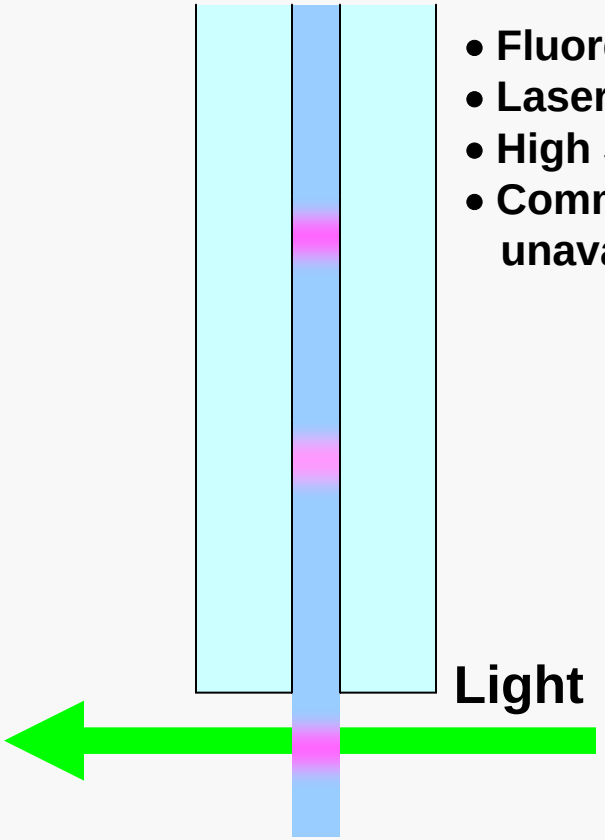
Detection

On-Column
Light Scattering



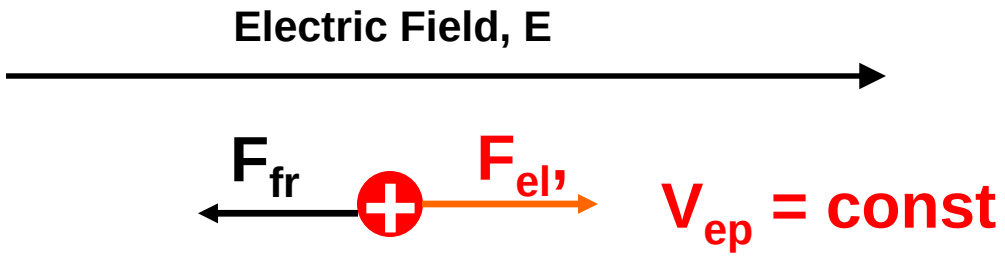
- Absorbance
- UV lamp
- Low Sensitivity
- Commercially available

Post-Column
No Light scattering



- Fluorescence
- Laser
- High Sensitivity
- Commercially unavailable

Electrophoretic Mobility



$$v_{ep} = const$$

$$\sum_i F_i = F_{el} + F_{fr} = qE - 6\pi\eta r v_{ep} = 0$$

$$v_{ep} = \frac{qE}{6\pi\eta r}$$

v_{ep} is determined experimentally as L/t_m,
where L is the capillary length, t_m is migration time

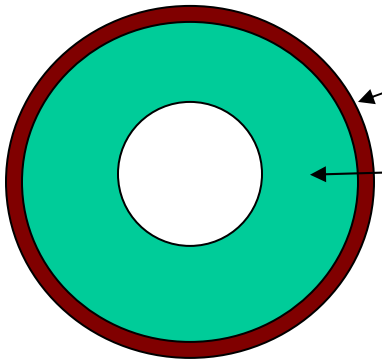
$$\mu = \frac{v_{ep}}{E} = \frac{q}{6\pi\eta r}$$

Electrophoretic Mobility = Electric Charge / Friction Coefficient

Does not depend on electric field

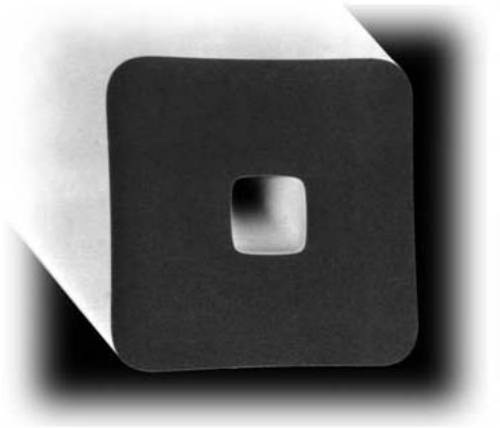
Difference between Electrophoresis and Mass Spectrometry

Capillaries



Polyimide Coating

Material: fused silica, Teflon, quartz
Inner bore, diameter 2 – 150 μm
Outer diameter 150 – 375 μm

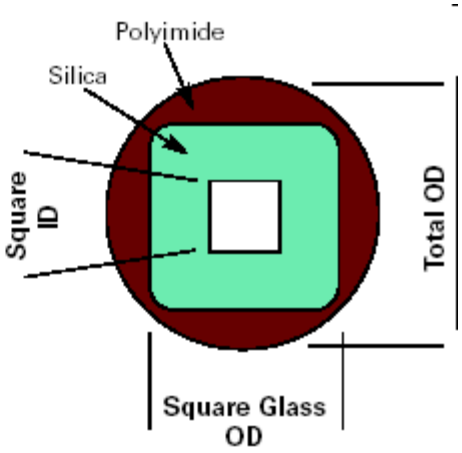


Square capillaries are available

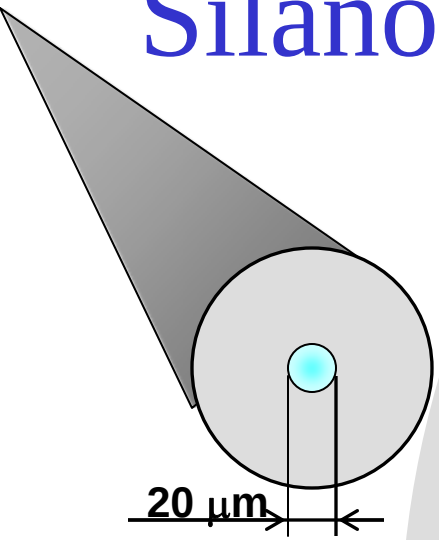
Square Capillary Tubing Sizes

Product Descriptor	ID* (μm)	Glass OD* (μm)	Total OD* (μm)
WWP050375	050 ± 05	300	363 ± 15
WWP075375	075 ± 05	300	363 ± 15
WWP100375	100 ± 05	300	363 ± 15

* Measured flat-to-flat.



Ionization of Silanols



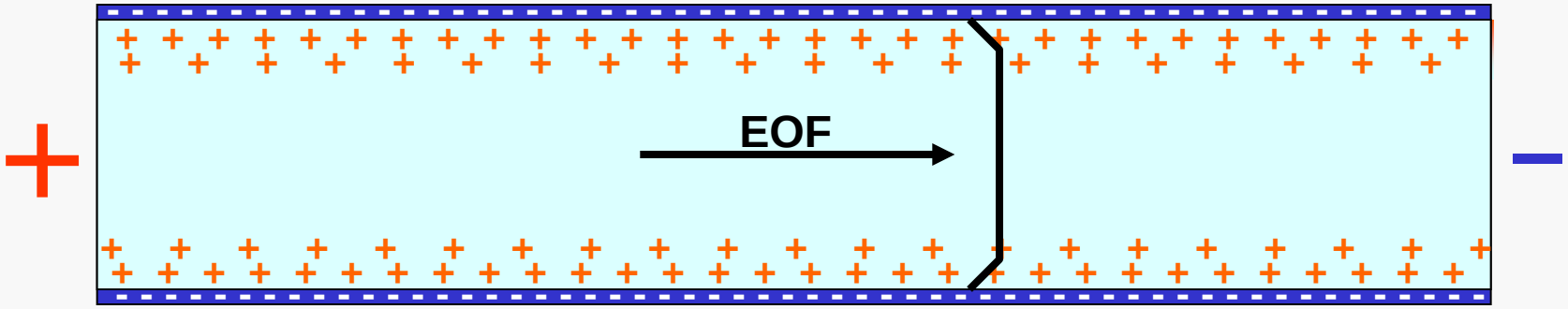
Negatively charged inner surface

Electroneutral entire system

Positively charged buffer solution

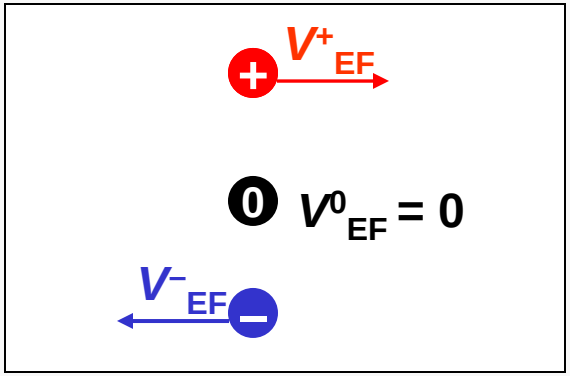
Electroosmotic Flow (EOF)

Flow of positively charged buffer from anode to cathode



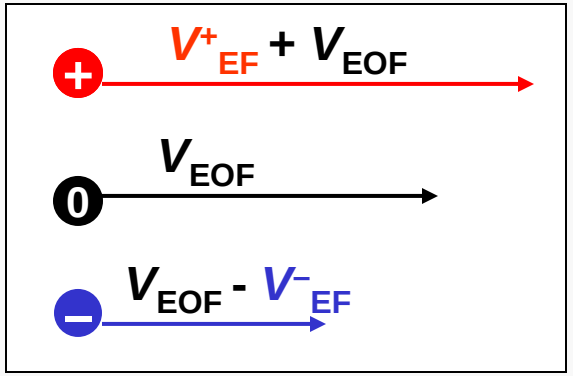
EOF allows simultaneous separation of positive, neutral and negative species

Electrophoretic Velocities



Electrophoretic Velocity

Resulting Velocities



Measuring the velocity of EOF, v_{eo}

Theoretically v_{eo} is defined through the Smoluchovski equation:

$$v_{EOF} = \frac{\varepsilon\zeta}{4\pi\eta}E$$

Where ε is the dielectric constant of the background electrolite (run buffer), ζ is the zeta potential, η is viscosity, E is electric filed.

It is hard to calculate $\boldsymbol{v_{eo}}$ since ε of mixtures is rarely available and ζ is hard to accurately determine

Therefore, v_{EOF} is typically measured experimentally using a neutral molecule ($v_{ep} = 0$):

$v_{app} = v_{ep} + v_{EOF} = 0 + v_{EOF} = v_{EOF}$

Control of EOF

1. Viscosity
2. Electric field
3. Temperature (decreases viscosity)

Obvious due to:

$$v_{EOF} = \frac{\varepsilon \zeta}{4\pi\eta} E$$

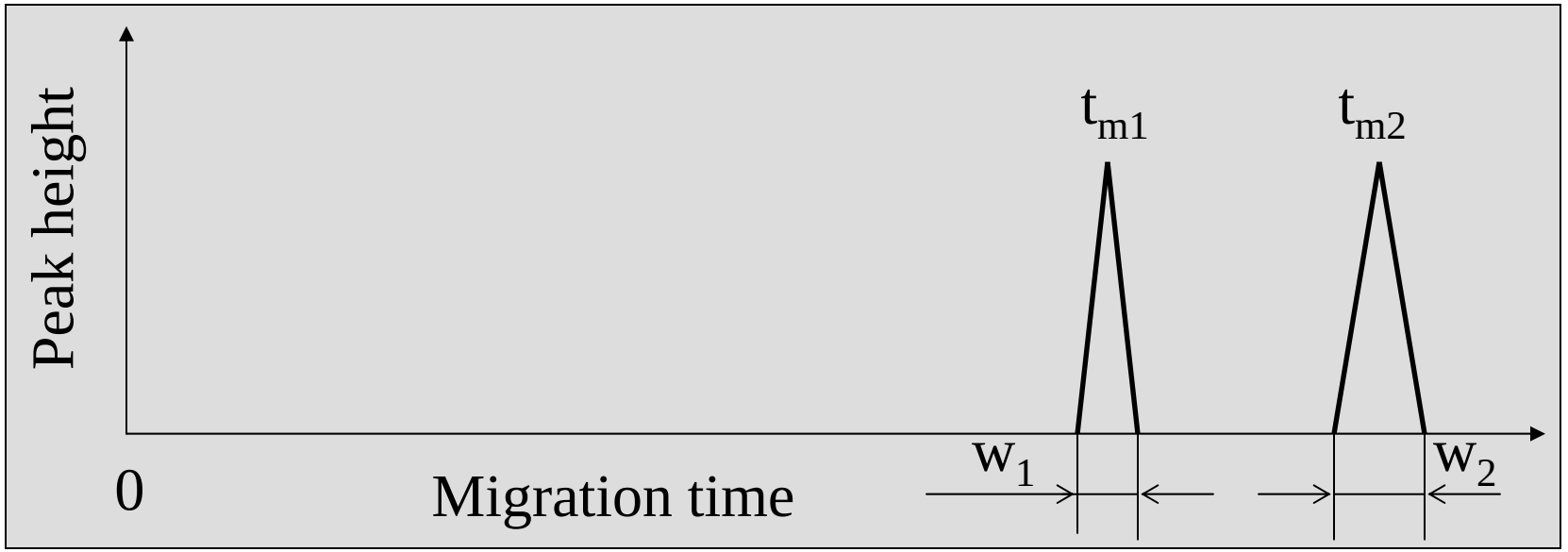
4. pH: increasing pH increases deprotonation of silanol groups and, thus, increases v_{EOF}

5. Buffer concentration: $v_{EOF} \approx 1/(ZC^{1/2})$

6. Capillary surface:

- static coating (covalent silanol modifications or physical coatings with cross-linked polymers)
- dynamic coatings with buffer additives e.g. surfactants

Efficiency and Resolution



Efficiency is characterized by the number of theoretical plates:

$$N = 16(t_m/w_{1/2})^2$$

where t_m is migration time, $w_{1/2}$ is peak width on the half height

Resolution is characterized by the distance between the peaks relative to the sum of peak widths:

$$R = \frac{2(t_{m1} - t_{m2})}{(w_1 + w_2)}$$

Calculation of Injected Plug Parameters for Suction/Pressure Injection

The velocity of sample injection is described by the Poiseulle eqn:

$$v_{inj} = \frac{r^2 P}{8\eta L}$$

Where r is the radius of the capillary, P is the differential pressure between the inlet and outlet of the capillary, η is the viscosity (9×10^{-4} kg m^{-1} s^{-1} for aqueous solutions), and L is the total length of the capillary

The length of the injected plug is: $L_{inj} = v_{inj} \times t_{inj} = \frac{r^2 P}{8\eta L} t_{inj}$

The volume of the injected plug is: $V_{inj} = L_{inj} \times \pi r^2 = \frac{\pi r^4 P}{8\eta L} t_{inj}$

Calculation of Injected Plug Parameters for Electrokinetic Injection

- First, the apparent velocity of the analyte ($v_{app} = v_{EOF} + v_{ep}$) should be determined in experiment:

$$v_{app} = \frac{L}{t_m}$$

where L is the total length of the capillary and t_m is the experimentally determined migration time from the inlet to the detector

- Injection is often performed at a voltage, U_{inj}, lower than that used for separation, U. Therefore, the velocity of injection will be lower: $v_{inj} = v_{app}\frac{U_{inj}}{U} = \frac{L}{t_m}\frac{U_{inj}}{U}$

The length of the injected plug is: $\boldsymbol{L_{inj} = v_{inj} \times t_{inj} = \frac{LU_{inj}}{t_m U} t_{inj}}$

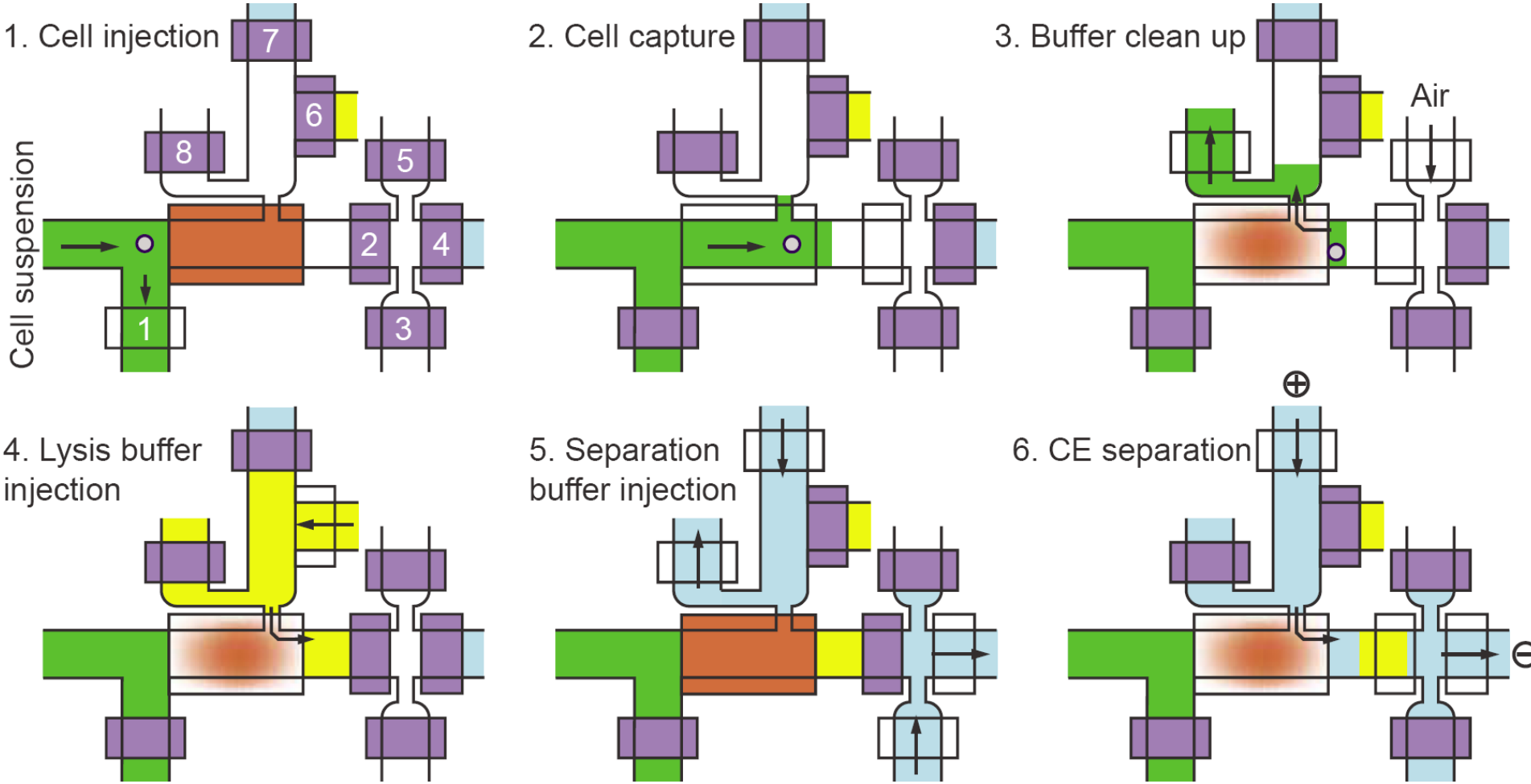
The volume of the injected plug is: $V_{inj} = L_{inj} \times \pi r^2 = \pi\ r^2 \frac{LU_{inj}}{t_m U} t_{inj}$

Counting Low–Copy Number Proteins in a Single Cell

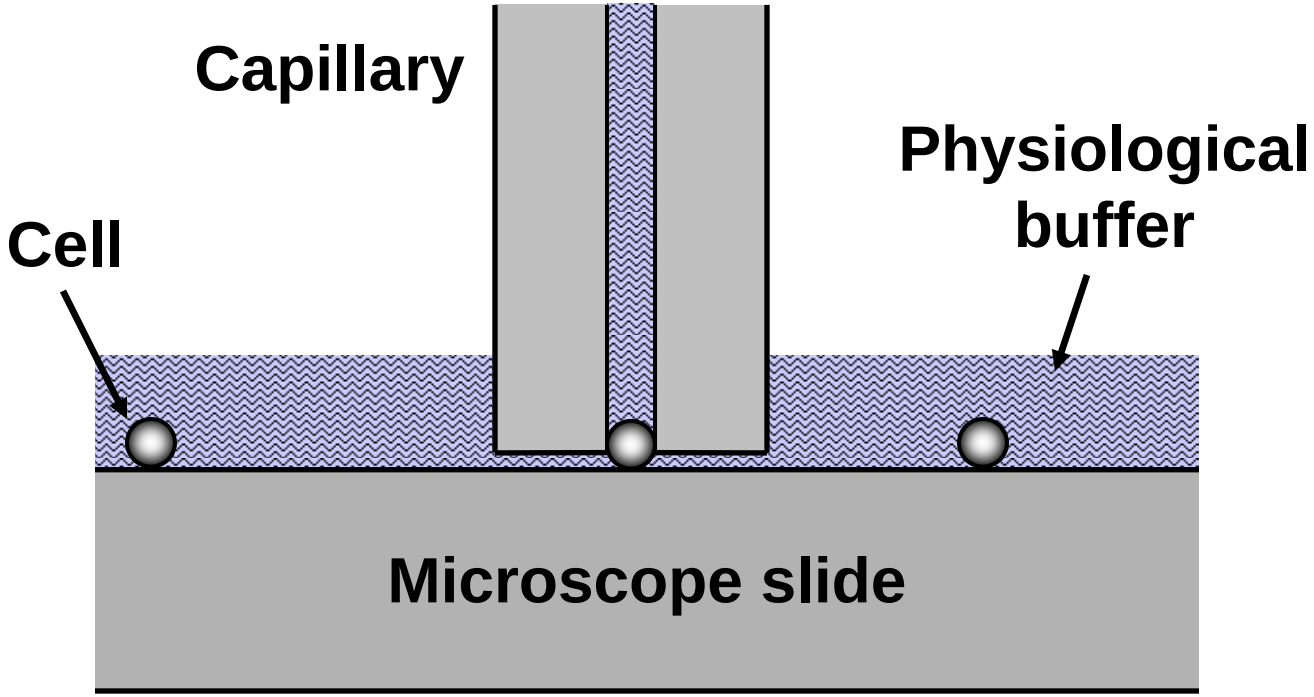
Zare's group

Science 2007, 315, 81-94

The single-cell analysis chip



Sampling cells for chemical cytometry



Injection of Cells into the Capillary by Drug Force $F_{drag} = av$

The fluid flow converges into the capillary. The flow continuity condition and the convergent nature of the flow require that the velocity increase with decreasing distance, *R*, from the capillary entrance:

$$v \cong v_0 \frac{r^2}{(R+r)^2}$$

where v_0 is the fluid velocity inside the capillary and *r* is the radius of the capillary. The flow can be created by suction or by Electrosmosis:

$$\boldsymbol{v}_0^{\boldsymbol{suction}} = \frac{\boldsymbol{r}^2 \boldsymbol{P}}{8\boldsymbol{\eta L}} \qquad v_0^{\mathrm{EOF}} = \frac{\varepsilon \zeta U}{4\pi \eta L}$$

$$F_{drag}^{suction} \cong \frac{a}{8} \frac{r^2}{(R+r)^2} \frac{r^2}{L} \frac{1}{\eta} P \qquad F_{drag}^{EOF} \cong \frac{a}{4\pi} \frac{r^2}{(R+r)^2} \frac{1}{L} \frac{\varepsilon \zeta}{\eta} U$$

Drag Force Properties

Thus, the drag force caused by siphoning can be increased by (i) reducing the distance, R, between the cell and the entrance to the capillary, (ii) decreasing the viscosity, η, of the buffer, (iii) increasing the capillary inner radius, r, (iv) reducing the length, L, of the capillary, and (v) increasing the pressure drop, P, across the capillary. The drag force caused by electroosmotic flow is affected by all the above factors except for capillary inner radius and pressure. Instead, it depends on zeta potential, ζ, which is a function of the density of negative charge on the inner capillary wall, and applied voltage, U.

Cell Adhesion Force, F_{adh}

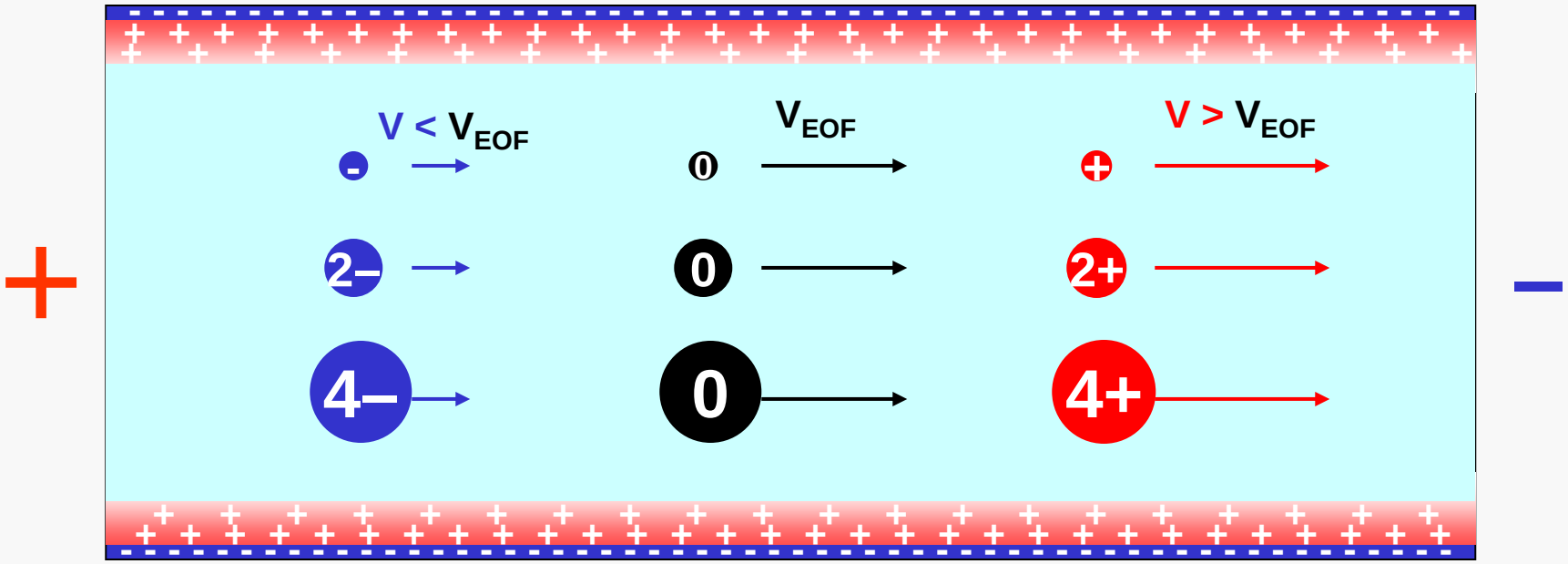
The force of cell adhesion to the surface, F_{adh}, can prevent cell injection. Most cell types, including the tissue culture cells used in this study, tend to adhere to the surface of a plate or microscope slide. This adhesion is caused mainly by interactions between cell surface proteins and the surface of plastic or glass material. Live cells are usually negatively charged; therefore, electrostatic interaction can also influence adhesion if the surface is charged. That is, F_{adh} is determined by both cellular status and surface parameters such as the chemical nature, charge, and hydrophobicity of the surface. Obviously, the cell can be injected only if $F_{drag} > F_{adh}$.

Major Modes of Capillary Electrophoresis

Mode	Medium	Separates	Does not Separate
1. Free Zone Electrophoresis	No additives	Small molecules Peptides, Proteins	Neutral molecules Oligo-nucleotides DNA
2. Micellar Electrokinetic Chromatography	Micelles	Small Molecules Peptides Oligo-nucleotides	Proteins DNA
3. Capillary Isoelectric Focusing	Ampholytes	Peptides Proteins	Molecules without pI
4. Capillary Gel Electrophoresis	Gel	Oligo-nucleotides Proteins	Small Molecules
5. Affinity Capillary Electrophoresis	Affinity molecules	Proteins, DNA, RNA	Small molecules

Capillary Zone Electrophoresis (CZE)

separates molecules with different electrophoretic mobilities



Separates

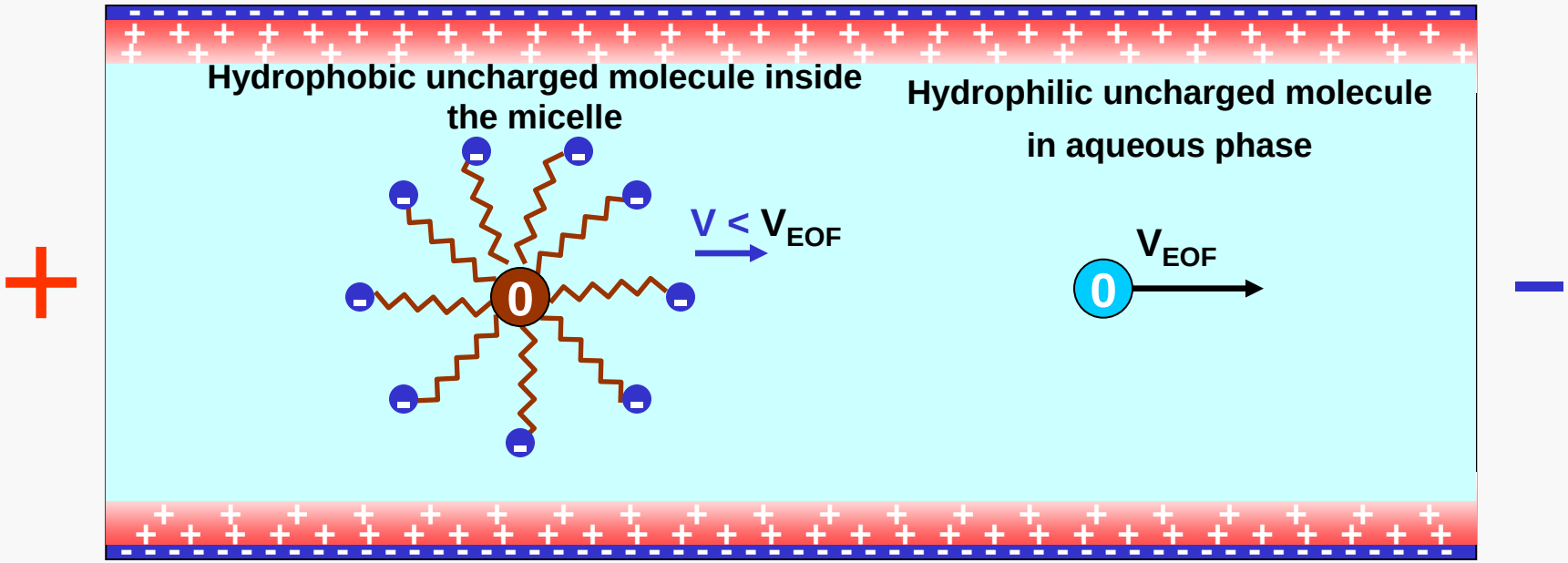
Small ions and molecules
Peptides
Proteins

Does not separate

neutral molecules
molecules with similar electrophoretic mobilities (oligo-nucleotides and DNA)

Capillary Micellar Electrokinetic Chromatography (CMEC)

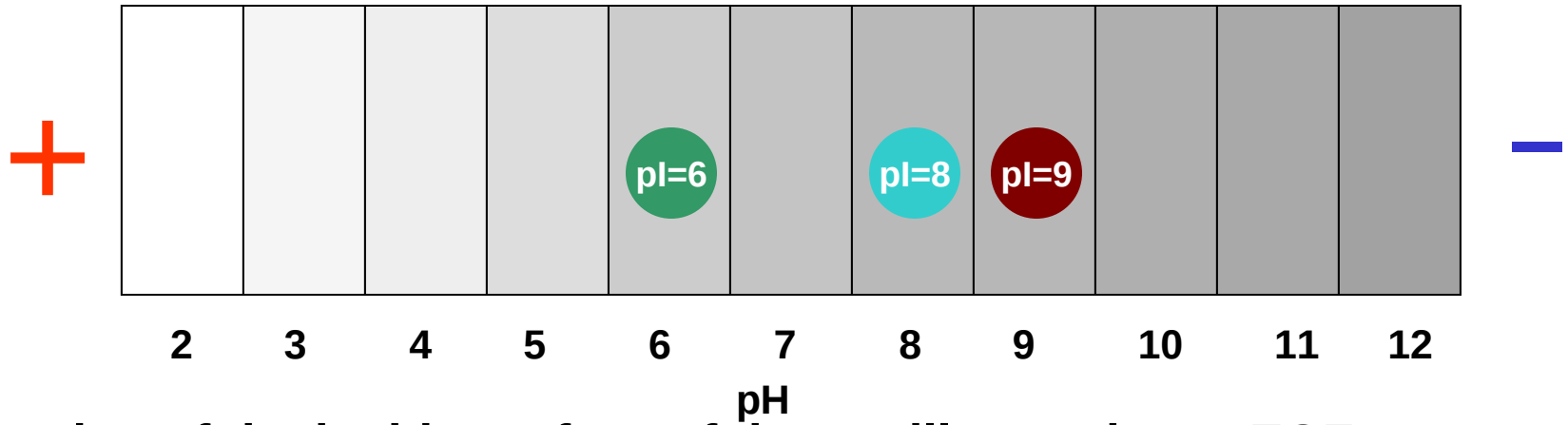
separates molecules with similar electrophoretic mobilities (e.g. neutral molecules)



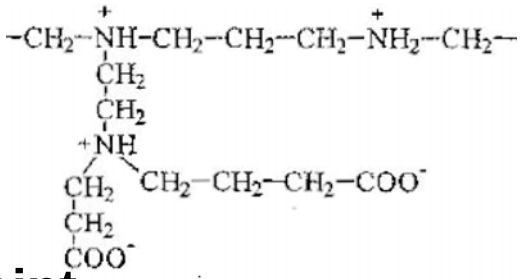
Different partitioning between the micellar and aqueous phases
Micellar phase is a pseudo-stationary phase

Separates: **Small Molecules, Peptides, Oligo-nucleatides**

Capillary IsoElectric Focusing (CIEF)



- Coating of the inside surface of the capillary reduces EOF
- Mixture of carrier ampholytes creates a linear pH gradient (Carrier ampholites are polyprotic amino-carboxylic acids with pIs within the pH range to be covered and high buffering capacity near their pIs which is achieved by having pKa of ionizable groups close to pI)
- Elution by hydraulic pressure



Separates: Peptides, Proteins
Does not separate: Molecules without isoelectric point

CIEF with whole column imaging that allows avoiding a mobilization step

Home | Our Promise | Our Technology | Our Products | Our Support | Our Education | What's New | Site Map | Our Contacts

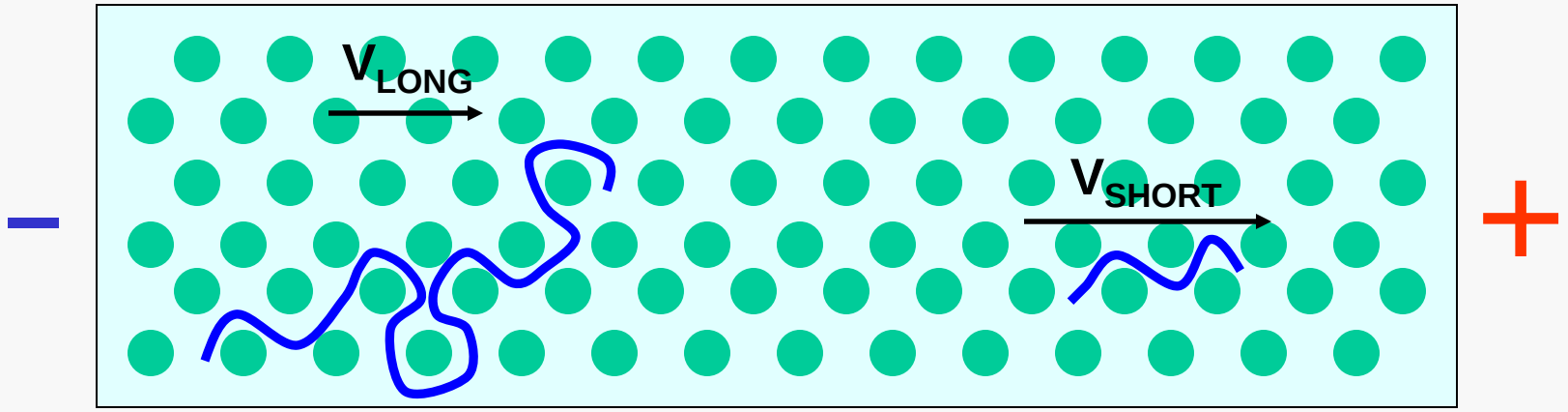
CONVERGENT BIOSCIENCE

PRODUCT SPECIFICATIONS

The iCE280, with whole column detection, completely eliminates the need for a mobilization phase. This results in a unique combination of speed, resolution and reproducibility.

- Separation Technique: Free solution whole column detection capillary isoelectric focusing (cIEF)
- Separation column: 100 µm ID x 50 mm (2″) long fluorocarbon coated capillary
- Detection Method: whole column light absorption at 280 nm, Detection Linear Range > 100
- Focusing voltage: constant voltage, stepwise adjustable from 100 to 600 volts/cm
- Automation: choice of PrinCE MicroInjector or Alcott 719AL Autosampler
- Sample Throughput: up to 7 injections per hour
- Ambient Temperature: 18°C to 23°C
- Relative Humidity: 40% - 80% non-condensing
- Electrical: 115/220 VAC +/- 10%, 50 – 60 Hz +/- 0.5%
- Size: 28.25 cm (11 1/8″) Wide x 60.5 cm (23 3/4″) High x 31 cm (12 1/4″) Deep
  - the iCE280 sits on a stand (14.6 cm or 6 5/8″ high) when used with the PrinCE MicroInjector
- Weight: 20 kg (45 pounds)

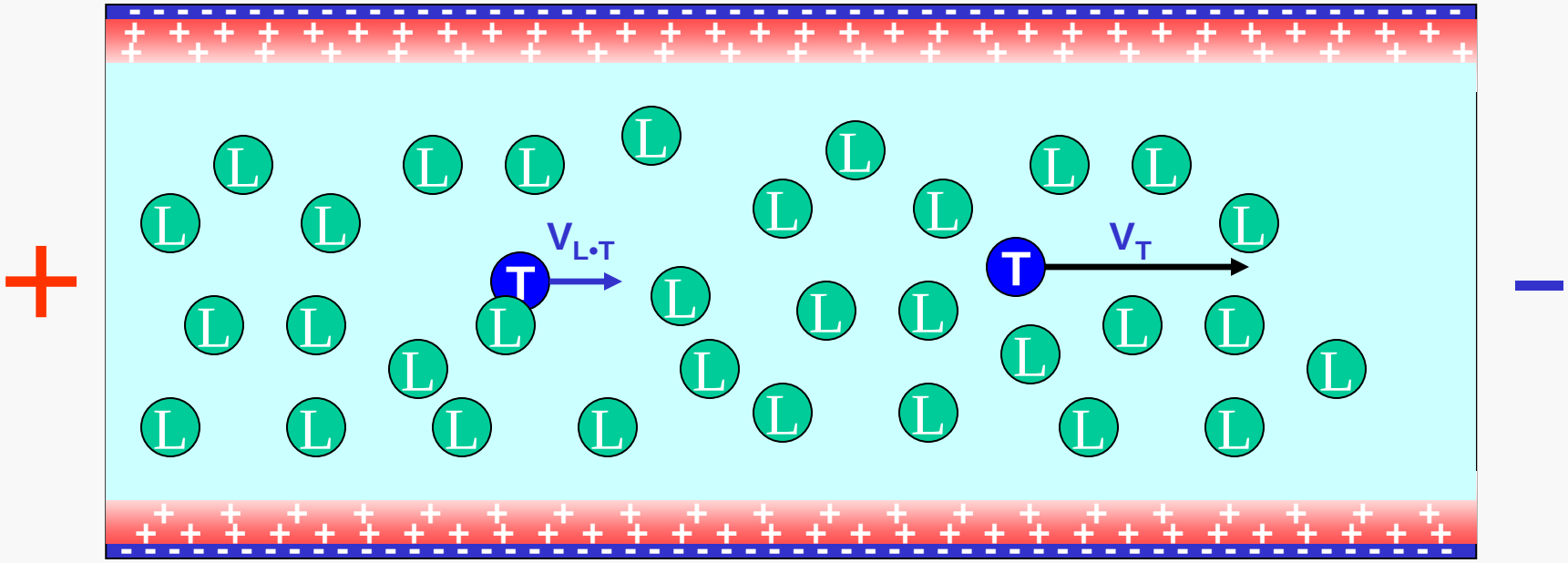
Capillary Gel Electrophoresis (CGE)



Coated capillary does not generate EOF
Sieving matrix allows separation of similar electrophoretic mobilities
Shorter oligo-nucleotides move faster because of less friction

Separates: **Oligo-nucleotides, DNA**
Proteins in SDS-CGE

Affinity Capillary Electrophoresis (ACE)



The target-ligand complex has mobility different from that of the free target

Separates: Everything

Separation of Proteins by CE

- There is no single recipe for separation of proteins. The reason is structural diversity of proteins that leads to variety of physical-chemical properties
- The choice of the right mode and right buffer is critical

Separation of Proteins by CZE

- Major problem: protein interaction with capillary walls results in peak broadening
- Solutions:

1) Extreme buffer pHs that make capillary walls and proteins bare identical charges. High pH (>9, borate buffer): proteins and capillary walls are negatively charged -> electrostatic repulsion. Low pH (< 3, phosphate and acetate buffers): proteins and capillary walls are positively charged -> electrostatic repulsion. Not applicable to proteins with extreme pIs (<3 or >9). Example: Histons require pH 13 to become negatively charged, which can be achieved by using 0.1M NaOH as a run buffer but 0.1 M NaOH generates too much heat.
2) Capillary coating (a hundred of recipes). Very important.

Separation of Proteins by Other Modes

CIEF: different from other techniques since it results to equilibrium

Problems:

- Ampholytes (*ampho*teric electro*lytes*) absorb light
- For detection the focused proteins have to be delivered to the detection window (this stage is called mobilization). Mobilization reduces the resolution.

CGE: SDS-polyacrylamide gel electrophoresis (SDS-PAGE). SDS binds to proteins proportionally to their molecular weight (1.4 g of SDS per 1 g of protein). The charge of the SDS-protein complex is determined by the number of SDS molecules ->

charge/mass ratio is constant

The SDS-protein complexes are separated according to their sizes by electrophoresis through polyacrylamide gels.

In CGE non-cross-linked polyacrylamide are used

Conditions for CZE of Proteins

Table 1. Separation of Proteins by CE[a]

mode[b]	proteins[c]	capillary	buffer	detection	LOD
CZE	proteins from single cells	uncoated	50 mM phosphate + 11 mM pentasulfate pH 6.8	LIF, covalent labeling with FQ	
CZE	model proteins, cell lysate	uncoated	2.5 mM Borax and 5 mM SDS pH 9 50 mM phosphate, 11 mM SPS pH 6.8	LIF, covalent labeling with FQ	0.3 pM
CZE	model protein: Ova	uncoated	2.5 mM borate + 5 mM SDS pH 9.5	LIF, covalent labeling with FQ	0.1 nM, 0.2 fmol
CZE	model proteins: heme-free globin chains	uncoated	50 mM IDA + 6 M urea and 0.5% HEC pH 3.2	UV absorbance	
CZE	model protein: HSA	uncoated	25 mM borate pH 8.9	LIF, noncovalent labeling with indocyanine green	1.4 nM
CZE	whey proteins	methyl deactivated	(i) 50 mM borate pH 9.3; (ii) 120 mM borate pH 9.2	UV absorbance	
CZE	model proteins: α-Lal, β-Lgl A, β-Lgl B, BSA	uncoated	20 mM CHES + 10 mM KCl pH 10.2	LIF, noncovalent labeling with 1,8-ANS, bis-ANS, and 2,6-TNS	0.2−4 μM
CZE	monoclonal IgG + impurities	uncoated	20 mM borate pH 9.3	UV absorbance	
CZE	proteins of cerebrospinal fluid	uncoated	129 mM borate + 0.5% PEG (MW 8 kDa) + 75 mM O-PEA pH 9.25	UV absorbance	
CZE	wheat proteins	uncoated	100 mM phosphate/glycine (Na-free) + 20% CAN + 0.05% HPMC pH 2.5	UV absorbance	
CZE	model proteins: α-Lal, β-Lgl B, CA, HAT, TI	(i) ethylene glycol-coated; (ii) methyl deactivated	50 mM borate pH 9.5	LIF, covalent postcolumn in sheath-flow labeling with OPA	0.7−10 nM; 2−20 amol
CZE	proteins from aqueous humor	polybrene-coated	5% acetic acid + 2 mM ammonium acetate	UV; ESI-MS	
CZE	model proteins: α-Chy, Cyt, Lys, Rnase,	uncoated	50 mM phosphate pH 2.5 + 20 mM CMBCD	UV absorbance	

Conditions for CZE of Proteins contd.

CZE	trancated growth factor	Beckman eCap amine-coated	Beckman eCap 50 mM Tris-HCl pH 8.0	UV absorbance	
CZE	meat proteins	uncoated	20 mM phosphate + 0.1% g/mL anionic dendrimer	UV absorbance	
CZE	model proteins: heme-free globin chains	uncoated	50 mM IDA + 7 M urea and 0.5% HEC pH 3.2 + 2% Tween 20 in some cases	UV absorbance	
CZE	model proteins: β-Lgl A, Cyt, Lys, Myo, Rnase	aminopropylsilane-coated	10 mM acetic acid pH 3.4	UV absorbance ESI-MS	0.6 fmol
CZE	model proteins: CA, HSA, Lys, Myo	Polybrene-coated	5% acetic acid + 2% ammonium acetate	UV absorbance	
CZE	food proteins	uncoated	(i) 5 mM phosphate + 0.01% PDDACl +10 mM sodium octanesulfonate pH 3.7; (ii) 25 mM phosphate + 0.05% DSA pH 7.2	UV absorbance	
CZE	model proteins: α-Chy, Cyt, α-Lal, β-Lgl A, β-Lgl B, Lys, Myo, Ova, Rnase, TI	polystyrene–ethylenediamine diol-coated	40 mM phosphate pH 3.1	UV absorbance	
CZE	protein extract from krill	uncoated	50 mM phosphate + 149 μg/mL FC128 + 1 μg/mL FC134 pH 5.5	UV absorbance	
CZE	proteins of rat airway surface fluid	uncoated	(i) 100 mM borate pH 9.1; (ii) 100 mM phosphate + 0.5 mM spermine pH 2.5	UV absorbance	6 μM
CZE	model proteins: Cyt, Myo, BI	uncoated	1% acetic acid	ESI-TOF-MS	
CZE	human plasma proteins	uncoated	borate	UV absorbance	
CZE	human plasma proteins	uncoated	borate	UV absorbance	

Conditions for CIEF and SDS-CGE of Proteins

mode[b]	proteins[c]	capillary	buffer	detection	LOD
ITP−CZE	model proteins: Cyt, Lys A, Lys B, Rnase, rhIL-3, rhIL-6	75 μm × 50 cm, coated neutrally	20 mM triethylamine/ acetate pH 4.2	UV absorbance	25−50 nM
CIEF	model proteins: Cyt, Myo, Per	uncoated	Pharmalyte 3−10 ampholyte + HPMC	chemiluminescence	6 nM
CIEF	complexes of *scr* SH2 domain with six peptides.	PVA-coated	2% carrier ampholyte (Ampholine 3.5−10)	UV absorbance, ESI-ITMS	
CIEF	phosphorylated forms of Ova	linear PA coated	0.2% carrier ampholyte (Pharmalyte 4−6.5)	UV absorbance, ESI-MS	
CIEF	six isoforms of monoclonal antibody	PA- or fluorocarbon-coated	4% carrier ampholyte + 4 M urea	UV absorbance, whole capillary imaging	
CIEF	scrapie prion protein	neutrally coated	carrier ampholyte 3−10	UV absorbance	
SDS−CGE	model proteins: α-Lal, β-Gal, BSA, Cal, Ova, Pep	(i) uncoated; (ii) e-CAP-coated; (iii) uncoated Borofloat glass chip	SDS-14−200 gel buffer (Beckman)	LIF, covalent labeling with FMAL and FNHS	
SDS−CGE	human plasma proteins	non-cross-linked PA coated	0.05 M Tris + 0.05 M tricine + 0.1% SDS + 4% linear PA pH 8.3	UV absorbance	
SDS−CGE	model proteins: BSA, CA, Con, Ova	uncoated with very viscous 8% linear polyacrylamide	0.1 M Tris + 0.25 M borate + 0.05% SDS + 8% linear PA pH 8.1	LIF, noncovalent labeling with Cypro Red	1.5 nM
SDS−CGE	β-trace protein from cerebrospinal fluid	eCAP kit (Beckman)	eCAP kit (Beckman)	UV absorbance	
SDS−CGE	monoclonal antibody + impurities	uncoated	Bio-Rad SDS running bufer	LIF, covalent labeling with TMR−NHS	
SDS−CGE	α-amylase and glucoamylase from sake rice koji	uncoated	Bio-Rad SDS running bufer	UV absorbance	
SDS−CGE	lysates from different bacteria	uncoated	Bio-Rad SDS running bufer	UV absorbance	

The diversity of separation regimes is much greater that in chromatography

Separation of Nucleic Acids

- Due to structural similarity of nucleic acids there are good "recipes" for nucleic acids separation
- The mode of choice is capillary gel electrophoresis
- Depending on the gel, nucleic acids of different lengths can be separated.
- Gels must NOT be too viscous so that the capillaries could be refilled with the gel after every run.

Gels used for CGE of nucleic acids:

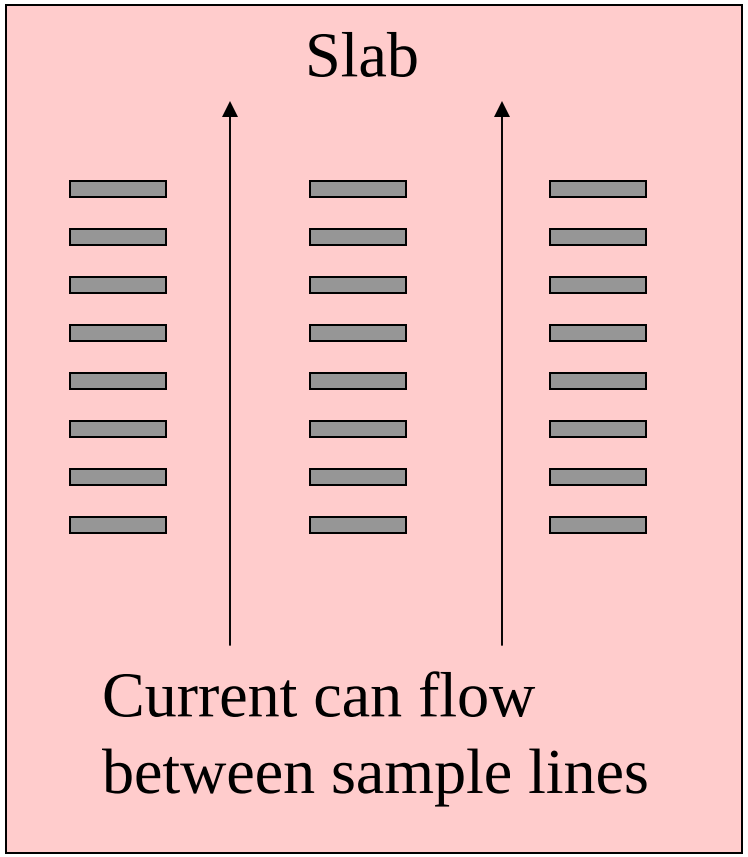
- High molecular mass polyacrylamide exhibits the best performance (>1000 bases are separated in 2 h). The static viscosity is high but once pumped into a capillary, the viscosity dramatically decreases.
- Dimethylacrylamide (commercial instruments)
- Tens of other polymers

Importance of Sample Purity

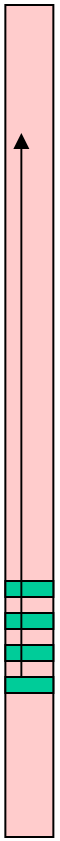
Impurities, e.g. proteins and salts result in low quality separations.

The reason is the anti-stacking effect: the sample of a higher conductivity than the run buffer defocuses

The effect of impurities is more severe in slab gel electrophoresis as the current flows not only through the sample



Capillary



Current always flows through the sample

DNA- and RNA-binding proteins can be used instead of gels

DNA- and RNA-binding Proteins

Nature

DNA replication (e.g. DNA polymerases, methylases)
DNA damage control (e.g. Mut S ,Mut L , Mut H)
DNA repair (e.g. DNA photolyase, Polymerase β, Rec A, Rec F, Rec B)
Gene expression (e.g. transcription factors, RNA polymerases)
Defense from viral infections (e.g. nucleases)

Technology

PCR (e.g. DNA polymerases, ssDNA-binding protein)
RT-PCR (e.g. reverse transcriptases, polymerases)
Cloning (e.g. restriction enzymes, ligases)
Sensors (e.g. Mut S)

DNA- and RNA-binding are an analytical tool in Separation Sciences

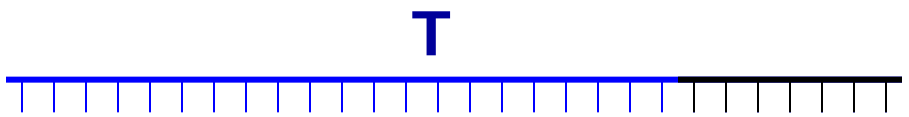
DNA-Binding Proteins as Mediators of Gel-Free Hybridization Analyses of DNA and RNA

Hybridization analyses such as Southern and Northern Blotting are the major tool for quantitation of DNA and RNA (another one is quantitative PCR)

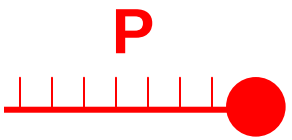
DNA (RNA) Hybridization

Hybridization is affinity interaction of DNA-DNA or DNA-RNA, which can facilitate specific detection DNA or RNA

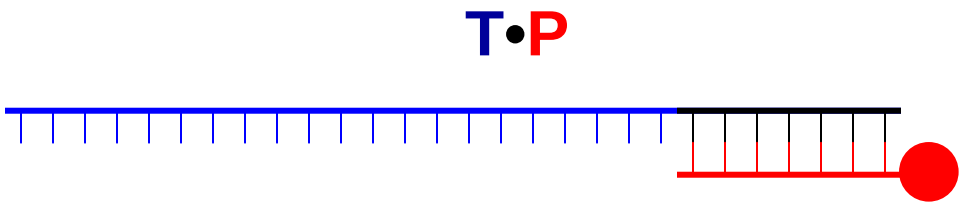
Target (T) is a ssDNA or ssRNA to be detected; it is typically either a gene or its RNA complement (mRNA)



Hybridization Probe (**P**) is a short strand of DNA (18-22 bases) complementary to a part of T and typically labeled for detection

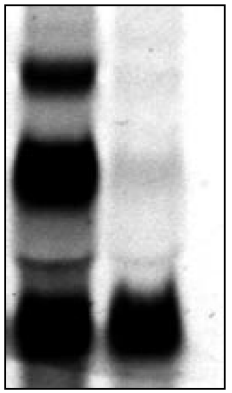


Hybridization complex (**T•P**) is formed by annealing T and P at high temperature (50-60 °C) to prevent non-specific hybridization



Classical Hybridization Analyses:
Southern blotting for DNA and Northern blotting for mRNA

Step 1: Slab-gel electrophoresis separation of DNAs (or RNAs) of different sizes (1-6 hours)
Step 2: Transfer of DNA to a nylon membrane (overnight)
Step 3: Cross-linking of DNA by UV (1 hour)
Step 4: Hybridization with the excess of a radioactively-labeled probe (overnight)
Step 5: Washing out the excess of the probe (1 hour)
Step 6: Detection of T•P with by radiography (overnight)



Limitations

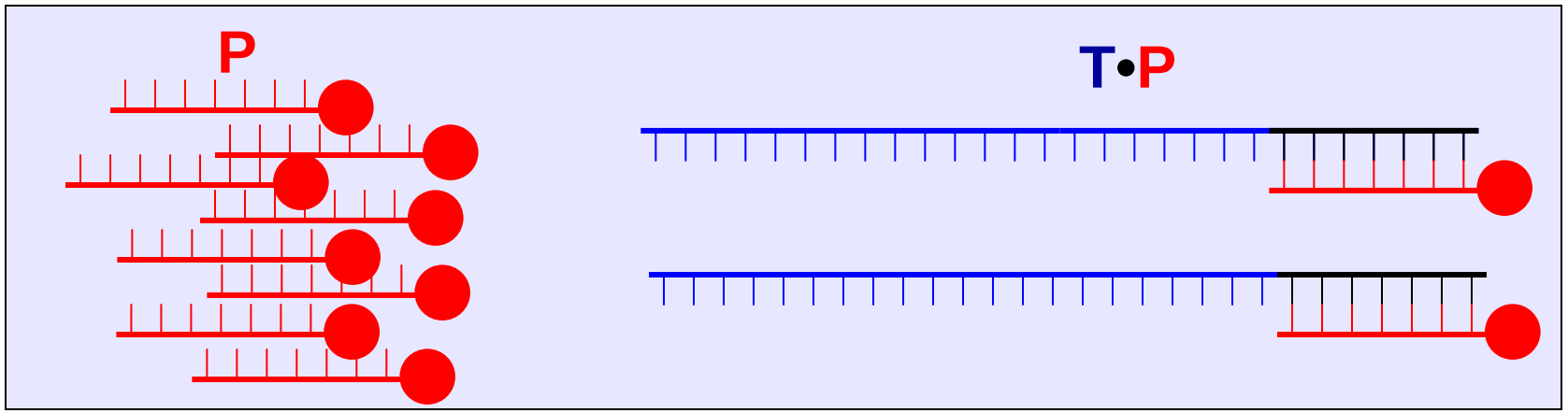
1. Time consumption (1 week)
2. Low sensitivity (requires 10-50 μg of DNA or RNA)
3. Radioactive hazard (uses ^{32}P)
4. Semi-quantitative (requires calibration with known amounts of T)

"Ideal" Hybridization Analysis by Gel-Free Capillary Electrophoresis (CE)

Step 1 (1 h): Reacting **T** with excess of fluorescently labeled **P** *in solution*:

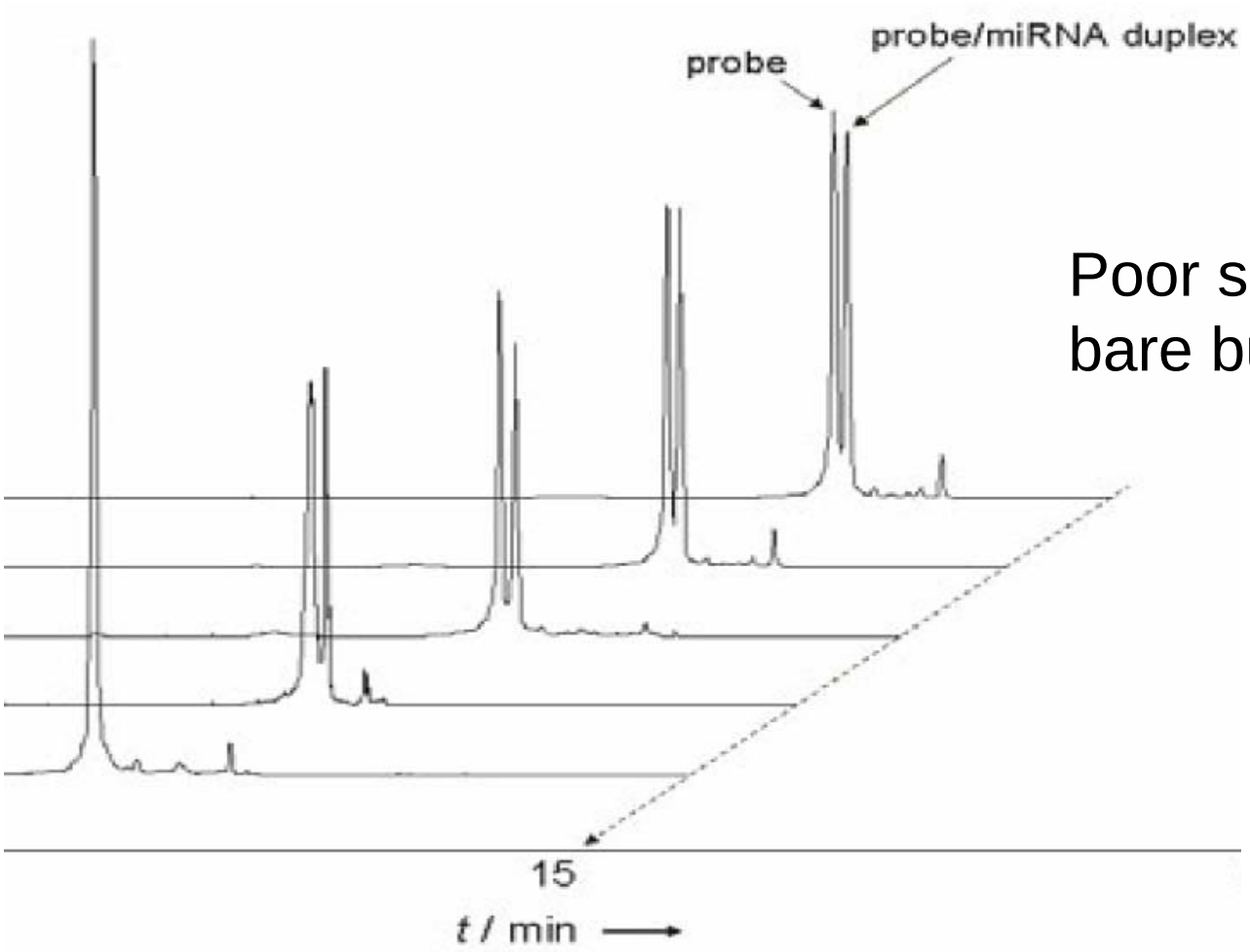
$$\mathbf{T + P(excess) \longrightarrow T{\bullet}P + P(excess)}$$

Step 2 (10 min): Separating **T•P** from excess of **P** **by gel-free CE**



Step 3 (0 min): Detecting **T•P** by on-line fluorescence detection

How to arrange gel-free CE separation of T•P from the excess of P?



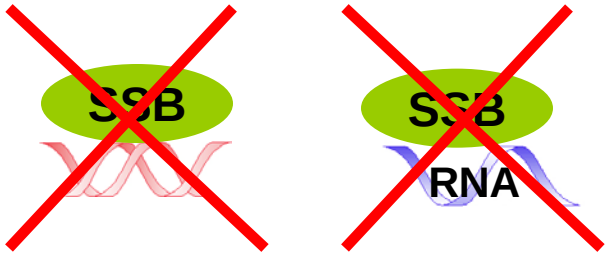
Poor separation in a bare buffer

Gel could help but lets try to use DNA-Binding Proteins Instead of Gel

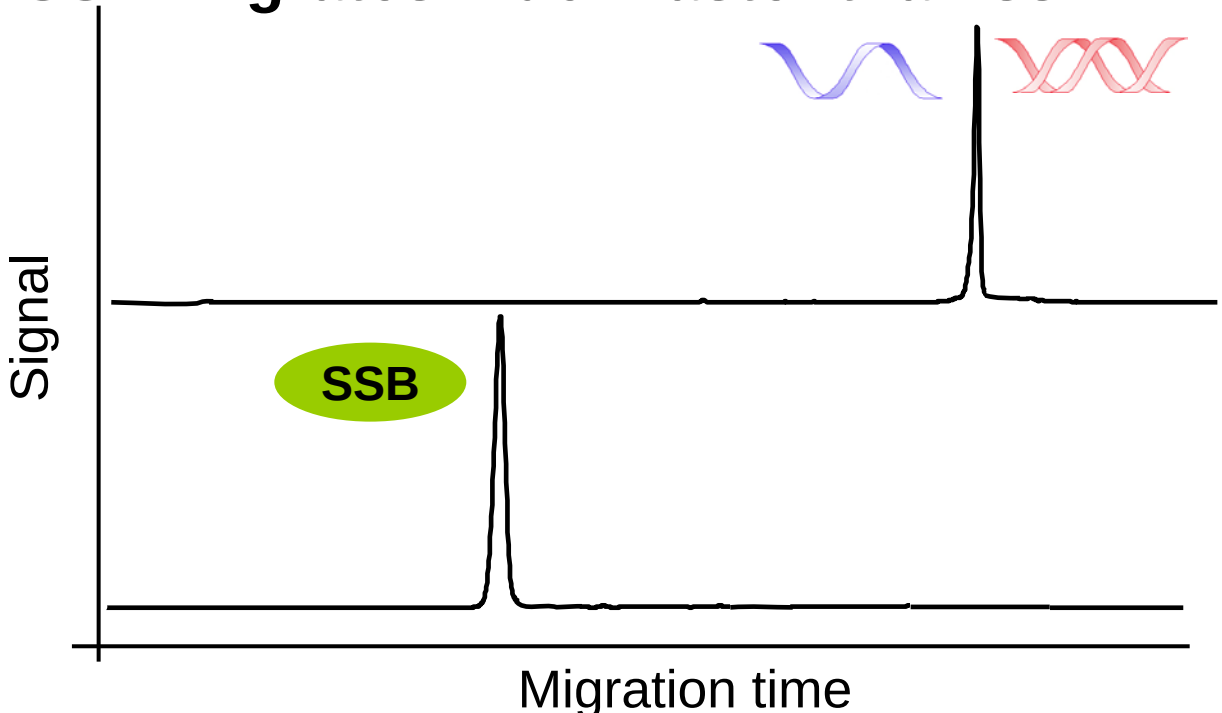
ssDNA-Binding Protein (SSB)

SSB binds ssDNA sequence-non-specifically

SSB does not bind dsDNA or RNA
or double-stranded DNA-RNA hybrids



SSB migrates much faster than ssDNA and dsDNA



Note: dsDNA is not separated from ssDNA without in a gel-free media

JACS 2003, 125, 13451

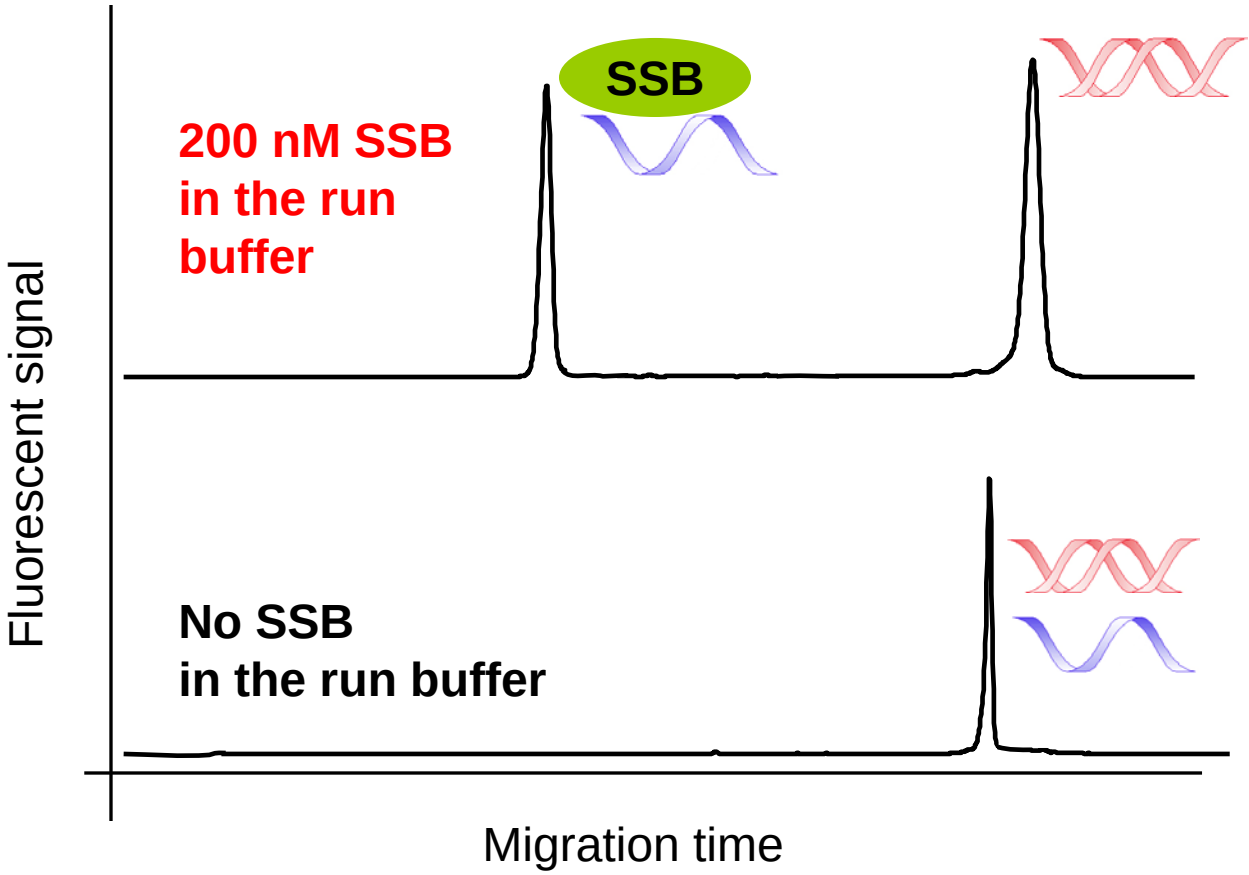
SSB facilitates gel-free separation of ssDNA from dsDNA if added to the run buffer

SSB ssDNA-binding protein

dsDNA

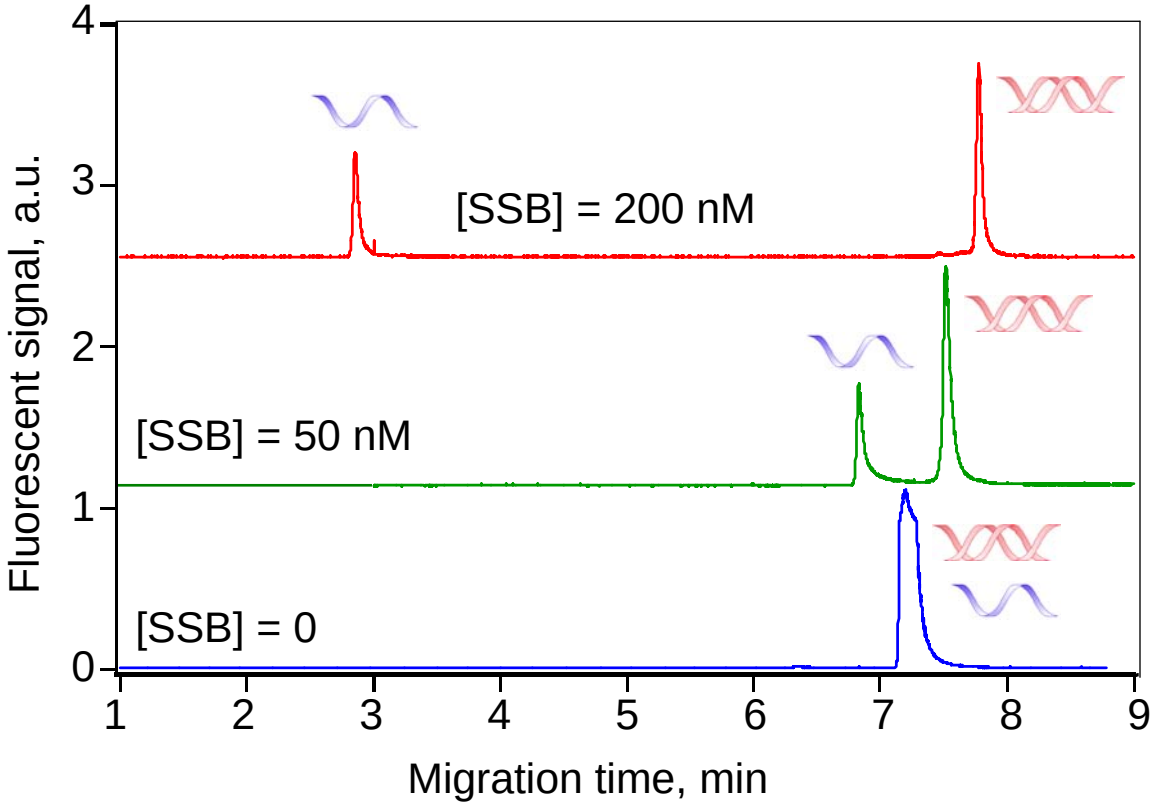
ssDNA

Note: DNA is fluorescently labeled



JACS 2003, 125, 13451

Separation Quality Depends on the Concentration of SSB



JACS 2003, 125, 13451

Example: Analysis of mRNA

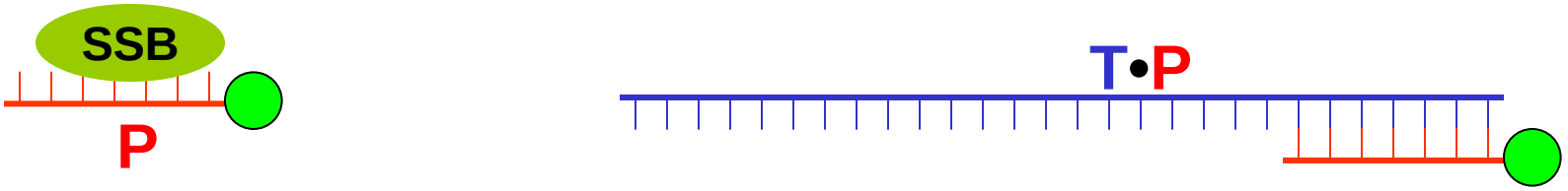
Target: mRNA of GFP (~750 bases long)
Probe: ssDNA (20 bases) labeled with fluorescein

Step 1 (1 h): Reacting **T** with excess of fluorescently labeled **P** *in solution*:

$$\mathbf{T} + \mathbf{P(excess)} \longrightarrow \mathbf{T{\bullet}P} + \mathbf{P(excess)}$$

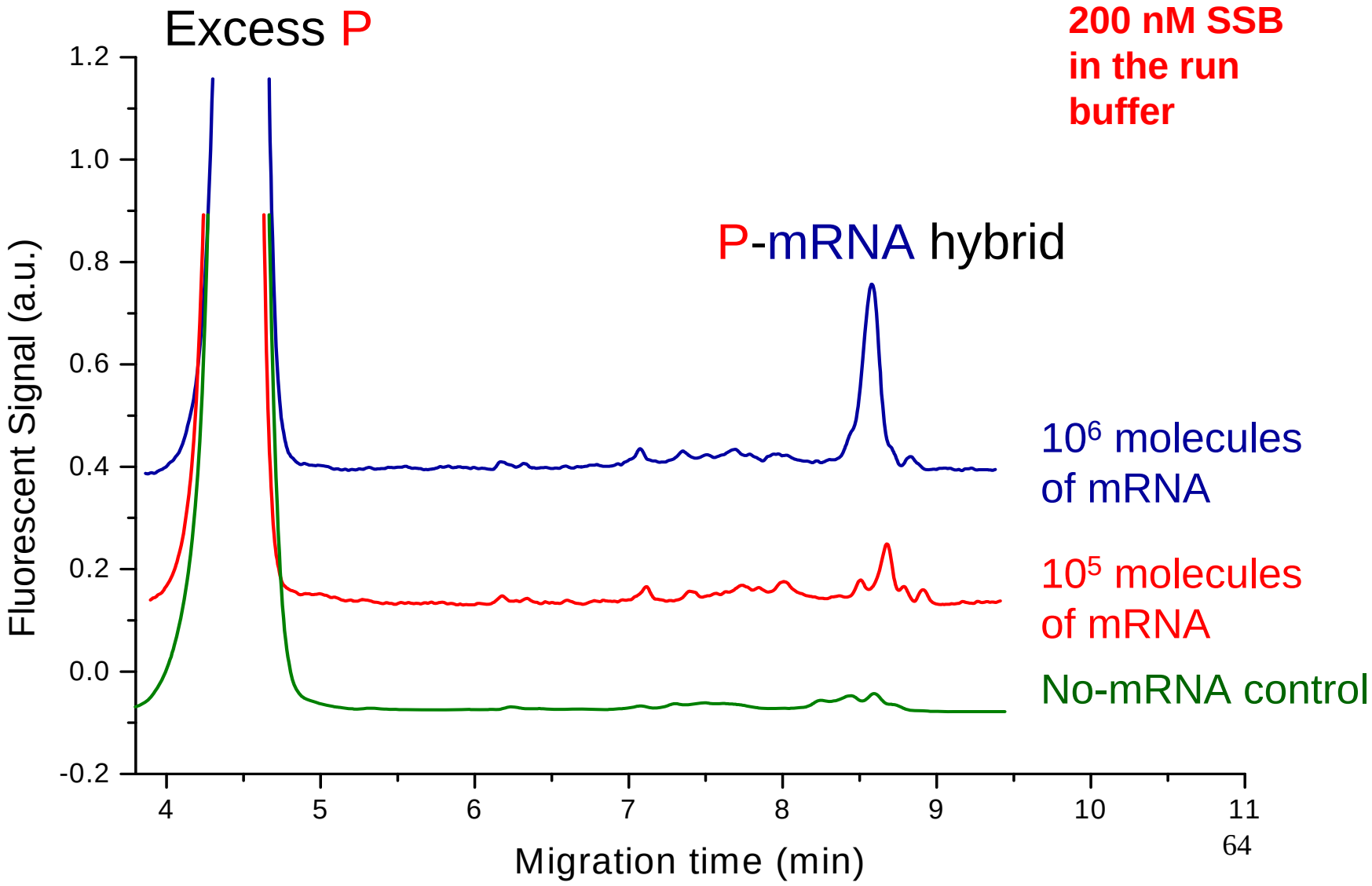
Step 2 (10 min): Separating **T•P** from excess of **P** by SSB-mediated CE.

Note: SSB does not bind to a long ssRNA overhang:

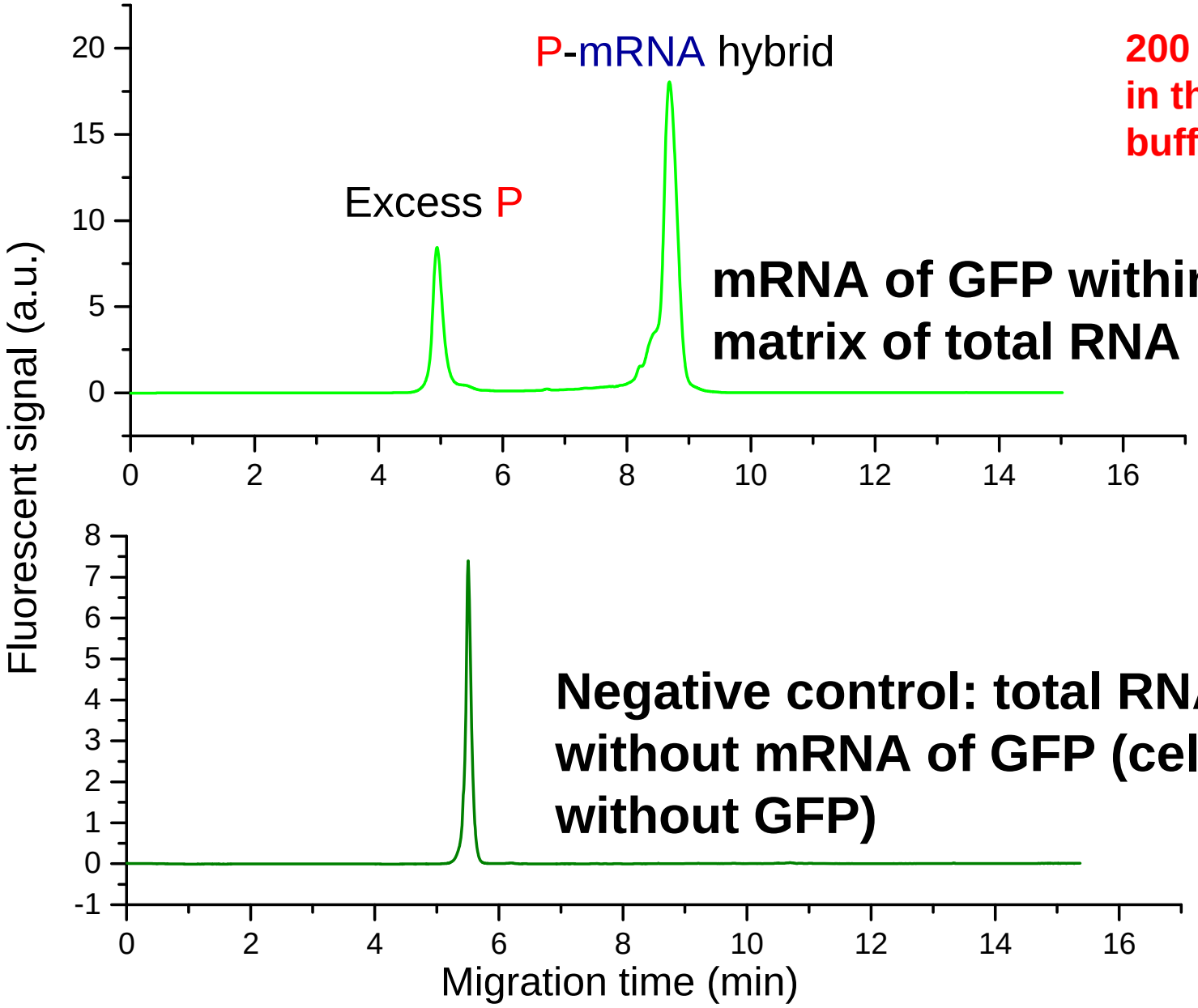


Step 3 (0 min): Detecting **T•P** by on-line fluorescence detection

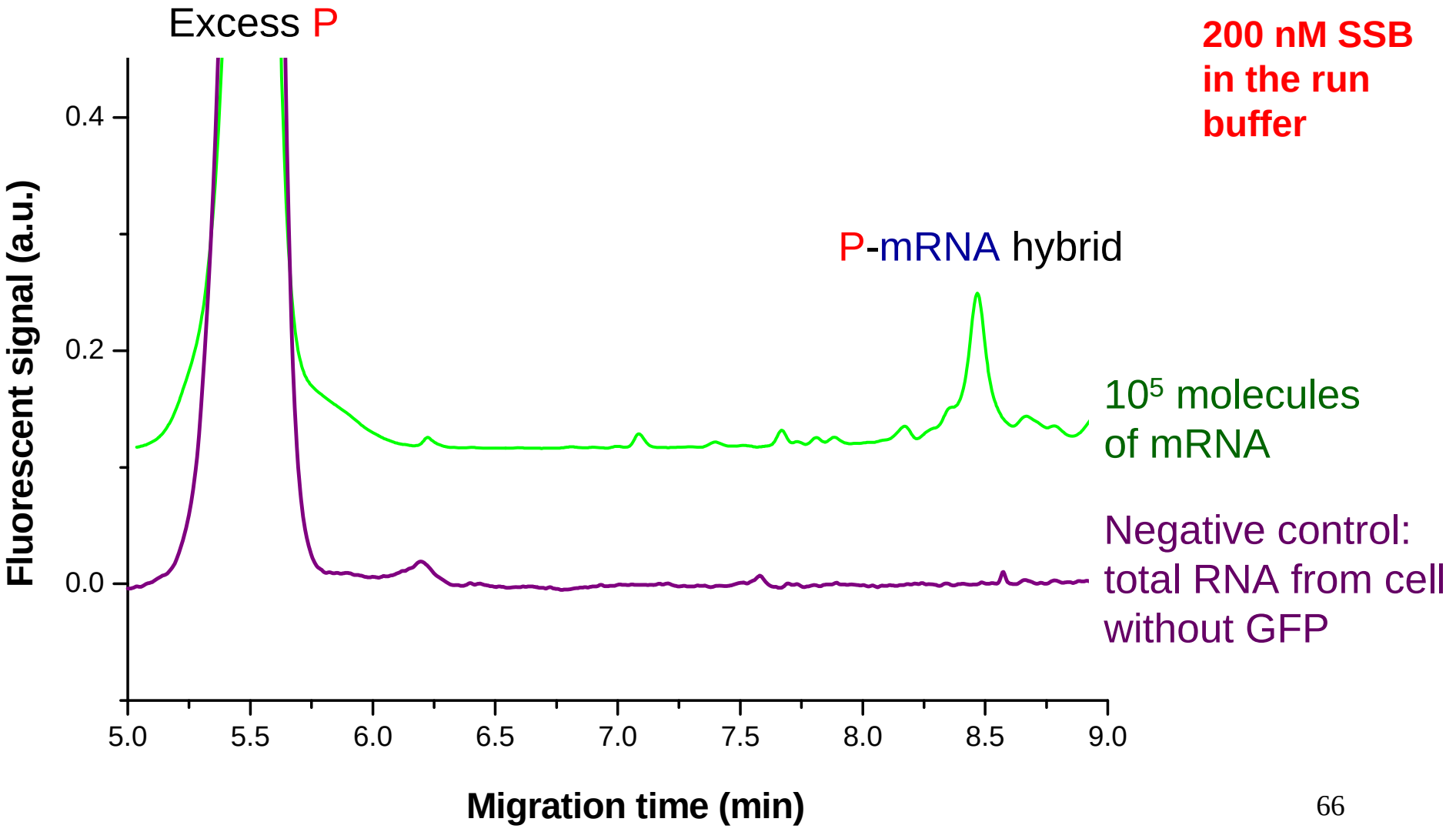
Analysis of Pure GFP mRNA (no other RNA)



Analysis of GFP mRNA in the Matrix of Total RNA



Analysis of GFP mRNA in the Matrix of Total RNA: Limit of Detection



Single-Nucleotide Polymorphisms (SNPs)

Single-nucleotide polymorphism (SNP) is an abundant form of genomic variations, which involves the variability of a single base pair with a frequency of the least copious allele of no less than 1% of the entire population. SNPs have been found to cause health disorders and drug resistance. The major approach to the analysis of SNPs is based on the detection of mismatches. The DNA in question is mixed with a reference allele, melted, and re-annealed. If the SNP is present, the re-annealed DNA will contain single-nucleotide mismatches, which can be detected by a number of methods. None of the published mismatch-based methods, however, can recognize the mismatches.

Example of Mismatches for an A-T Reference Sequence

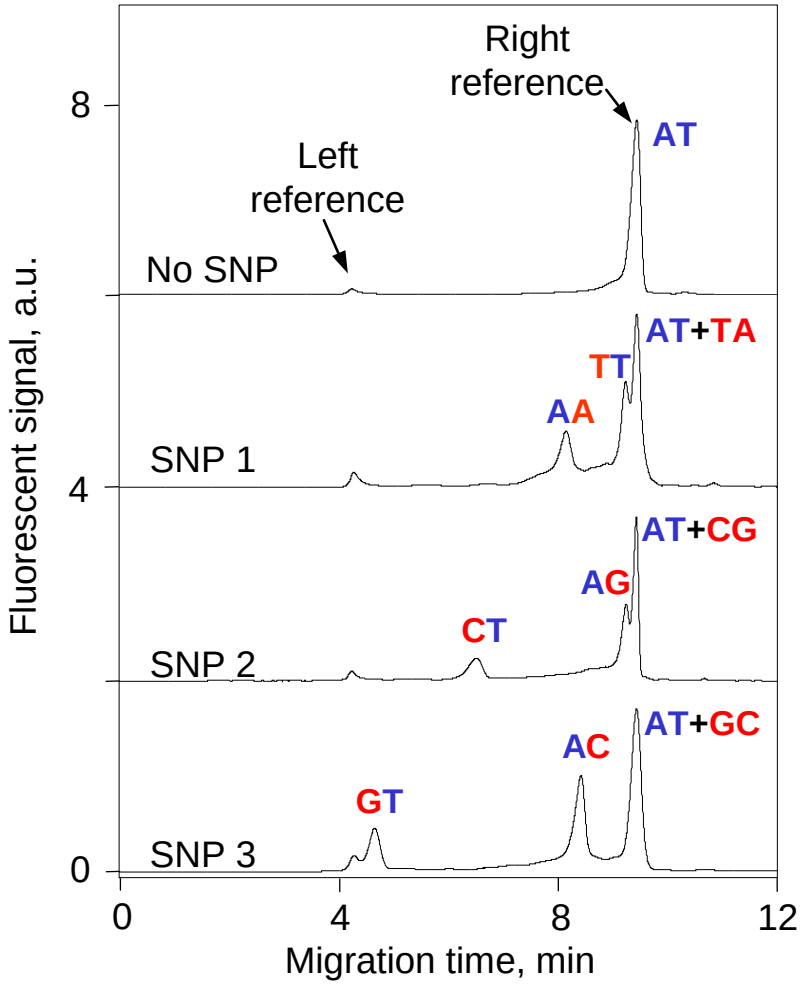
SNP status	Sequence mixed with the reference	Sequences after mixing, melting, and re-annealing of the sequence in question with the reference sequence			
		Matching sequences		Mismatching sequences	
No SNP	5′—A—3′ 3′—T—5′	5′—A—3′ 3′—T—5′			
SNP 1	5′—T—3′ 3′—A—5′	5′—A—3′ 3′—T—5′	5′—T—3′ 3′—A—5′	5′—A—3′ 3′—A—5′	5′—T—3′ 3′—T—5′
SNP 2	5′—C—3′ 3′—G—5′	5′—A—3′ 3′—T—5′	5′—C—3′ 3′—G—5′	5′—A—3′ 3′—G—5′	5′—C—3′ 3′—T—5′
SNP 3	5′—G—3′ 3′—C—5′	5′—A—3′ 3′—T—5′	5′—G—3′ 3′—C—5′	5′—A—3′ 3′—C—5′	5′—G—3′ 3′—T—5′

Conventional methods do not distinguish between SNP 1, SNP2, and SNP 3

MutS protein

MutS protein binds single-nucleotide mismatches with different affinities.

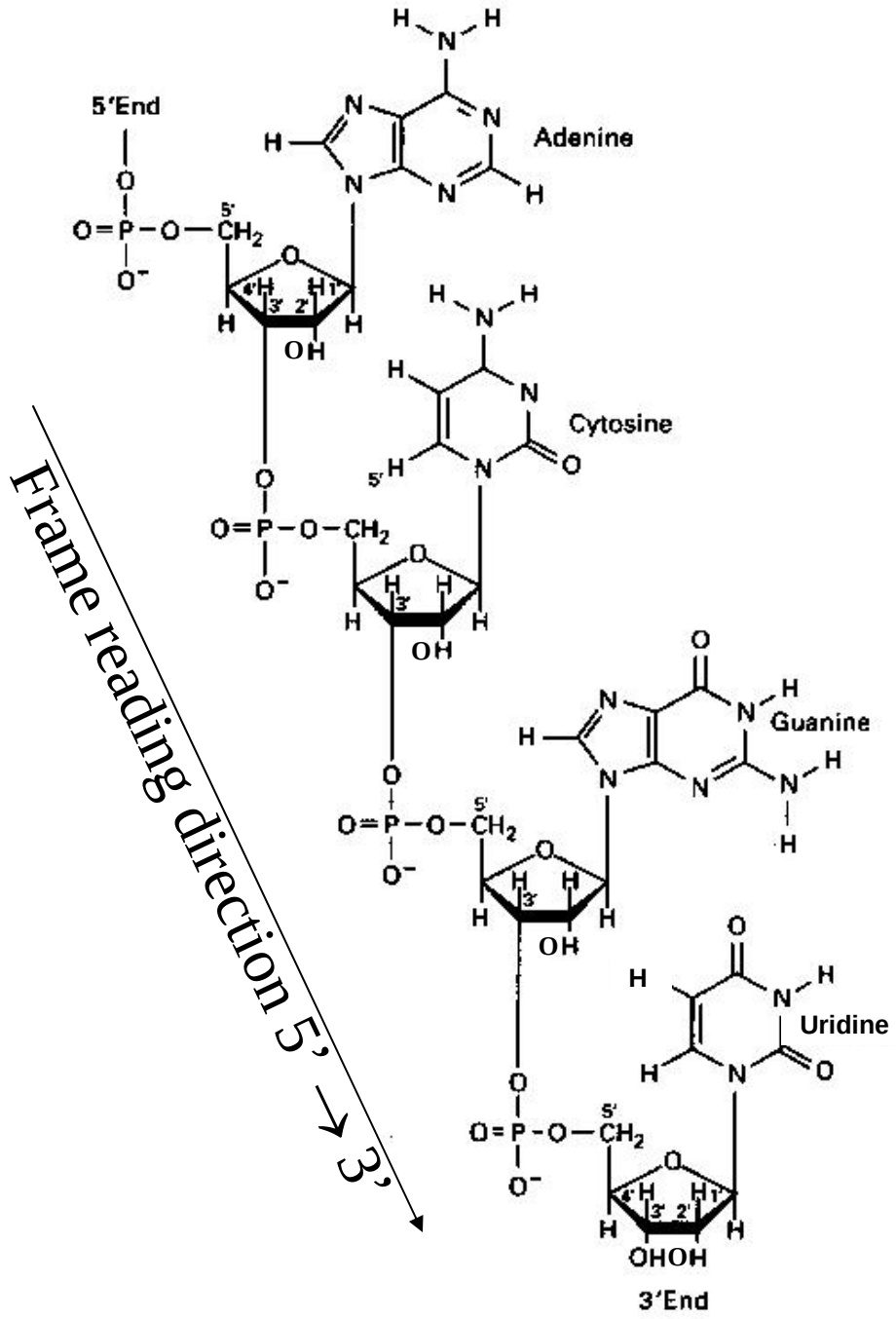
Can it be used for base-pair-selective analysis of SNPs?



SNP status	Sequence mixed with the reference	Sequences after mixing, melting, and re-annealing of the sequence in question with the reference sequence	
		Matching sequences	Mismatching sequences
No SNP	5′—A—3′ 3′—T—5′	5′—A—3′ 3′—T—5′	
SNP 1	5′—T—3′ 3′—A—5′	5′—A—3′ 3′—T—5′ 5′—T—3′ 3′—A—5′	5′—A—3′ 3′—A—5′ 5′—T—3′ 3′—T—5′
SNP 2	5′—C—3′ 3′—G—5′	5′—A—3′ 3′—T—5′ 5′—C—3′ 3′—G—5′	5′—A—3′ 3′—G—5′ 5′—C—3′ 3′—T—5′
SNP 3	5′—G—3′ 3′—C—5′	5′—A—3′ 3′—T—5′ 5′—G—3′ 3′—C—5′	5′—A—3′ 3′—C—5′ 5′—G—3′ 3′—T—5′

DNA Sequencing – Approach to Genome Discovery

Genetic Code – RNA to Protein



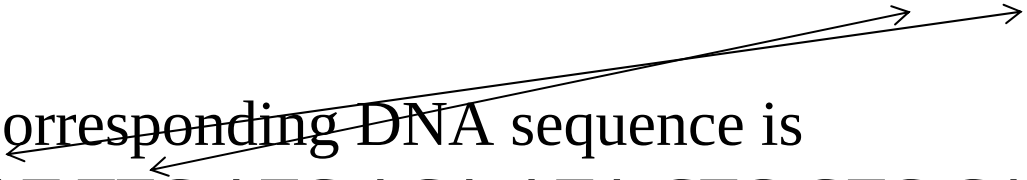
RNA sequence

Genetic Code – 3 bases are needed to encode every one of 20 amino acids (number of combinations = 4^N, $4^2 = 16$ is not enough, $4^3 = 64$ is more than we need)

First Letter (5' end)	U	C	A	G	Third Letter (3' end)
		Second letter			
U	UUU-phe	UCU-ser	UAU-tyr	UGU-cys	**U**
U	UUC-phe	UCC-ser	UAC-tyr	UGC-cys	**C**
U	UUA-leu	UCA-ser	**UAA**-stop	**UGA**-stop	**A**
U	UUG-leu	UCG-ser	**UAG**-stop	**UGG**-trp	**G**
C	CUU-leu	CCU-pro	CAU-hys	CGU-arg	**U**
C	CUC-leu	CCC-pro	CAC-hys	CGC-arg	**C**
C	CUA-leu	CCA-pro	CAA-gln	CGA-arg	**A**
C	CUG-leu	CCG-pro	CAG-gln	CGG-arg	**G**
A	AUU-ile	ACU-thr	AAU-asn	AGU-ser	**U**
A	AUC-ile	ACC-thr	AAC-asn	AGC-ser	**C**
A	AUA-ile	ACA-thr	AAA-lys	AGA-arg	**A**
A	**AUG-met** ← **Also the start codon**	ACG-thr	AAG-lys	AGG-arg	**G**
G	GUU-val	GCU-ala	GAU-asp	GGU-gly	**U**
G	GUC-val	GCC-ala	GAC-asp	GGC-gly	**C**
G	GUA-val	GCA-ala	GAA-glu	GGA-gly	**A**
G	GUG-val	GCG-ala	GAG-glu	GGG-gly	**G**

Example: What might be the genetic sequence for the peptide met-his-gln-tyr-cys-asp-glu-met?

The RNA sequence could be:
5’-AUG CAC CAG UAU UGU GAU GAA AUG



The corresponding DNA sequence is
5’-CAT TTC ATC ACA ATA CTG GTG CAT

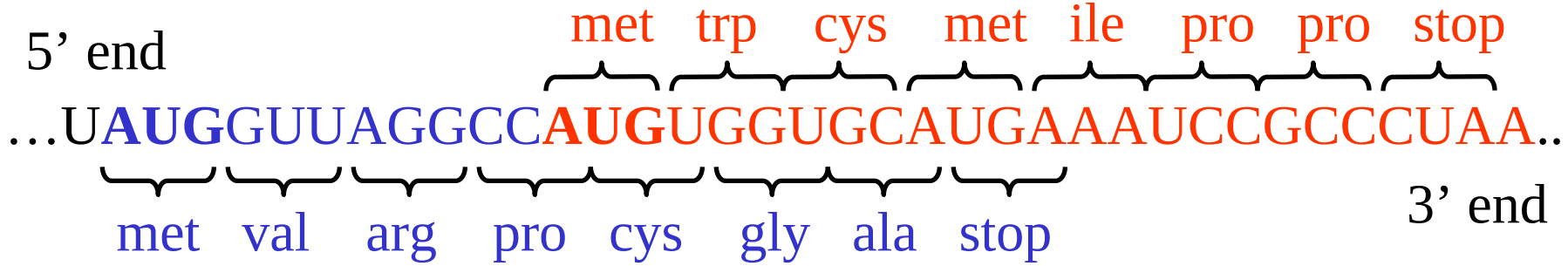
Remember that complementary strands are generated in the opposite direction

Any portion of the genome that contains this sequence is likely to contain the protein in which this peptide is present.
Note: There are other DNA sequences that could also code for this peptide.

Condensed Information

If the start codon AUG appears in several places in the gene but out of alignment, then several peptides can be coded from the same gene. Viruses are in particular known for condensing much information in a small amount of DNA. The hepatitis B genome is highly compact, with certain regions coding for three different peptides.

Example:



Primers and Probes

Short regions of DNA that are complimentary to a single strand of DNA are called primers or probes. The primers are used in DNA sequencing and in the polymerase chain reaction. Probes are used in DNA fingerprinting and in Southern blotting

If a double stranded DNA is heated, it will denature, forming two single strands of DNA. These single strands can bind to a primer or probe with a complementary sequence. Probes and primers can be radioactively, fluorescently or chemiluminescently labeled for visualization of their complex with loner pieces of DNA. The probe is valuable if it is complementary to only one location of the genome. On average (statistically), an n-base sequence will be found every 4^n bases in the genome.

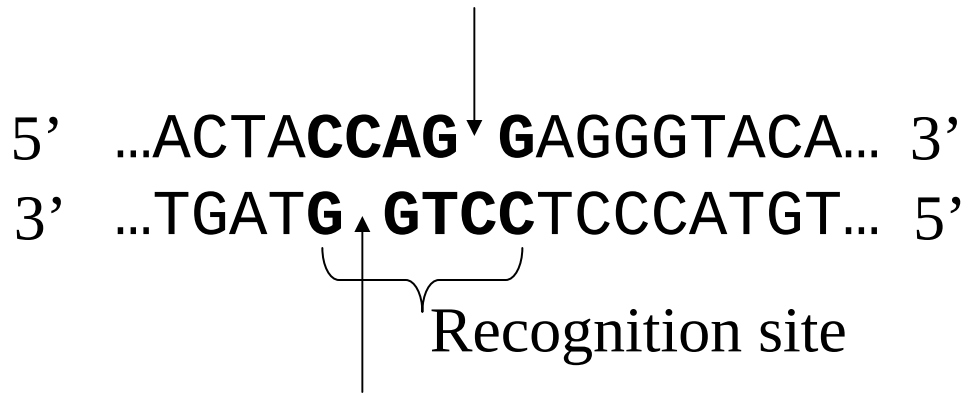
Example: The human genome has 3×10^9 bases of DNA. If the sequence of nucleotides is random, which is not, then an oligonucleotide of 16 bases ($4^{16} = 4.3 \times 10^9$) will have a ~50:50 chance of being unique in the genome. An 18-mer ($4^{18} = 6.9 \times 10^{10}$) is ~95% likely to be unique, and a 22-mer ($4^{22} = 1.8 \times 10^{13}$) is ~99.98% likely to be unique.

Home assignment: prove that percentages are correct

Restriction Enzymes

Restriction endonucleases are enzymes that cut dsDNA at specific locations. They are found in bacteria, where they serve as a defense mechanism. When the bacteria is infected by a virus, the restriction enzyme can cut the viral DNA preventing it from functioning within the bacteria. These enzymes are very valuable in molecular biology. More than 3,000 of them have been studied and more than 600 are commercially available.

Example: the enzyme *Ava* II cuts dsDNA that has the sequence 3'GGTCC5' (and 3'GGACC5' on the complementary strand). The enzyme does not cut the DNA at the center but instead cuts the DNA between the two G's:



Into how many pieces of DNA will AVA II cut the human genome?

Restriction Enzymes Contd.

After the DNA is cut, *Ava* II leaves a three-base piece of ssDNA at the 3' end of the cut. This piece of DNA is called **overhang** or a **sticky end**:

5'…ACTA**CCAG** 3'
3'…TGAT**G** 5'

+

5' **G**AGGGTACA… 3'
3' **GTCC**TCCCATGT… 5'

A short oligonucleotide can be **ligated** onto the sticky end through the use of ligase enzyme. The oligonucleotide can contain a primer site or a fluorescent label for analytical uses.

5'…ACTA**CCAG** 3'
3'…TGAT**G** 5'

+

3' **GTCC**ATGT… 5'

⟶

5'…ACTA**CCAG** 3'
3'…TGAT**GGTCC**ATGT… 5'

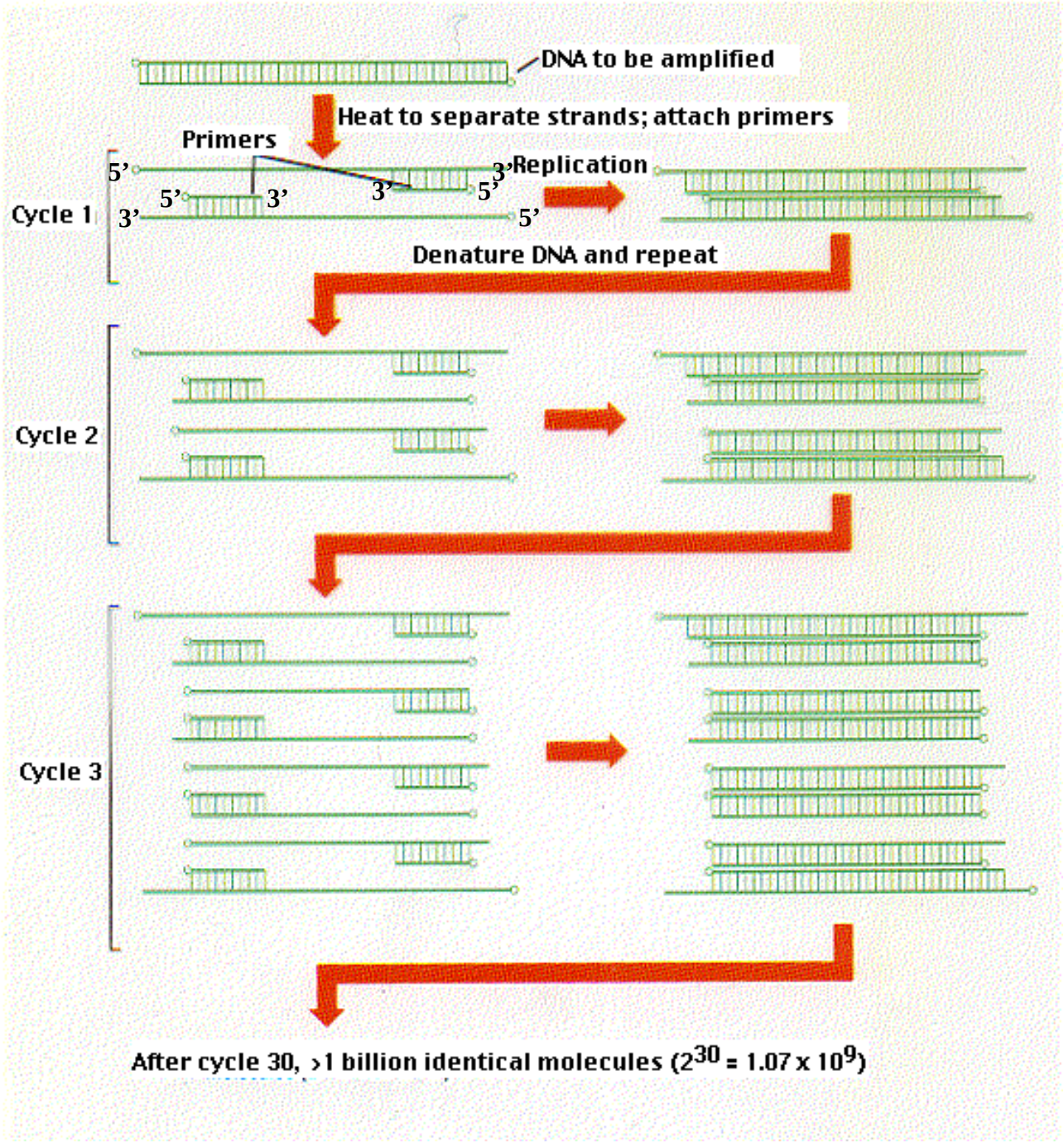
There are restriction enzymes that leave blunt ends with no overhang, such as *Rsa* I:

5' …ACTA**GT** ↓ **AC**ACCTC… 3'
3' …TGAT**CA** ↑ **TG**TGGAG… 5'

Polymerase Chain Reaction

- In order to perform PCR, you must know at least a portion of the sequence of the DNA molecule that you wish to replicate.
- You must then synthesize **primers**: short oligonucleotides (containing about two dozen nucleotides) that are precisely complementary to the sequence at the 3' end of each strand of the DNA you wish to amplify.
- The DNA sample is heated to separate its strands and mixed with the primers.
- If the primers find their complementary sequences in the DNA, they bind to them.
- Synthesis begins (always 5' -> 3') using the original strand as the template.
- The reaction mixture must contain
    - all four deoxynucleotide triphosphates (dATP, dCTP, dGTP, dTTP)
    - a DNA polymerase. DNA polymerase should be stable to high temperature needed to separate the DNA strands.
- Polymerization continues until each newly-synthesized strand has proceeded far enough to contain the site recognized by the **other primer**.
- Now you have two DNA molecules identical to the original molecule.
- You take these two molecules, heat them to separate their strands, and repeat the process.
- Each cycle doubles the number of DNA molecules.

Using automated equipment, each cycle of replication can be completed in less than 5 minutes. After 30 cycles, what began as a single molecule of DNA has been amplified into more than a billion copies ($2^{30} = 1.02 \times 10^9$).



Polymerase Chain Reaction

Taq polymerase. *Taq* is a nickname for *Thermus aquaticus* a bacterium that lives in hot springs

Synthesis proceeds in the 5' → 3' direction

Sequencing of the Genomic DNA

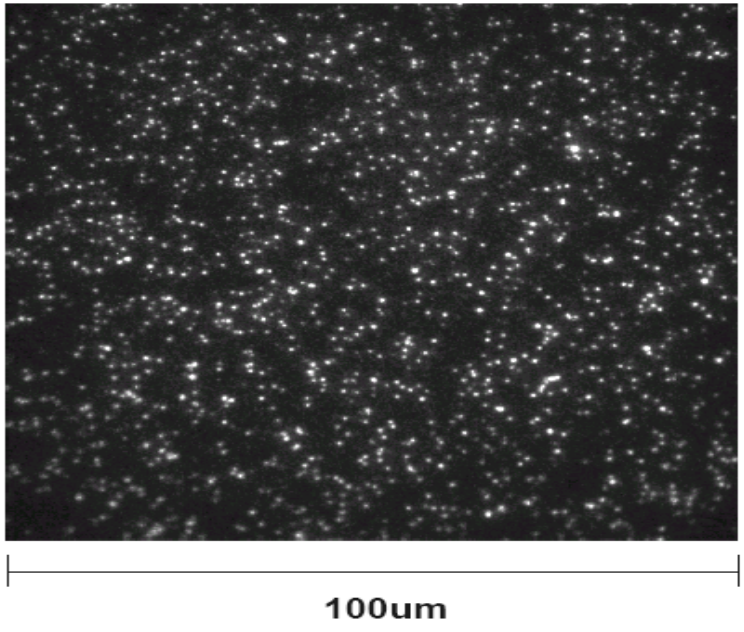
1. Isolation of chromosomes by flow cytometry
2. The chromosome is randomly disrupted and cut into shorter pieces
3. The shorter pieces of DNA are then sequenced
4. Computer screening of sequences allows to find neighboring pieces

Sequencing-by-synthesis of shorter pieces

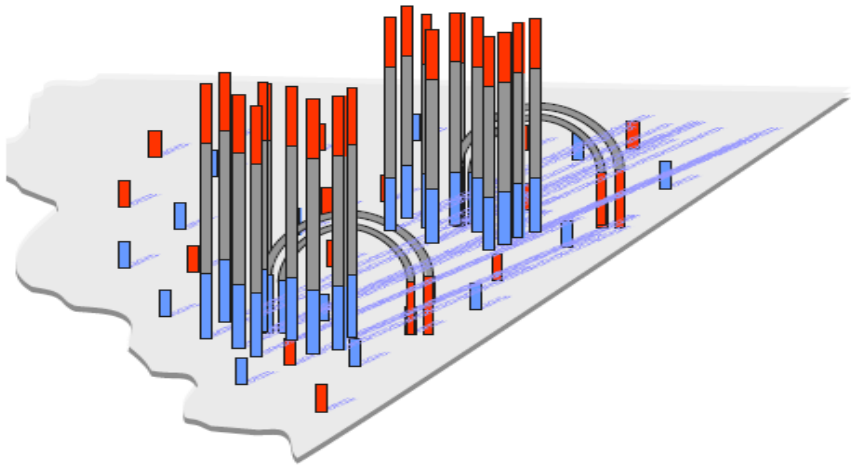
http://www.illumina.com/media.ilmn?Title=Sequencing-By-Synthesis%20Demo&Cap=&PageName=solexa%20technology&PageURL=203&Media=1

Sequencing-by-Synthesis

Next generation sequencing on clonal arrays



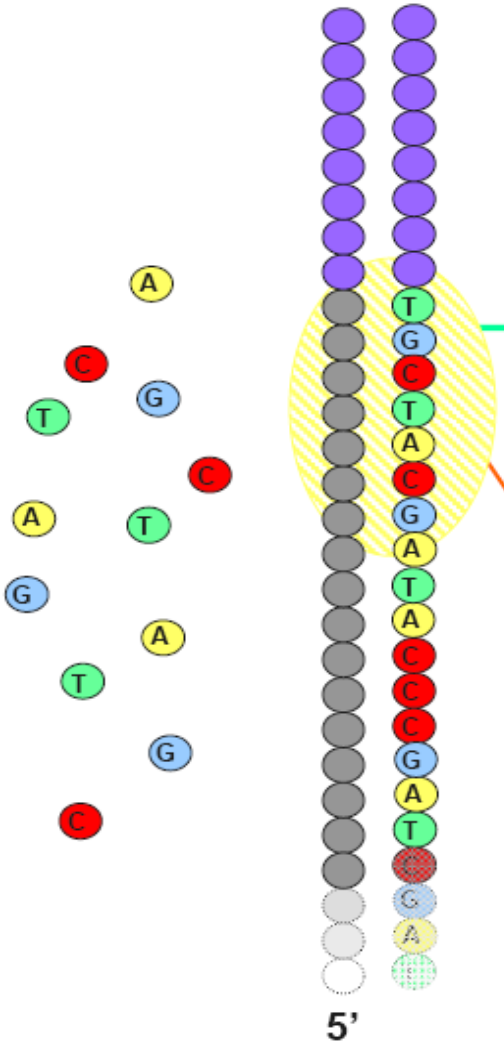
Attach single molecules to surface
Amplify to form clusters



~1000 molecules per ~ 1 um cluster
~30-40 million clusters per experiment

Random array of clusters

Sequencing by synthesis



Cycle 1: Add sequencing reagents

First base incorporated

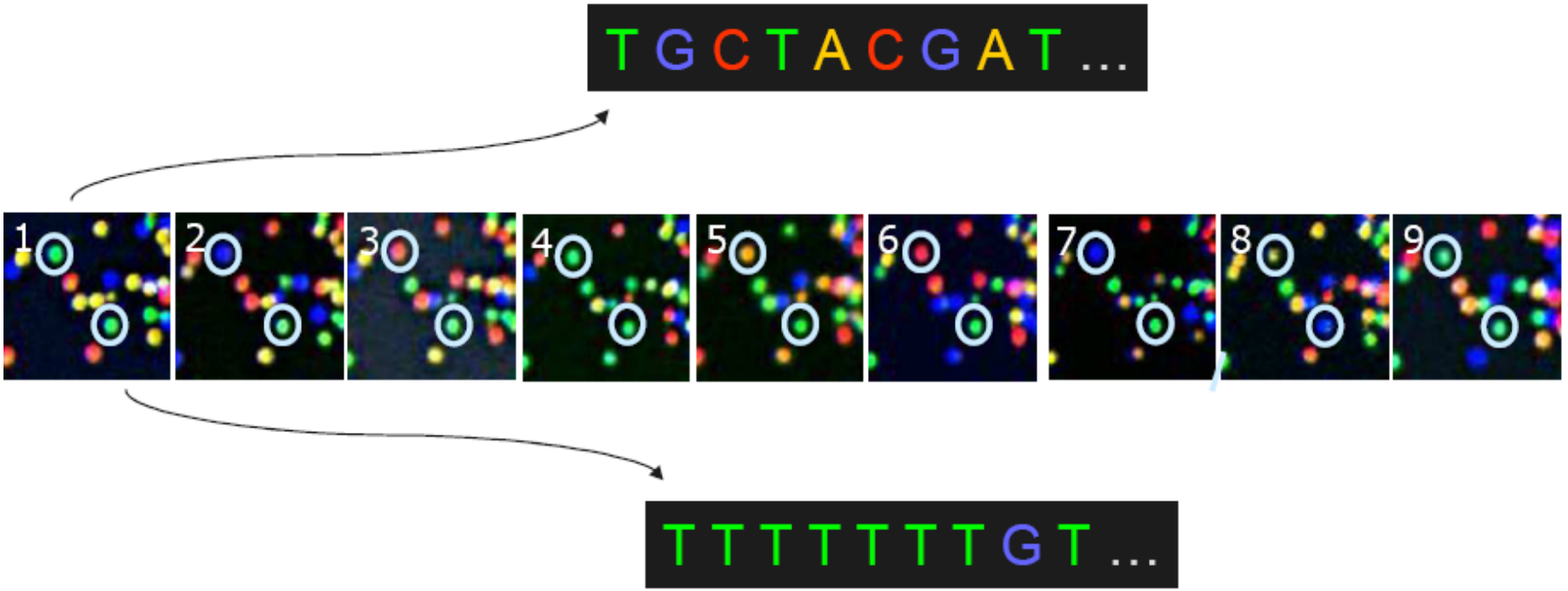
Remove unincorporated bases

Detect signal

Cycle 2-n: Add sequencing reagents and repeat

- All four labelled nucleotides in one reaction
- High accuracy
- Base-by-base sequencing
- No problems with homopolymer repeats

Sequencing by sequential imaging



Each nucleotide of a cluster is identified a single image. Multiple images gives the entire sequence