

# Non-covalent interactions and how macromolecules fold

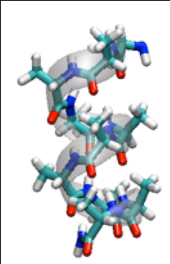
Dr Philip Fowler

First-year Biophysics course

- 1 Do we understand how proteins fold? Attractive forces (electrostatics).
- 2 More attractive forces. Repulsive forces. Torsion angles and the Ramachandran plot
- 3 Hydrogen bonds and water
- 4 The hydrophobic effect
- 5 Cooperativity of non-covalent interactions
- 6 Protein folding and misfolding

`philip.fowler@bioch.ox.ac.uk`

`@philipwfowler`



# Non-covalent interactions and how macromolecules fold

## Lecture 1: Do we understand how proteins fold? Attractive forces (electrostatics).

Dr Philip Fowler

First-year Biophysics course

### Objective:

Briefly examine the protein folding problem

What force do two electrically charged particles experience?

What are the differences between the Coulombic force, the electrostatic potential energy, the electric field and the electrostatic potential?

How do we treat situations where the charges are surrounded by other atoms (a solvent)?

Look at some simple example from proteins

### Summary:

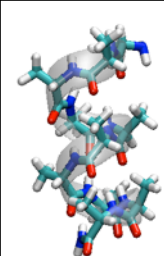
Proteins fold because they experience simple physical forces.

It is difficult to predict the structure (and therefore function) of a protein from its primary sequence

Electrostatics is the fundamental attractive force between atoms

Coulomb's Law, electrostatic potential energy, the electric field and the electrostatic potential allow us to characterise the interactions between permanent electrical point charges.

The dielectric constant of water is much higher than that of protein or other biological macromolecules



# Practicalities

## Handouts

The handouts may be slightly different to what is shown in the lectures.

The handouts will be put on WebLearn soon after each lecture both as PDFs and QuickTime movies (the animations and embedded movies will work in the latter). The electronic versions will be in colour.

## Lectures

week 7: Tu 10, Th 10

week 8: M 10, Tu 10, Th 10, Fri 9

## Questions / Suggestions

`philip.fowler@bioch.ox.ac.uk`

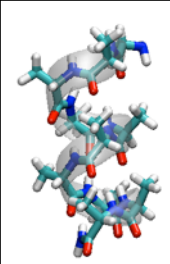
M, Tu: Sherrington LTT

Thu, Fri: Zoology LRB

## Overlap with previous courses

Biological Chemistry: Structure and Properties of Proteins

Biophysical Chemistry: Quantum Mechanics, Thermodynamics



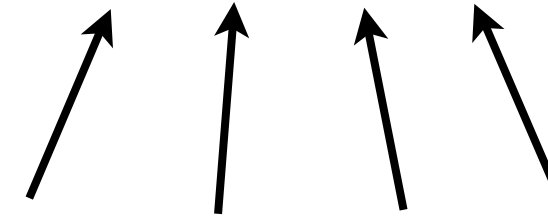
# Mendeley & References

throughout the lectures I will refer to either scientific papers or the textbooks in case you want to read some more

if there is a only a reference to one of the textbooks it doesn't mean the others don't have anything to say on the subject!

## Papers

PNAS (2007) 104 11963



name of journal      year      volume      first page

You can download the papers yourself\* or you can look in the Mendeley group I've created.



## What is Mendeley?

This is a free cloud-based reference management system. It works in browsers, on PCs, Macs, linux and also on iOS devices like iPhones etc. Once you have a free\*\* account your paper database is automatically synced "with the cloud" on all your devices.

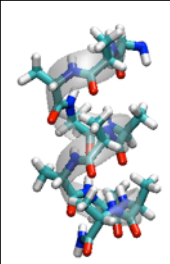
The other way of describing it is: "Facebook for Scientists" (you can even create an account using your Facebook login)

<http://www.mendeley.com/groups/1095611/prelim-biophysics-non-covalent-interactions-and-protein-folding/>

Or search for me and then look at my groups or search for key words in the group name. You can then "follow" the group or just take the references you want.

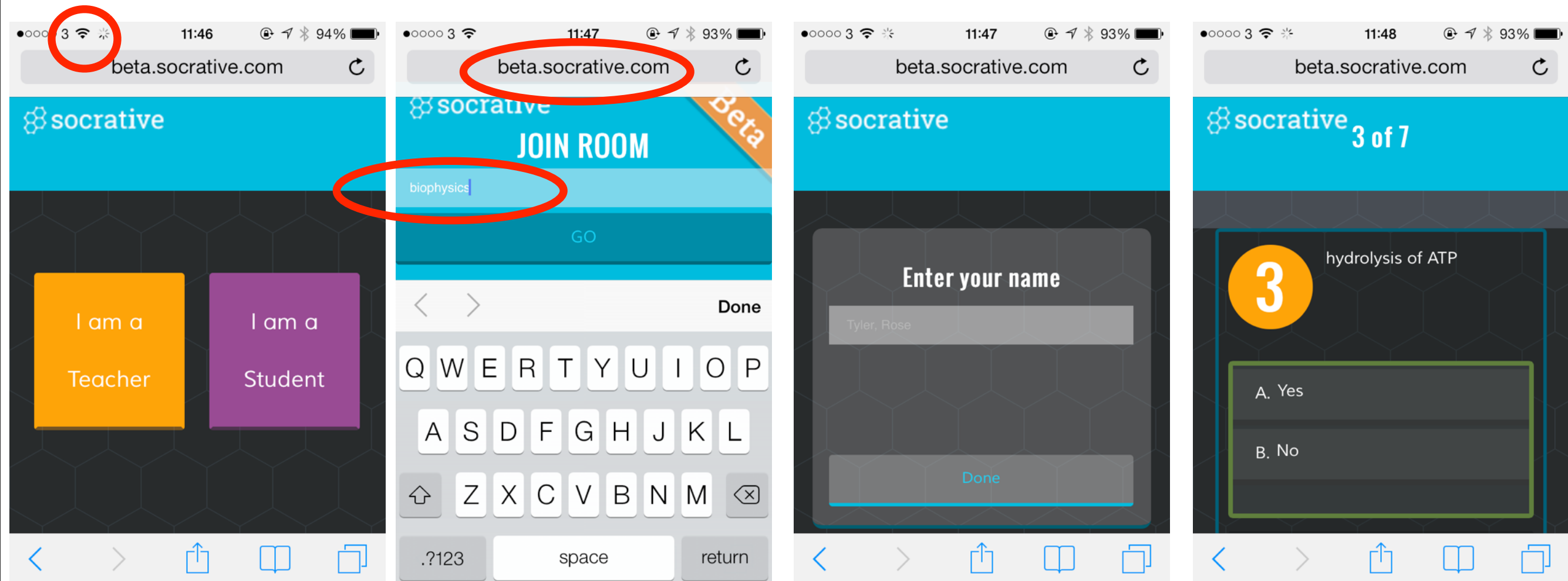
\* as long as the computer has a University of Oxford IP address. In other words you won't be able to do this at home *unless* you have setup your computer to use the Oxford VPN. This requires a *remote access password* which you can generate online and is different to your SSO password. \*\* up to 500Mb storage





# Live Quizzes

- We will do some **live quizzes** during the lectures using a website called Socrative (our room will be “biophysics”)
- We will do them in small groups, ideally pairs
- You will need a smartphone, tablet or laptop (can connect to webpages) connected to eduroam (using your OUCS Remote Access password) or OWL (using the Oxford VPN). Eduroam is better as once it is setup you don’t need to do anything; for OWL you have to turn on the VPN each time.

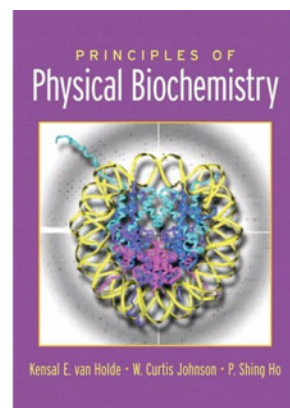
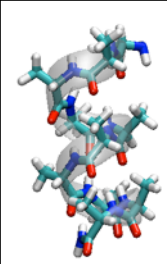


<http://beta.socrative.com>

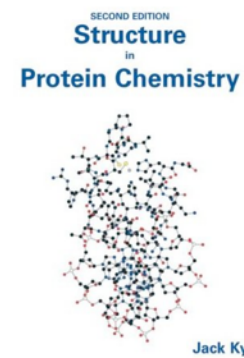
<http://www.oucs.ox.ac.uk/network/wireless/services/eduroam/>

<http://www.oucs.ox.ac.uk/network/wireless/services/owl/>

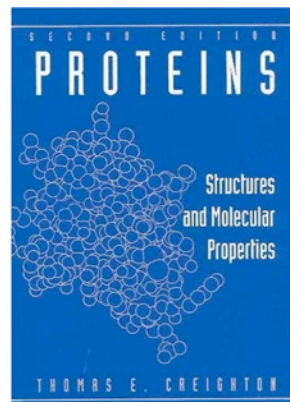




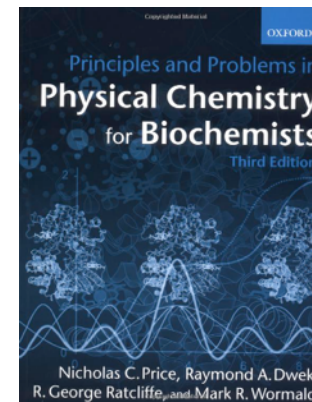
Principles of Physical Biochemistry  
van Holde, Johnson and Ho  
2nd ed



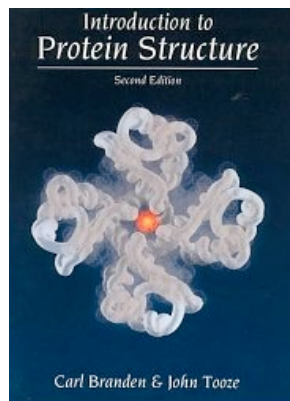
Structure in Protein Chemistry  
Kyte  
2nd ed



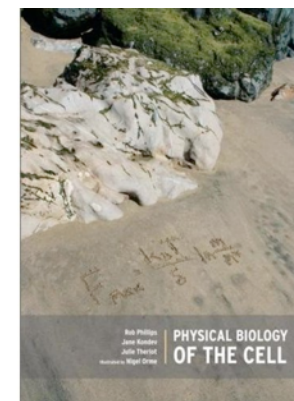
Proteins  
Creighton  
2nd ed



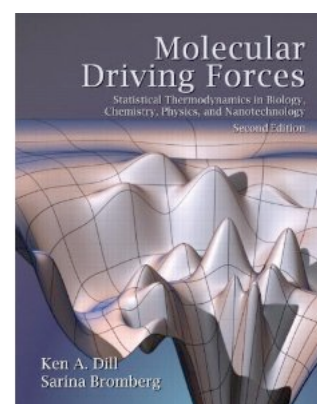
Physical Chemistry for Biochemists  
Price, Dwek, Ratcliffe and Wormald  
3rd ed



Introduction to Protein Structure  
Branden & Tooze  
2nd ed



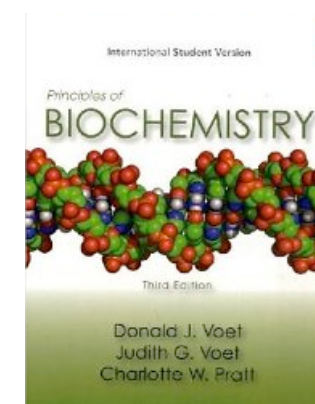
Physical Biology of the Cell  
Phillips, Kondev & Theriot  
1st ed

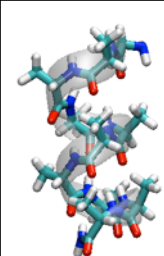


Molecular Driving Forces  
Dill and Bromberg  
2nd ed

....and of course

<http://www.youtube.com/watch?v=gFcp2Xpd29I>





# Anfinsen (1957) showed that proteins fold reversibly



Christian Anfinsen

Nobel Prize in Chemistry  
(1972)

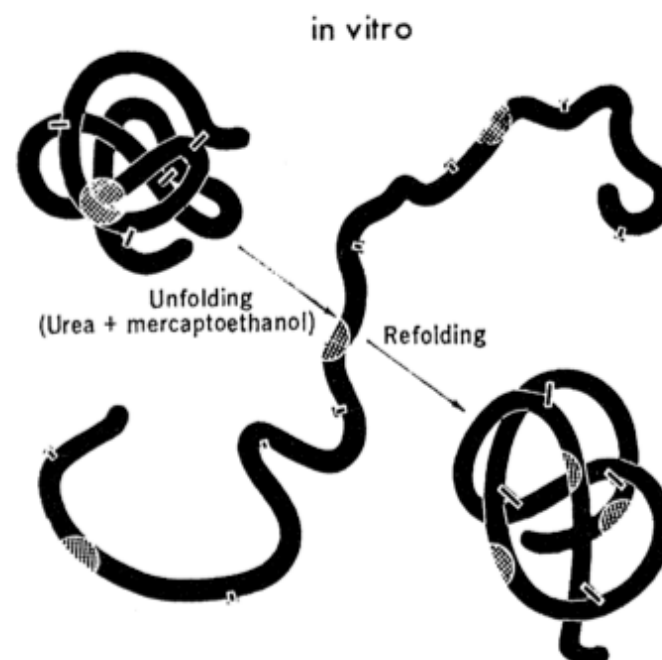
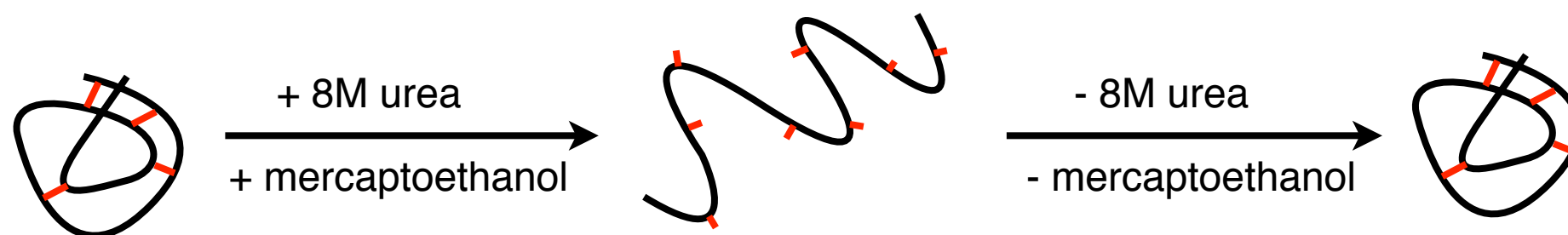


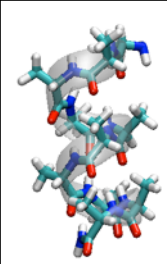
Fig. 2. Schematic representation of the reductive denaturation, in 8M urea solution containing 2-mercaptoethanol, of a disulfide-cross-linked protein. The conversion of the extended, denatured form to a randomly cross-linked, "scrambled" set of isomers is depicted at the lower right.

Science (1973) **181** 223



Ribonuclease A

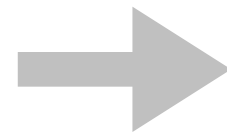
(in the right order - see Lecture 5)



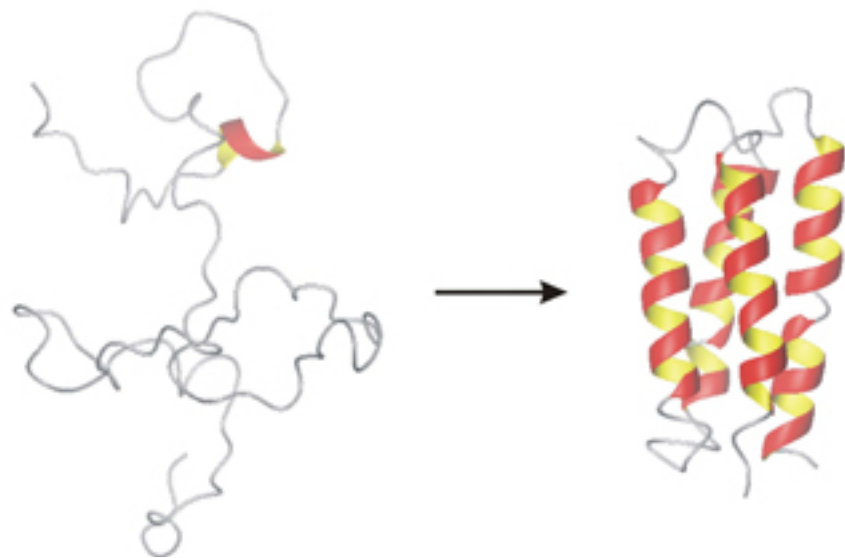
# Proteins fold and form compact structures in vivo

## Anfinsen's dogma / thermodynamic hypothesis:

*"The native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment."*



proteins fold to a unique tertiary structure under physiological conditions  
(this is dynamic (moves) (e.g. 'breathes'))



extended or unfolded structures  
(many possible)

native state

snapshots of this structure (usually at very low temperatures) can be obtained by X-ray crystallography

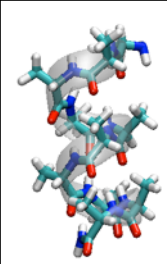
NMR often produces ensembles (collections) of structures that are all consistent with the experimental data

Why is structure important?

sequence  $\longrightarrow$  structure  $\longrightarrow$  function

...some proteins can misfold causing disease





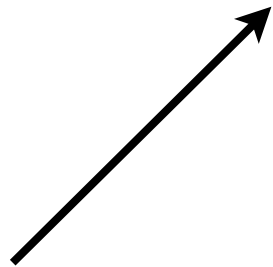
# Taking Anfinsen too far: a common misconception

“How does the amino acid sequence of a protein **specify** its 3D structure?”

-Stryer, 4th edition, pg 417

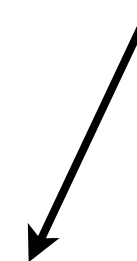
“..the **information** needed to fold a macromolecule into its native 3D structure is **contained** within its sequence.”

-van Holde, 2nd edition, pg 107



These descriptions are misleading...

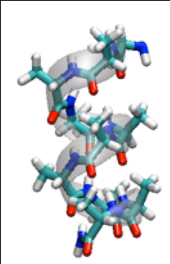
..don't confuse these two!



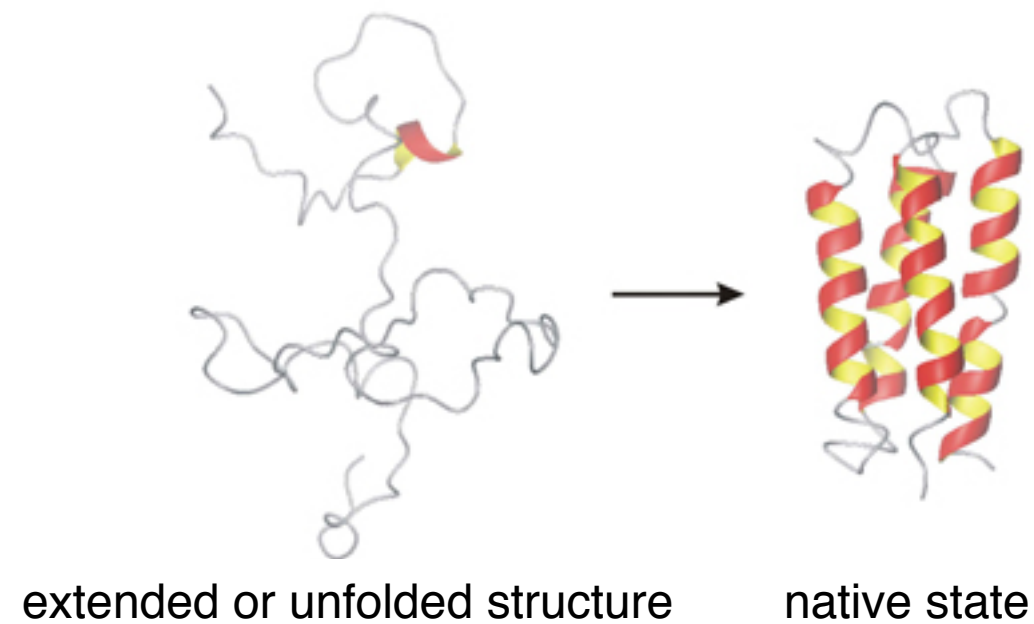
Instead:

The **primary sequence** of a protein is **encoded** by DNA.

The protein experiences **physical forces** that cause it to fold into its **native state**.



# The protein folding problem

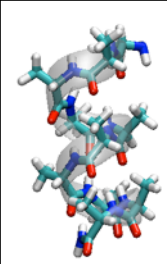


**these lectures:** Can we understand the physics that cause proteins to fold and adopt complex, compact structures in vivo?

**protein structure prediction:** Can we predict the native state (structure) of a protein from its primary sequence?

**protein folding problem:** Can we predict how a protein folds from its primary sequence into its native state?

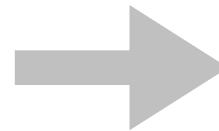
—  
more difficult  
↓



# The Levinthal paradox

Anfinsen's dogma / thermodynamic hypothesis: *"The native conformation is determined by the **totality of interatomic interactions** and hence by the amino acid sequence, in a given environment."*

How does a protein find its native state?



Assume it samples all conformations and chooses the one with the lowest (free) energy

## Calculation

If we have a protein with  $n$  residues and

...each residue has 2 torsion angles  $\phi$  and  $\psi$

...assume  $\phi$  and  $\psi$  can each exist in 1 of in 3 stable conformations

Then the number of conformations is

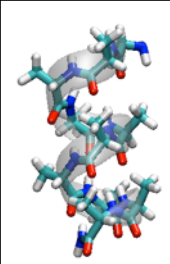
$$3^{2n} \approx 10^n$$

If we have a 100 residue protein and assume it takes 100 fs to convert between each conformation then the protein will fold in

$$10^{100} \times 10^{-13} \text{ s} = 10^{87} \text{ s} = 10^{79} \text{ years}$$



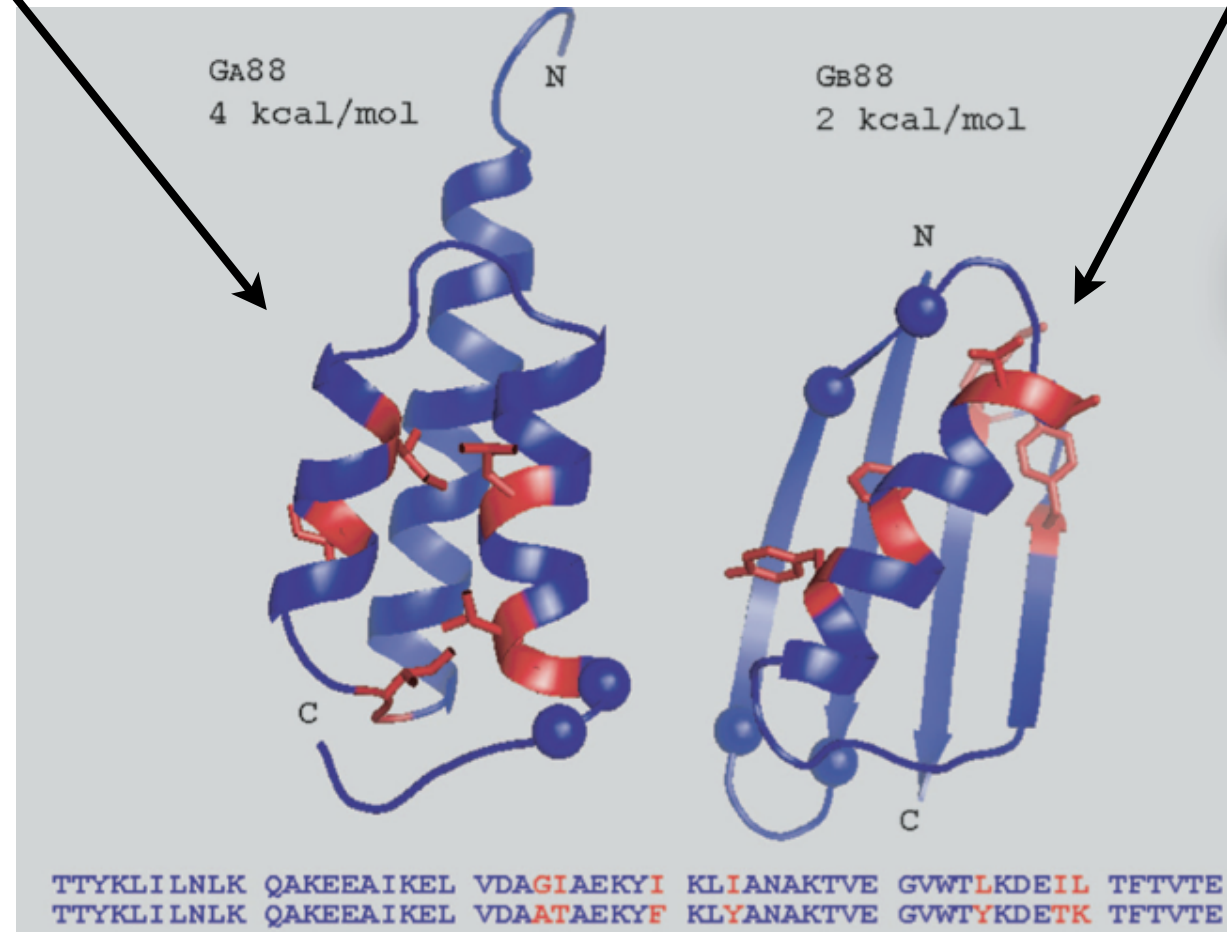
This is not possible (most proteins fold in ms - s)



# Small differences in sequence can lead to large differences in structure and function

3- $\alpha$  helix fold with albumin-binding epitope

$\alpha/\beta$  fold with IgG epitope

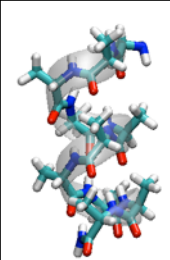


PNAS (2007) **104** 11963

49 residues (88 %) are identical

7 residues (88 %) are different

...the reverse can also be true. Large differences in sequence may only lead to small differences in structure and function.



It is the free energy that determines whether a reaction proceeds\*

Gibb's free energy  
(i.e. constant temperature and pressure)

Temperature

$$\Delta G = \Delta H - T\Delta S$$

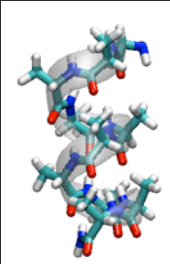
Entropy = *energy dispersal*

Enthalpy

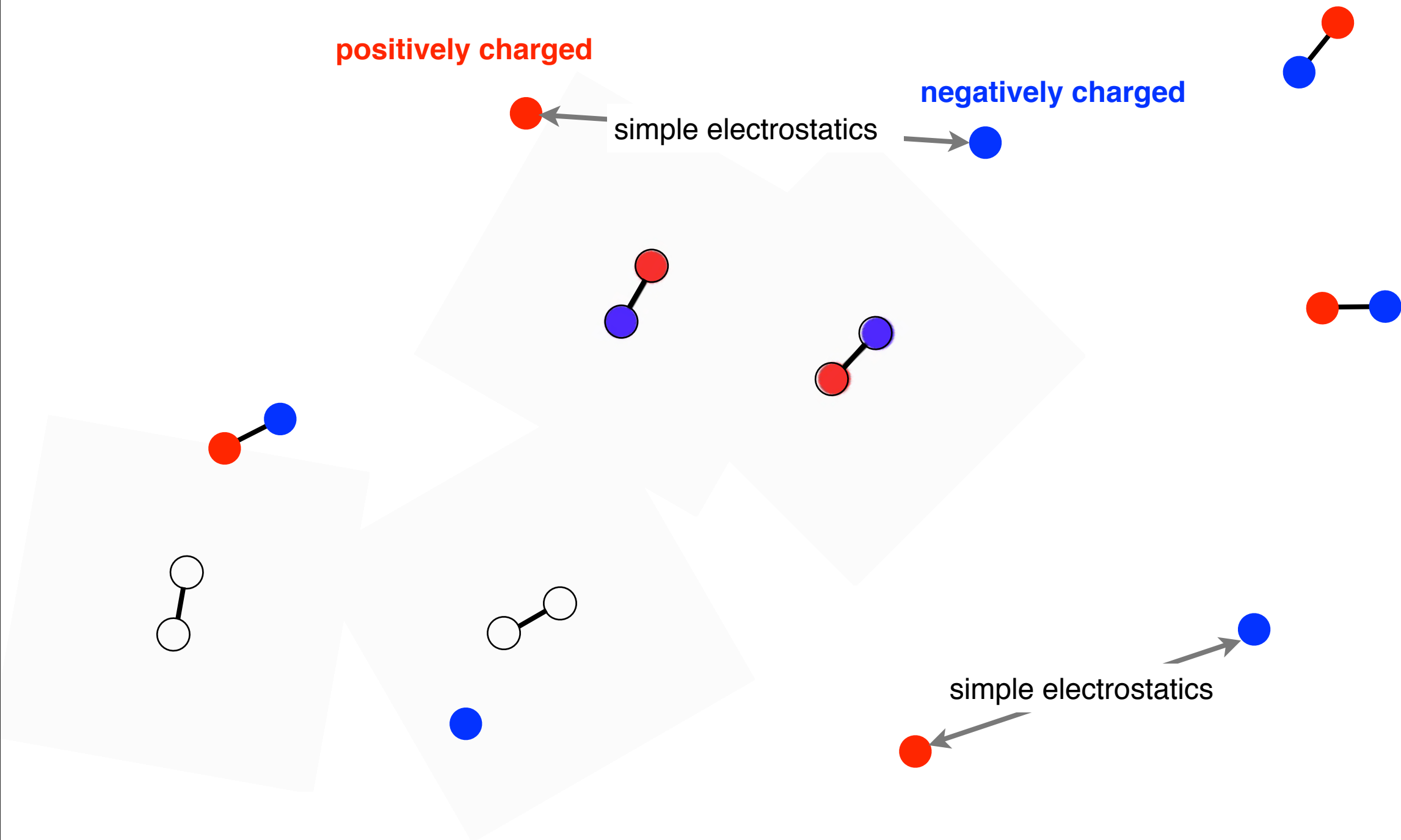
$$H = U + pV$$

\* we are ignoring the possibility of a large kinetic barrier here



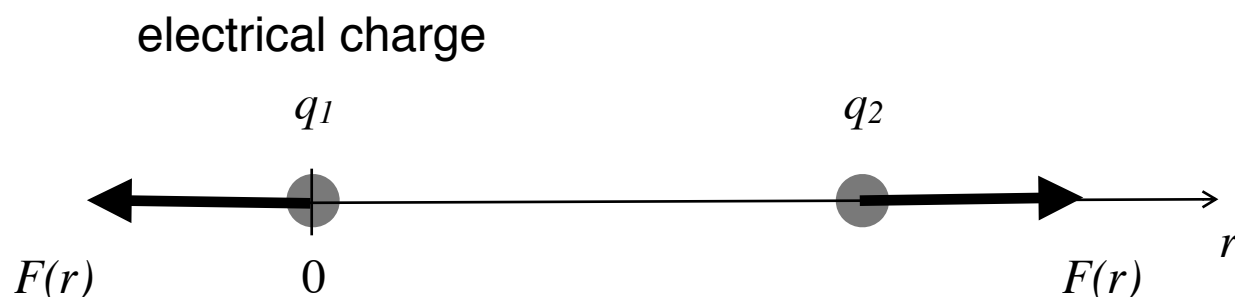


# The fundamental classical\* attractive force is electrostatics



\* i.e. none of the descriptions we shall use involve quantum mechanics

# Simple Electrostatics



since we usually deal with “chemical” charges e.g. we think of  $\text{Na}^+$  as having a charge of +1 when it really each  $\text{Na}^+$  ion has a charge of  $+1e$

$$q_i = Q_i e$$

+  $1.6 \times 10^{-19} \text{ C}$

+1

$1.6 \times 10^{-19} \text{ C}$

elementary charge  
(charge on an electron)

## Coulomb's Law

assumes **point** charges

Force on  $q_2$  due to  $q_1$   
( $q_1$  experiences an equal and opposite force)

$$F(r) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2}$$

electrical charges

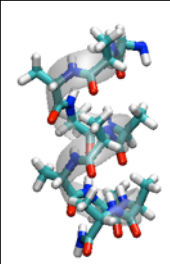
distance between ions

permittivity of free space

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2 \text{ m}^{-2} \text{ N}^{-1} \text{ (F m}^{-1}\text{)}$$

$$F(r) = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2 e^2}{r^2}$$

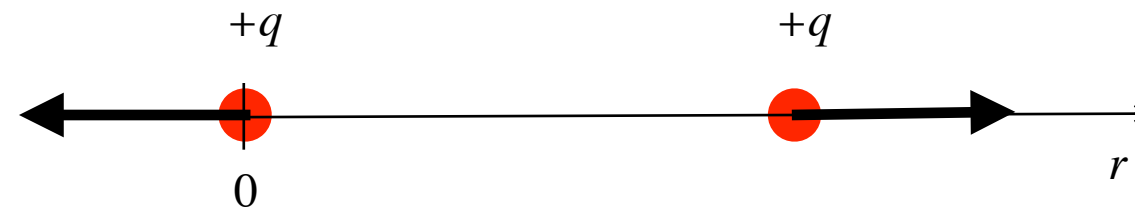
elementary charge (C)



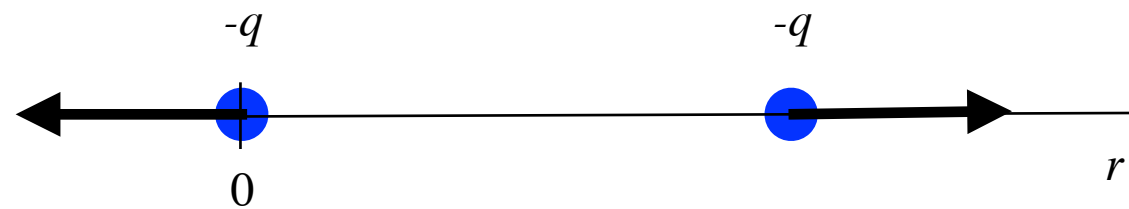
# Sign conventions

Like charges **repel**

if  $q_1 = q_2 = q > 0$   
i.e. both are **positive** charges



if  $q_1 = q_2 = q < 0$   
i.e. both are **negative** charges

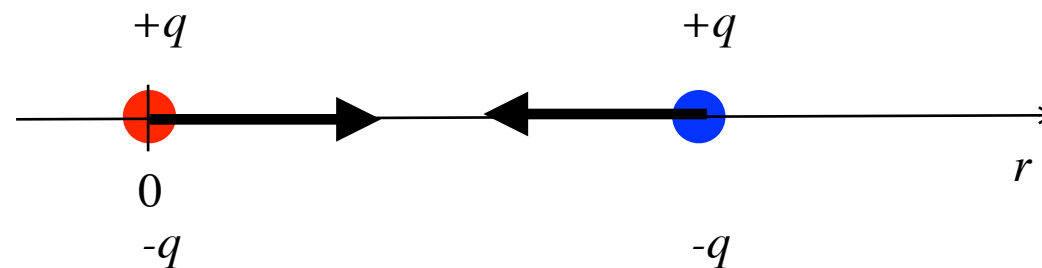


$$F(r) = \frac{1}{4\pi\epsilon_0} \frac{q^2}{r^2}$$

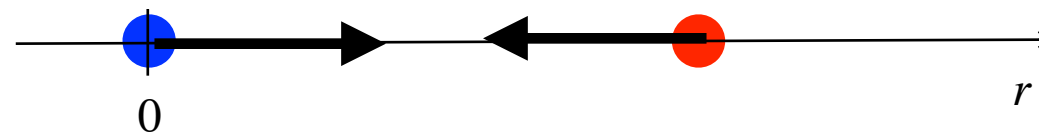
$F(r)$  is positive and therefore is directed with  $r$  for  $q_2$   
(remember  $q_1$  experiences an equal and opposite force and therefore this is directed against  $r$ )

Unlike charges **attract**

if  $q_1 = -q_2 = q > 0$

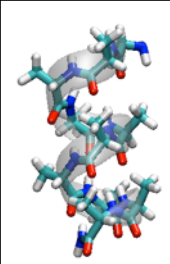


if  $q_1 = -q_2 = q < 0$



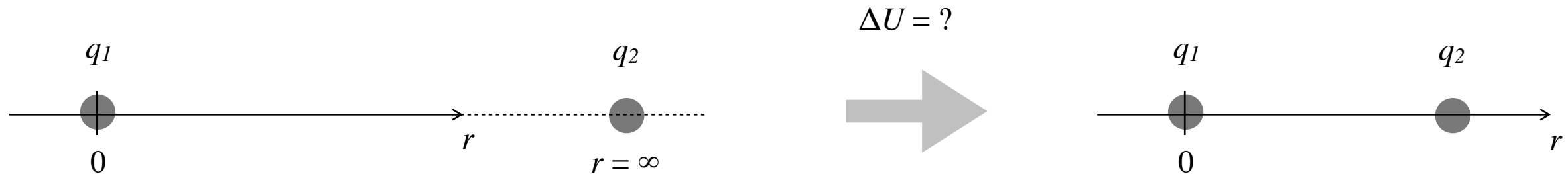
$$F(r) = -\frac{1}{4\pi\epsilon_0} \frac{q^2}{r^2}$$

$F(r)$  is negative and therefore is directed against  $r$  for  $q_2$   
(remember  $q_1$  experiences an equal and opposite force and therefore this is directed with  $r$ )



# Electrostatic potential energy

Consider the change in energy in bringing a charge  $q_2$  from infinity to a distance  $r$  away from  $q_1$



we know what the force at each separation  $r$  is, but how is the force related to the change in potential energy?

Work = Force x distance

related to

$$F(r) = -\frac{dU}{dr}$$

this is a first-order differential equation so we can integrate w.r.t.  $r$

$$\int_{U_\infty}^{U_r} dU = \int_{\infty}^r -F dr$$

substitute in  
Coulomb's Law

define:  $\Delta U = U_r - U_\infty$  (by convention)  $U_\infty = 0$

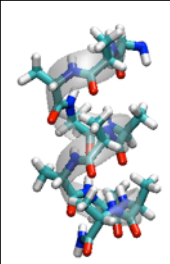
$$\Delta U = \int_{\infty}^r -F dr$$

$$\Delta U = \int_{\infty}^r -\frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} dr$$

integrating...

$$\Delta U = \left[ \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r} \right]_{\infty}^r$$

$$\Delta U = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r}$$



# Electrostatic potential energy ... cont

**Potential energy of two charges separated by a distance  $r$**

$$\Delta U = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r}$$

This is an energy so the units are Joules (J)

For proteins this is usually quoted as kJ mol<sup>-1</sup> or kcal mol<sup>-1</sup>  
(1 kJ = 1000 J, 1 cal = 4.18 J)

Like we did earlier we can write it in terms of “chemical” charges. This helps us to remember that we cannot forget the elementary charge,  $e$

$$\Delta U = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2 e^2}{r} \quad \text{where} \quad q_i = Q_i e$$

$$F(r) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \quad \begin{array}{c} \xrightarrow{\Delta U = \int_{\infty}^r -F dr} \\ \xleftarrow{F(r) = -\frac{dU}{dr}} \end{array} \quad \Delta U = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r}$$



# Example: a Na<sup>+</sup> ion and a Cl<sup>-</sup> ion

Na<sup>+</sup>



Cl<sup>-</sup>



$r = 2.76 \text{ \AA}$

$$\Delta U = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2 e^2}{r}$$

$$\Delta U = 8.988 \times 10^9 \times \frac{1 \times -1 \times (1.6 \times 10^{-19})^2}{2.76 \times 10^{-10}}$$

$$\Delta U = -8.34 \times 10^{-19} \text{ J}$$

$$\Delta U = -8.34 \times 10^{-19} \times N_A \text{ J mol}^{-1}$$

$$\Delta U = -502 \text{ kJ mol}^{-1}$$

Avogadro's number

negative therefore preferred

$$F(r) = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2 e^2}{r^2}$$

$$F = 8.988 \times 10^9 \times \frac{1 \times -1 \times (1.6 \times 10^{-19})^2}{(2.76 \times 10^{-10})^2}$$

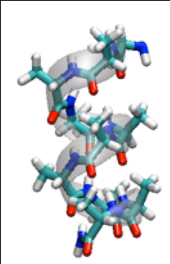
$$F = -3.02 \times 10^{-9} \text{ N}$$

$$F = -3.02 \text{ pN}$$

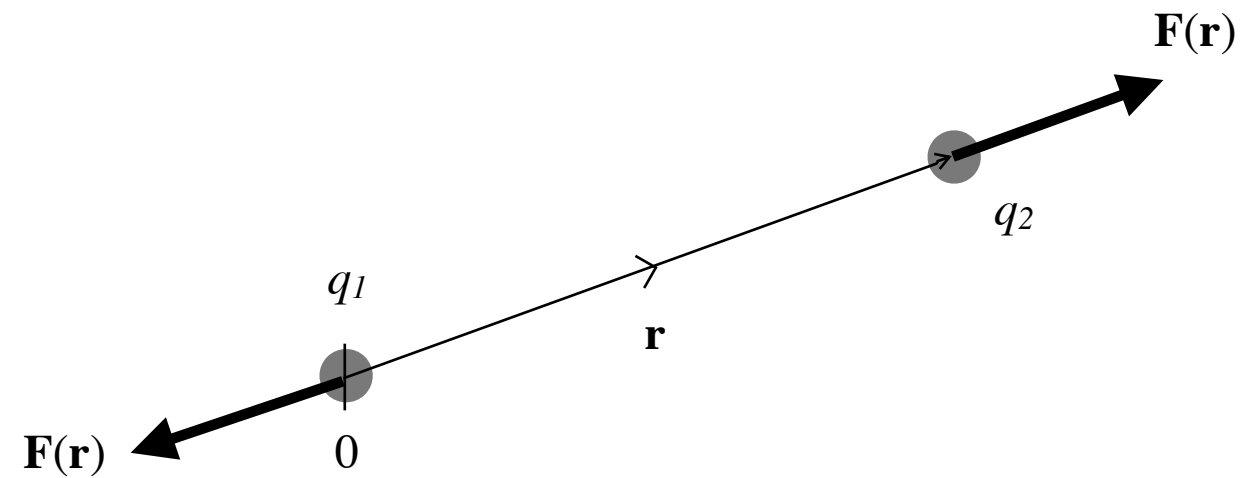
negative i.e. for  $q_2$  directed against  $r$   
i.e. this is attractive

We can calculate the potential energy of and force on two **isolated stationary** charges

$$\frac{1}{4 \times \pi \times 8.854 \times 10^{-12}} = 8.988 \times 10^9 \text{ F}^{-1} \text{ m}$$



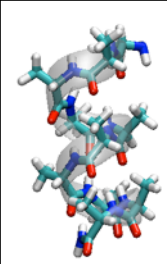
# The electrostatic force is a vector



$$\mathbf{F}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \hat{\mathbf{r}}$$

now the force is a vector quantity and so is written  $\mathbf{F}(\mathbf{r})$

this is a unit vector (i.e. has length 1) directed along  $\mathbf{r}$

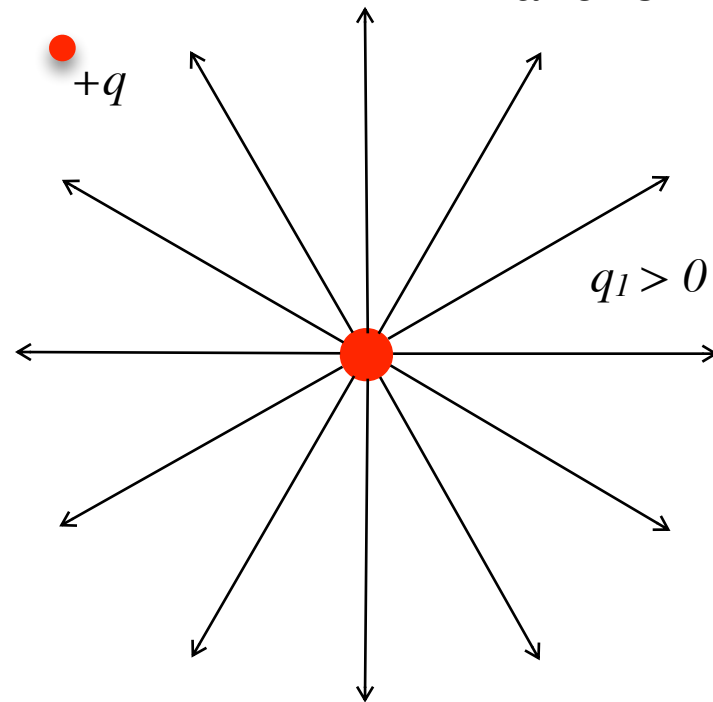


# Consider how a small positive charge would move

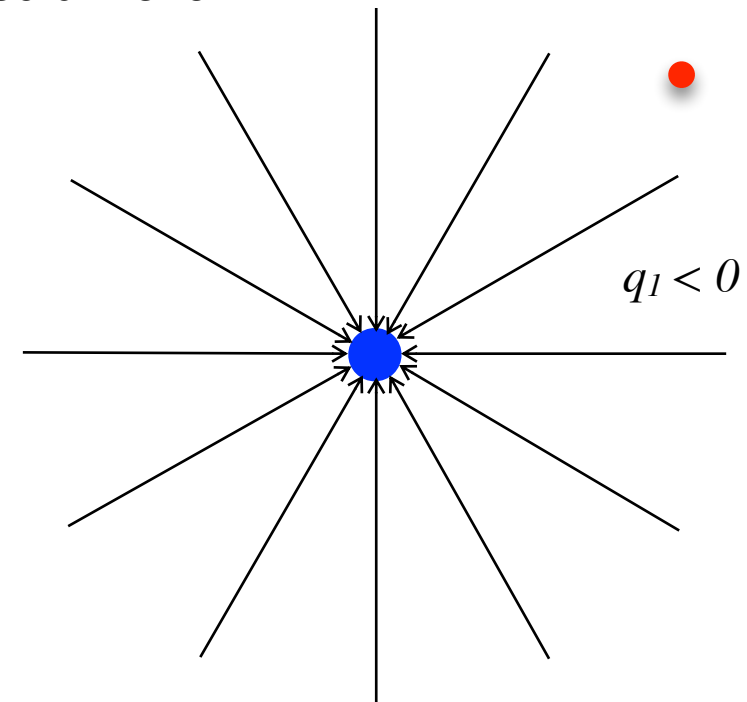
test charge

(i.e. does not exert a force on  $q_1$ )

arrows indicate how the probe would move



probe would be radially repelled from another positive charge



probe would be radially attracted towards a negative charge

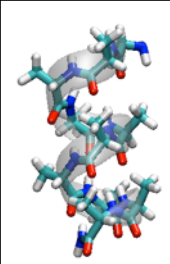
The **electric field strength** due to  $q_1$  is the force that would be exerted on a positive test charge if it were placed at  $\mathbf{r}$

$$\mathbf{E} = \frac{\mathbf{F}}{q} \quad \mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q_1}{r^2} \hat{\mathbf{r}}$$

units of  $\text{N C}^{-1}$  or  $\text{V m}^{-1}$  (equivalent SI units)

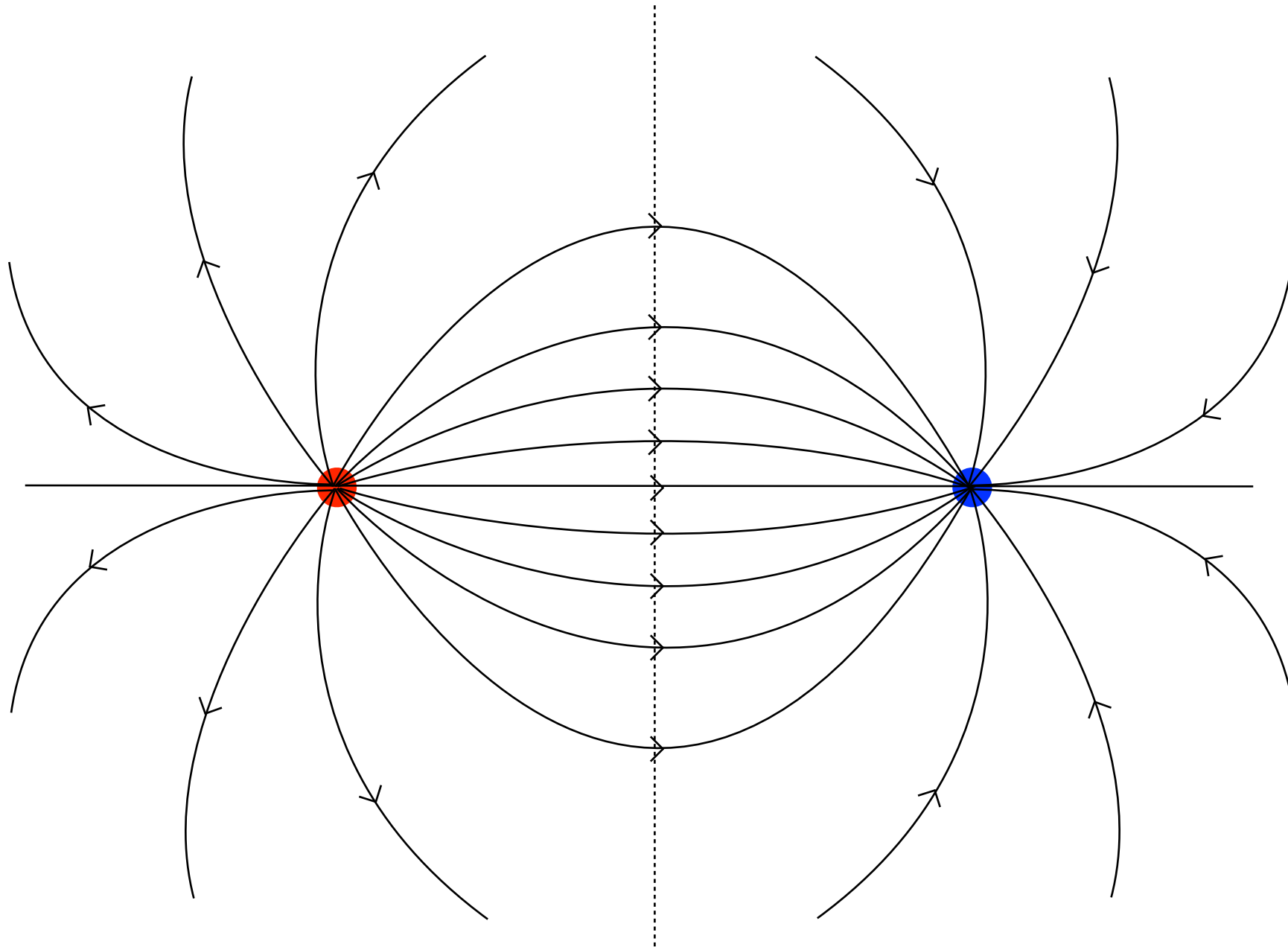
since the test charge does not itself exert a force,  $\mathbf{E}$  is strictly defined as:

$$\mathbf{E} = \lim_{q \rightarrow 0} \frac{\mathbf{F}}{q}$$



## Electric field when $q_1 = +q$ and $q_2 = -q$

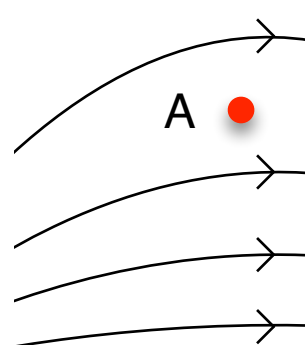
Since  $\mathbf{E}$  is a vector you can add the electric fields due to two separate charges



What would the electric field look like for two charges with the same sign?

# What is the work done in moving the test charge to a point A?

by analogy with the electrostatic force and the electrostatic potential energy



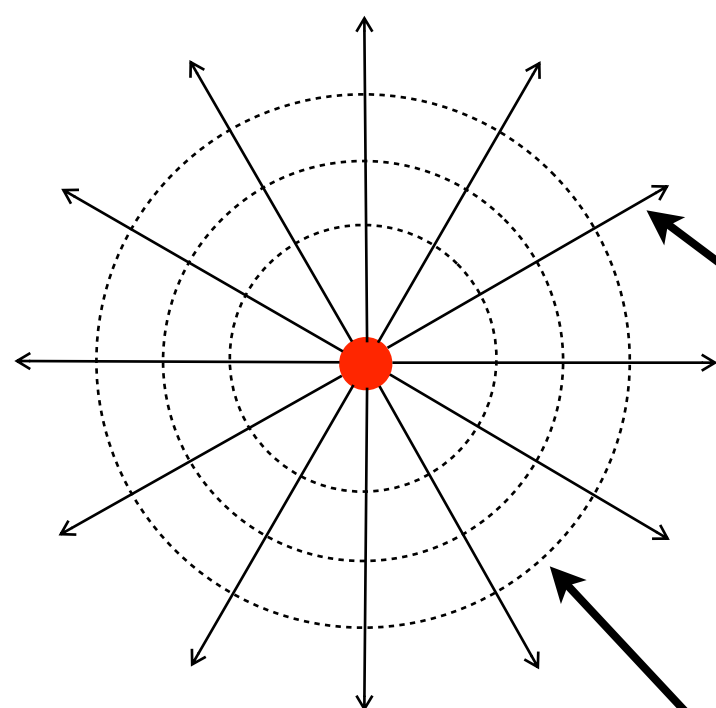
if  $\mathbf{E} = \frac{\mathbf{F}}{q}$  then  $\Phi = \frac{\Delta U}{q}$

electrostatic potential

$$\Phi = \frac{1}{4\pi\epsilon_0} \frac{q_1}{r}$$

units V or J C<sup>-1</sup> (equivalent SI units)

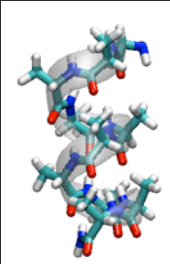
the **electrostatic potential** is the work done that would be done in bringing a test charge from infinity to a distance  $r$  away from a charge  $q_1$



direction a positive test charge would move  
(**electric field strength**)

positions at which the positive test charge  
has the same energy (**electric potential**)





# How are all these quantities related?

## electrostatic force

$$\mathbf{F}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \hat{\mathbf{r}} \quad (\text{units N})$$

the electrostatic force on  $q_2$  due to  $q_1$

a vector but can also write as a scalar since we know the direction the force is pointing

$$F(r) = -\frac{dU}{dr} \quad \Delta U = \int_{\infty}^r -F dr$$

## electrostatic potential energy

$$\Delta U = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r} \quad (\text{units J or N m}^{-1})$$

the work done in moving charge  $q_2$  from infinity to be a distance  $r$  away from charge  $q_1$

$$\mathbf{E} = \frac{\mathbf{F}}{q}$$

## electric field strength

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}} \quad (\text{units N C}^{-1} \text{ or V m}^{-1})$$

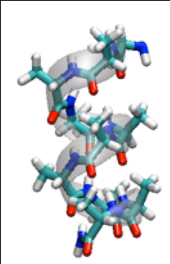
the electrostatic force a test charge would feel due to  $q_1$

How is the electric field strength related to the electrostatic potential?

## electrostatic potential

$$\Phi = \frac{1}{4\pi\epsilon_0} \frac{q_1}{r} \quad (\text{units J C}^{-1} \text{ or V})$$

$$\Phi = \frac{\Delta U}{q}$$



# Multiple charges

It is much easier to calculate the electrostatic potential energy at any point in space than to calculate the electrostatic forces due to each charge and then adding them vectorially.

To calculate the total potential energy we add the potential energies for each pair of interactions:

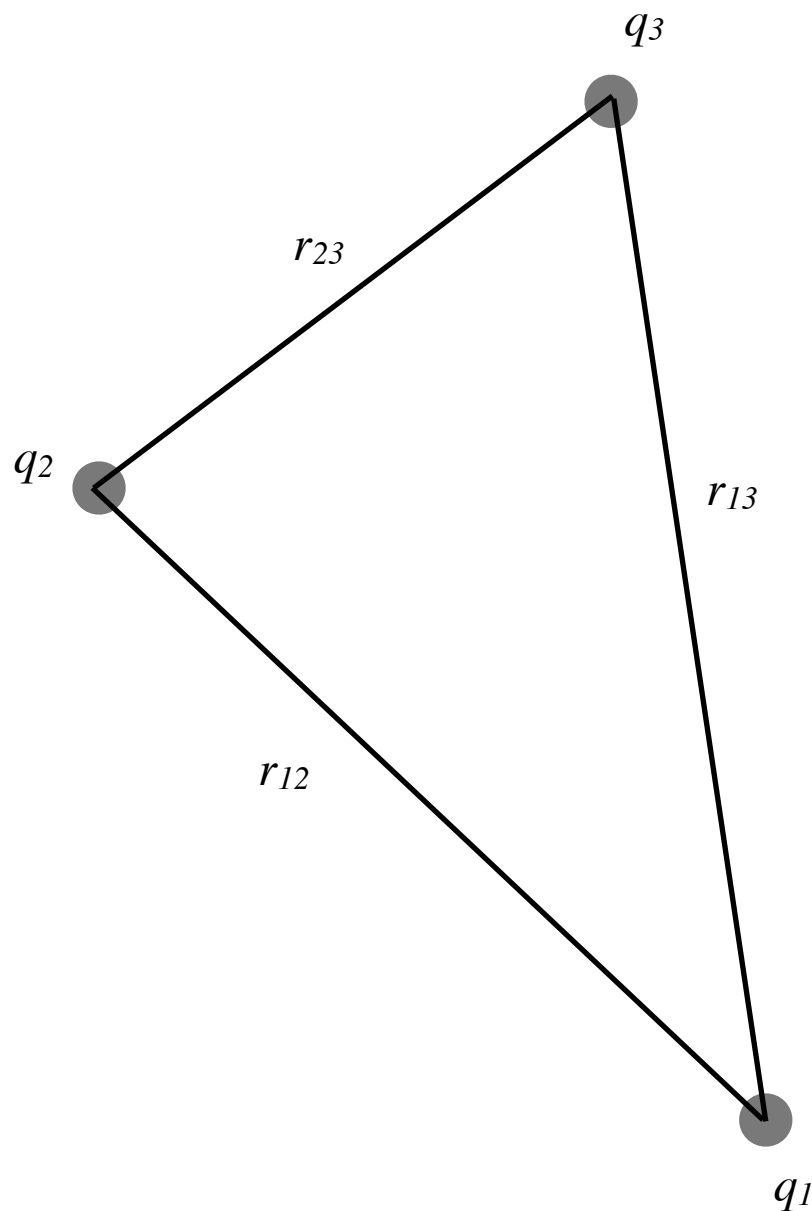
$$\Delta U = \frac{1}{4\pi\epsilon_0} \left[ \frac{q_1 q_2}{r_{12}} + \frac{q_2 q_3}{r_{23}} + \frac{q_1 q_3}{r_{13}} \right]$$

For  $N$  electrical charges:

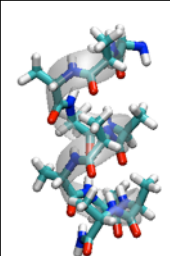
$$\Delta U = \sum_{i=1}^N \sum_{j<i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

sum over all charges  $i=1\dots N$

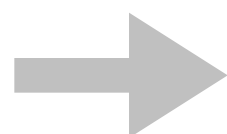
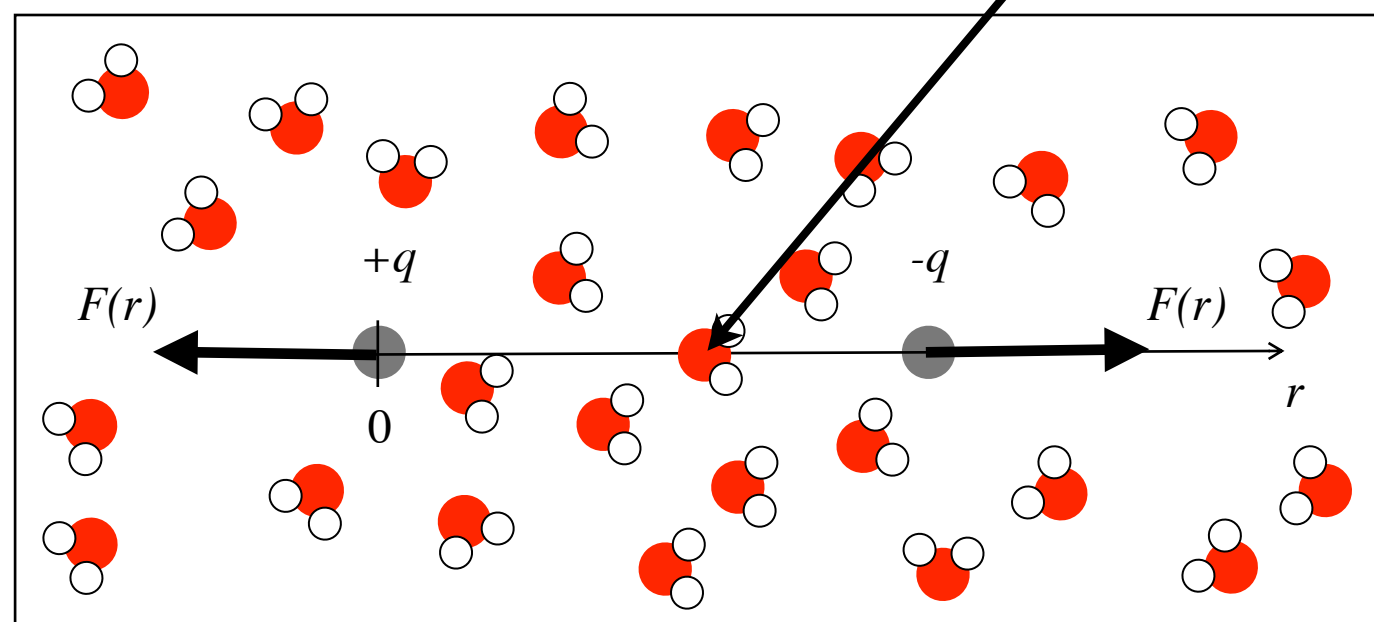
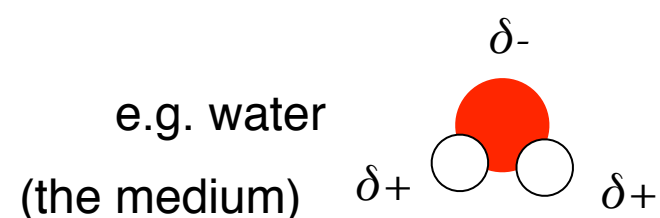
sum over all charges  $j=1\dots(i-1)$   
(i.e. don't double count)



# Dielectrics



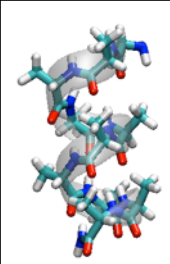
But what if the charges are not in a vacuum? What if there is a medium?



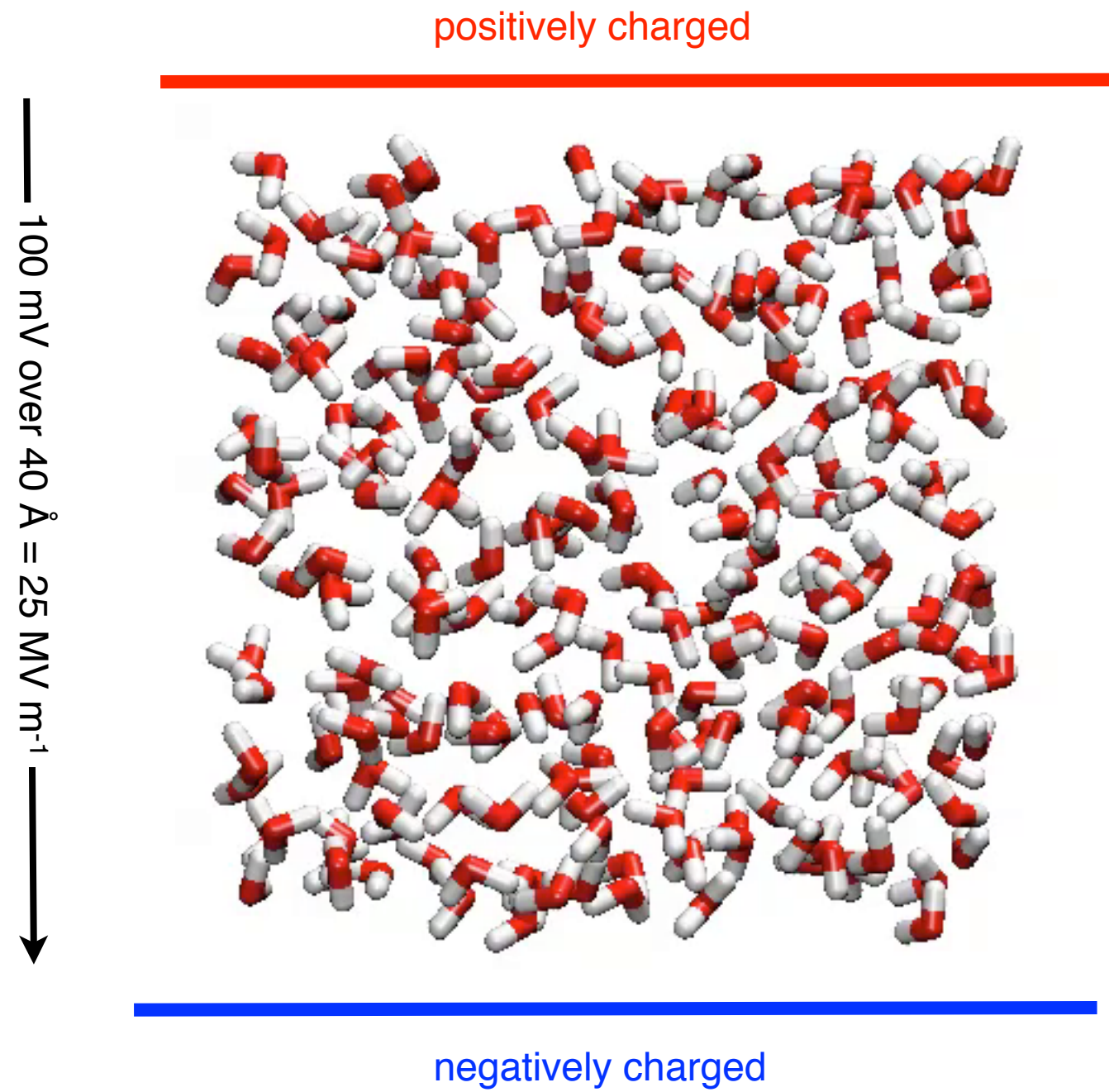
water molecules themselves feel forces due to the two charges and as a result align between the two charges  
this has the net effect of **reducing** the Coulombic force that  $q_2$  experiences due to  $q_1$

How much it is reduced depends on two factors:

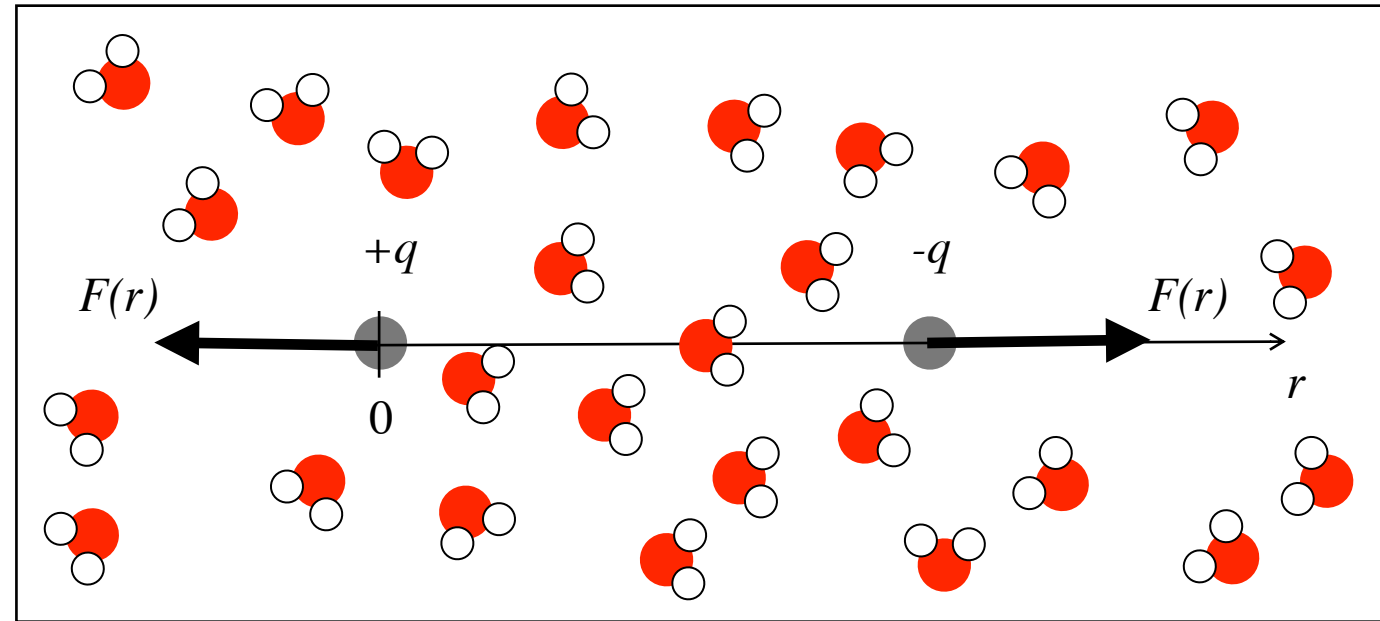
1. How easily can an applied electric field induce a separation of charge in the molecules of the medium  
(**induction effect**)
2. How easily can the molecules of the medium move to minimise the electrostatic force they experience?  
(**orientation effect**)



# Dielectrics



# The dielectric constant



## Options:

1. Calculate the forces between ALL molecules (microscopic approach)

This is difficult since there are many waters and they also move in response to the force so you would have to iterate the calculation and take an average

2. Introduce a **dimensionless factor** that takes into account the behaviour of the medium in which the charges exist (macroscopic approach)

$$F(r) = -\frac{1}{4\pi\epsilon\epsilon_0} \frac{q^2}{r^2}$$

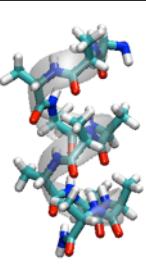
For a vacuum  $\epsilon = 1$

For everything else  $\epsilon > 1$

i.e. all media reduce the electrostatic forces and electrostatic interaction energies

$\epsilon$  is the **dielectric constant** of the medium (it has no units)  
(also called the relative permittivity)





# Typical values of dielectric constants

Reducing all the forces between the charges and the medium (e.g. the individual waters) to a dielectric constant assumes that:

the medium is homogenous

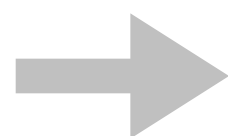
the electric charges are embedded deep within the medium

the charges are far apart

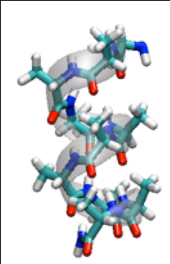
$$F(r) = \frac{1}{4\pi\epsilon\epsilon_0} \frac{q_1q_2}{r^2}$$

medium	$\epsilon$	
nothing (i.e. vacuum)	1	
air	1.0006	
octane	1.95	octane is <b>non-polar</b> ; it is therefore <b>poor</b> at reducing electrostatic forces
water	78.5	water is <b>polar</b> and <b>readily moves</b> in response to the presence of electric charges. It is therefore very <b>effective</b> at reducing (shielding) electrostatic forces.
lipid	~2	neither <b>proteins</b> nor <b>lipids</b> are <b>homogenous</b> so the value of the dielectric constant depends on e.g. the <b>orientation</b> and <b>precise composition</b> of e.g. the protein
protein	2-4	

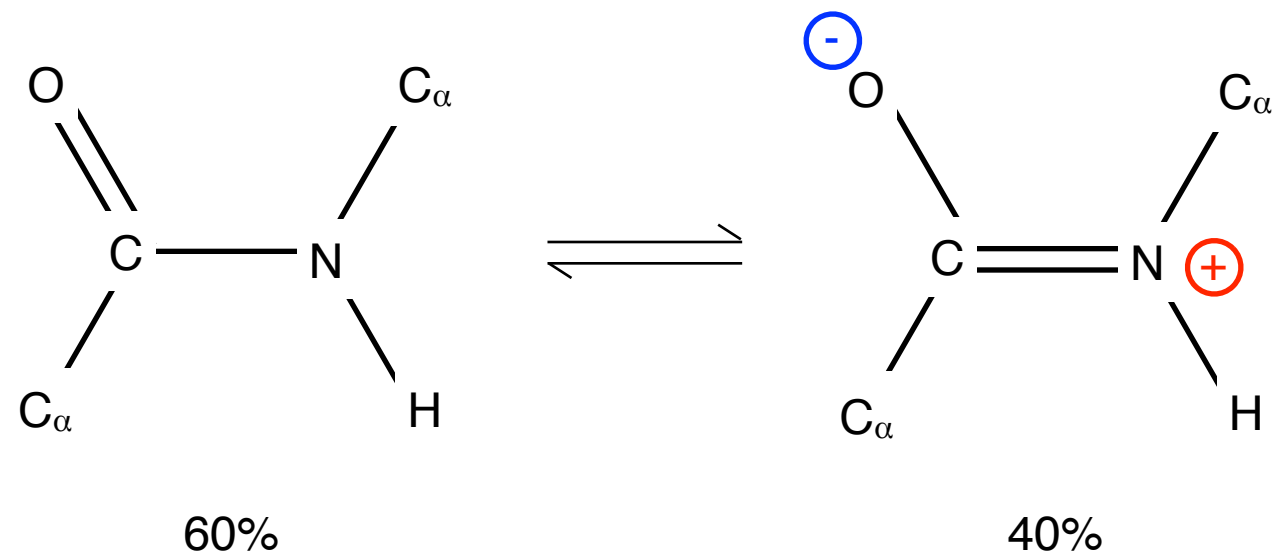
20-40 fold difference!



The large difference in the dielectric constants between the interior of a protein, a lipid bilayer and in water is very important for the behaviour of proteins and other biological macromolecules



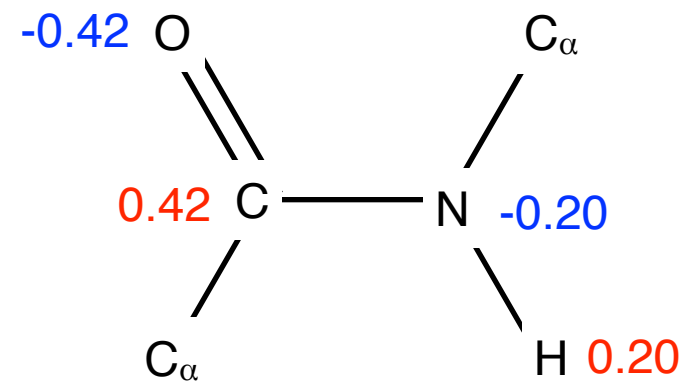
# The peptide bond



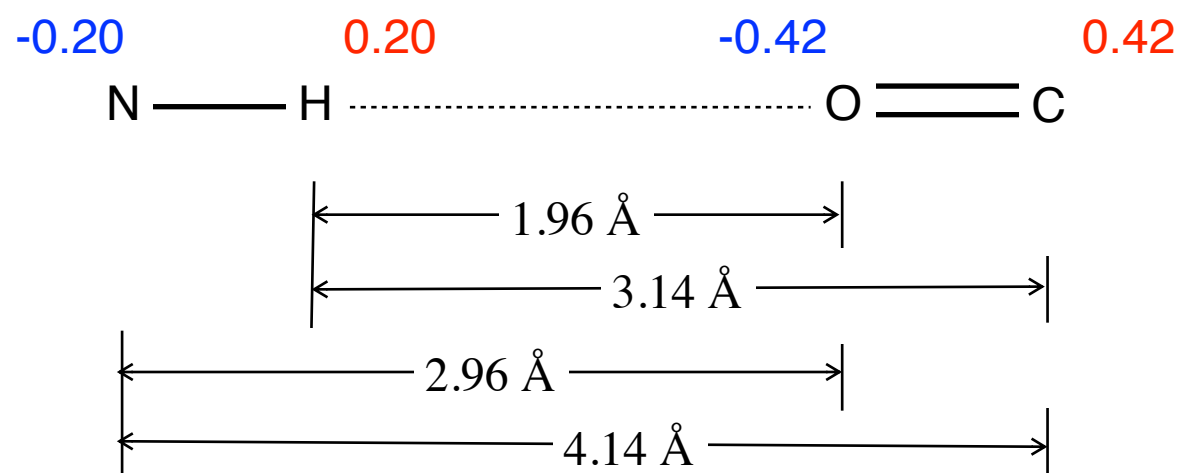
the peptide bond is planar

we can represent the atoms as having partial charges

as a first approximation these are  $\sim \pm 0.4e$



# Example



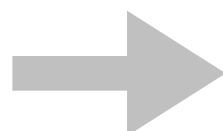
$$\Delta U = \frac{1}{4\pi\epsilon\epsilon_0} \left( \frac{q_H q_O}{r_{HO}} + \frac{q_H q_C}{r_{HC}} + \frac{q_N q_O}{r_{NO}} + \frac{q_N q_C}{r_{NC}} \right)$$

$$\Delta U = -10.4 \times 10^{-21} \text{ J}$$

(we have assumed  $\epsilon = 1$ )

$$\Delta U = -6.2 \text{ kJ mol}^{-1}$$

How do you think this would vary if the C=O group were tilted by an angle  $\theta$ ?



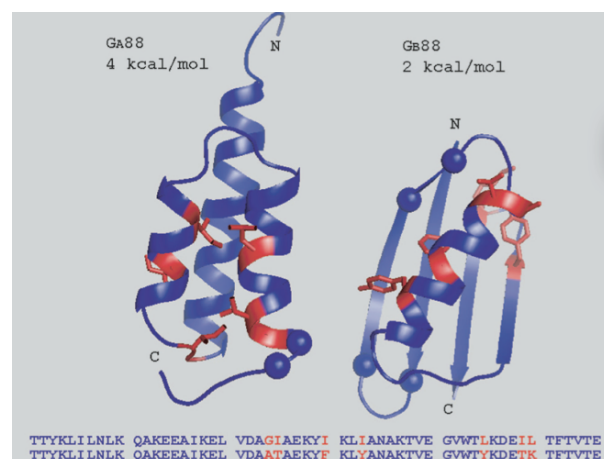
even though the net electrical charges is zero, there is still a net interaction between the two groups

Notice how the magnitudes of the interactions between each pair is large but when added together almost cancel one another out. e.g.

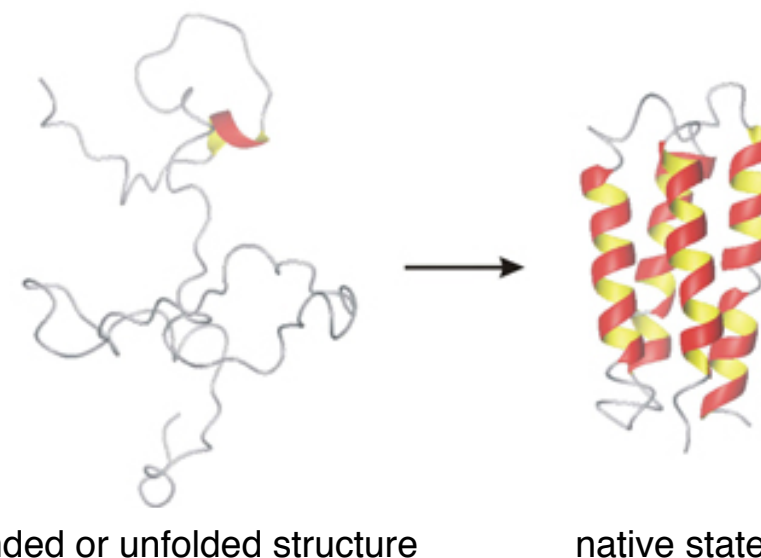
$$\Delta U_{HO} = -47.8 \text{ kJ mol}^{-1}$$

# Lecture 1: Summary

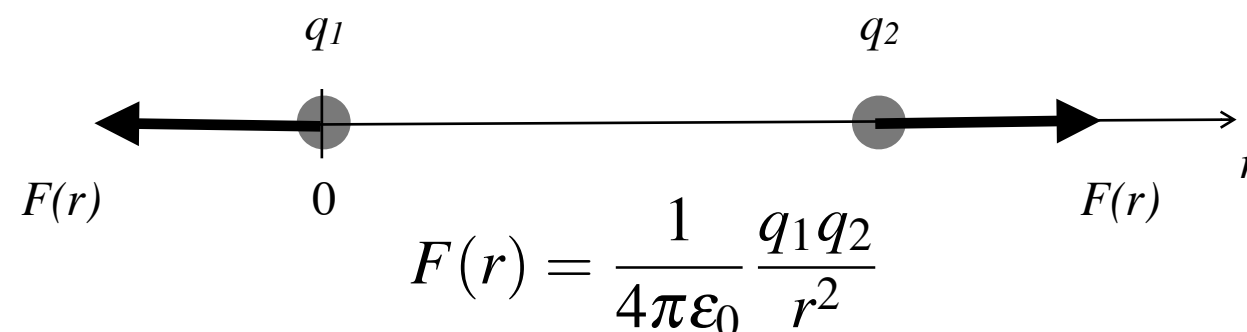
Proteins fold because they experience simple physical forces.



It is difficult to predict the structure (and therefore function) of a protein from its primary sequence

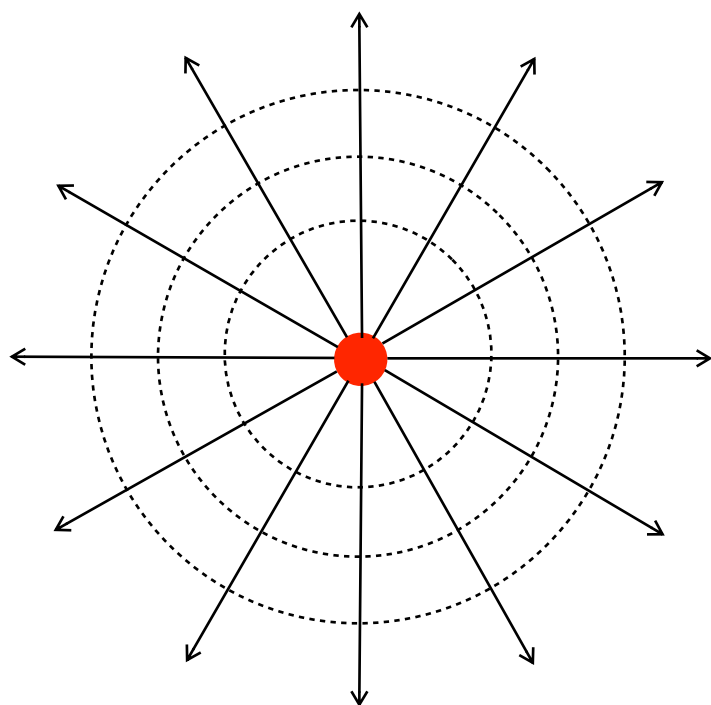


Electrostatics is the fundamental attractive force between atoms



$$F(r) = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2}$$

Coulomb's Law, electrostatic potential energy, the electric field and the electrostatic potential allow us to characterise the interactions between permanent electrical point charges.



The dielectric constant of water is much higher than that of protein or other biological macromolecules

