Intrinsically Disordered Proteins

Are proteins always structured ?

instrinsically disordered, natively disordered, natively unstructured..

Can a protein have no structure and be useful ?

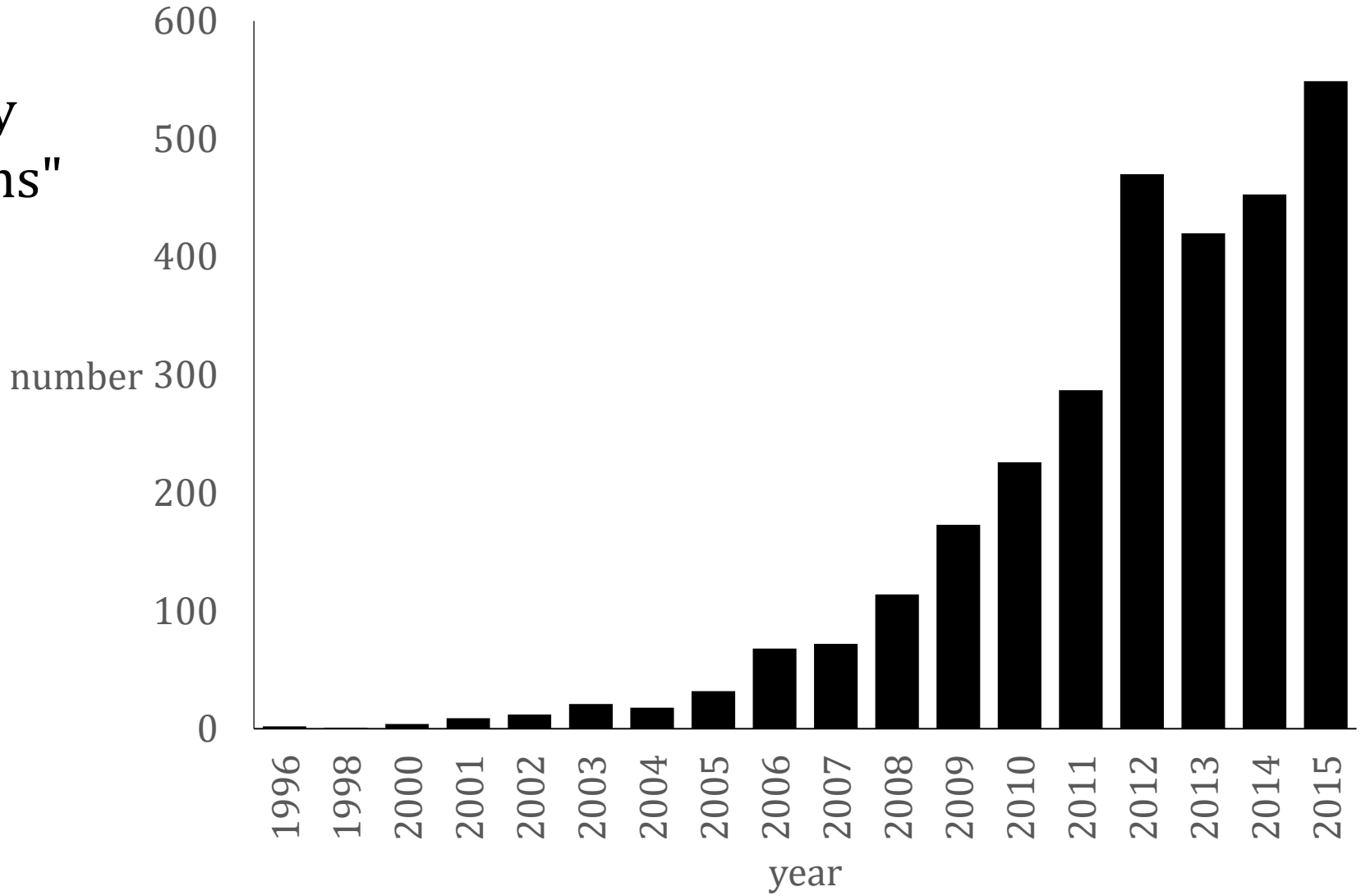
The real question

- are there proteins which do something (message, catalysis)
- have almost no fixed structure in solution
- maybe take on a structure when they meet a ligand or other protein ?

Fashion

A trendy discovery

How many papers with "instrinsically disordered proteins" each year ?



Why discuss them ?

Claim

- there are proteins whose native structures we will never see
- might adopt a conformation when they bind something
  - DNA, other protein, other ligand
- contradict classic belief – function requires well-defined structure

Maybe they are important - claim

- 44 % of human protein-coding genes have a disordered segment more than 30 residues long

What do they do ?

- molecular recognition, molecular assembly, protein modification, and entropic chains
- sites for modifications (phosphorylations, cleavage, ..)

What properties would you expect ?

The features that you see from non-native, unhappy proteins

- gel filtration – not sharp peaks
- proteases – very susceptible
- folding / unfolding curves – no specific unfolding temperature as with most proteins

The difference ?

- despite appearing unhappy, they still have some function

Example functions - enzymes

Could you have an enzyme with little structure ?

- normal picture
  - well-defined structure positions groups in space so they catalyse some reaction

- bad examples
  - RNase E – has structured domain + large unstructured region
  - hepatitis C virus NS3 protease – Zn^{2+} necessary to fold and make active

  - is there evidence that the unstructured protein does anything interesting ?

Example functions - binding

- DNA – binding proteins – take on structure only after binding
- targets for modification – maybe being exposed makes them easier to phosphorylate, acetylate, …
  - these changes lead to activation / inactivation
- small ligand storage.. Zn^{2+} example coming

What is disorder ? When is it significant ?

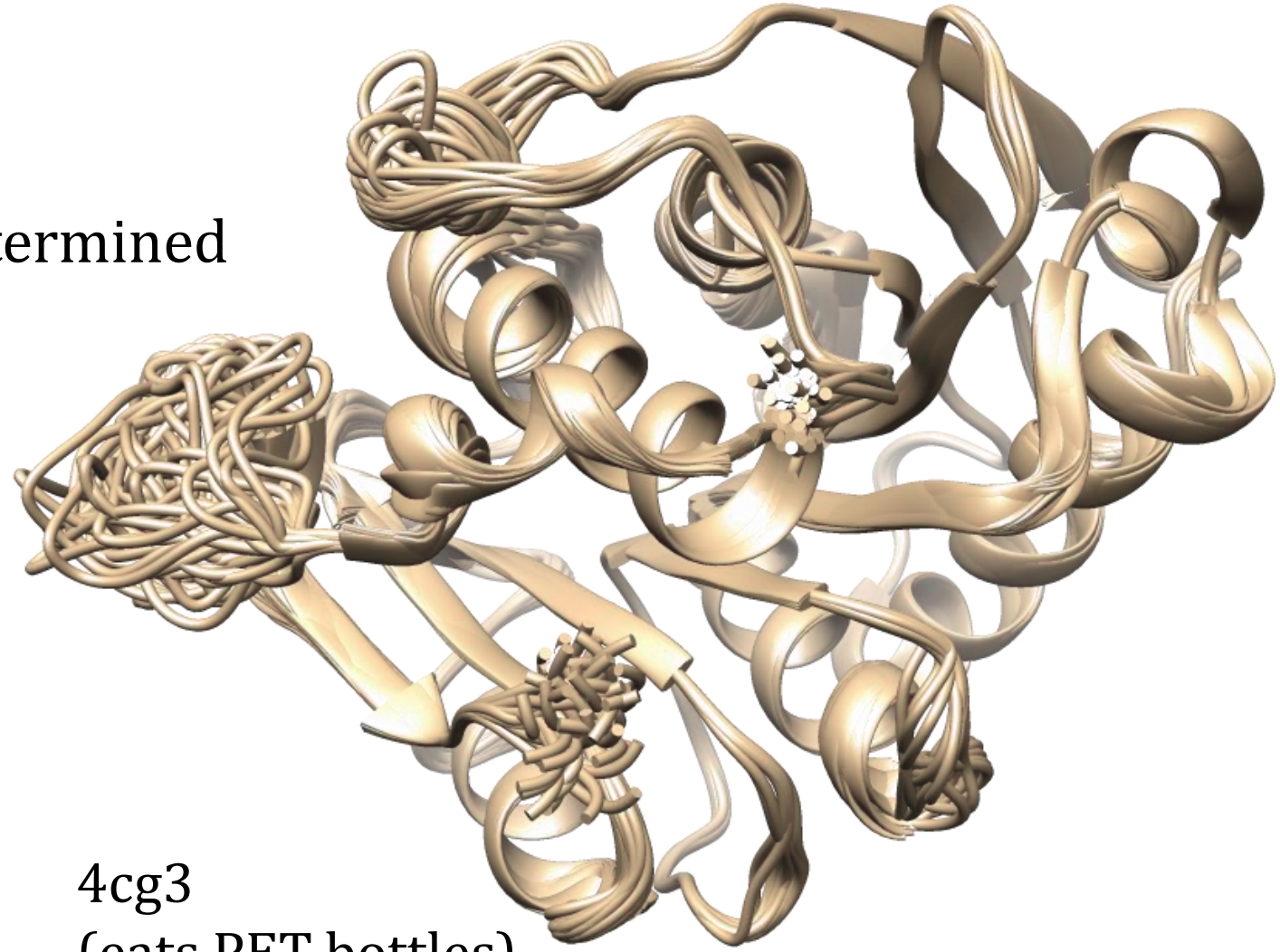
uncertainty

typical NMR structure

- 50 models
- some parts of structure well-determined

- some very mobile

- just a feature of NMR ?



4cg3
(eats PET bottles)
polyethylene therephtalate degrading hydrolase

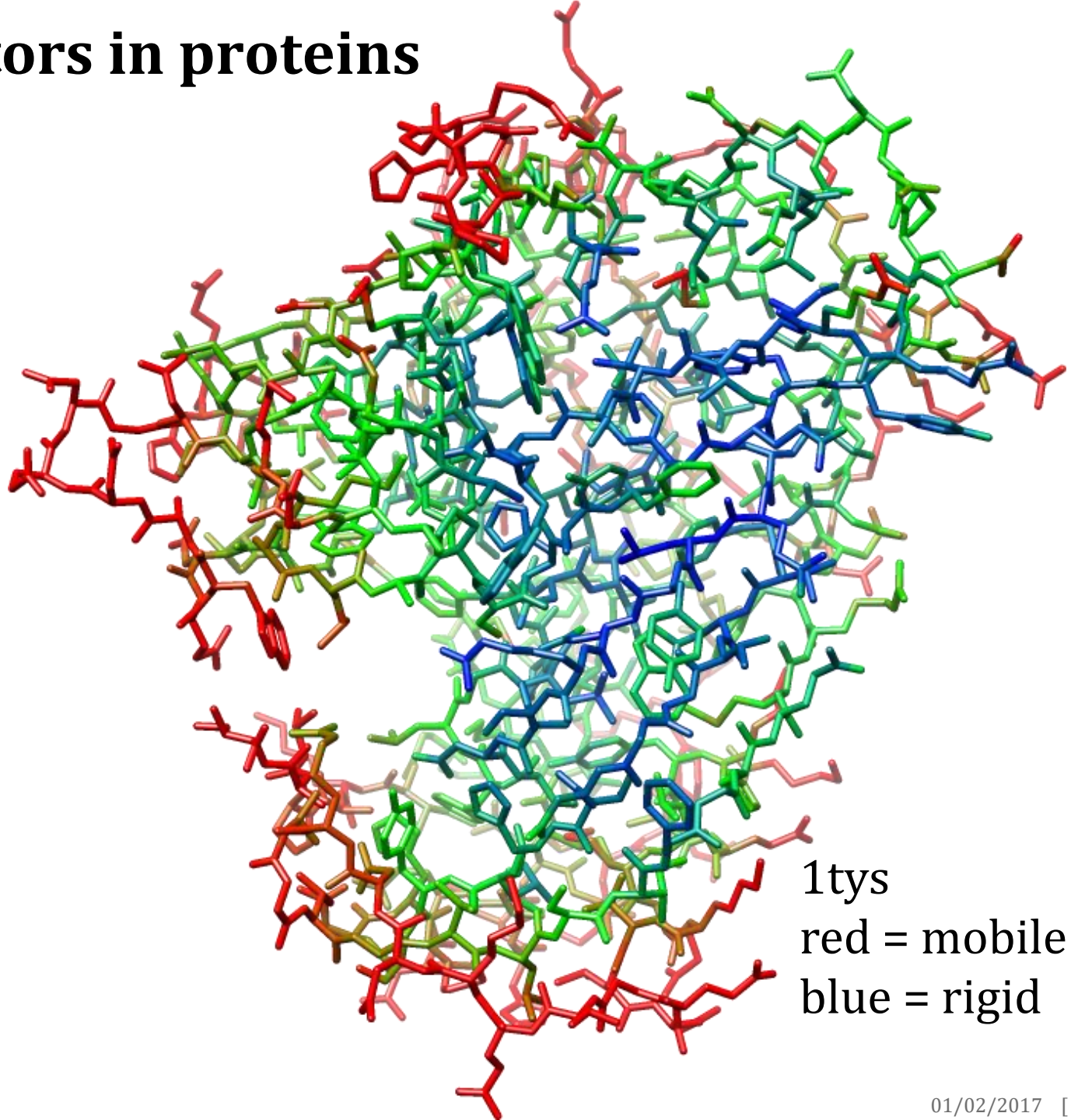
B factors in proteins

B-factors

- better measure of mobility

- crystallographers have always known that some parts of structure are mobile

- what about really mobile ?



1tys
red = mobile
blue = rigid

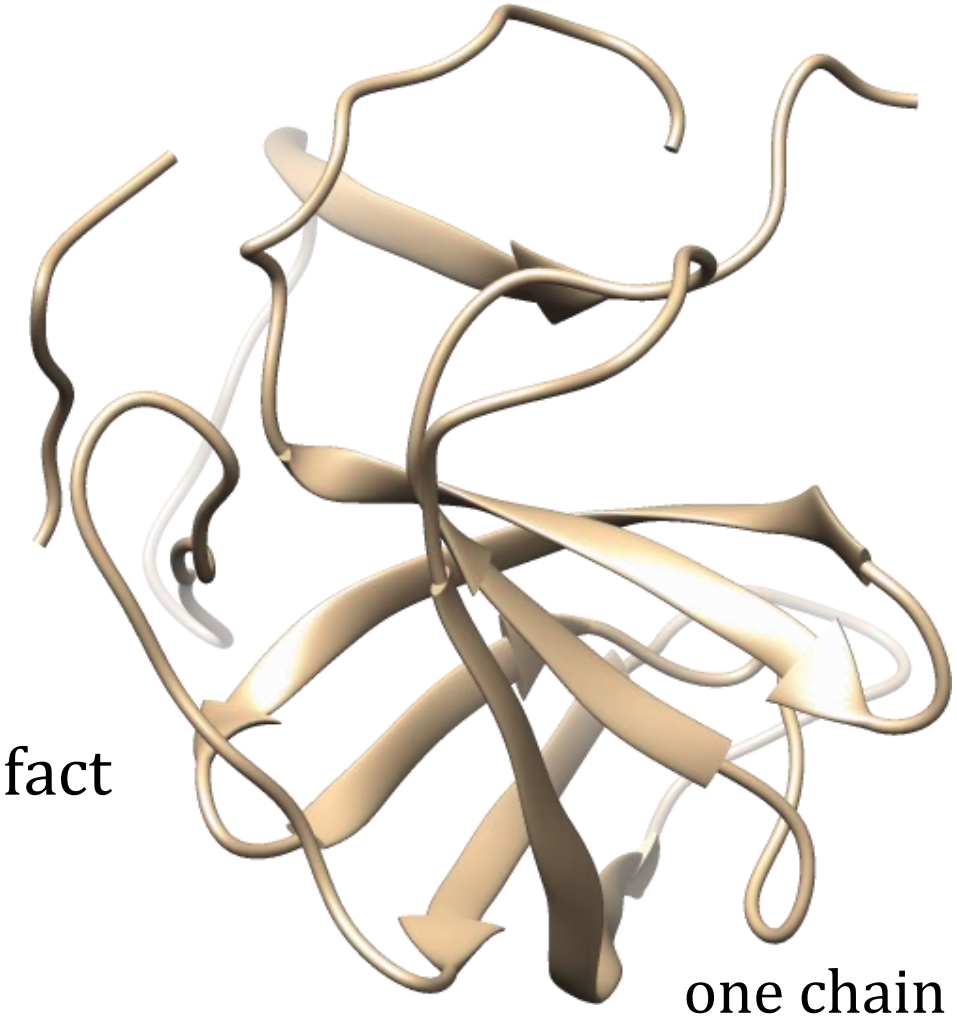
Mobile pieces

History of structures like chymotrypsin

- easy to obtain, crystallise
- lots of structures

- always missing regions

- almost all chymotrypsin structures have this region
  - does not seem to be an experimental artefact



one chain
from 2cha

do intrinsically disordered proteins exist ?

In perspective

- all atoms move at room temperature
- some parts of a structure are relatively unrestrained
- are there large parts which are unrestrained under normal conditions ?

Why be suspicious

Why do you not see structure ?

- It is easy to denature some proteins
- most proteins are marginally stable

- speak to a molecular biologist – my protein does not fold
  - because the protein is unfolded when native (intrinsically unfolded) or
  - you have wrong conditions ?

What are conditions ?

- cells are very crowded
- sometimes exotic
  - organelles (mitochondria – low pH)
  - electric fields near membranes (think of nerve cells)

What if you do not see structure

NMR

- NOE's – distances between residues
- not easy to measure
- signal to noise
- technical reasons for not seeing NOE (τ_c correlation time – bad luck)
- your sample is bad

Crystallography

- not easy – most proteins need slightly unusual conditions

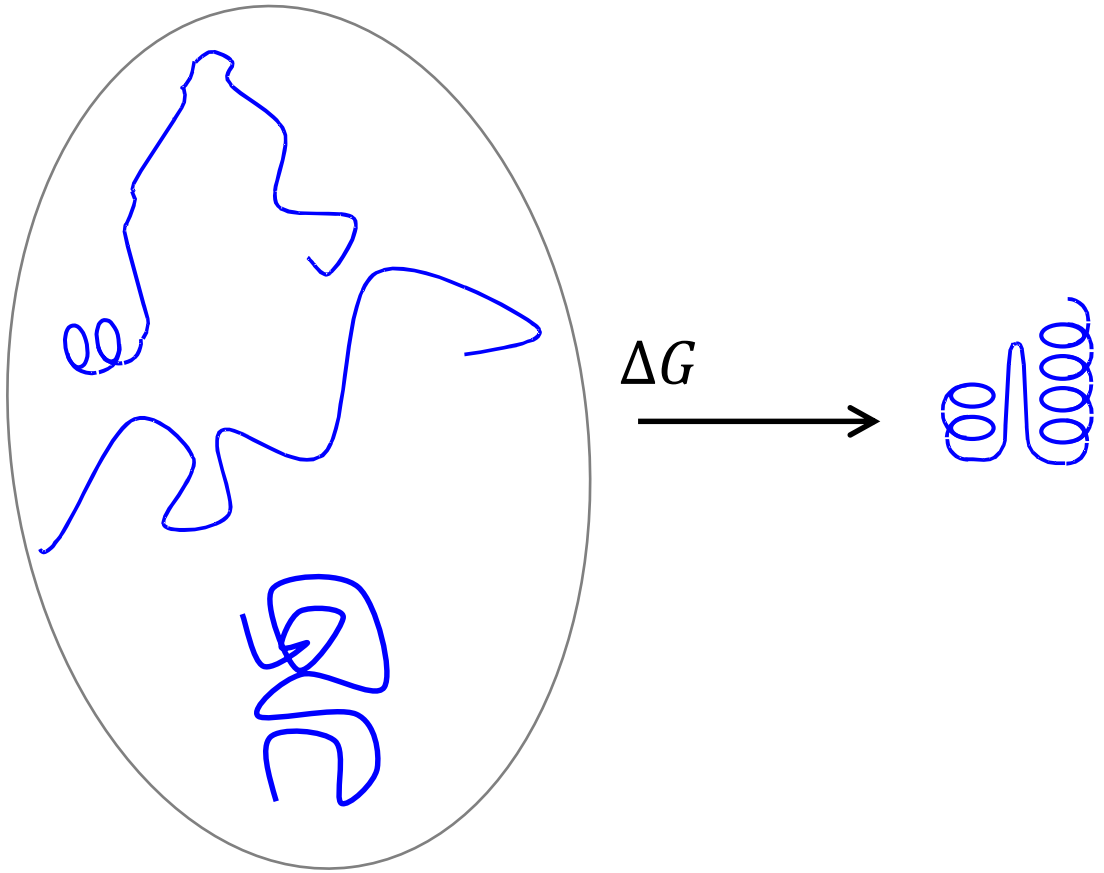
What if I do not see structure ?

- maybe measurements are under wrong conditions

What does unstructured mean ?

Most chemistry is an equilibrium

- folding $\Delta G = RT \ln \frac{[\text{folded}]}{[\text{unfolded}]}$
- ΔG is usually small



- if $K_{fold} = \frac{[\text{folded}]}{[\text{unfolded}]} = 10^3$
  the unfolded is like a 1 in 1000 inpurity

- if $K_{fold} = 1$
  - then your protein is 50% present, might be enough to work
  - may not look structured in NMR, may not crystallise
- does this explain “transient structure” ?

What would you see experimentally

- crystallography ? nothing – boring
- NMR ? What you see from denatured proteins
  - broad lines
  - no NOEs (distances)
  - evidence of mobility ?

Better

- measurements that look at overall shape
- radius of gyration $r_g = \left(\frac{1}{N}\left(\sum_{i=1}^{N}|\vec{r}_i - \vec{r}_c|^2\right)\right)^{1/2}$

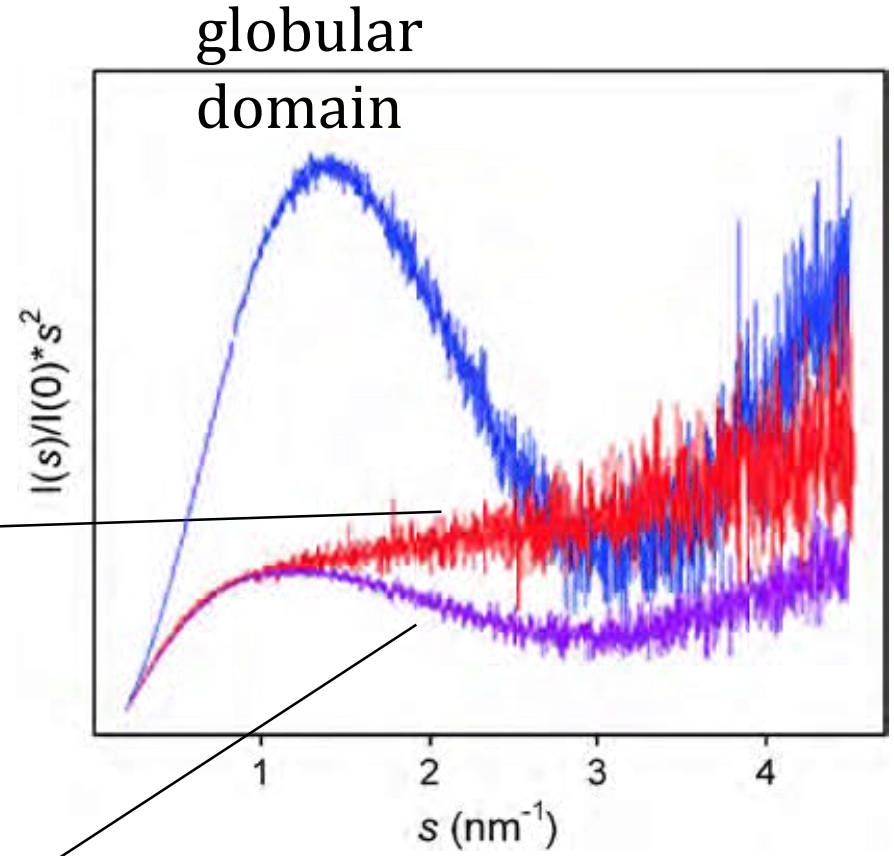
  where $\vec{r}_c$ is centre sum over N particles

- light scattering, small angle x-ray scattering (SAXS)

Example…

src-kinase

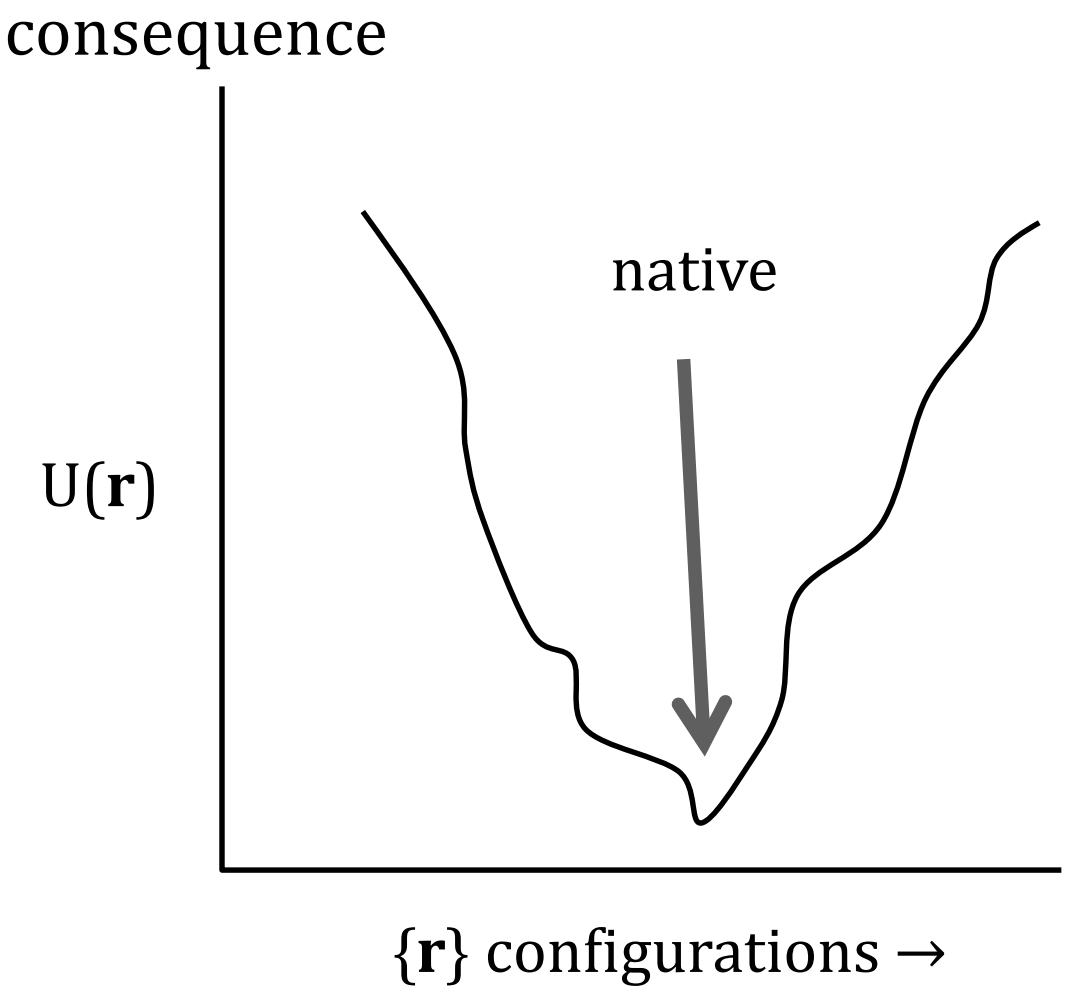
- certainly can find evidence of lack of structure
- Krattky plots for
  - globular domain
  - unstructured domain
  - the two in one chain

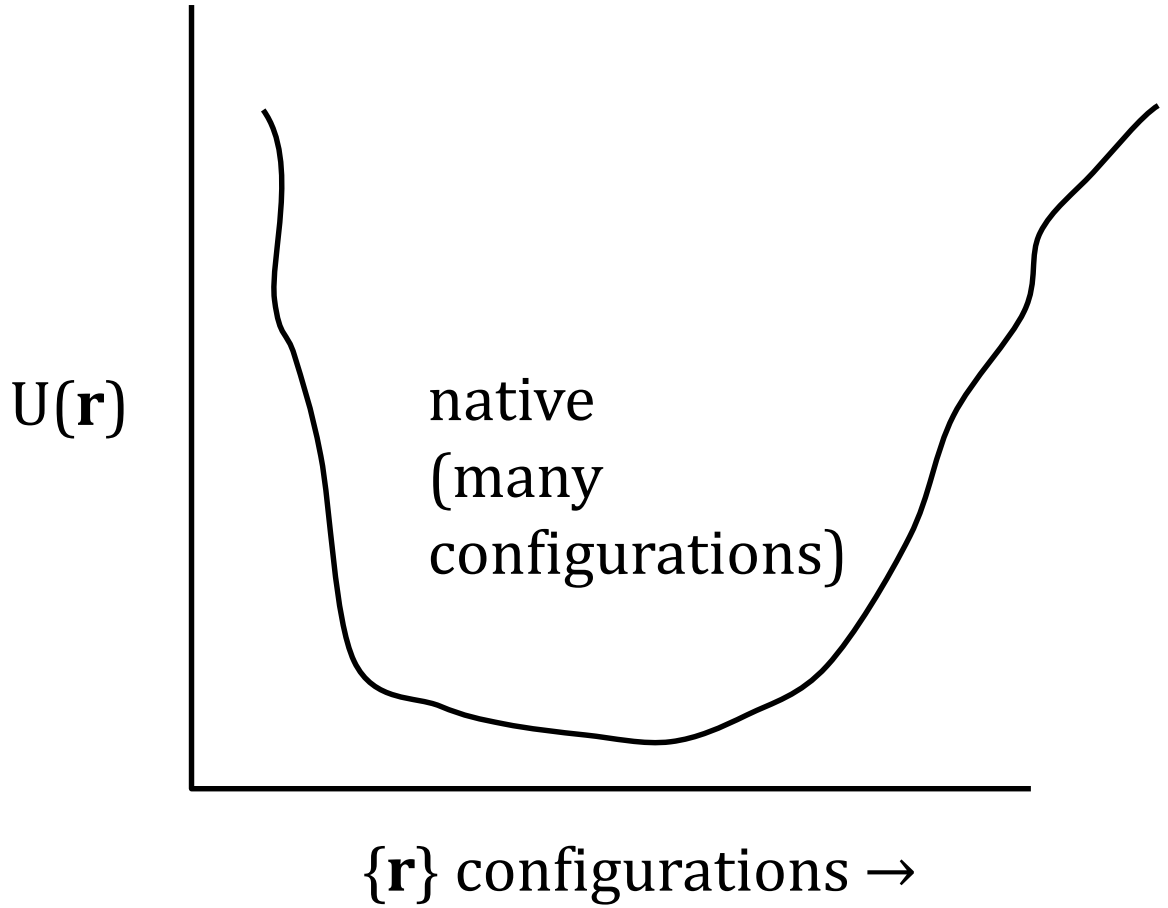


What must the energy landscape look like ?

Bernadó, P., Svergun, D.I. Mol. BioSyst, 2012, 8, 151-167

Energy landscape





what would one need for this kind of landscape ?

Flat energy landscape

Two possibilities

1. particles like to interact with each other, but they do not care with whom

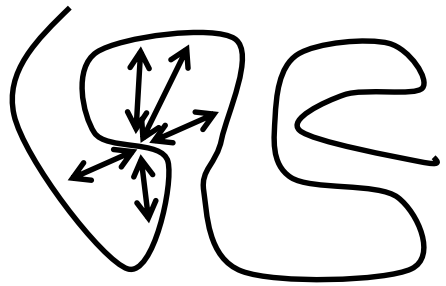
2. particles interact with solvent

   …. some detail

Particles interact with each other

Non-specific

- a particle interacts equally with many other sites
- simplest extreme example
  - $(leu)_n$ or $(trp)_n$ each hydrophobic residue likes to pack with other hydrophobic residues
  - would not be soluble
- maybe realistic
  - "low complexity"
  - similar sets of residues repeat
  - DSKDSKDSK.. DSKDSK…
  - lots of interactions of similar energy



  - statistically – many proteins have regions of low complexity

particles interact with solvent

if side chains are very soluble, there is no hydrophobic core

- does it happen ?

- prothymosin α (nuclear protein, interactions with chromatin, histones)
  - 109 residues, half are glu or asp, charge -54 at neutral pH
  - normal proteins have a small charge

  - binds lots of metal^{+} ions in a cell, has smaller r_g when Zn^{2+} present

  - other histone binding proteins have acidic domains

  - what will happen at low pH ?

r_g radius of gyration

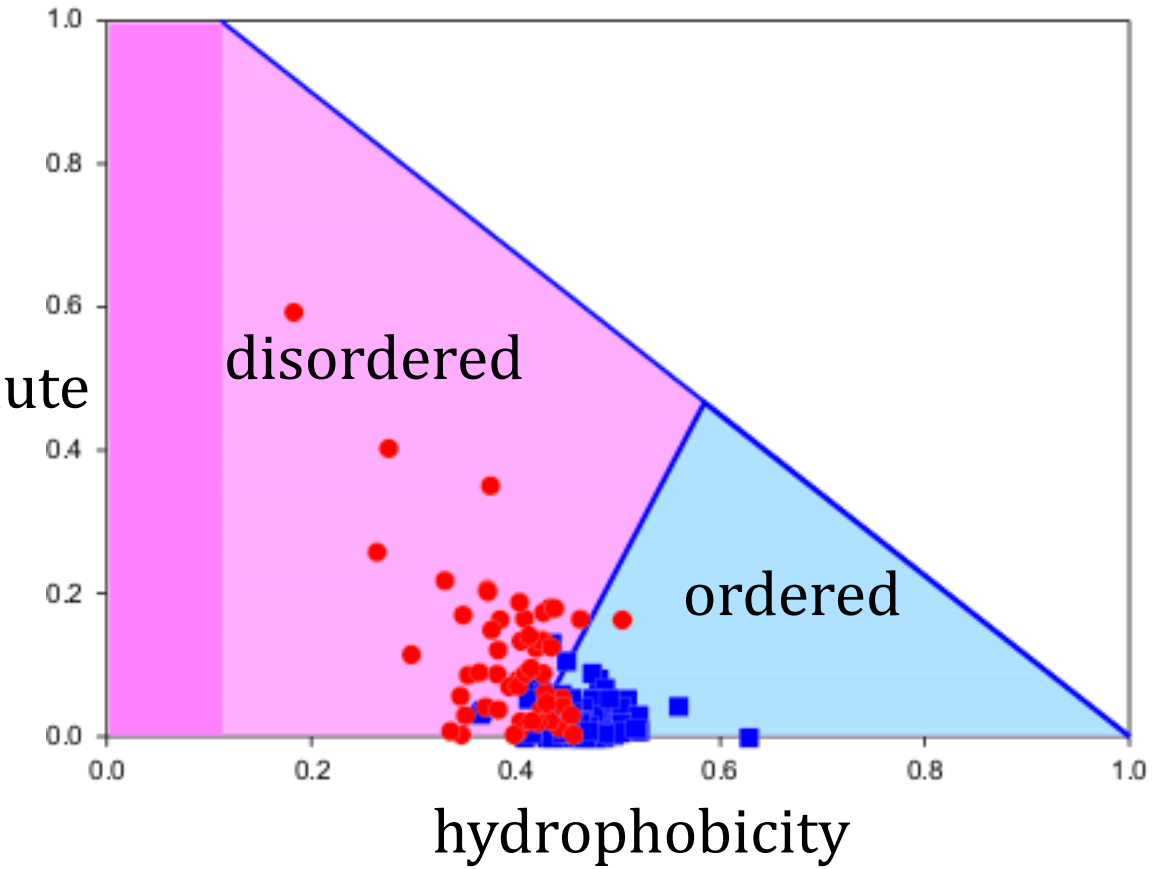
What else do we know ?

People collect lists – there are databases of disordered proteins

Sequence properties of intrinsically disordered proteins

- rich in polar and charged amino acids
- less than usual – bulky hydrophobic

Best property…



Uversky, Biochim Biophys Acta, 2013, 1834, 932-951

Can one predict disorder

For building and testing methods, one needs test set

- lists of largely disordered proteins *vs* normal proteins with known structures people collect these

OR

- take set of PDB crystal structures with B-factors
  - high B-factors = less rigid
  - try to predict B-factors

What descriptors would you use ?

Typical prediction method

Features

1. sequence complexity
2. hydrophobicity
3. charge
4. single residue propensity

a few words on each …

low complexity

example

something like

$$c = -\sum_i p_i \ln p_i$$ visiting each site over window of ≈ 40 residues

p_i is the frequency of each amino acid type i in the window

- summing over all overlapping windows gives per residue score

hydrophobicity and charge

- literature values for hydrophobicity (easy)
- charge – very easy

single residue propensity

- forget physics and chemistry just look at statistics
- visit all sites in protein data bank
- for each type of residue
  - collect average B factors
  - call this disorder propensity

  - residues missing in crystal structures are called "very disordered"

- end result ? table with big variation / error bars
- likelihood of an amino acid type to be disordered

predictions

Methods

- neural networks, support vector machines (SVMs)
- classifiers (trees, bagging + boosting)
- elegant statistics ?

What comes out ?

- first, how do you measure it

Predictors – ROC curves

How good is your predictor ?

- that a medication works, that a signal is real, that amino acid is disordered ?

A good predictor

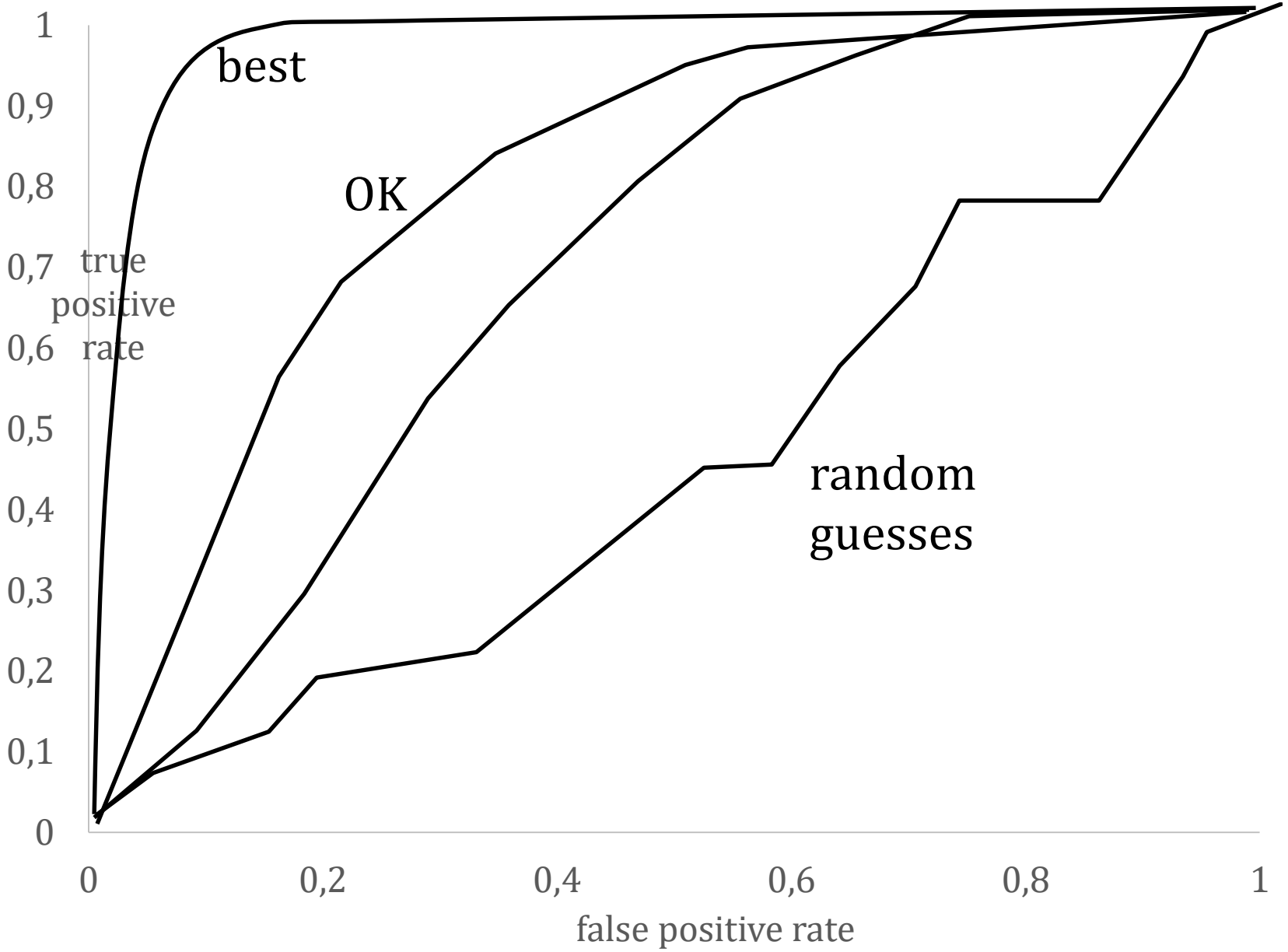
- the first 8 predictions are correct, then one wrong, then…

A bad predictor

- the first 1 is correct, next is wrong, then 1 correct, then one wrong..

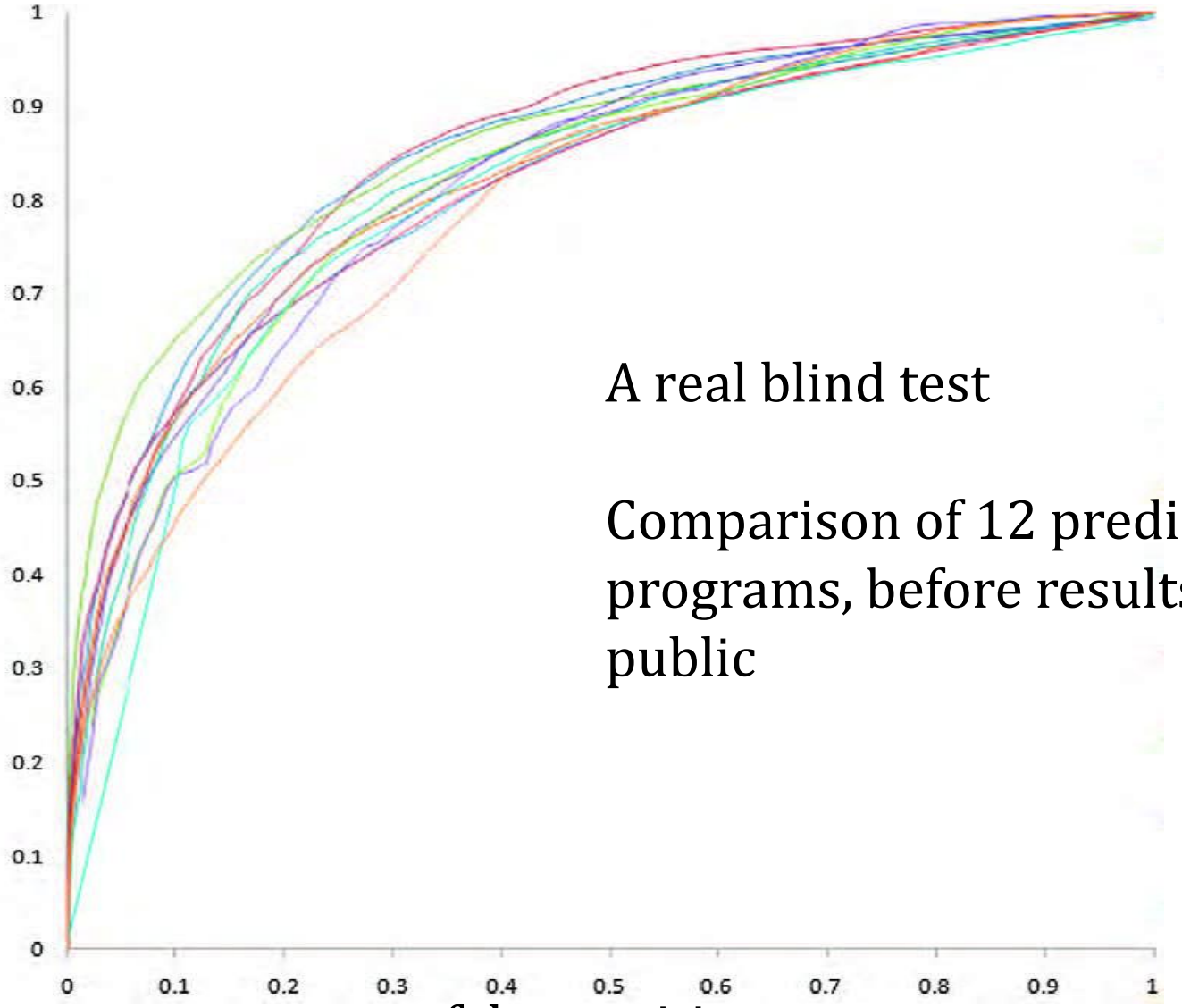
Turn this into a curve

- When 1 % of my predictions are wrong, how many are correct ?
  - when 2 % of my predictions …


ROC curve
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
true positive rate
best
OK
random guesses
0
0,2
0,4
0,6
0,8
1
false positive rate

How good are predictions ?

true positive



A real blind test

Comparison of 12 prediction programs, before results were public

12 of the best predictors

Deng, X., Eickholt, J., Cheng, J. Mol. BioSyst., 2012, **8**, 114–121

interpretation

No surprises

- we already knew that charge and hydrophobicity were good predictors

Is any of this real ?

- It is very fashionable (100s of literature articles)

Imagine a protein has

- little structure in test tube, finds structure when binding DNA, other protein
  - is this really different to a protein which is more structured in test tube + changes structure a bit upon binding ?

- What if a protein folds at low pH ?

For Klausur

- what is an intrinsically unstructured protein ?
- why is there no clear definition ?
- what experimental methods would recognise the property ?
- what are implications for the energy landscape ?
- how does a ROC curve work ?
- how would one estimate sequence complexity ?