Proteins: between order and disorder

SAPIENZA
UNIVERSITÀ DI ROMA

Proteins are interdisciplinary molecules

Proteins are at the border between Physics and Biology.
- They are complex and versatile molecular machines that obey the laws of statistical physics.
- They are evolving machines. Protein function emerges from molecular interactions and is the target of natural selection.

Proteins are essential for life as we know it, they are in almost all cellular processes and cellular structures

Proteins are fascinating because they bridge conceptual worlds:
- Quantum and classical physics
- Individual macromolecule and statistical description
- Dynamical ensemble and single structure
- Physics (statistical mechanics) and biology (evolution)

From a talk by Ugo Bastolla
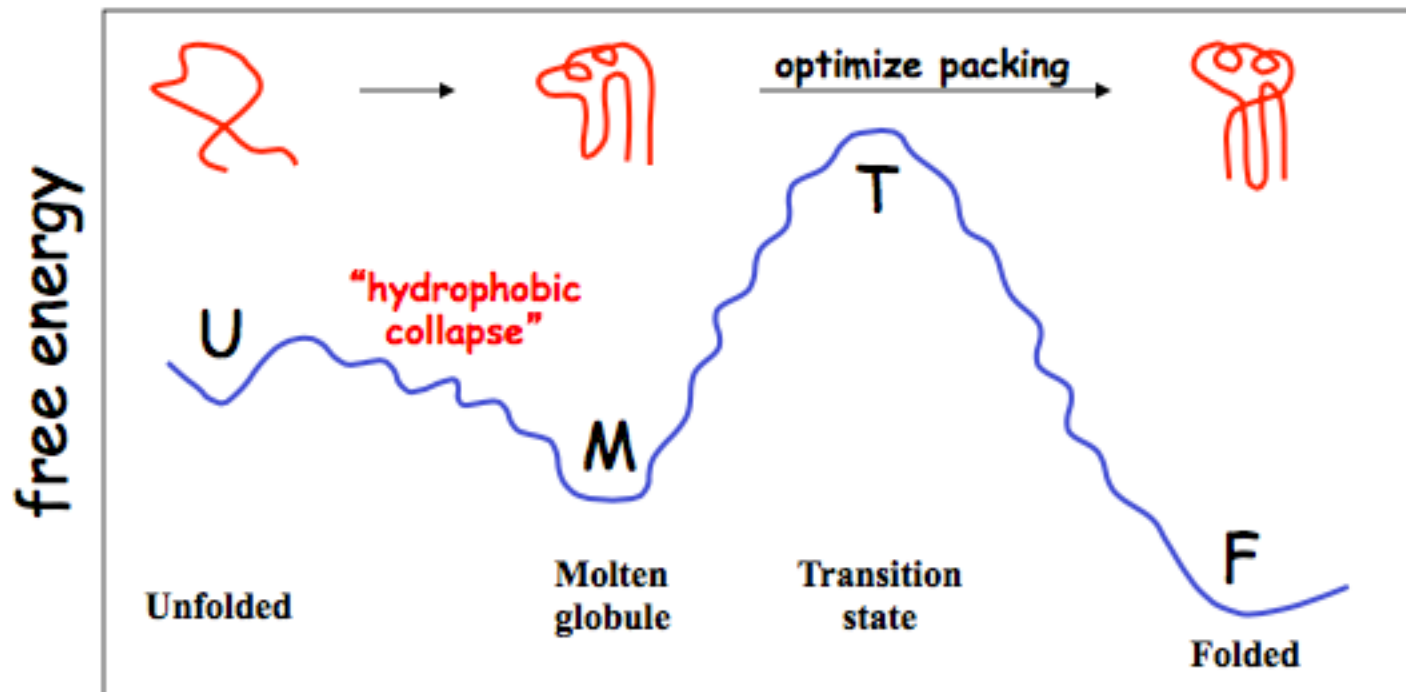
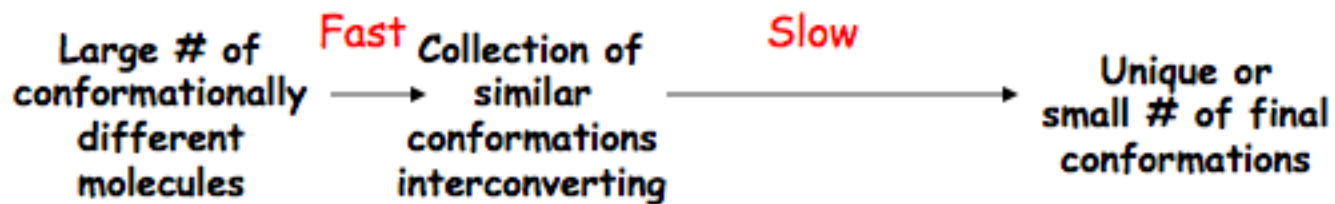## Clues on Protein Evolution

Proteins seem to evolve respecting two criteria (so to speak): minimally perturbing the solvent and contrasting aggregation

The evolutionary logic of proteins sequences/molecules
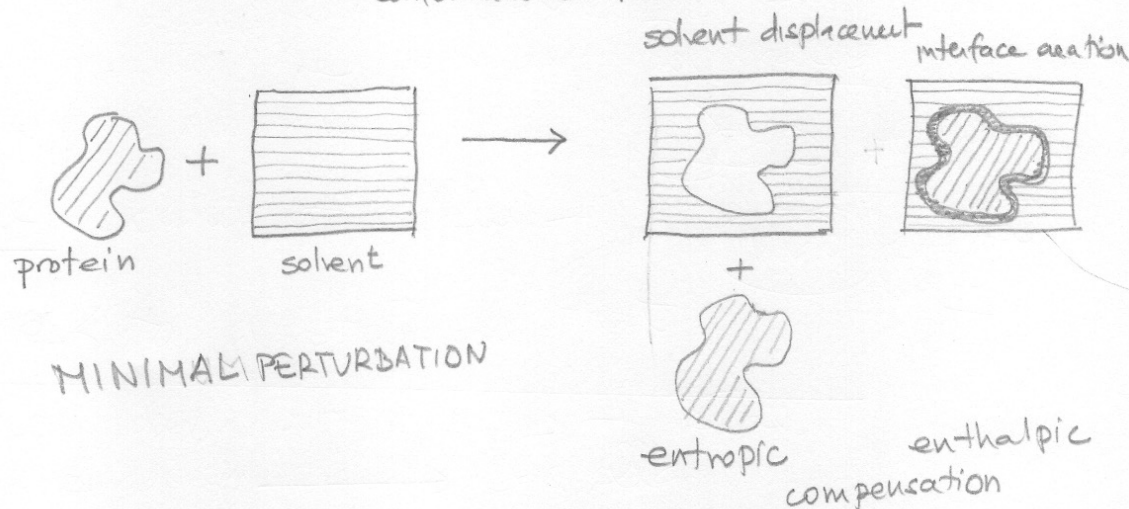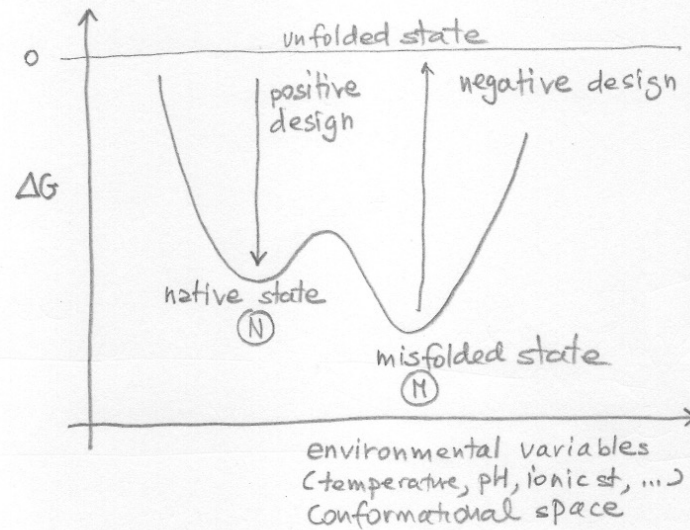
Proteins seem to evolve respecting two criteria (so to speak): minimally perturbing the solvent and contrasting aggregation

An energetic view of the folding process

Large # of conformationally different molecules → **Fast** → Collection of similar conformations interconverting → **Slow** → Unique or small # of final conformations

free energy

"hydrophobic collapse"

U — Unfolded

M — Molten globule

T — Transition state

F — Folded

optimize packing

folding trajectory

↑ Local secondary structures form first

Adapted from Branden & Tooze

POSITIVE/NEGATIVE DESIGN

Between Scylla (unfolding) and Cariddi (misfolding)

For **natively folded proteins**, the native state must be stable both against the **unfolded** state and against **wrongly folded** (misfolded) compact conformations.

Positive design: features of protein sequences and structures that increase the stability of the native state against the unfolded state

Negative design: Features that decrease the stability of wrongly folded structures.

ELSEVIER

CrossMark

References

A good introduction to structural biology =
Introduction to Protein Structure
        - Carl Branden & John Tooze

Web resources:
Protein Data Bank = > 24,000 protein structures, atomic
        coordinates, & the "protein of the month"
        http://www.rcsb.org/pdb

CATH/SCOP protein structure hierarchies:
        http://www.biochem.ucl.ac.uk/bsm/cath/
        http://scop.mrc-lmb.cam.ac.uk/scop/

Some Protein Interaction Resources on the Internet

Protein interaction databases

Biomolecular Interaction Network Database (BIND)

ae

http://www.blueprint.org/bind/bind.php

Currently 73,000 interactions

Database of Interacting Proteins (DIP)

http://dip.doe-mbi.ucla.edu

Currently 44,000 interactions

Protein Quaternary structure database (PSQ)

http://pqs.ebi.ac.uk

Atomic structures of interacting proteins

Interactive visualization of networks

Cytoscape:

http://www.cytoscape.org

Interactive display of protein networks

LGL (Large Graph Layout):

http://bioinformatics.icmb.utexas.edu/LGL

Visualization of networks with up millions of edges, 100,000's of vertices

Alpha-helices appear a lot in trans-membranal proteins
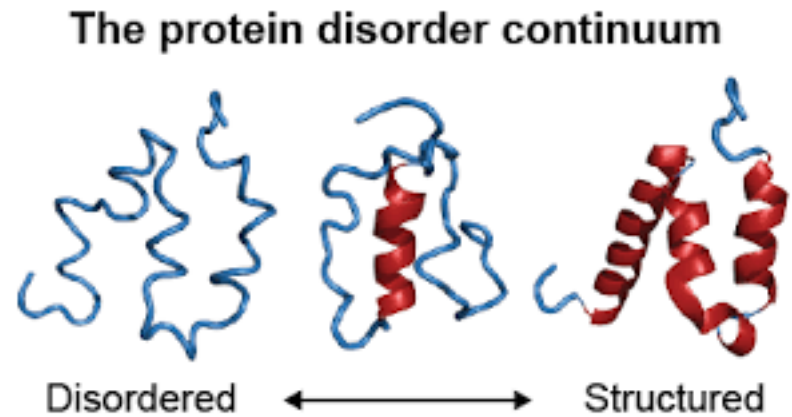

membrane
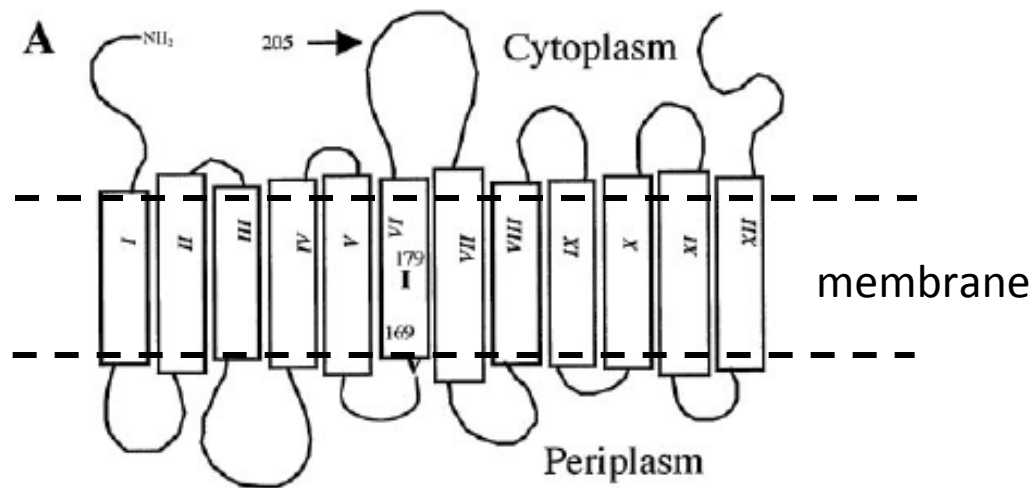
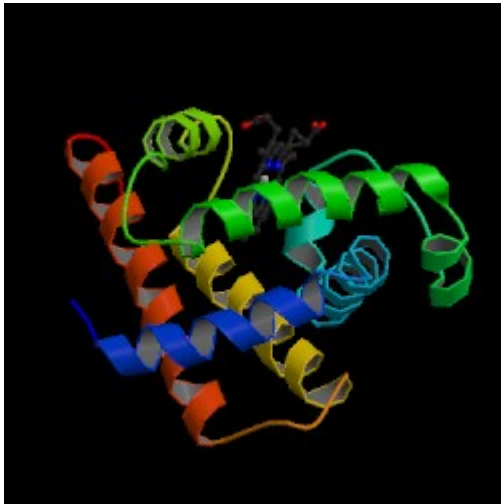1pv6.pdb

E.g. Lactose permease (LacY)

Structure/Unstructure
is quite basic in protein science

The protein disorder continuum

Disordered ⟷ Structured

Intrinsically disordered proteins

Myoglobin (1958) x-ray structure M.F. Perutz

Exercise go to PUBMED https://pubmed.ncbi.nlm.nih.gov
and search for " intrinsically disordered proteins"

See also:

Deiana A, Forcelloni S, Porrello A, Giansanti A (2019) Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. PLoS ONE 14(8): e0217889. https://doi.org/10.1371/journal.pone.0217889

A

B
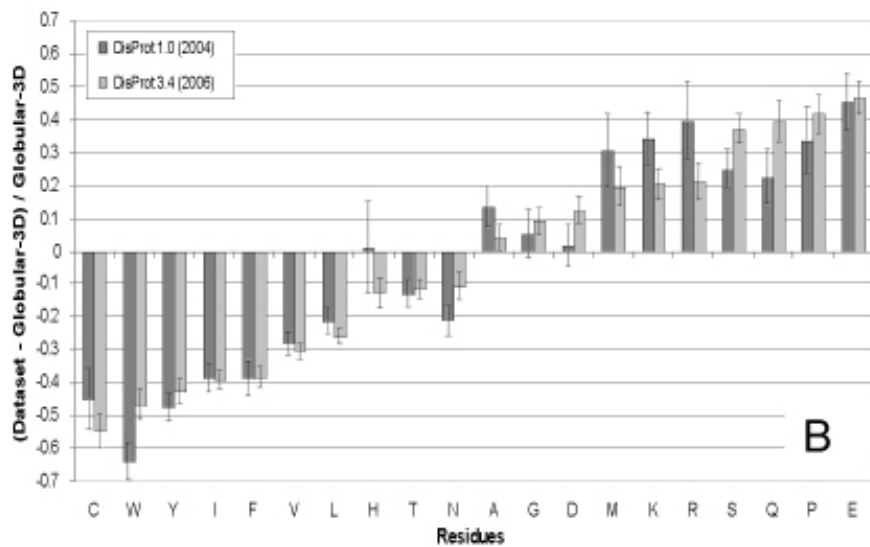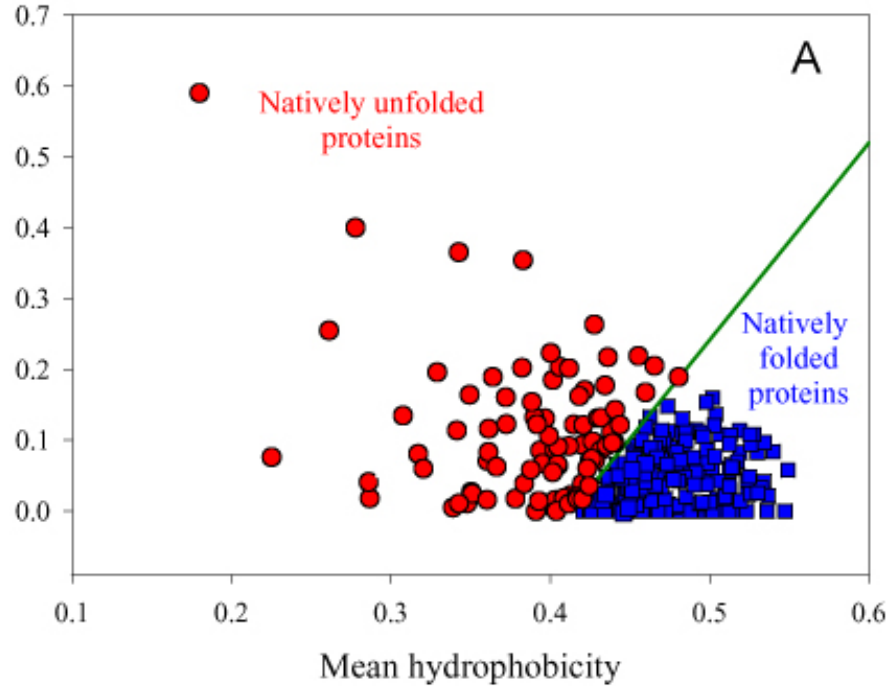
Dunker, A & Oldfield, Christopher & Meng, Jingwei & Romero, Pedro & Yang, Jack & Chen, Jessica & Vacic, Vladimir & Obradovic, Zoran & Uversky, Vladimir. (2008). **The unfoldomics decade: an update on intrinsically disordered proteins** <http://dx.doi.org/10.1186/1471-2164-9-S2-S1. BMC genomics. 9 Suppl 2. S1.

10.1186/1471-2164-9-S2-S1.

Protein Folding

Classic experiment from 1960's (Chris Anfinsen):
    Purified small protein RNaseA,
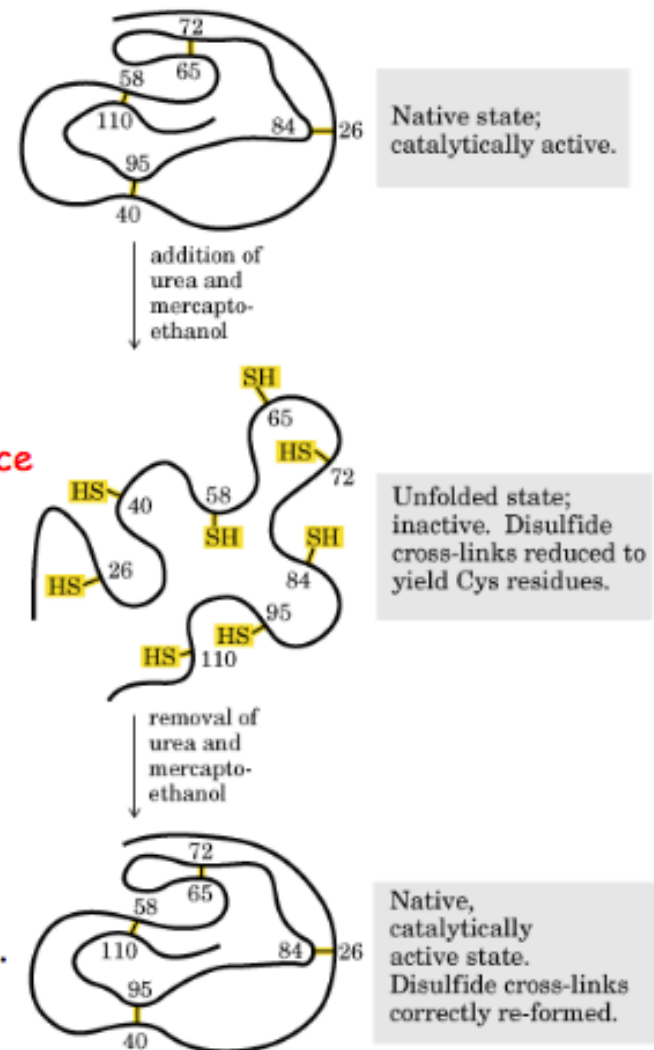   Refolded in a few minutes in solution
==> all information necessary for correct folding
        was captured in the linear amino acids sequence

Native state; catalytically active.

addition of urea and mercapto-ethanol

Unfolded state; inactive. Disulfide cross-links reduced to yield Cys residues.

removal of urea and mercapto-ethanol

Native, catalytically active state. Disulfide cross-links correctly re-formed.

Corollary:
Proteins do **not** fold by randomly testing conformations.

Given a 100 amino acid protein,
        & 10 possible conformations / amino acids
                = 10^{100} possible conformations for the protein
==> not possible to randomly sample, clearly constrained search

    See Anfinsen1973

**N.M. Luscombe,
D. Greenbaum,
M. Gerstein**

Department of Molecular Biophysics
and Biochemistry
Yale University
New Haven, USA

Review

What is bioinformatics? An introduction and overview

Abstract: A flood of data means that many of the challenges in biology are now challenges in computing. Bioinformatics, the application of computational techniques to analyse the information associated with biomolecules on a large-scale, has now firmly established itself as a discipline in molecular biology, and encompasses a wide range of subject areas from structural biology, genomics to gene expression studies.

In this review we provide an introduction and overview of the current state of the field. We discuss the main principles that underpin bioinformatics analyses, look at the types of biological information and databases that are commonly used, and finally examine some of the studies that are being conducted, particularly with reference to transcription regulatory systems.

**N.M. Luscombe,
D. Greenbaum,
M. Gerstein**

Department of Molecular Biophysics
and Biochemistry
Yale University
New Haven, USA

Review

What is bioinformatics? An introduction and overview

Abstract: A flood of data means that many of the challenges in biology are now challenges in computing. Bioinformatics, the application of computational techniques to analyse the information associated with biomolecules on a large-scale, has now firmly established itself as a discipline in molecular biology, and encompasses a wide range of subject areas from structural biology, genomics to gene expression studies.

In this review we provide an introduction and overview of the current state of the field. We discuss the main principles that underpin bioinformatics analyses, look at the types of biological information and databases that are commonly used, and finally examine some of the studies that are being conducted, particularly with reference to transcription regulatory systems.