BCHS 6229
Protein Structure and Function

Lecture 4 (October 20, 2011)

From Structure to Function (I):

Flexibility and Protein Function
Intrinsically Unstructured Proteins and their Functions

The folded state has a flexible structure

The protein molecule does not have a static rigid structure at normal temperature. Instead, all the atoms are subject to small T-dependent fluctuation.

Breathing of the molecule - random or collective.

Motions are usually small (a few tenths of an Å); but sometimes large and very significant.

How such large collective movements are reflected in x-ray studies?

Insight into these individual and collective motions has been obtained by theoretical studies - MD

~picosecond for individual residues
~nanosecond for loop regions

Such movements: very important for the functions of many proteins.

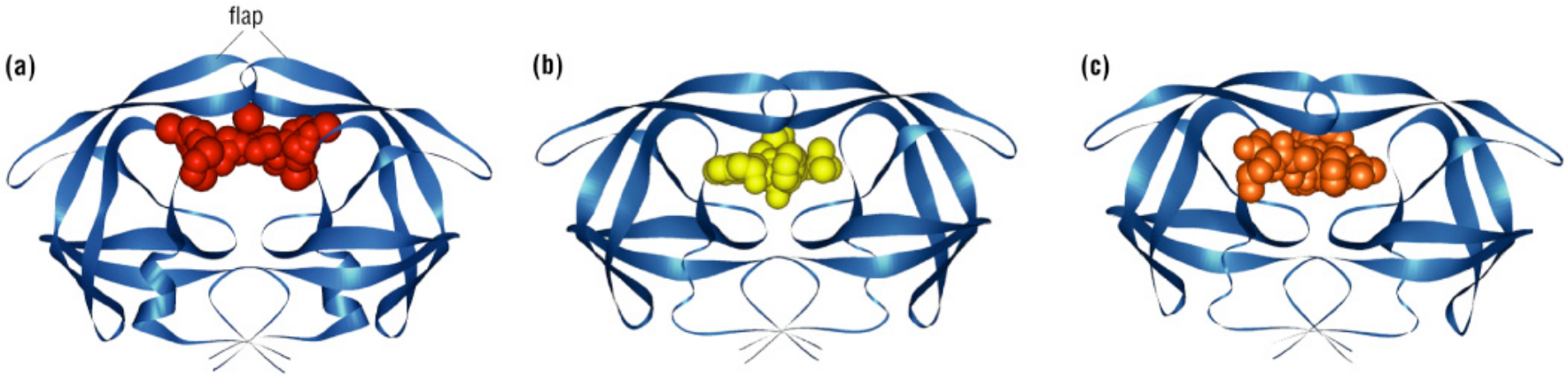
There can also be larger conformational changes between different functional states of the molecule.

(pH, ligands)

The flexibility of tertiary structure allows proteins to adapt to their ligands.

Induced fit: originally, the change in the structure of an enzyme, induced by binding of the substrate, that brings the catalytic groups into proper alignment. Now generalized to the idea that specific ligands can induce the protein conformation that results in optimal binding interactions.

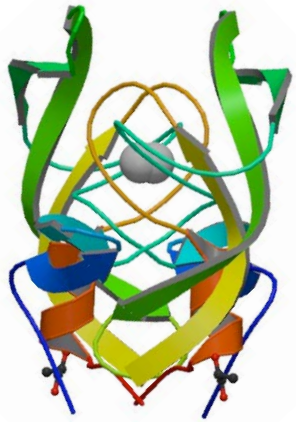
Inherent flexibility of proteins → Conformational changes



Haloperidol Crixivan Peptide analog

HIV protease bound to three different inhibitors.

Each inhibitor has a quite different structure, yet all bind tightly to the active site and induce closure of a flap that covers it.



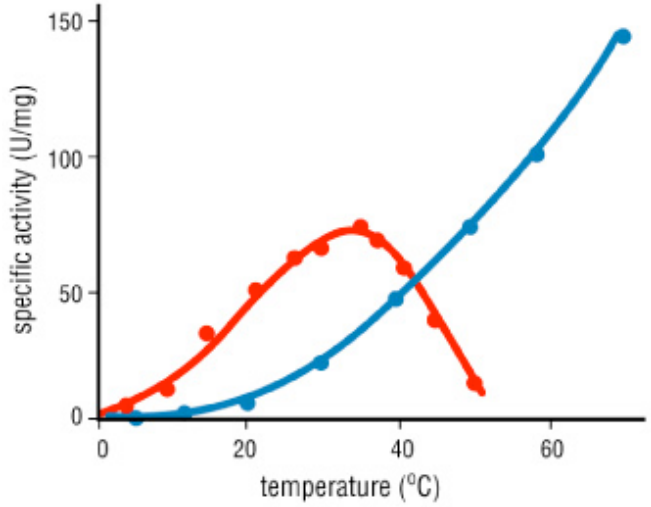
Protein flexibility is essential for biochemical function

In most proteins, binding is followed by some action (chemical transformation of L, conformational change in proteins, translocation of the protein, etc.)

In all cases, the structure of the protein must be flexible enough that the net free energy released by binding and/or chemical transformation of the L can derive the required changes in the protein's structure and function

Binding: K_d = ~10^{-3} M (weak) to ~10^{-12} M(extremely strong)

Protein flexibility is essential for biochemical function



Difference in T-dependence of the specific activity of glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Some mutant enzymes are stable at higher T than normal T and rigid (can abolish functions).

```
Yeast       MAITVGINGFGRIGRLVLRIALSRK--DIQIVAINDPFIAPEYASYMFKYDSTHGRYSGEVSHEGENIVIDGKKIRVYQERDPVNIPWGKDGVDYVIDST
Tmari       MAR-VAINGFGRIGRLVYRIIYERKNPDIEVVAIND-LTDTKTLAHLLKYDSVHKKFPGKVEYTENSLIVDGKEIKVFAEPDPSKLPWKDLGVDFVIEST
            **  *.*********** **  .**  **::***** :  .:  :::::****.* ::.*:*.:  :.:::***:*:*: * ** ::** . ***:**:**

                      110       120       130       140       150       160       170       180       190       200
                       |         |         |         |         |         |         |         |         |         |
Yeast       GVFKELDSAQKHIDAGAKKVVITAPS-STAPMFVVGVNEDKYTPDLNIISNASCTTNCLAPLAKIINNKFGIEEGLMTTVHSITATQKTVDGPSHKDWRS
Tmari       GVFRNREKAELHLQAGAKKVIITAPAKGEDITVVIGCNEDQLKPEHTIISCASCTTNSIAPIVKVLHEKFGIVSGMLTTVHSYTNDQRVLDLP-HKDLRR
             ***:: :.*: *::******:****: .    .*:* ***: .*: .*** ******.:**:.*::::**** .*::***** *  *:.:* * *** *

                      210       220       230       240       250       260       270       280       290       300
                       |         |         |         |         |         |         |         |         |         |
Yeast       GPTASGNIIPSSTGAAKAVGKVIPELAGKLTGMSLRVPTVDVSVVDLTVKLLKDATYDEIKAAVKEAAEGPLKGVVGYTEDQVVSSDFLTDNRSSIFDAE
Tmari       ARAAAVNIIPTTTGAAKAVALVVPEVKGKLDGMAIRVPTPDGSITDLTVLVEKETTVEEVNAVMKEATEGRLKGIIGYNDEPIVSSDIIGTTFSGIFDAT
            . :*: ****::*******. *:***: *** **::**** * *:.**** : *:** :*:**.:***:** ***:**.:: :*****::  . *.****

                      310       320       330
                       |         |         |
Yeast       AGIWLSPRFVKLIAWYDNEYGYSTRVVDLLEYVASKN
Tmari       ITNVIGGKLVKVASWYDNEYGYSNRVVDTLELLLKM-
                 :. ::**: :*********.**** ** : .
```

Identity (*) : 170 is 50.45 %
Strongly similar (:) : 74 is 21.96 %
Weakly similar (.) : 27 is 8.01 %
Different : 66 is 19.58 %
yeast (334 residues).
Thermotoga maritima (333 residues).

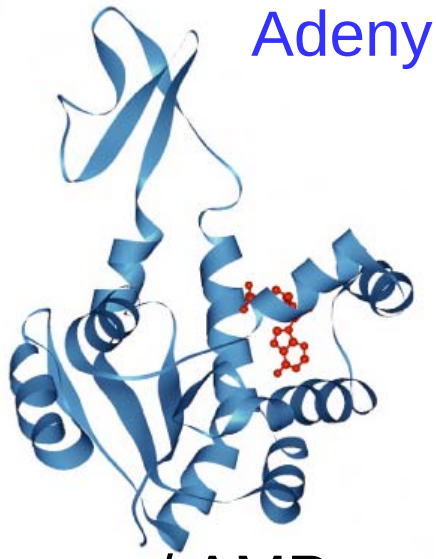
The degree of flexibility varies in proteins with different functions:

Not all proteins are equally flexible.

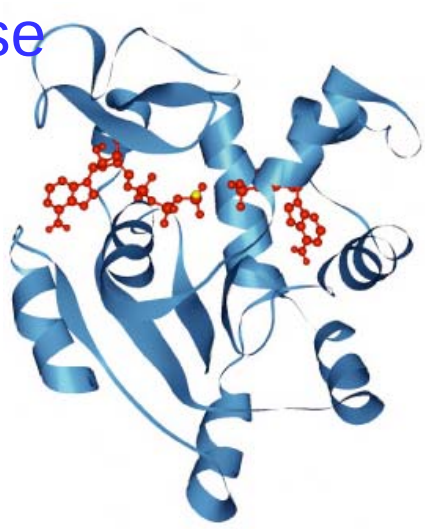
A number of proteins: relatively rigid (extracellular proteins)

Other proteins: very large shape changes when the correct ligand binds.

Adenylate kinase



w/ AMP



w/ AMP + ATP (AMPPNP)

Conformational changes in proteins:

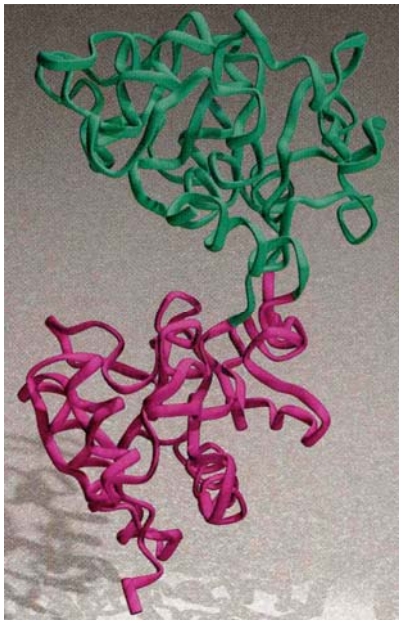
- Small specific & localized changes: enzyme catalytic sites.
- Myoglobin: oxy- & deoxy forms are similar.
- Hemoglobin: oxygen binding leads to changes in 3° and 4° structure.
- Allosteric transitions involve long-range integrated conformational changes: Hb, aspartate carbamoyl-transferase, etc.
- Some very large proteins function as motors or pumps through changes in conformation:

Hinge motions in proteins:

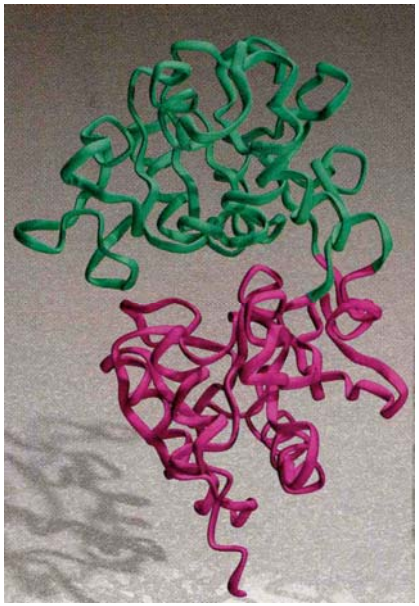
The simplest mechanism of conformational change in proteins

Hinge motion in lactoferrin

Residues 89-92; 249-252, 54 Å rotations as rigid bodies



Open



closed

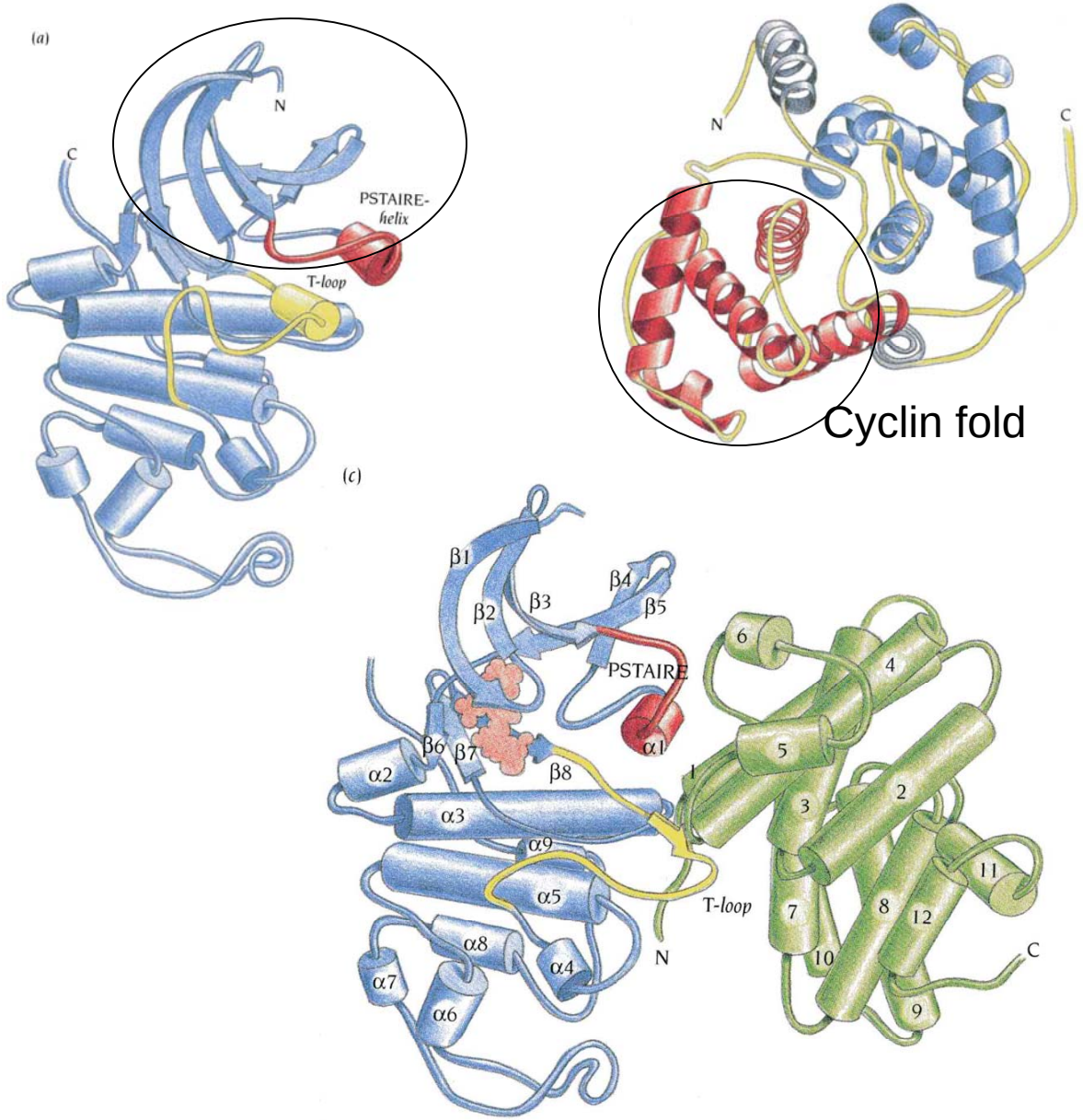
Residues superposed

D1 alone: 0.82 Å

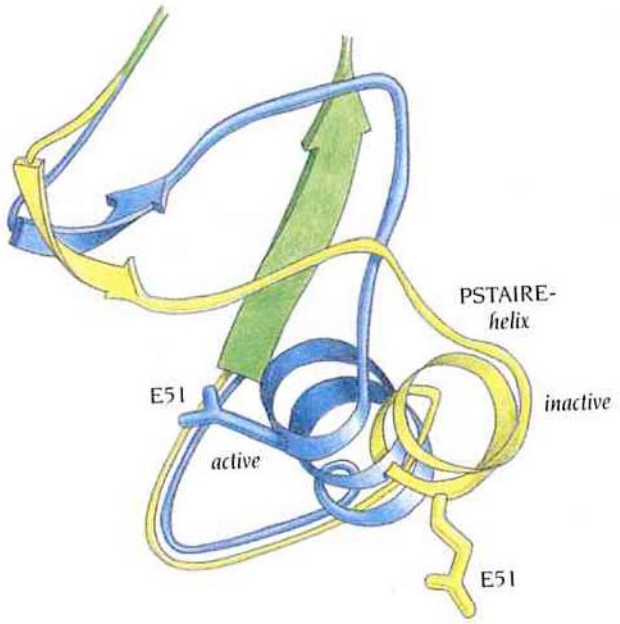
D2 alone: 0.58 Å

Both domains together: 6.1 Å

CDK2-cyclin A structure



Jeffrey PD et al. Nature 376: 313-320, 1995

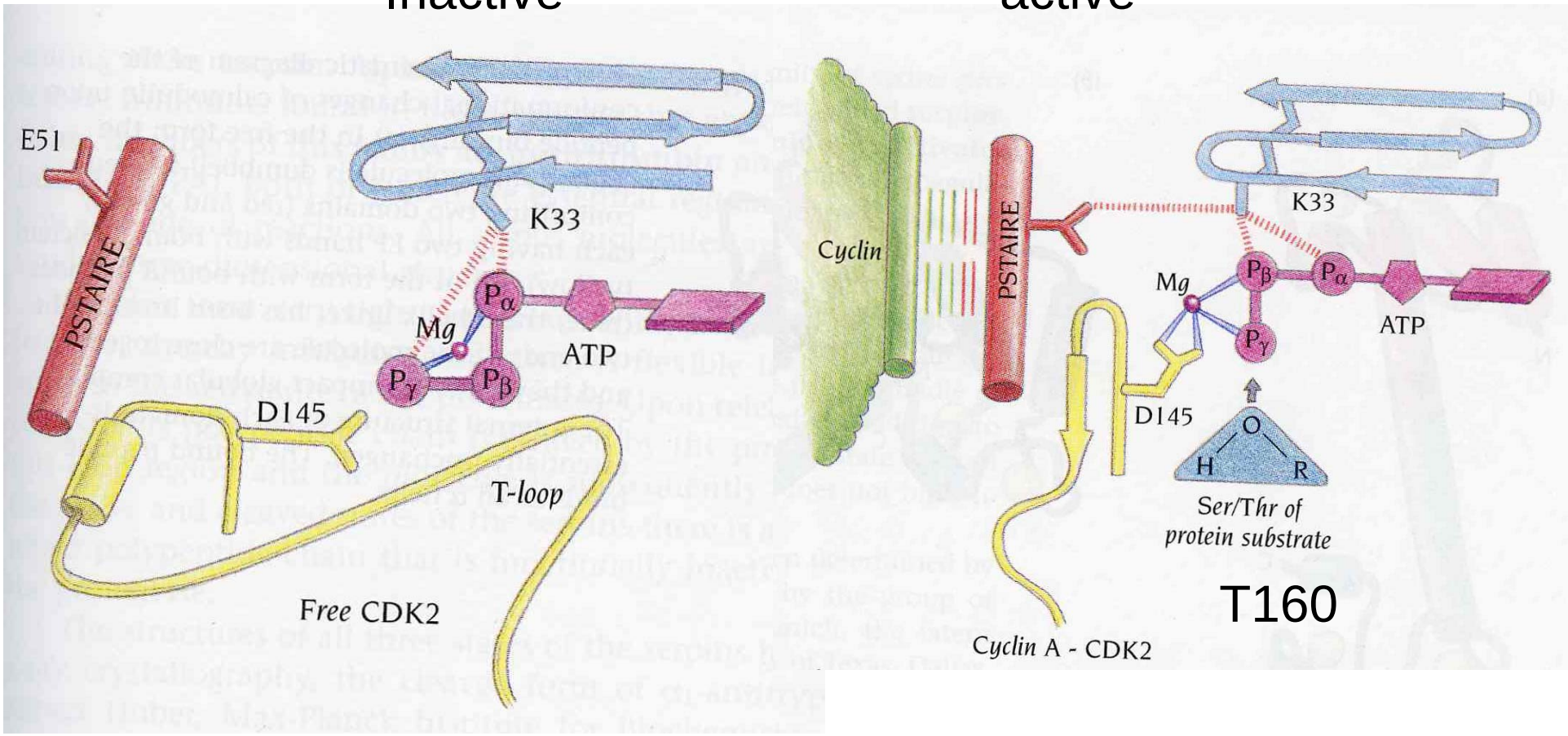


Upon cyclin A binding, the PSTAIRE helix undergoes a major conformational change 90°.

Consequence?

Inactive

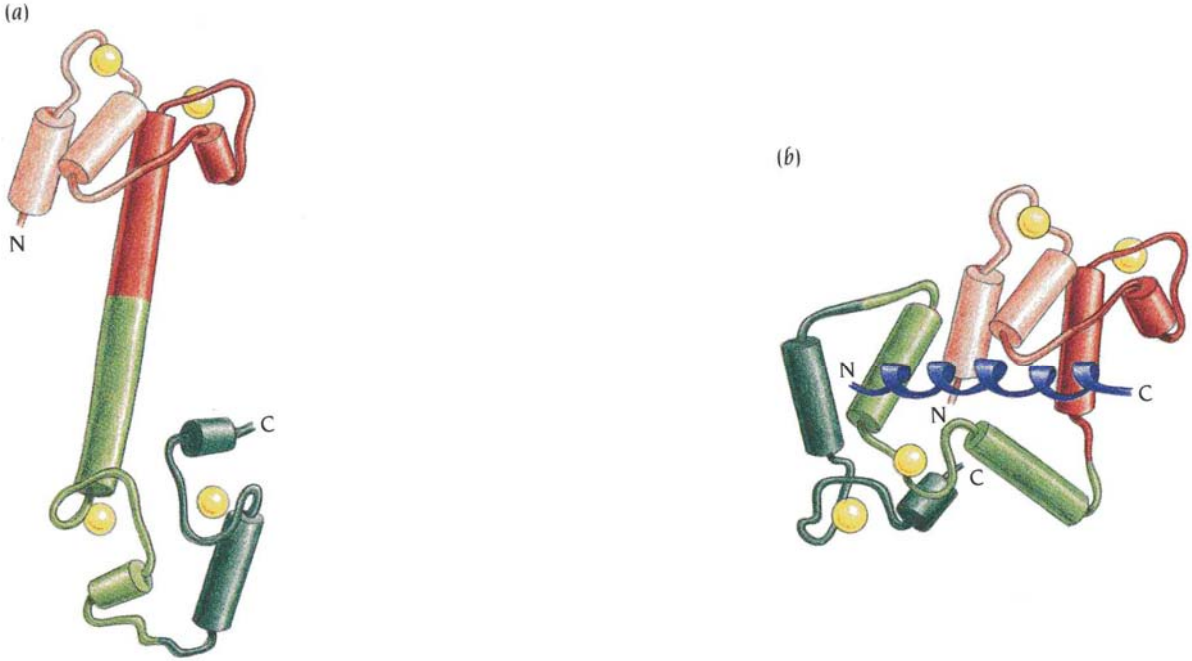
active



Drastic change in T-loop:
Substrate binding site open

Peptide binding to calmodulin induces a large interdomain movement

Calmodulin: Ca- dep signaling pathway



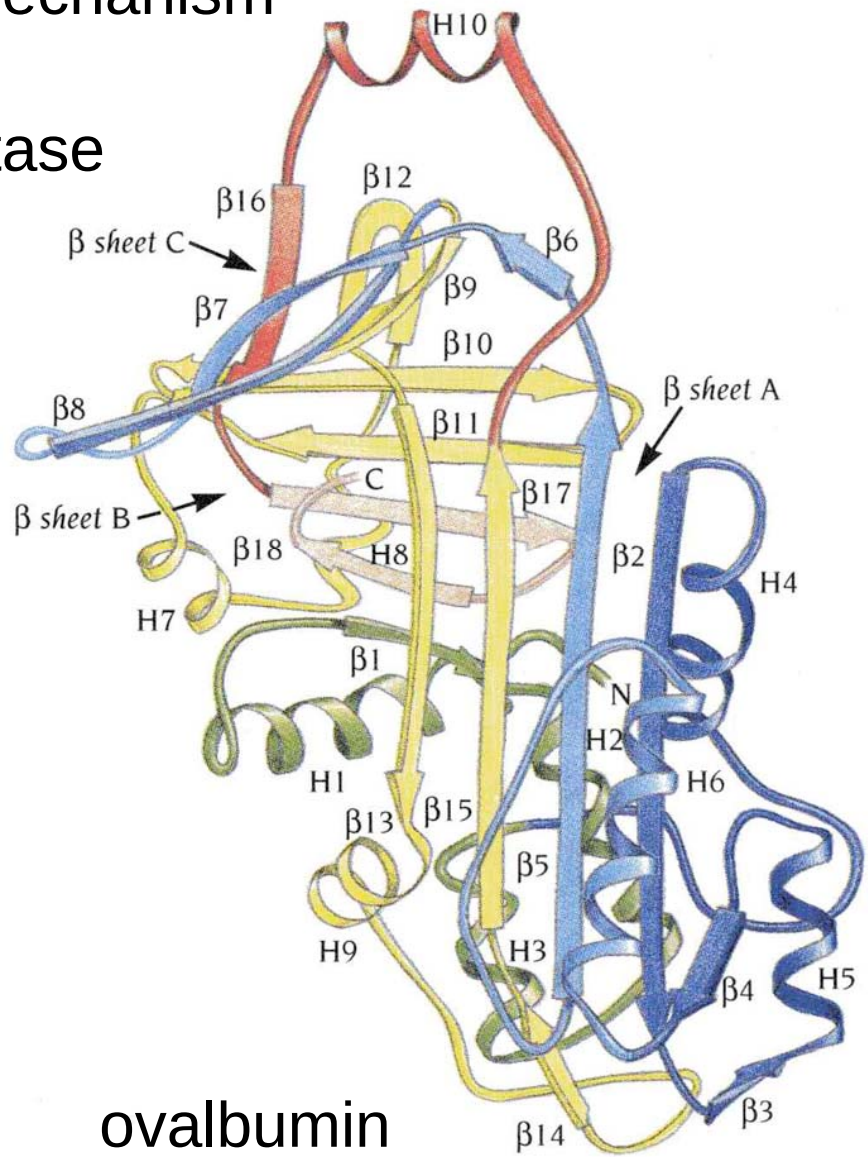
Serpins (Serine protease inhibitors):

Spring-loaded safety catch mechanism

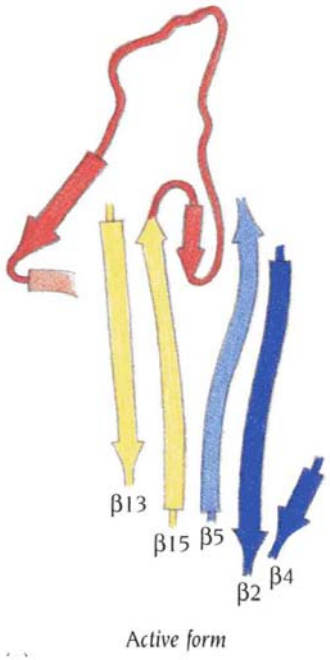
The lung health - neutrofil elastase

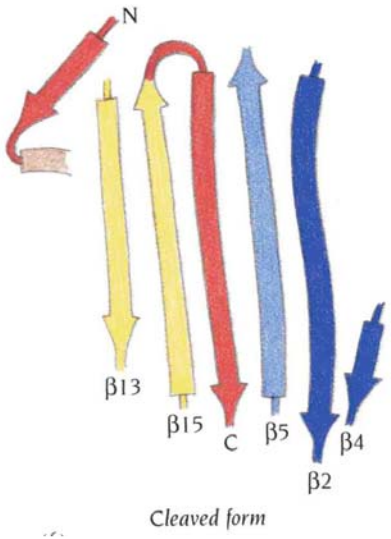
α1-antitrypsin (a blood plasma proteinase)

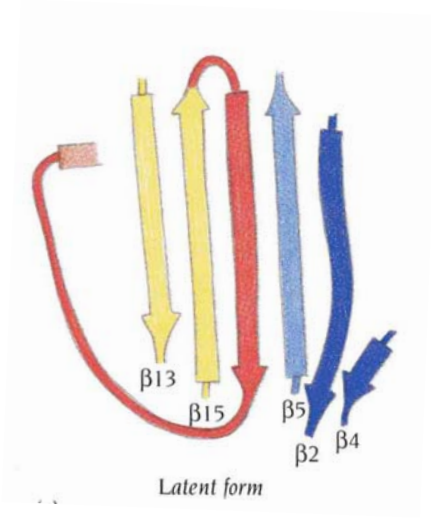
Serpins - antithrombin, plasminogen activator inhibitor: all very similar 3D structures, and form tight complexes with their corresponding S.P.



ovalbumin







Antithrombin α1-antitrypsin PAI

Which of these forms is most stable?

In vivo PAI and antithrombin are stabilized in their active forms by binding vitronectin and heparin.

"Spring-loaded safety catch mechanism" : revert to their latent, stable form unless the catch is kept in a loaded position by another molecule.

Intrinsically Unstructured Proteins and their Functions

NATURE REVIEWS | MOLECULAR CELL BIOLOGY (2005)

H. Jane Dyson and Peter E.Wright

Many gene sequences in eukaryotic genomes encode entire proteins or large segments of proteins that lack a well-structured three-dimensional fold. Disordered regions can be highly conserved between species in both composition and sequence and, contrary to the traditional view that protein function equates with a stable three-dimensional structure, disordered regions are often functional, in ways that we are only beginning to discover. Many disordered segments fold on binding to their biological targets (coupled folding and binding), whereas others constitute flexible linkers that have a role in the assembly of macromolecular arrays.

Why are we discovering unfolded proteins now?

Classic biochemical methods are strongly biased towards the production and characterization of folded, active proteins.

The standard preparation methods:

- Plant or animal tissues, or bacterial cells, are isolated and homogenized.
- The homogenate is assayed for the activity of interest.
- The homogenate is subjected to $(NH_4)_2SO_4$ fractionation and chromatography and/or gel filtration.
- The fractions are assayed for activity and the active protein is purified.
- The pure, active protein is sequenced.
- The three-dimensional structure is determined.

This methodology automatically selects folded proteins, because the formation of a homogenate invariably releases proteases and unfolded proteins are much more sensitive than folded proteins to degradation under these conditions. Also, if the unfolded domains are part of regulatory proteins, there might be only a few copies per cell, and they might not have a convenient activity to assay.

So, why are we discovering unfolded proteins now?

We have only begun to understand the presence and roles of unfolded proteins since the advent of new paradigms in biochemical methodology.

Instead of discovering a detectable activity and isolating it by purifying the protein, we now have access to a vast library of gene sequences. The use of genetic methods to isolate function, using mutants and knockouts, has been the way that most unfolded proteins have been identified so far:

- Formulating a function (for example, control of transcription).
- Mapping the function to a particular gene and to a particular area in the gene.
- Transcribing the gene, producing the protein and purifying the protein.
- Examining its structure in solution by circular dichroism and NMR spectroscopy.
- If the protein is unfolded, trying to co-express it with a binding partner (also identified through genetic mapping).

Will a domain be folded or unfolded?

Predicting the 3D structures of globular proteins from sequence data alone - key challenge

Identifying sequences that are likely to be intrinsically disordered - relatively straightforward

Sequence signatures of intrinsic disorder?

The presence of low sequence complexity and amino-acid compositional bias,with a low content of bulky hydrophobic amino acids (Val, Leu, Ile, Met, Phe,Trp, Tyr), and a high proportion of particular polar and charged amino acids (Gln, Ser, Pro, Glu, Lys and, on occasion, Gly and Ala).

Computer programs: prediction of unstructured regions from amino acid sequences.
PONDR, FoldIndex, DisEMBL, GLOBPLOT 2, and DISOPRED2

Intrinsically disordered proteins are highly prevalent

Proportion of proteins containing such segments increases with the increasing complexity of an organism

Proteins involved in eukaryotic signal transduction or associated with cancer: an increased propensity for intrinsic disorder

Experimental characterization of disordered proteins.

Nuclear magnetic resonance (NMR) spectroscopy
Information on the three dimensional structure and dynamics of biological molecules in solution.

CIRCULAR DICHROISM
The UV-CD spectrum uses the chirality or ‘handedness’ of biological molecules to provide information on secondary structure in solution.

Fluorescence spectroscopy
Information on the environment of aromatic rings, and can be used in conjunction with external probes to determine the distances between atoms in a molecule.

Experimental characterization of disordered proteins.

Vibrational CD Spectroscopy

the chiroptical version of infra-red spectroscopy
information on the vibrations of individual bonds in a molecule.

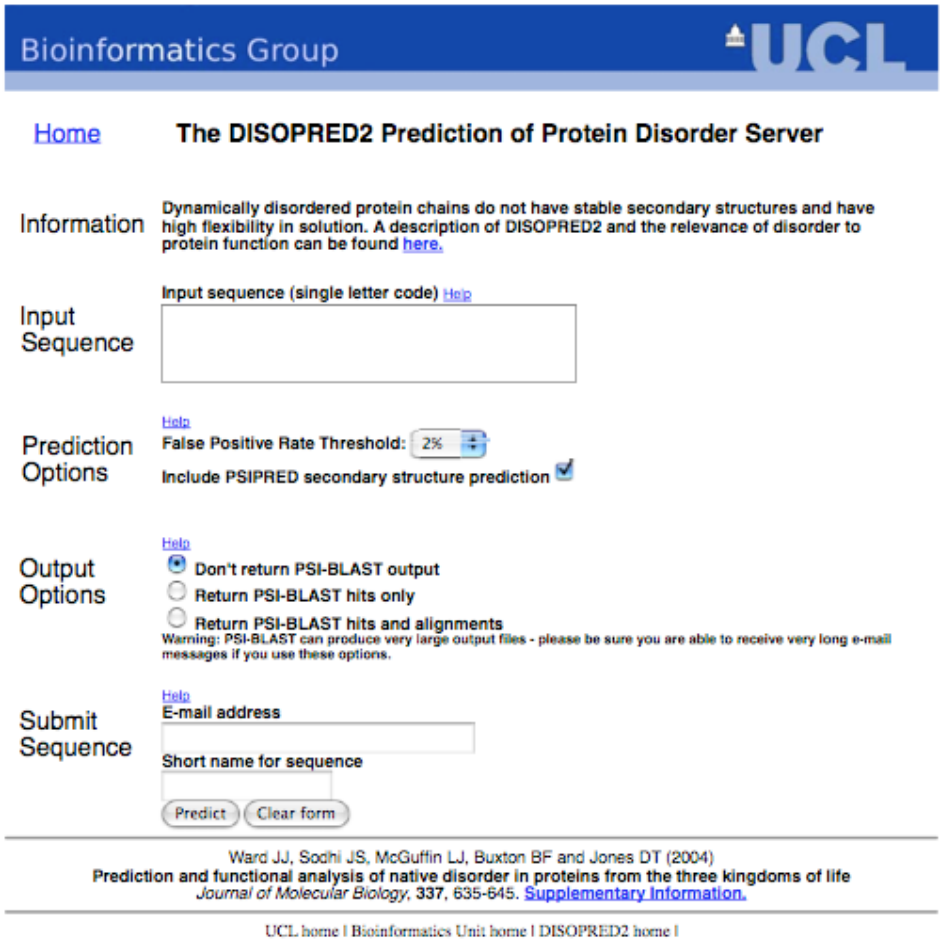
Raman spectroscopy

Information on bond vibrations that is complementary to that provided by infra-red spectroscopy.

Biochemical assays

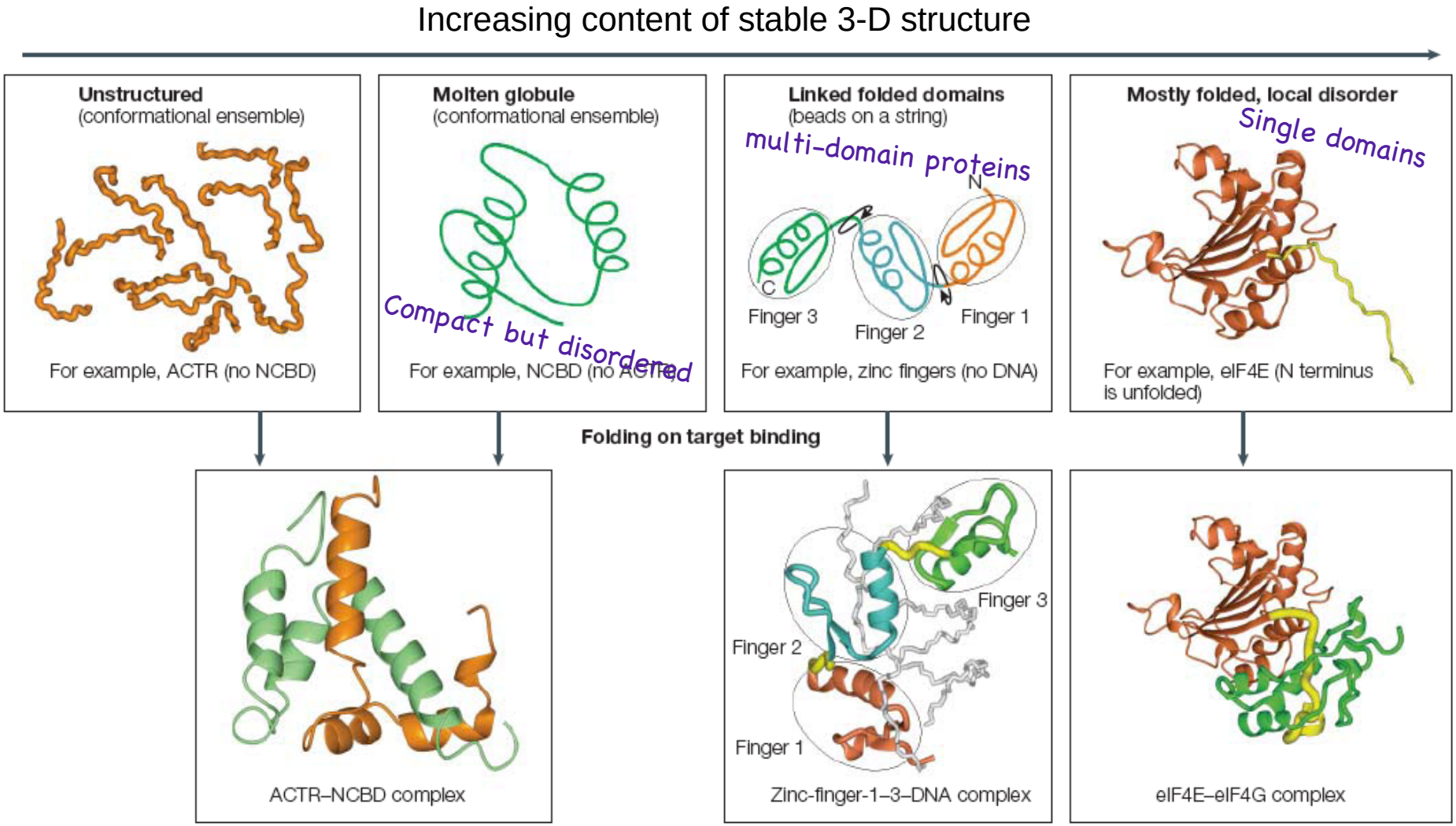
Proteolysis susceptibility

The DISOPRED2 Prediction of Protein Disorder Server



Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF and **Jones DT** (2004)
Prediction and functional analysis of native disorder in proteins from the three kingdoms of life
Journal of Molecular Biology, 337, 635-645.

The continuum of protein structure



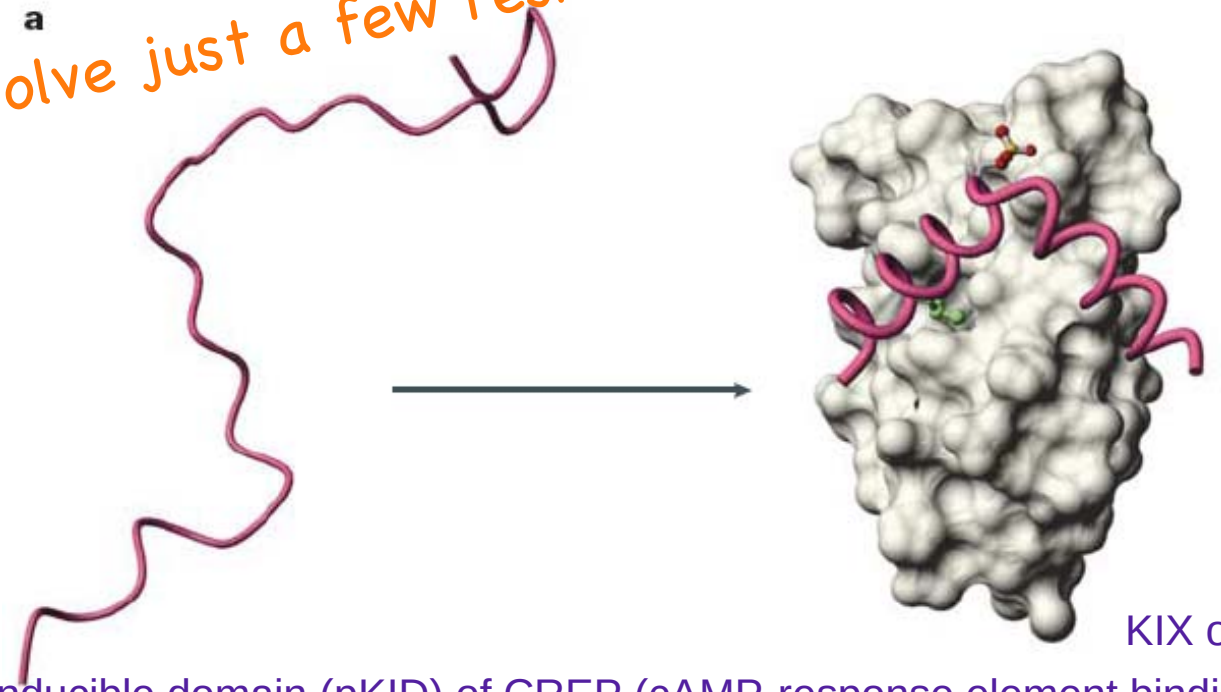
Functions of intrinsic disorder in proteins

- regulation of transcription and translation
- cellular signal transduction,
- protein phosphorylation,
- storage of small molecules,
- regulation of the self-assembly of large multiprotein complexes

Coupled folding and binding

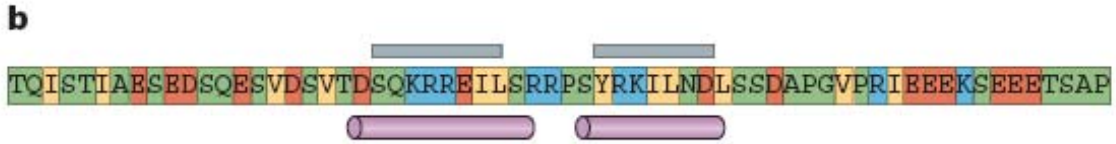
an intrinsically disordered protein (region) folds into an ordered structure concomitant with binding to its target.

might involve just a few residues

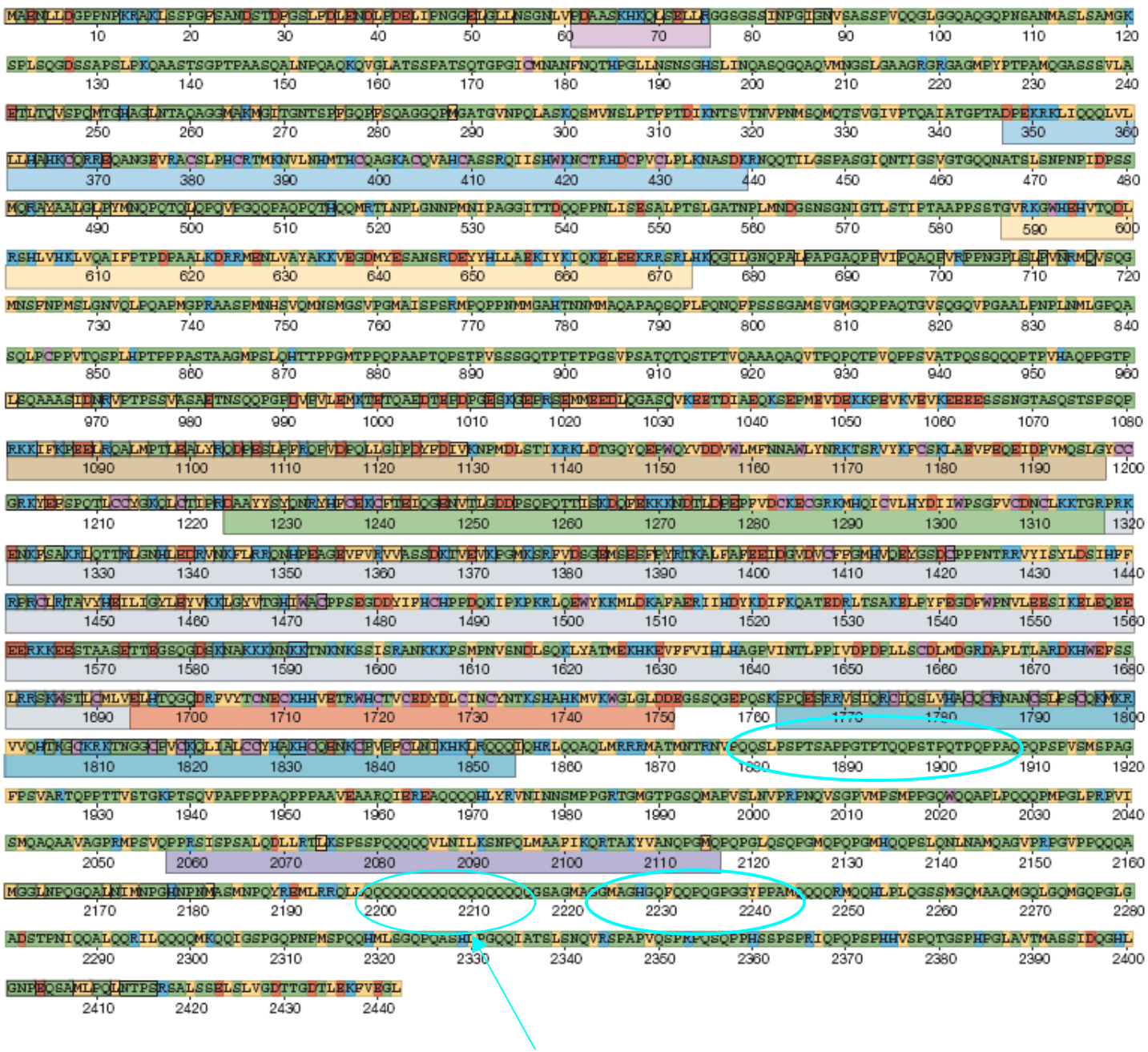


KIX of CBP

p-kinase inducible domain (pKID) of CREP (cAMP-response element binding P)



A case study: transcriptional co-activator CBP/p300



Domain structure of CBP

The biological ‘cost’ of disordered proteins

Intrinsically disordered regions are the sites of many chromosomal translocations that are associated with disease.
e.g.translocations that fuse regions of CBP or p300 to segments of MOZ (monocytic zinc-finger leukaemia protein) associated with human leukaemias.

Disordered regions can also have a biological cost in terms of the promotion and proliferation of protein folding diseases.
Neurodegenerative diseases such as prion diseases or Parkinson’s disease are associated with intrinsically disordered proteins (Q-rich motifs)

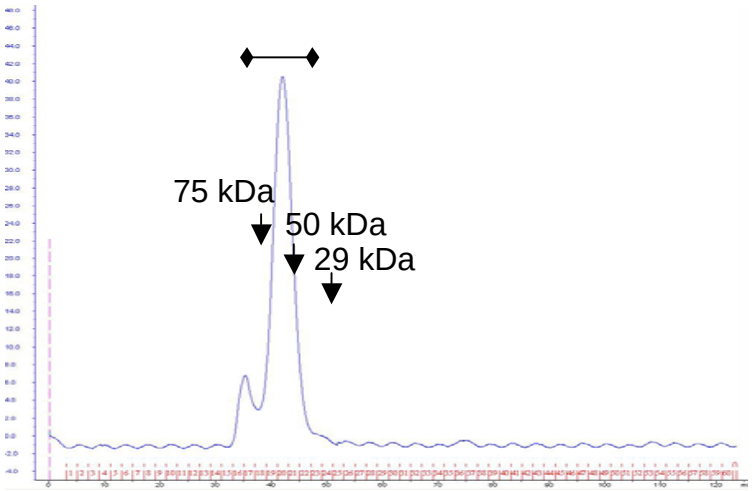
Amino acid runs in eukaryotic proteomes and disease associations

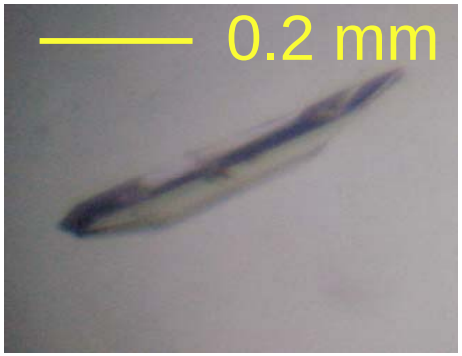
Samuel Karlin*†, Luciano Brocchieri*, Aviv Bergman‡, Jan Mrázek*, and Andrew J. Gentles*

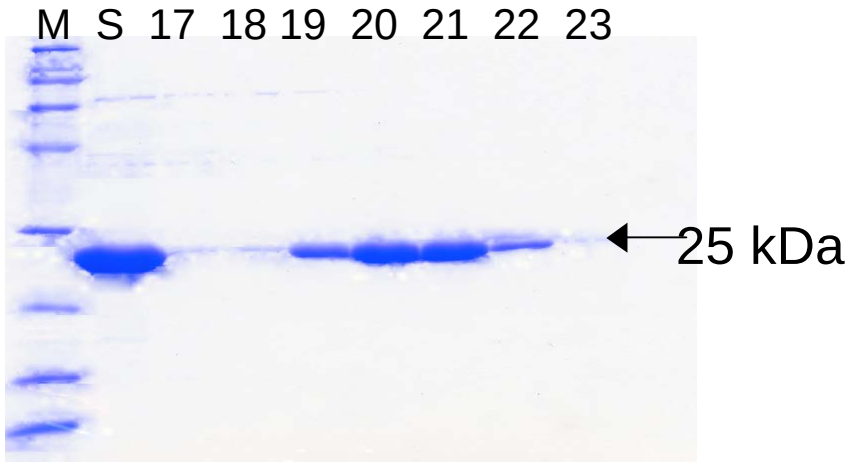
PNAS | January 8, 2002 | vol. 99 | no. 1 | 333–338

A case study: a small C-ter tail of Cj0977

Cj0977 is highly soluble and a dimer in solution



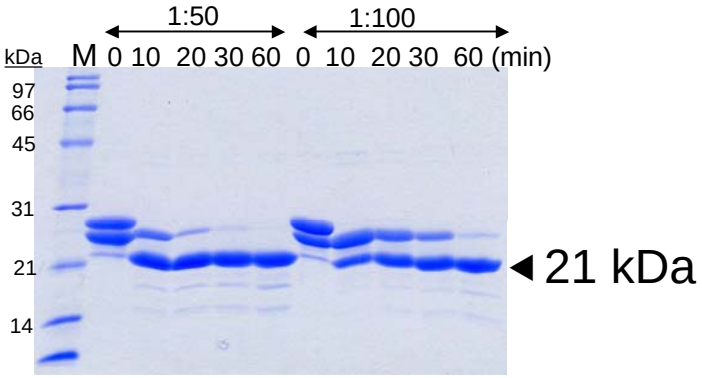


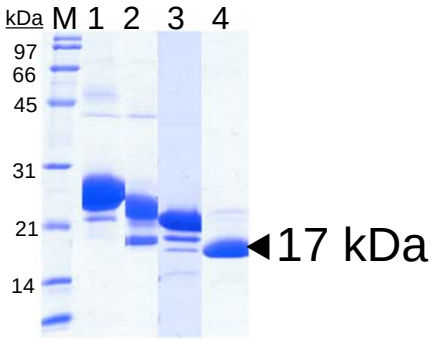


His-Cj0977 crystals obtained *with* $(NH_4)_2SO_4$

But, no diffraction!

Limited Proteolysis of His-Cj0977





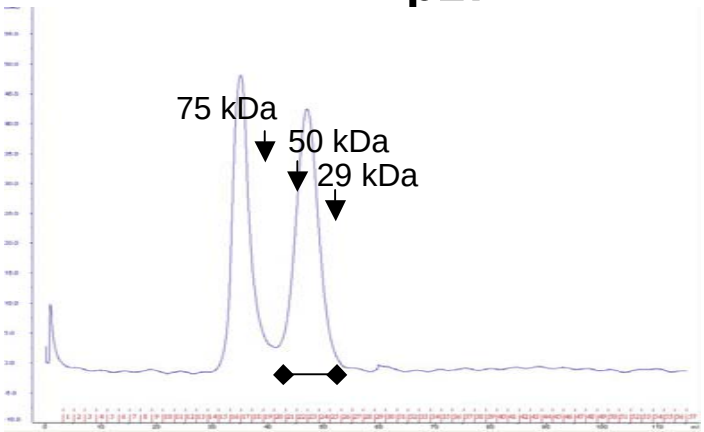

```
                 N-terminus
Cj0977   1  MAKKDNFEEYAQLEEYASAEDISRVRAELLTCPELNTSLAGTIIEIDKNYAKSILITTSEMVADDQGLIFDAFIFAAANYVAQASINKEFSVIIGSKCFF
Ws0669   1  ---MDTLNSPEEIDLPKTQEEFVPQRVYTDADPAF-----GELKELGKGVAKVWLVTTKAMASDSMGLVHAGFIFSSANFSAMCAINDPNAILIGADSKF
Hp0420   1  ------------MQESVVHVDYDSLETCKNFKPSVG----TELVVLEKNIAHVRFKGNENMVYEEN-FVHAGFVLLACNYAALCALNKKYSVVVSNNINF

                                                 Cjp17          Cjp19          Cjp21
Cj0977      YAPLKLGDVLELEAHALFDETSKKRDVKVVGHVKEIKMFEGTIQVVSTDEHIFKLKRPPLNTVKPTENQEQGTKINNPEAVAAALMASVGGK  192
Ws0669      LAPIEVGNEIEFRARALQNDT-KKREVKVEGYVLDIKVFDAMFYVAMFDHHIFKLR--------ITERKKPS--------------------  155
Hp0420      YAPLELNQEALIKAQVIQDGV-KKAEIKIEAFVLDIQVLEGMIEIVVFDKKPFKFN--------FKEE----------------------  142
```

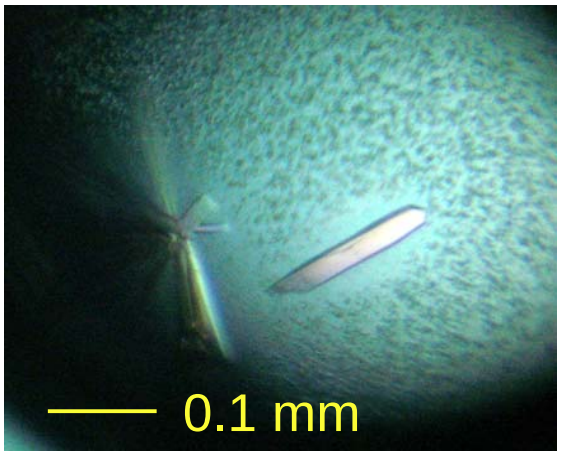
Crystallization of $Cj0977_{p17}$

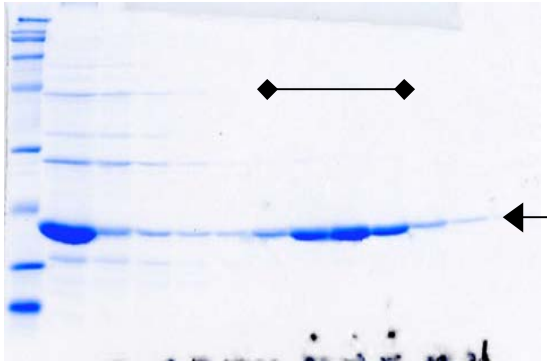
$GST::Cj0977_{p17}$

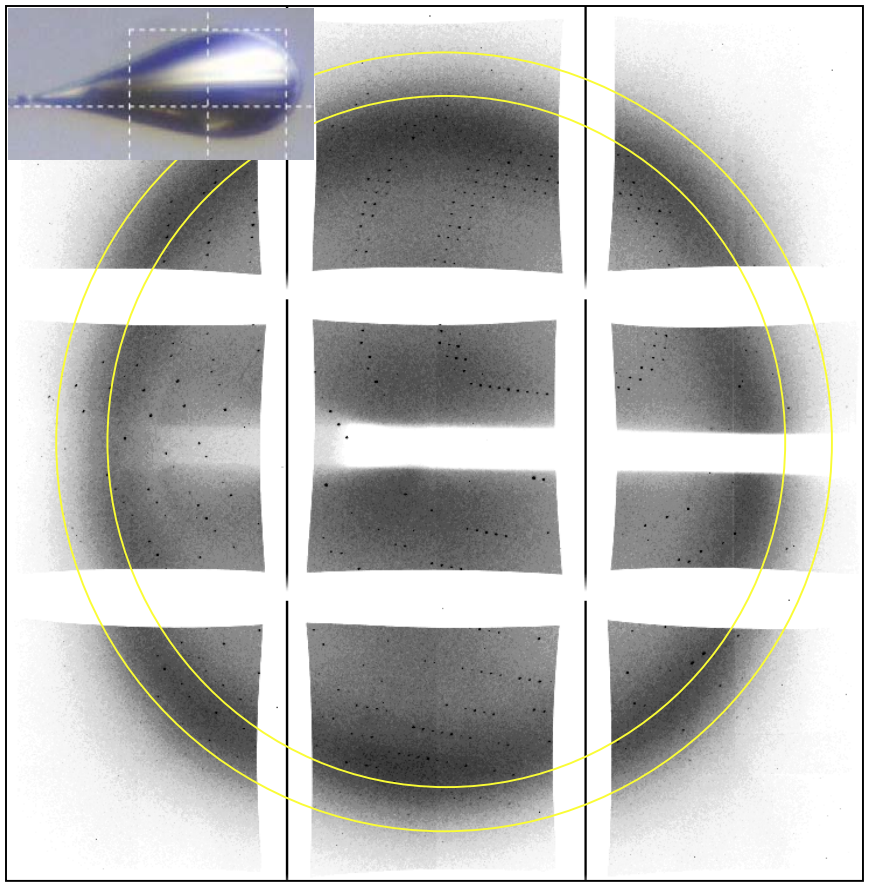
↓

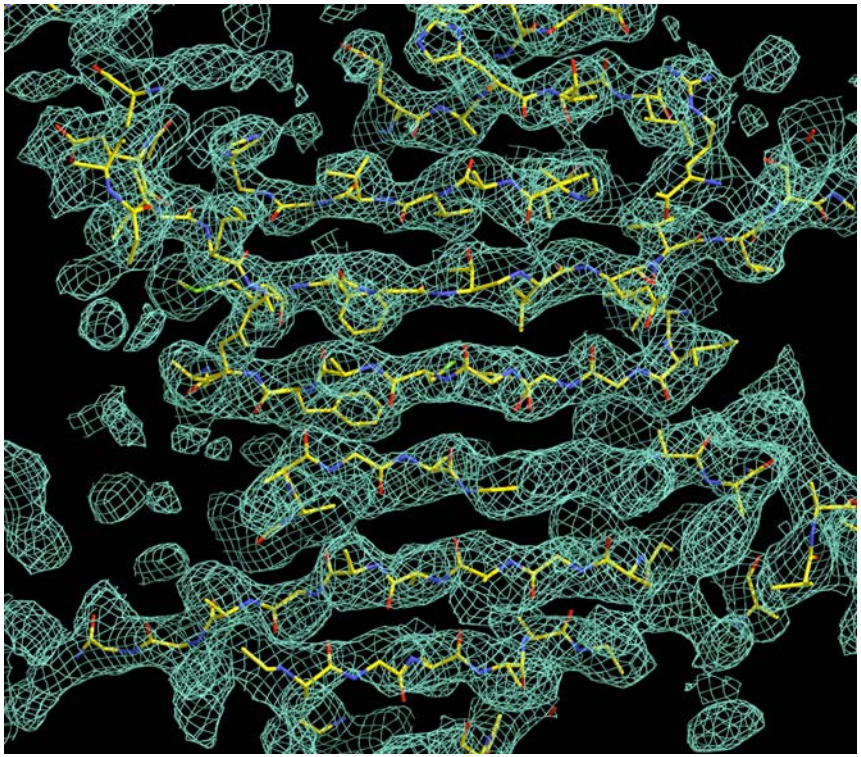
$Cj0977_{p17}$







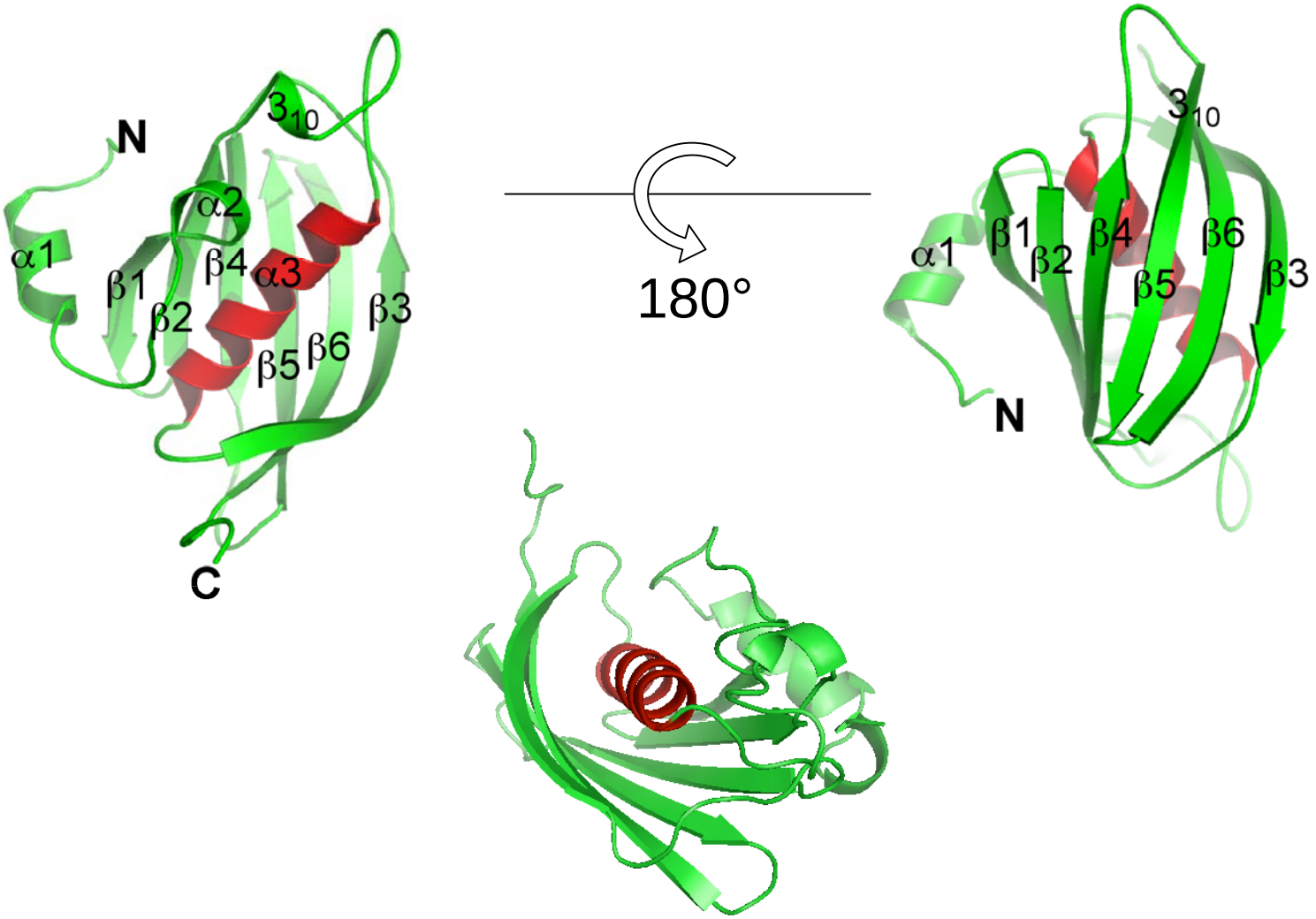




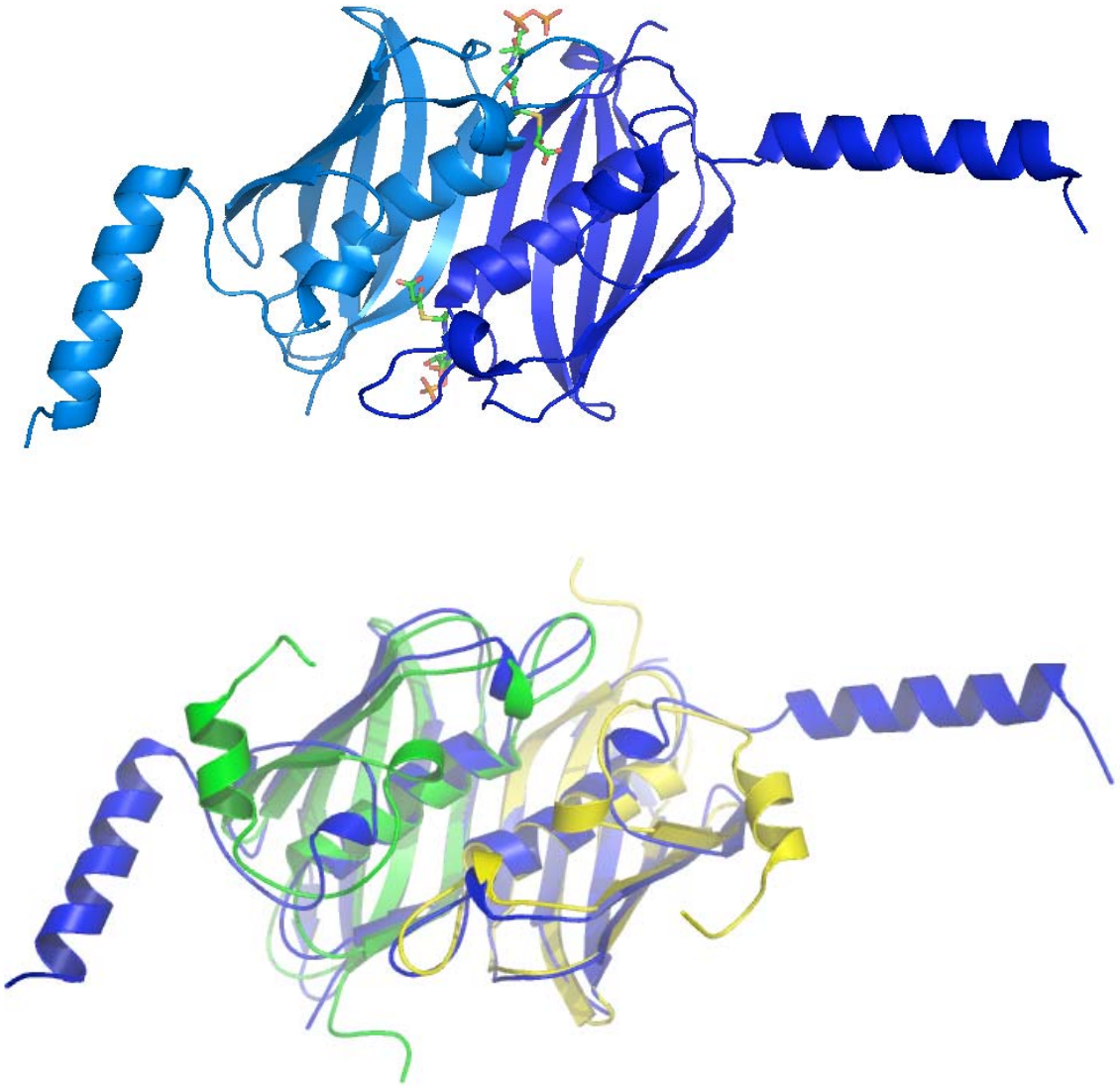
MAD phasing at 2.8 Å.

$Cj0977_{p17}$ electron density map.

Cj0977 adopts a Hotdog fold

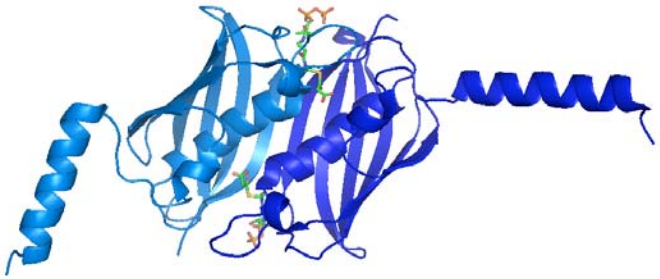


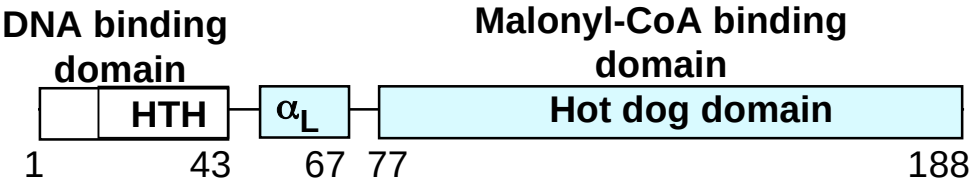
FapR: a structural homolog of Cj0977

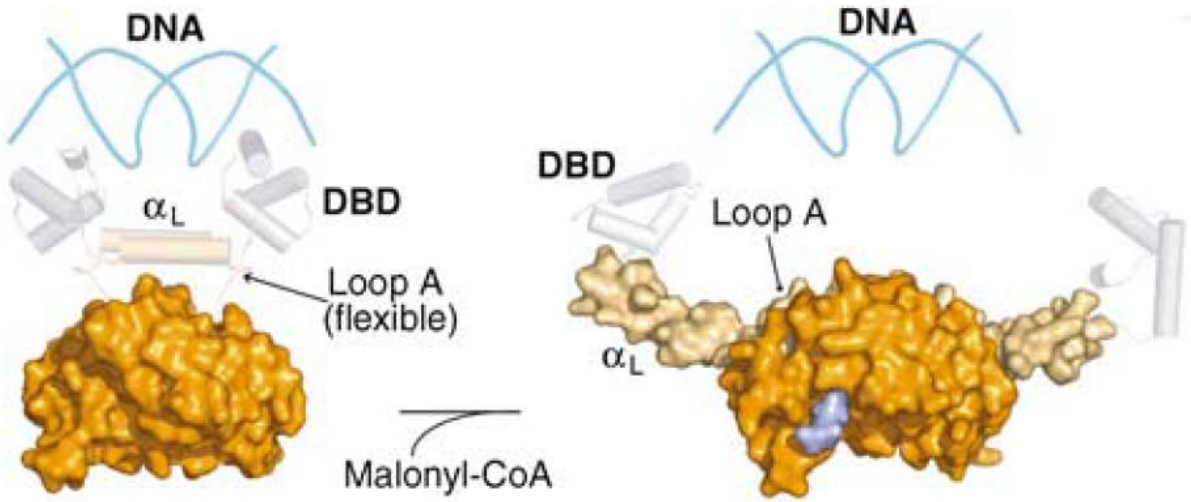


Lipid Biosynthesis Regulation in Gram-negative Bacteria

FapR



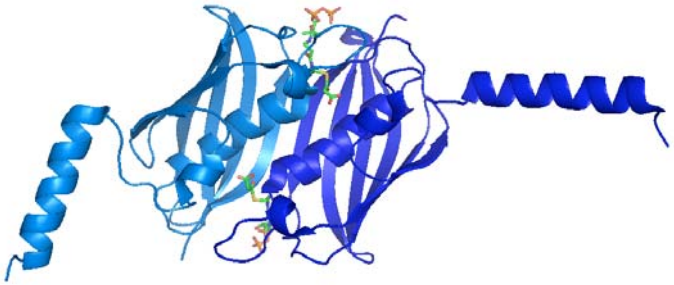


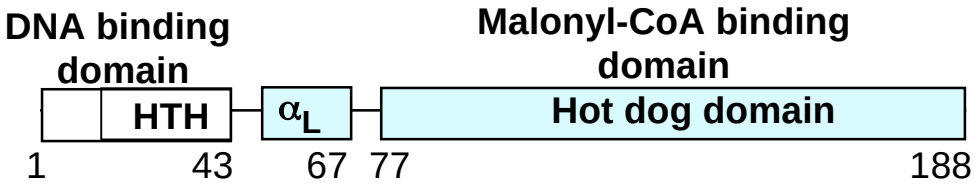


Schujman et al.
EMBO J. (2006) 25(17):4074-83.

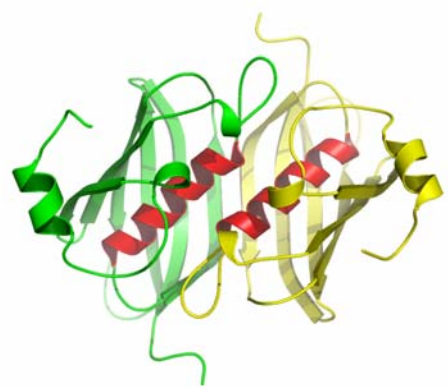
Function of Cj0977?

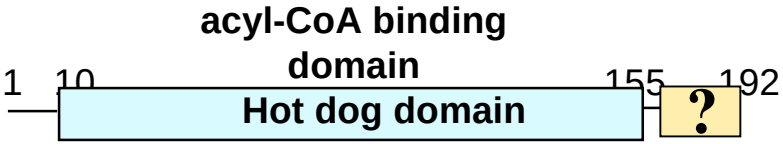
FapR





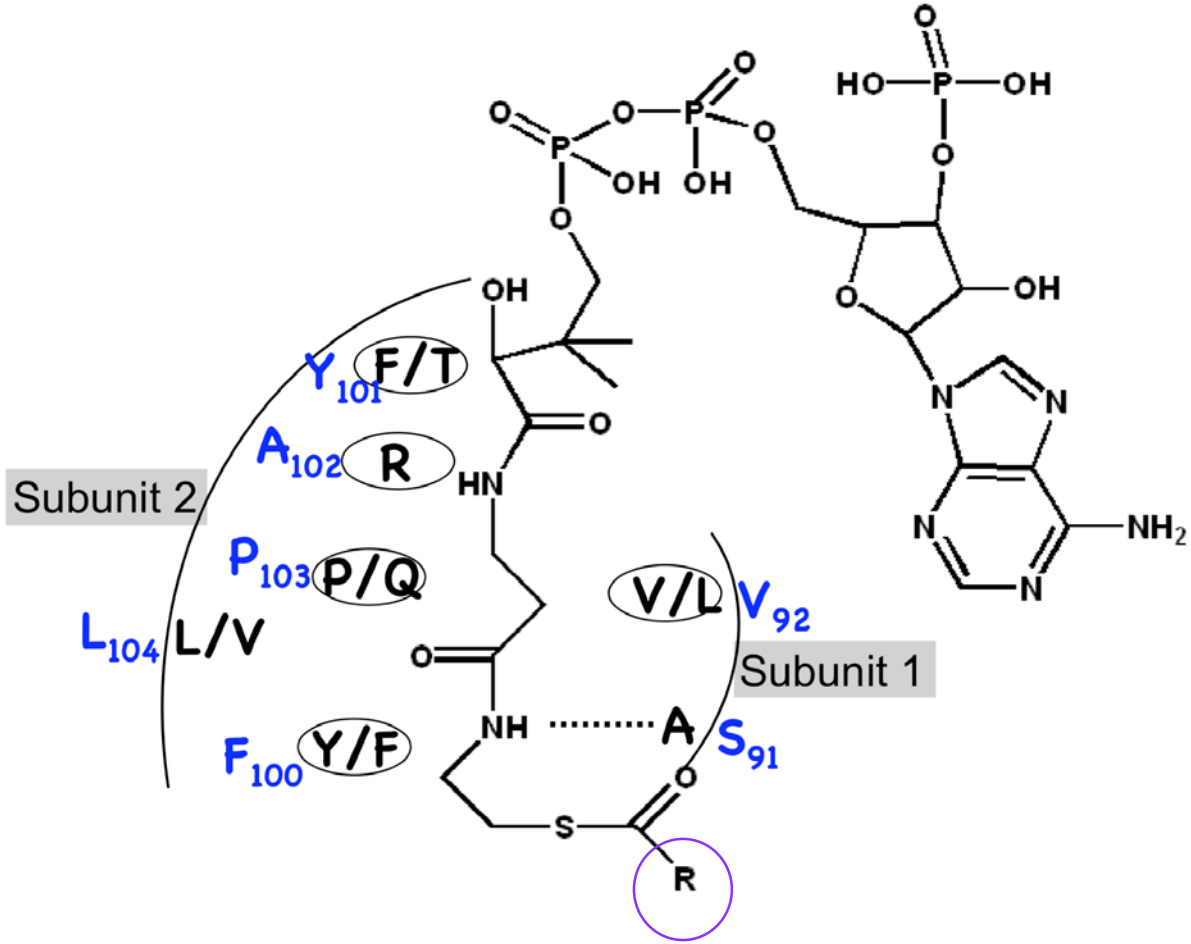
Cj0977



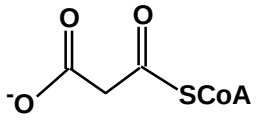


Yokoyama *et al.*
J. Mol. Biol. (2008) 384(2):364-76

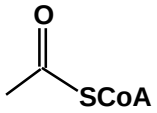
Ligand recognition site of Hotdog fold proteins



Malonyl CoA



Acetyl CoA



Proposed model of coupled folding and binding of Cj0977