Protein Folding *in vivo*

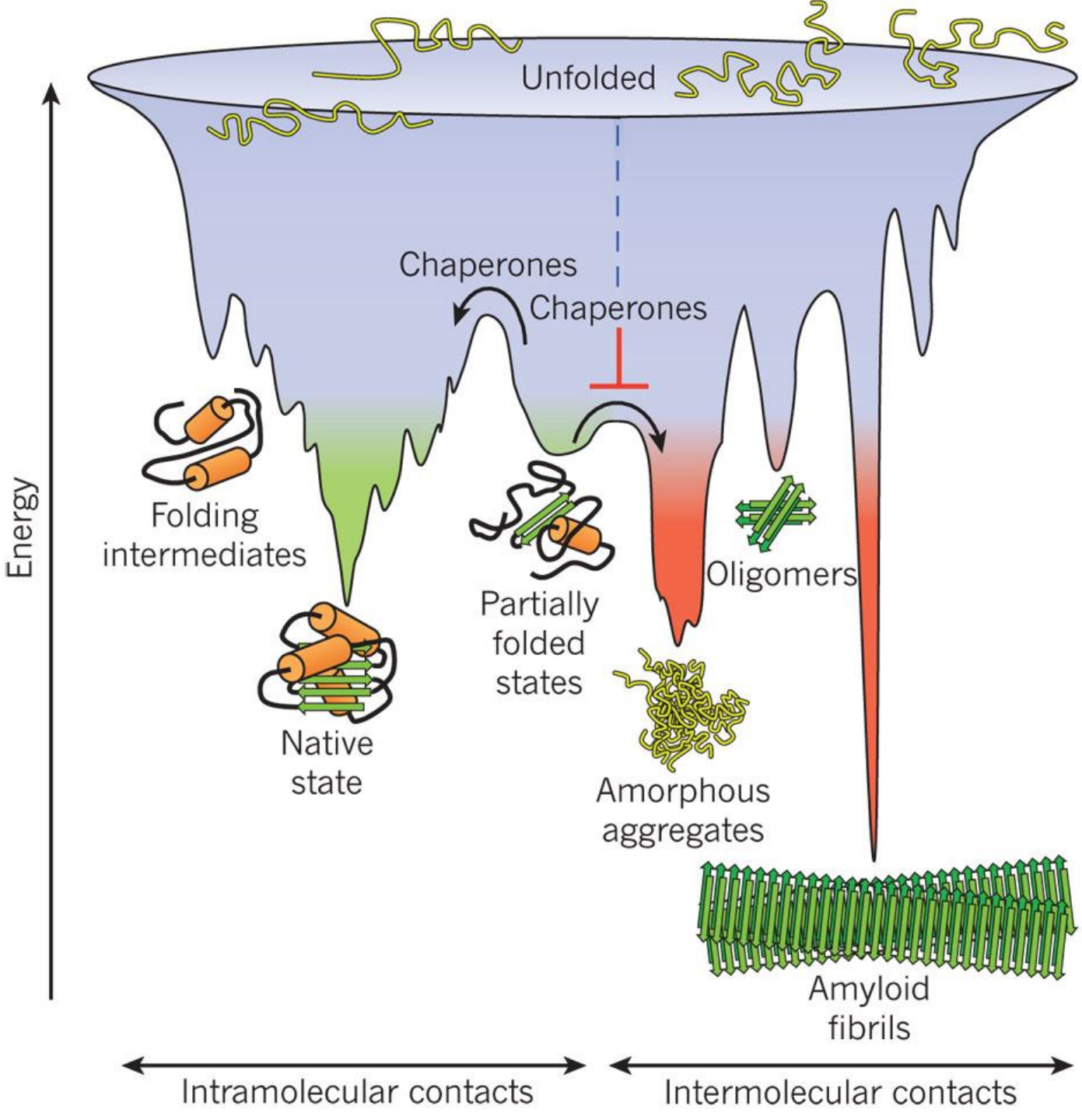
Biochemistry 530

David Baker

- I. Protein folding in vivo: Molecular Chaperones
- II. Protein unfolding in vivo: ATP driven unfolding in mitochondrial import and the proteasome.
- III. Protein folding reactions under kinetic control

Molecular Chaperonins: Function

- **since proteins fold spontaneously in vitro, why are additional factors required for proper folding in vivo?**
  - high protein concentration in cytoplasm (~200mg/ml).
  - compared to *in vitro* experiments (~0.1mg/ml)
- At high concentrations, unfolded polypeptide chains aggregate instead of refolding.
- regulation of protein folding important in a variety of in vivo situations:
  - the subunits of oligomeric proteins may not be soluble in isolation; proper assembly may require subunit binding proteins.
  - translocation of proteins accross membranes requires that premature folding be prevented
  - function of proteins can be regulated by timing the completion of folding


Unfolded
Chaperones
Chaperones
Energy
Folding intermediates
Native state
Partially folded states
Oligomers
Amorphous aggregates
Amyloid fibrils
Intramolecular contacts
Intermolecular contacts

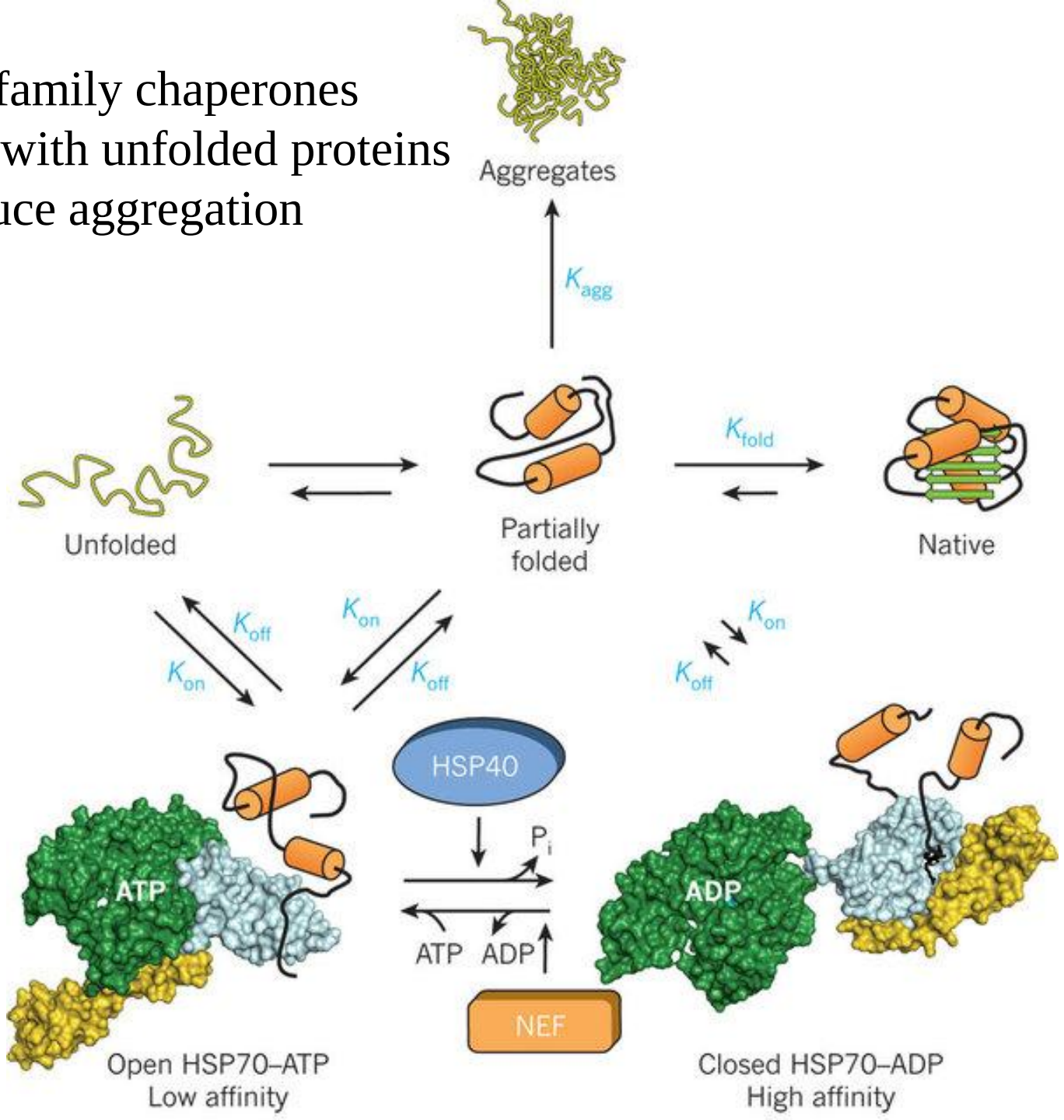
I. Molecular Chaperones

"a protein that binds to and stabilizes an otherwise unstable conformer of another protein--and by controlled binding and release of the substrate protein, facilitates its correct fate in vivo: be it folding, oligomeric assembly, transport to a particular subcellular compartment, or controlled switching between active/inactive conformations"

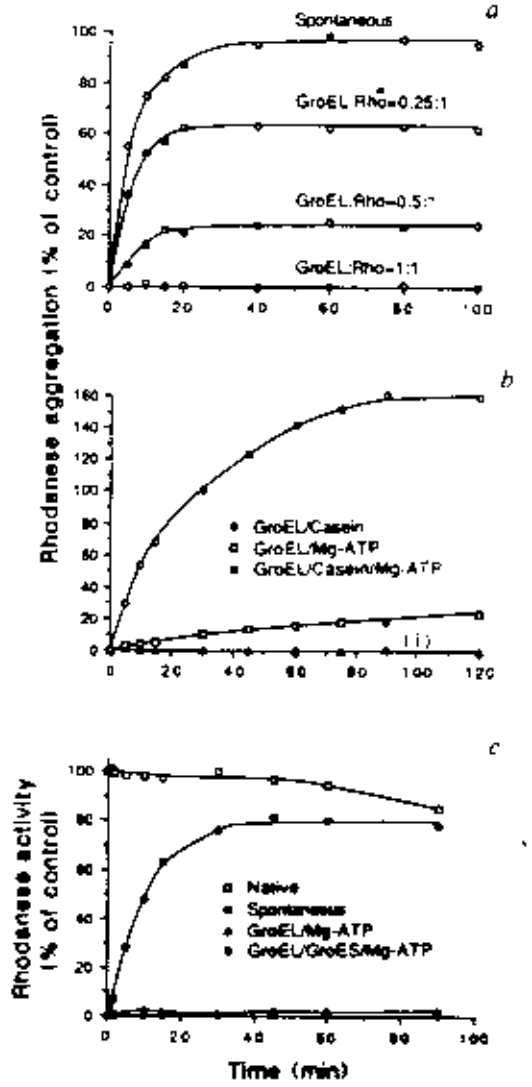
Table 1 Molecular chaperone functions of heat shock proteins hsp70, hsp60, and hsp90

Subcellular localization	Organism	Chaperone	Cooperating factors	Activity
Hsp70 family				
Prokaryotes				
Cytosol	*E. coli*	DnaK	DnaJ, GrpE	Stabilizes newly made proteins in vivo (28); preserves folding competence of proteins in vitro (29); stimulates protein export (30); promotes assembly/disassembly of replication complexes (24, 150–152); reactivates heat-inactivated RNA polymerase (15); facilitates degradation of abnormal proteins (20); controls heat-shock response (167).
Eukaryotes				
Cytosol	*S. cerevisiae*	Ssa1-4p	NEM-sensitive factor, YDJ1p	Stimulates protein transport into ER (7, 8), mitochondria (8, 9, 122), chloroplasts (244), and nucleus (106); binds to nascent polypeptides (12), and to polypeptides containing abnormal amino acids (133); dissociates clathrin from clathrin coats (6); promotes lysosomal degradation of cytosolic proteins (21); interacts with HSF transcription factor (167).
	Rabbit	Hsc73	NEM-sensitive factor	
	Human	Hsc73	None identified	
Endoplasmic reticulum	*S. cerevisiae*	Kar2p	Sec63p	Promotes protein translocation into ER (16, 19, 158); binds unassembled or misfolded subunits of multisubunit ER proteins (2, 4, 5, 159–162).
	Mammals	BiP/Grp78	None identified	
Mitochondria	*S. cerevisiae*	Ssc1p		Promotes protein translocation into mitochondria and subsequent folding (13, 14, 93).
Chloroplasts	*P. sativum*	cpHsp70	None identified	Promotes insertion of light-harvesting complex protein into thylakoid membrane (245, 246).
	E. gracilis			

HSP70 family chaperones interact with unfolded proteins and reduce aggregation



Experimental data: GroEL dependent folding of aggregation prone rhodanase



- stoichiometric amounts of GroEL suppress rhodanase aggregation
- rhodanase is tightly bound to GroEL, but is released into solution (where it aggregates), upon addition of ATP
- GroEL bound rhodanase is converted into the native state upon addition of GroEL and ATP

Structure of GroEL

1) Electron Microscopy

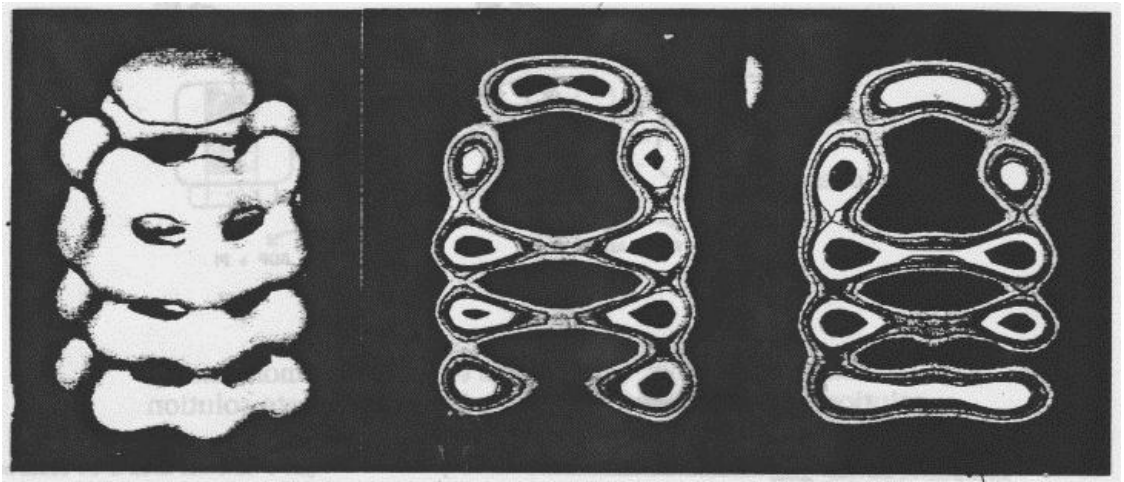


Fig. 4 Complexes with GroES and substrate. *a*, Surface-rendered view of the GroEL-GroES-ATP complex at 30 Å resolution determined from cryo EM. The GroES ring is seen as a disk above the GroEL. *b*, Section through this complex showing the large reorientation of the apical domains in contact with GroES, forming the contact with the same region as the substrate binding site. *c*, Section of the folding complex GroEL-MDH-GroES-ATP trapped by vitrification after 15 sec of folding and imaged by cryo EM. The MDH substrate density is found in the opposite ring to GroES (based on work in ref. 9)

2) X-ray crystallography

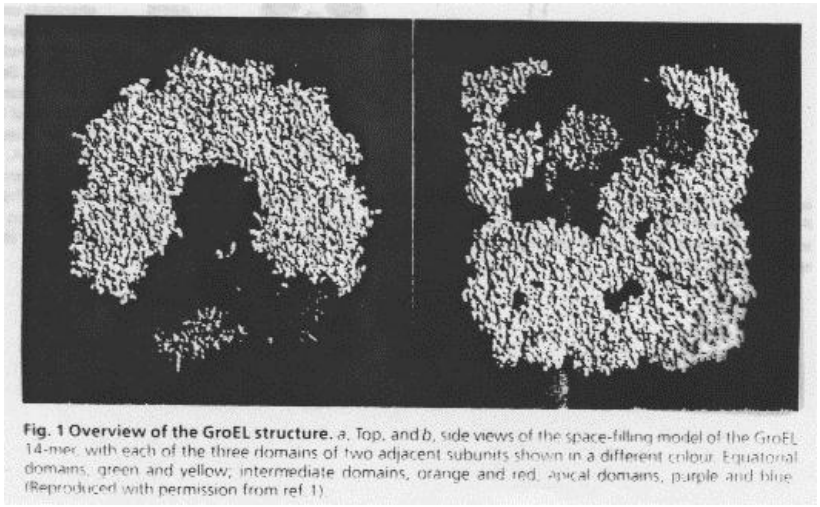


Fig. 1 Overview of the GroEL structure. *a*, Top, and *b*, side views of the space-filling model of the GroEL 14-mer, with each of the three domains of two adjacent subunits shown in a different colour. Equatorial domains, green and yellow; intermediate domains, orange and red; apical domains, purple and blue. (Reproduced with permission from ref. 1)

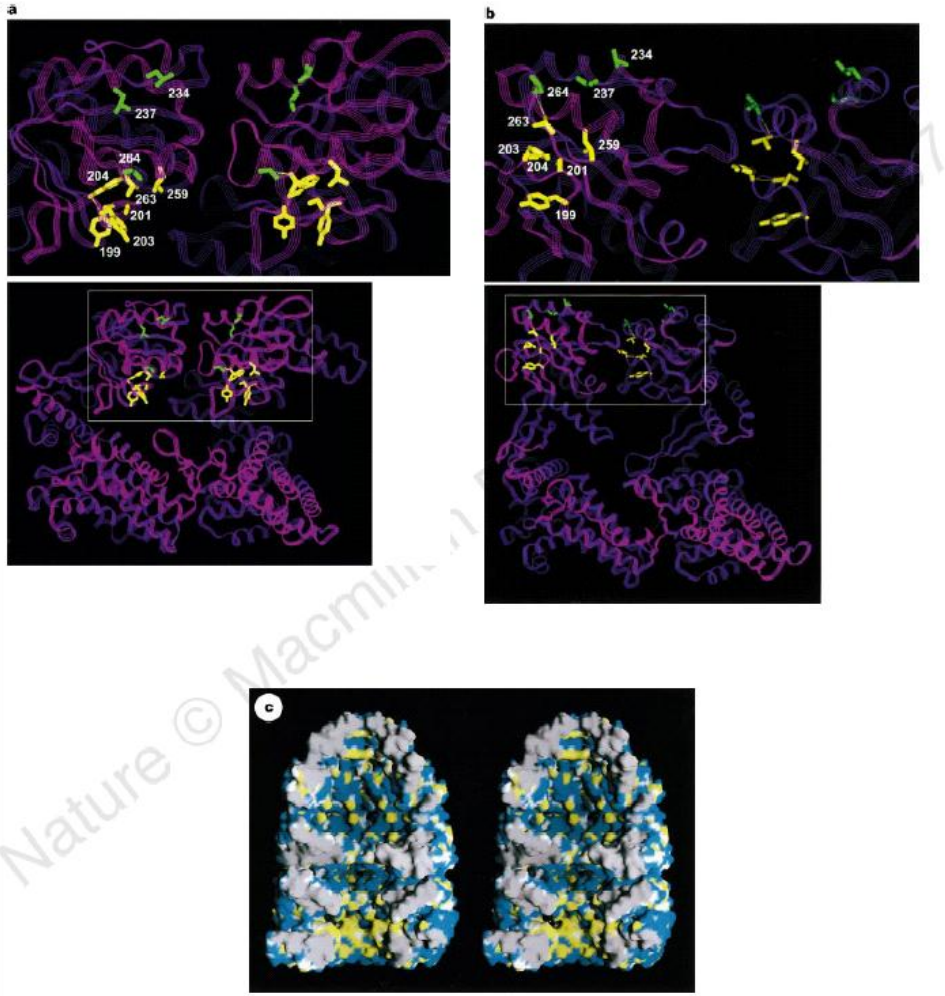
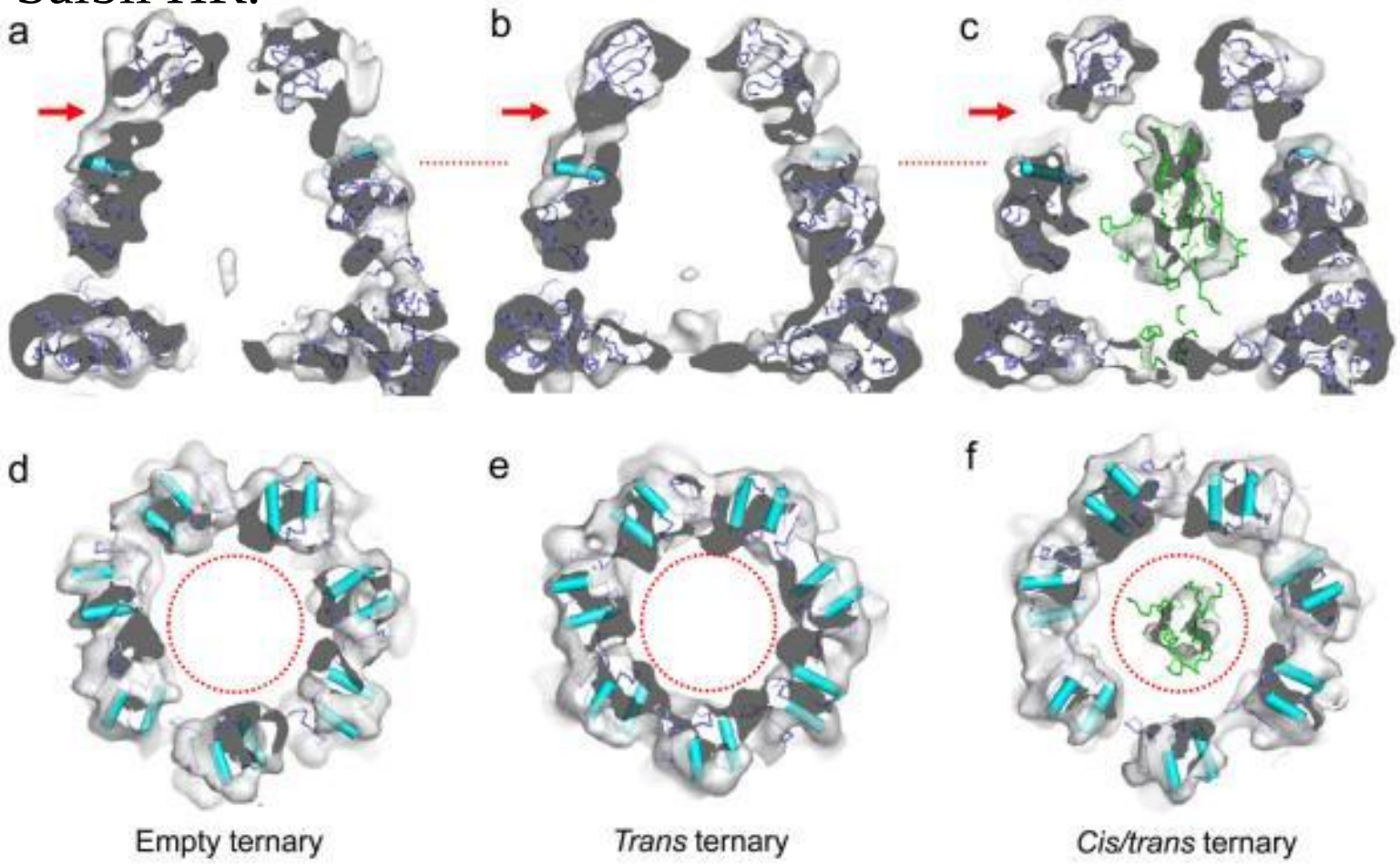


Figure 5 The change in the central cavity. **a**, Bottom, coiled-line ribbon drawing of two neighbouring subunits of the *trans* ring viewed from the central cavity; top, a close-up view of the rectangular area. Skeletal side chains denote residues involved in polypeptide binding as derived by mutagenesis[12]. These residues, with the exception of Ser 201, have hydrophobic side chains. **b**, Coiled-line ribbon drawing of two neighbouring subunits of the *cis* ring viewed from the central cavity in the same orientation and highlighted as in **a**. Note that these residues are moved away from the cavity surface. Three residues now form the GroES interface (green; Leu 234, Leu 237 and Val 264), and the rest form the new apical GroEL subunit interface (yellow; Tyr 199, Ser 201, Tyr 203, Phe 204, Leu 259 and Val 263). **c**, Stereo pair of the accessible surface of the central cavity of the GroEL–GroES complex. The complex is in the same orientation as in Fig. 1b. The three subunits from each of the rings nearest the viewer were removed to show the inside of the assembly. The solvent-exposed surface (assuming the assembly is complete) is coloured based on the underlying atoms: all backbone atoms, white; all hydrophobic side-chain atoms (Ala, Val, Leu, Ile, Met, Phe, Pro and Tyr), yellow; all polar and charged side-chain atoms (Ser, Thr, His, Cys, Asn, Gln, Lys, Arg, Asp and Glu), blue. All solvent-excluded surface at the subunit interfaces are grey. Note the yellow hydrophobic patches on the surface of the *trans* GroEL cavity and blue polar patches on the surface of *cis* GroEL cavity. **a**, **b**, Produced using InsightII (BioSym Technology); **c**, produced using Grasp[58].

Nature. 2009 Jan 1; 457(7225):107-10. Chaperonin complex with a newly folded protein encapsulated in the folding chamber. Saibil HR.



Horwich AL, Fenton WA. Chaperonin-mediated protein folding: using a centralcavity to kinetically assist polypeptide chain folding. Q Rev Biophys. 2009May;42(2):83-116.

The chaperonin ring assembly GroEL provides kinetic assistance to protein folding in the cell by binding non-native protein in the hydrophobic central cavity of anopen ring and subsequently, upon binding ATP and the co-chaperonin GroES to thesame ring, releasing polypeptide into a now hydrophilic encapsulated cavity whereproductive folding occurs in isolation. The fate of polypeptide during binding,encapsulation, and folding in the chamber has been the subject of recent experimental studies. We conclude that GroEL,in general, behaves passively with respect to its substrate proteins during thesesteps. While binding appears to be able to rescue non-native polypeptides from kinetic traps, such rescue is most likely exerted at the level of maximizing hydrophobic contact, effecting alteration of the topology of weakly structured states. Encapsulation does not appear to involve 'forced unfolding', and if anything, polypeptide topology is compacted during this step. Finally, chamber-mediated folding appears to resemble folding in solution, except that major kinetic complications of multimolecular association are prevented.

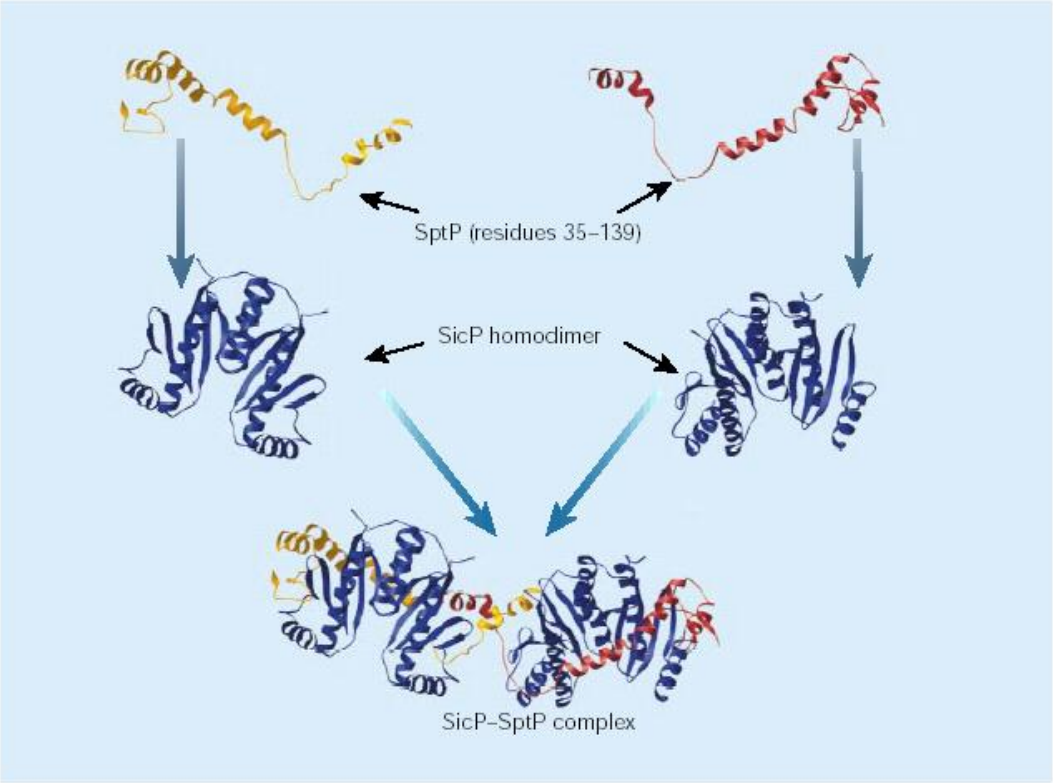


Figure 1 Formation of the complex between SicP (the chaperone) and SptP (the *Salmonella* effector protein). The non-globular, chaperone-binding domain of SptP (yellow and red) wraps around a SicP homodimer (blue). The crystal structure reported by Stebbins and Galán[1] is a 4:2 complex, created by the combination of two 2:1 SicP–SptP complexes.

Chaperones associated with the type III secretion system maintain their substrates in a secretion -competent state that is capable of engaging the secretion machinery to travel through the type III apparatus in an unfolded or partially folded manner.

There appears to be general molecular chaperone akin to GroEL In the eukaryotic cytosol. Was loss of chaperone machinery important for eukaryotic evolution??

Recombination of protein domains facilitated by co-translational folding in eukaryotes

William J. Netzer* & F. Ulrich Hartl*†

* *Cellular Biochemistry & Biophysics Program and Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA*

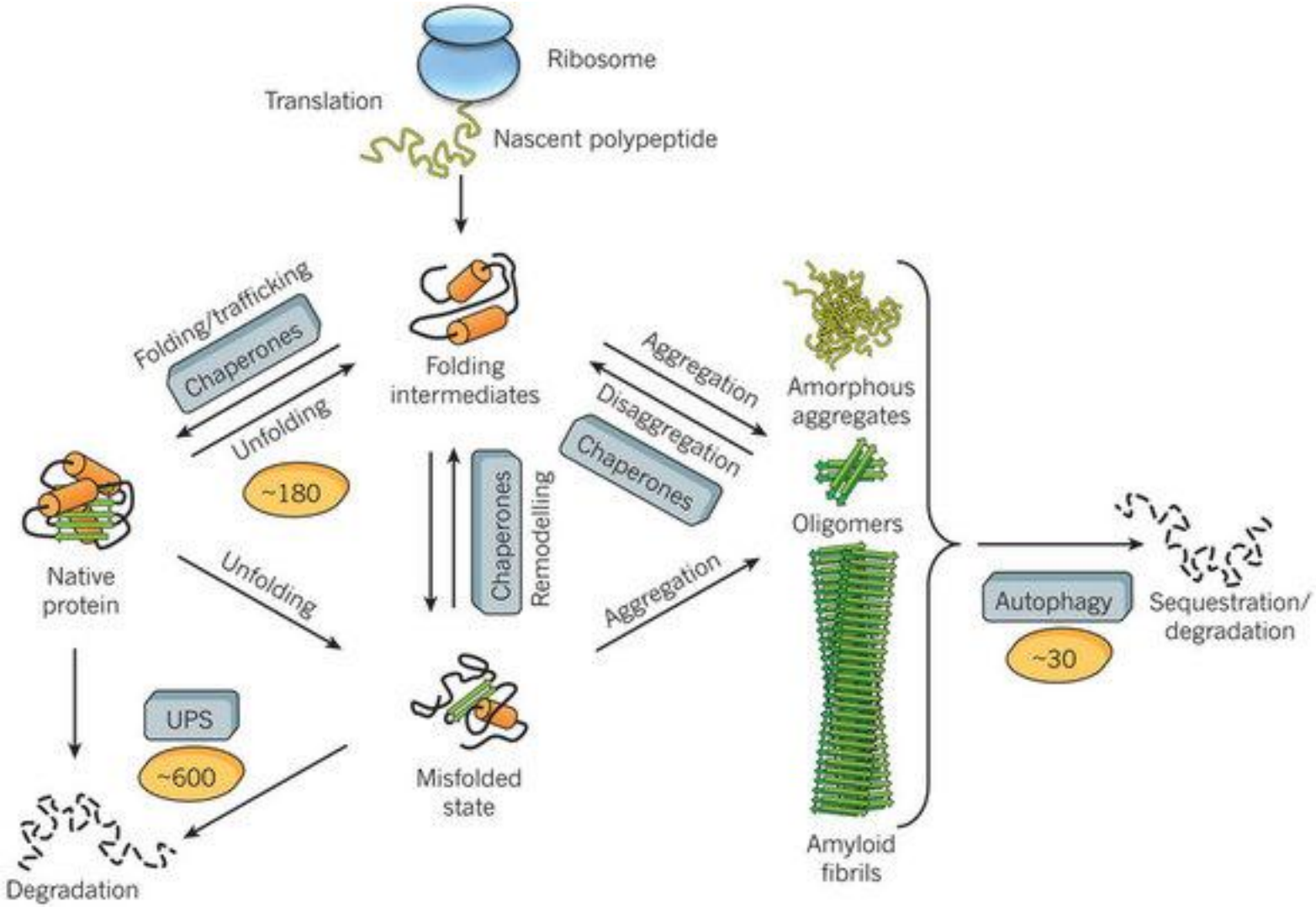
† *Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, 82152 Martinsried, Germany*

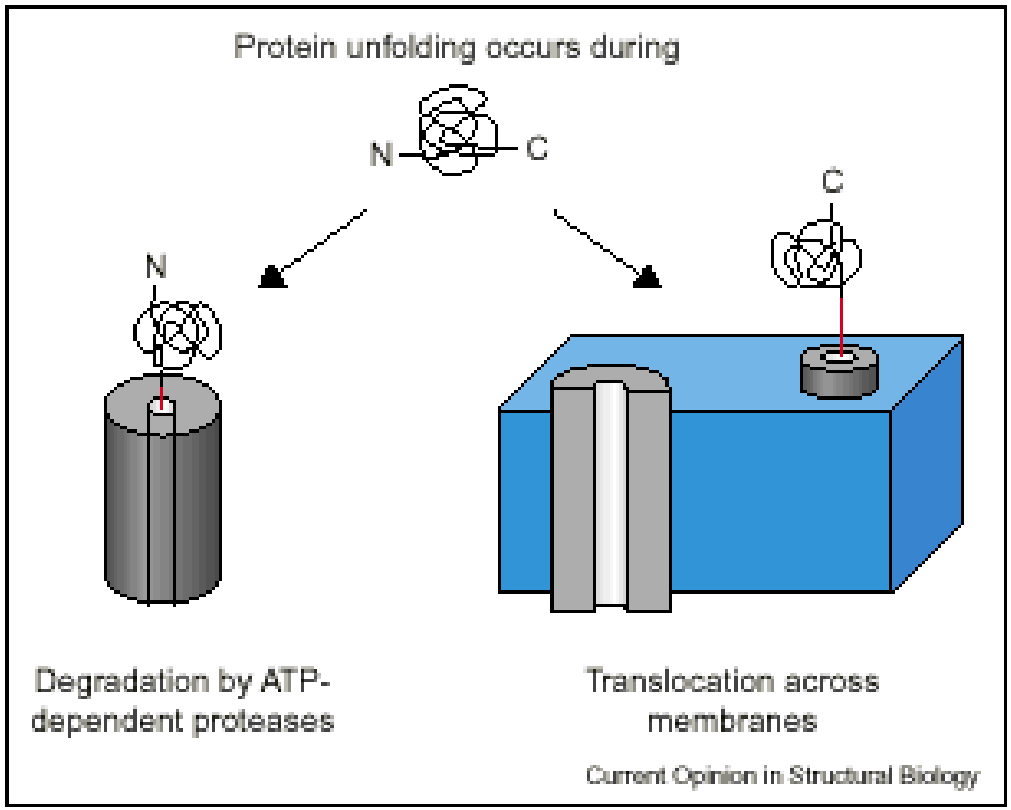
The evolution of complex genomes requires that new combinations of pre-existing protein domains successfully fold into modular polypeptides. During eukaryotic translation model two-domain polypeptides fold efficiently by sequential and co-translational folding of their domains. In contrast, folding of the same proteins in *Escherichia coli* is post-translational, and leads to intramolecular misfolding of concurrently folding domains. Sequential domain folding in eukaryotes may have been critical in the evolution of modular polypeptides, by increasing the probability that random gene-fusion events resulted in immediately foldable protein structures.

The evolution of eukaryotes has been characterized by a dramatic increase in the diversity and structural complexity of proteins. This

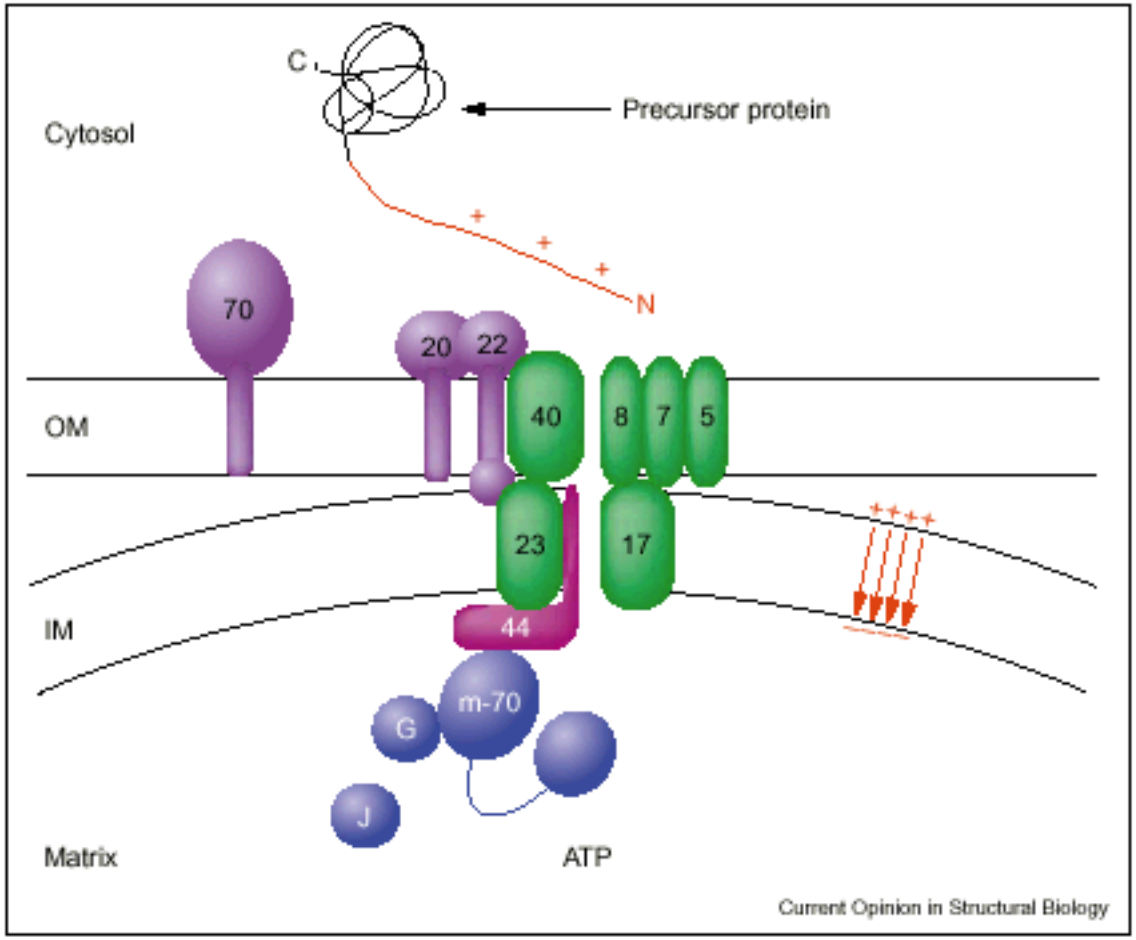
evolved. This process has occurred with much higher frequency in eukaryotes than in prokaryotes, as is evident from a simple

Ubiquitin-proteaseome system degrades misfolded proteins

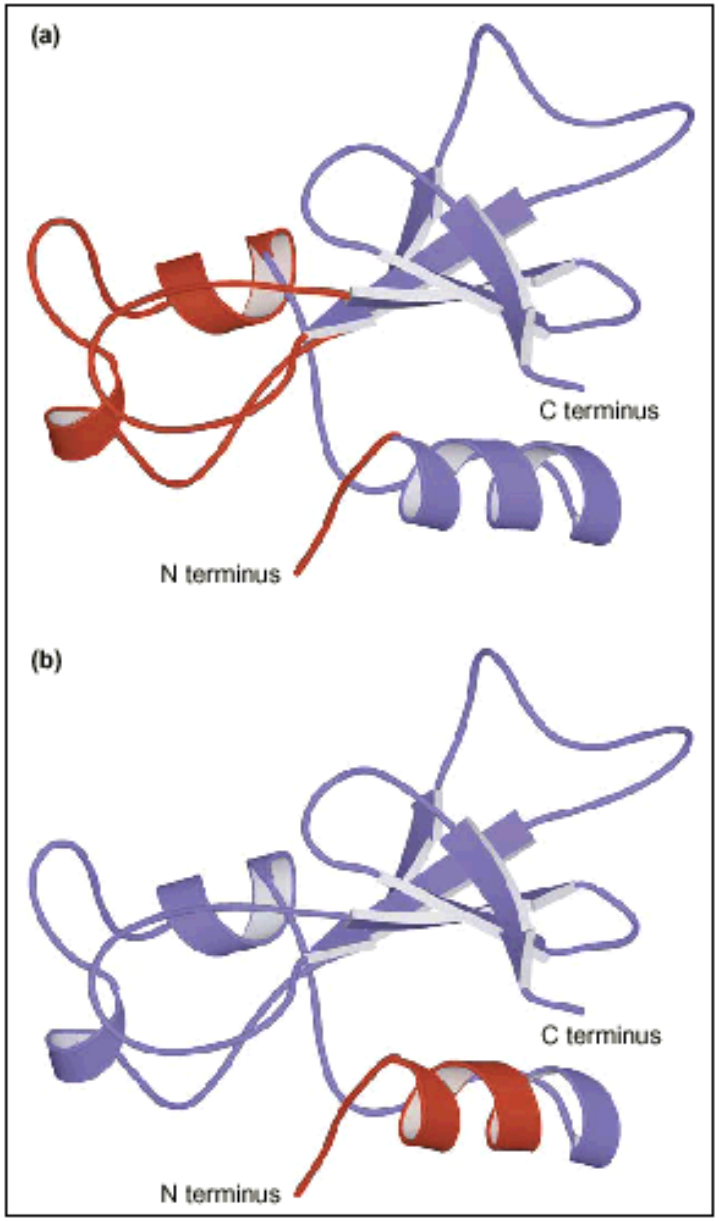




Protein unfolding occurs during translocation across some membranes and during degradation by ATP-dependent proteases (the targeting signal is shown in red).

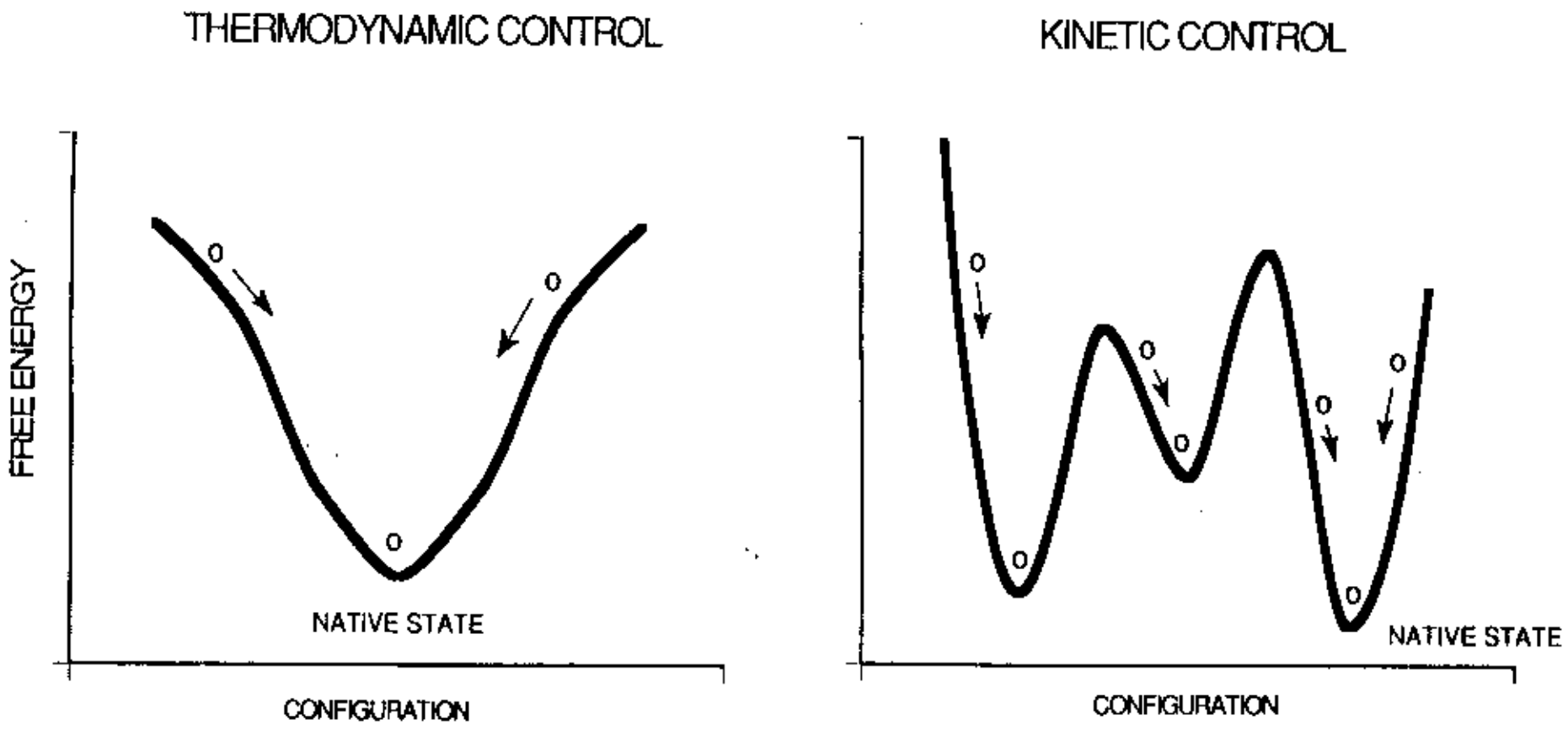


The mitochondrial protein import machinery. Proteins in the outer/inner membrane are called Tom/Tim, followed by the number indicated in the figure. The number reflects their approximate molecular weight. During import, precursor proteins first interact with the Tom20 and Tom22 receptors through their targeting sequence. The Tom70 receptor binds precursors associated with cytosolic chaperones. Targeting sequences insert into the Tom40 channel and pass through the Tom23 complex into the matrix. Import into the matrix always requires an electrical potential across the inner membrane and the ATP-dependent action of mtHsp70. mtHsp70 is found bound to the import machinery through Tim44 and free in the matrix. Precursors begin to interact with mtHsp70 while they are still associated with the import channels. G, J: mGrpE, Mdj1 — two co-chaperones of Hsp70; IM/OM: inner/outer membrane; m-70: mtHsp70.



Unfolding pathways of barnase [18]. Sketches of the structure of barnase, color coded according to the order in which structure is lost **(a)** during spontaneous global unfolding *in vitro* and **(b)** during import into mitochondria. The parts of the structure shown in red unfold early, whereas those shown in blue unfold late. Figure reproduced from [18] with permission.

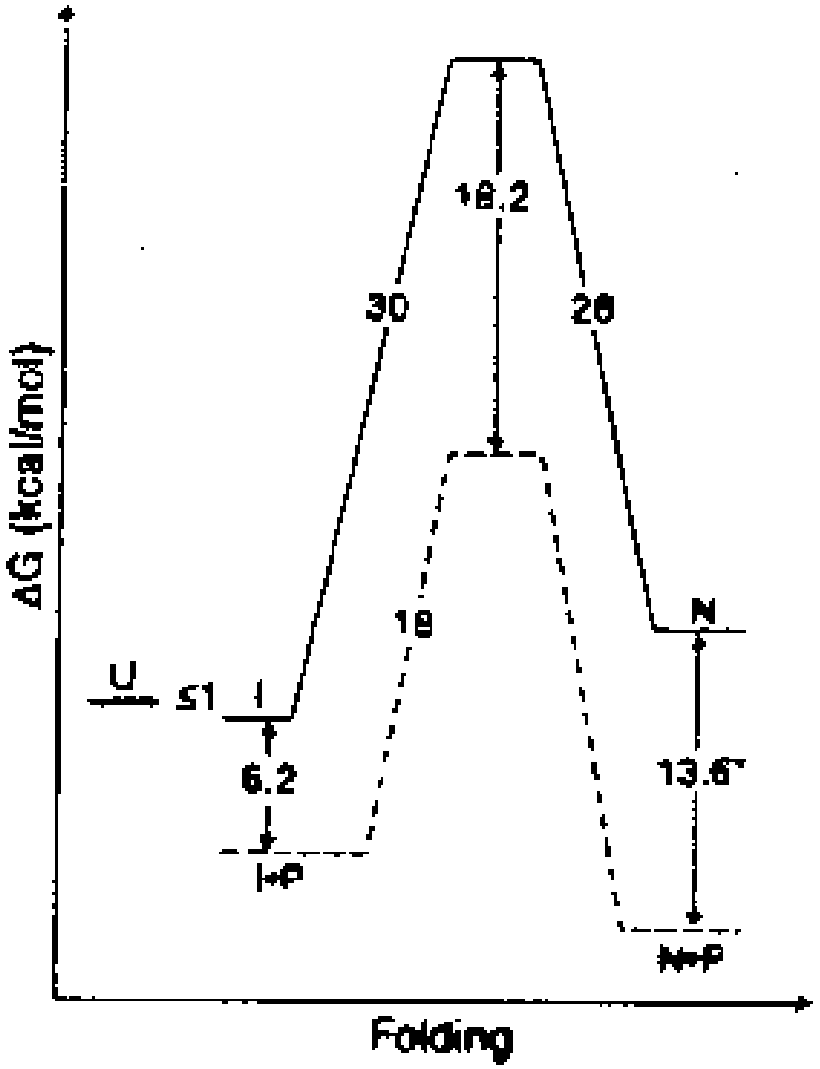
Two scenarios for protein folding



There are now several examples of protein folding reactions which are clearly under kinetic control:

- 1) Pro region dependent folding
- α-lytic protease and a number of other serine proteases are synthesized with large N-terminal pro regions
- The pro regions are required for the proper folding of the mature protease domains.
- Folding in the absence of the pro region results in an inactive conformation with completely different physical properties

Free energy diagram for alpha lytic protease folding. Note that in the absence of the pro region, N is higher than I



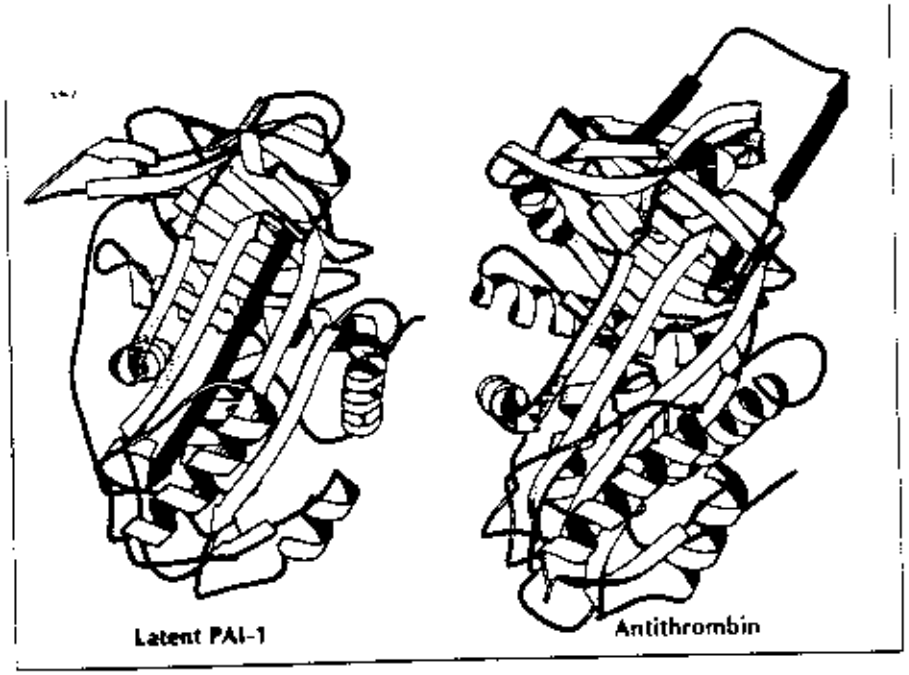
Serpins. (a large family of protease inhibitors)

Dramatic instance of kinetic control:

The initially synthesized form of PAI-1 is an active protease inhibitor.

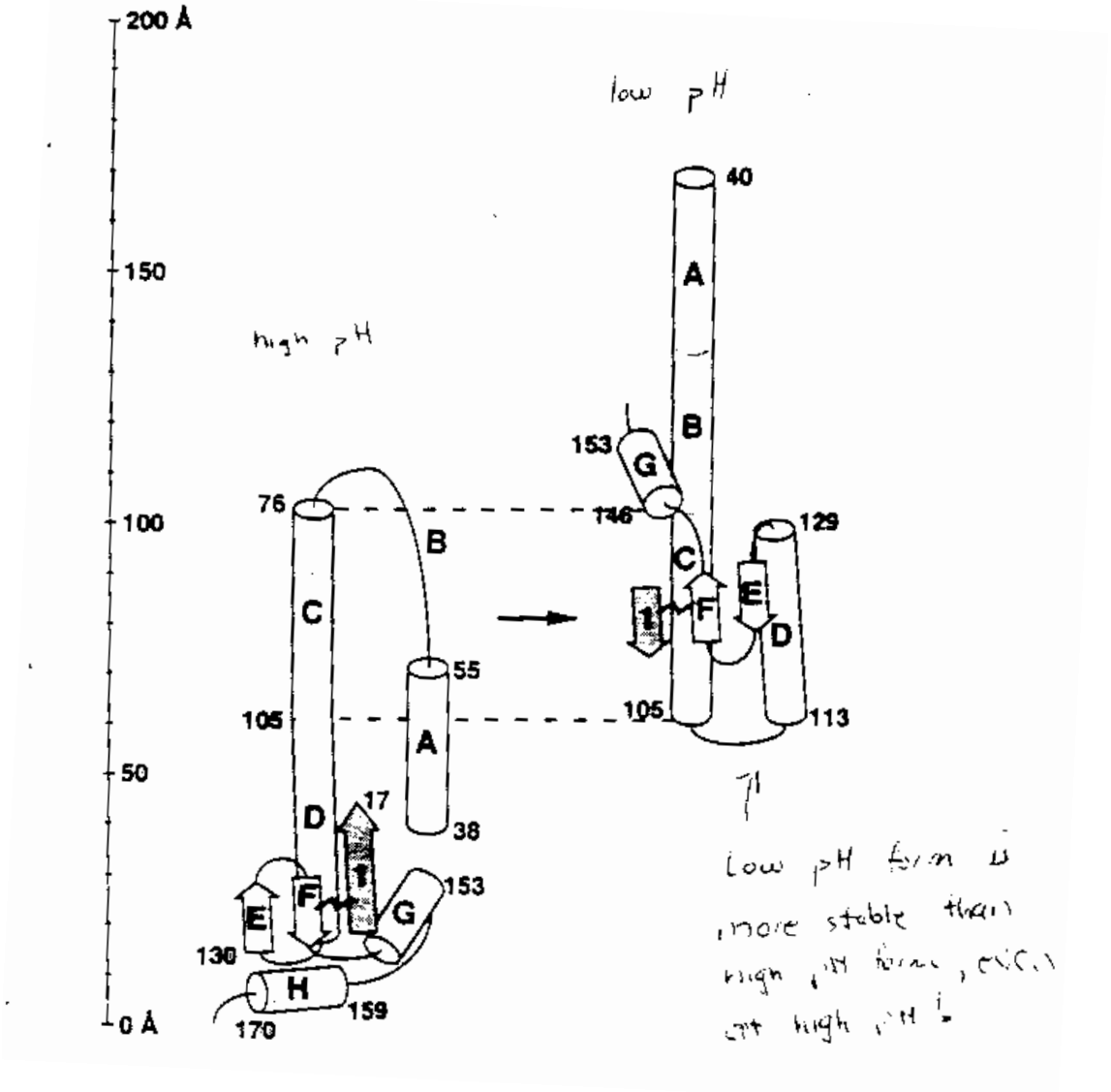
This active form slowly converts to an inactive, latent form over a period of several hours.

The latent form can be converted back to the active inhibitory form by denaturation and renaturation.



Influenza hemagglutinin (HA)

HA is a trimeric viral envelope glycoprotein which undergoes a dramatic conformational change at low pH, triggering the fusion of endocytosed virus with the endosomal membrane.



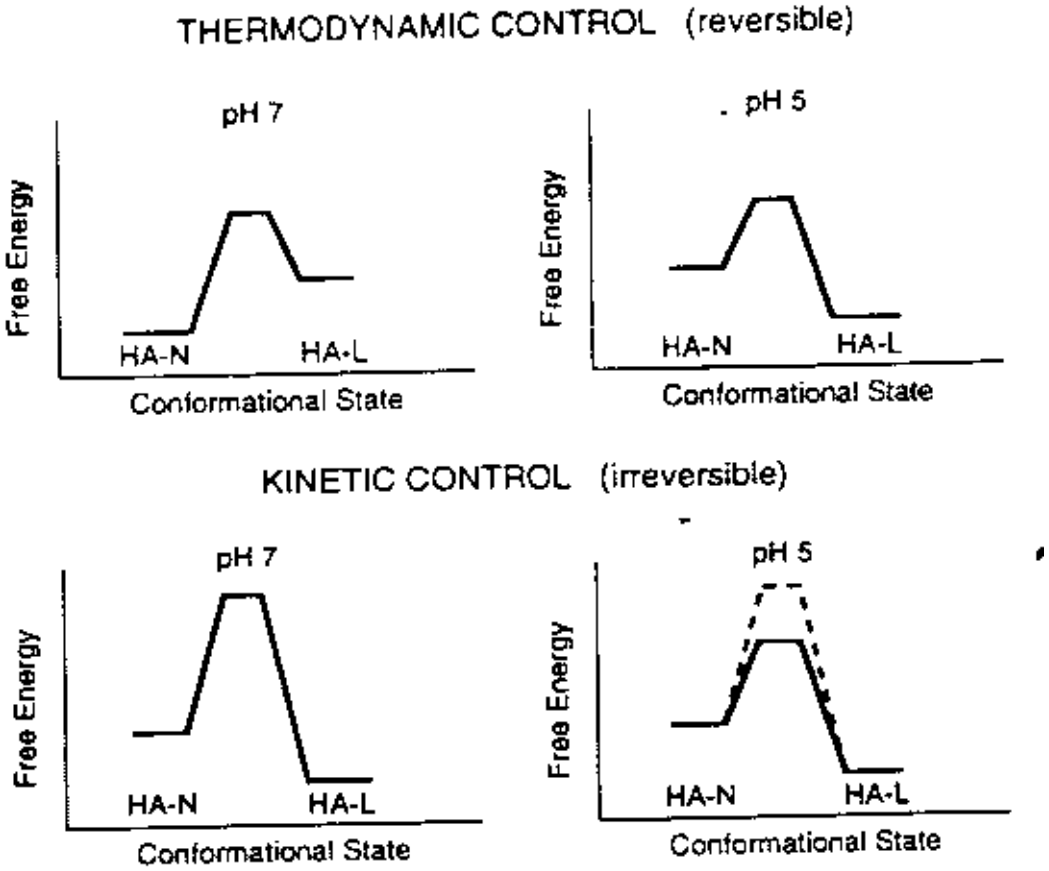


FIGURE 3: Possible free energy diagrams for the pH-induced transition in influenza hemagglutinin. The thermodynamic control situation is depicted in the top panel where the most stable state simply changes as a consequence of altering the pH. Such changes should be fully reversible. By contrast, the bottom panel depicts the system under kinetic control. Here, the initially formed structure at neutral pH (HA-N) is kinetically blocked from obtaining the most stable state (HA-L). Lowering the pH stabilizes the energy barrier, allowing the transition to HA-L. In this case, the processes would be irreversible.

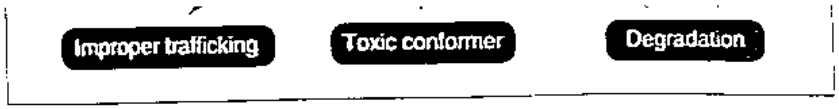


Figure 1

Diversity of protein misfolding in disease. The ability of a protein to perform its function in the cell depends upon its ability to assume a functional conformation. Thus, mutations and environmental changes that destabilize the native state, or divert proteins from their normal folding pathway, underlie several human diseases. For example, alteration of the thermodynamic stabilities of the native state and crucial folding intermediates may prevent folding into the functional conformation on a biological time scale. The cell's quality control apparatus must then recognize the misfolded and partially folded products and mark them for recycling. Off-pathway traps can be caused by aggregation, mistargeting into an inappropriate cellular location, or proteolysis of the polypeptide. Proteins and peptides that are aggregated into amyloid plaques are often resistant to degradation. Furthermore, the formation of these deposits, rather than the lack of native protein, may be responsible for the cellular pathology.

folding intermediate[19,22], indicating that the defect may be in the folding pathway itself (Fig. 2). These results agree with the finding that maturation of ΔF508 CFTR is altered at an early, perhaps cotranslational step[23,24], and that the ΔF508 mutant protein, if folded into a native state, is functional under physiological conditions[19,20]. This information suggests that correcting the defect in the folding pathway may ameliorate this form of the disease.

Mutation of fibrillin leads to the autosomal dominant disorder of connective

Table I. Some putative protein folding diseases

Disease	Mutant protein/protein involved	Molecular phenotype
Inability to fold		
Cystic fibrosis	CFTR	Misfolding/altered Hsp70 and calnexin interactions
Marfan syndrome	Fibrillin	Misfolding
Amyotrophic lateral sclerosis	Superoxide dismutase	Misfolding
Scurvy	Collagen	Misfolding
Maple syrup urine disease	α-Ketoacid dehydrogenase complex	Misassembly/misfolding
Cancer	p53	Misfolding/altered Hsp70 interaction
Osteogenesis imperfecta	Type I procollagen pro α	Misassembly/altered BiP expression
Toxic folds		
Scrapie/Creutzfeldt-Jakob/ familial insomnia	Prion protein	Aggregation
Alzheimer's disease	β-Amyloid	Aggregation
Familial amyloidosis	Transthyretin/lysozyme	Aggregation
Cataracts	Crystallins	Aggregation
Mislocalization owing to misfolding		
Familial hypercholesterolemia	LDL receptor	Improper trafficking
α_1-Antitrypsin Deficiency	α_1-Antitrypsin	Improper trafficking
Tay–Sachs disease	β-Hexosaminidase	Improper trafficking
Retinitis pigmentosa	Rhodopsin	Improper trafficking
Leprechaunism	Insulin receptor	Improper trafficking

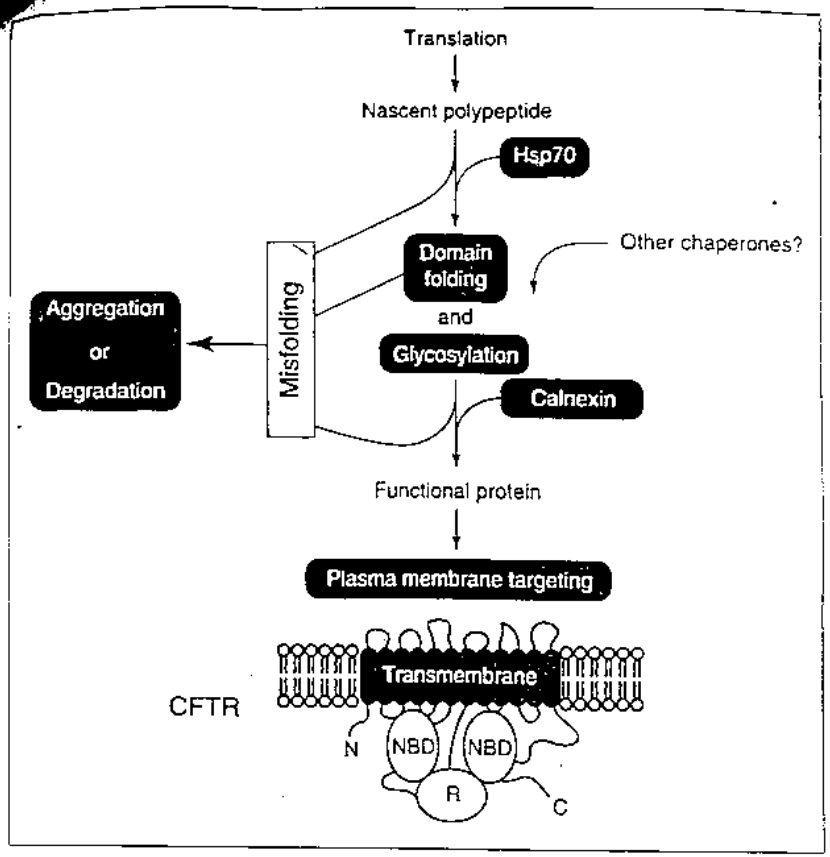


Figure 2

Many cases of cystic fibrosis are due to inappropriate protein folding. Cystic fibrosis (CF) provides a paradigm for understanding a folding disease. Over 400 CF-causing mutations have been identified; most of the small number assayed to date prevent maturation of the cystic fibrosis transmembrane conductance regulator (CFTR) protein[43]. Individual steps of the process that have been shown to be altered in the disease state are shown in red. The most common mutation of CFTR (ΔF508) prevents the folding and targeting of CFTR to the plasma membrane under physiological conditions; destabilizes a portion of the amino-terminal nucleotide binding domain (NBD)[19]; increases the temperature sensitivity of maturation but not of function[20]; and alters the kinetics of maturation[23,24] and of association of CFTR with Hsp70 and calnexin[35,37]. The order of the events depicted in this scheme is not completely established.

tissue called Marfan syndrome[25]. The bulk of disease-causing mutations occur in the numerous epidermal growth factor (EGF)-like domains that fibrillin contains. A 43 amino acid peptide corre-

population, impedes assembly of the complex[27]. Another disease, class 2 hypercholesterolemia, arises from mutations in the low density lipoprotein (LDL)-receptor locus, producing pro-

rather than the
may be respo
pathology. The
cussed below.
genetic diseas
ability of the m
its functional c

However, no
folding are due
for example, th
can affect the s
teins and ind
chaperones in
are known as
addition, nutri
often vital for
example, scurv
vitamin C in th
the inability to
collagen triple
proline is stab
peratures whi
absence of prol
at 24°C (Ref. 30

Toxic folds

Pathological
folding does n
inability of a
stable structu
extremely stab
itself toxic to t
tions that alte
prion protein
structure[31]. Po
protein is in
conformation
ation of the am
both the β-she
and its infectiv
oid plaques in
be a conseque
transition from
amino-terminal
β-peptide[32]. Si
amyloid fibrils
netic control[33]
to note, that
β-thalassemia

A large number of proteins have been observed to form amyloid fibrils

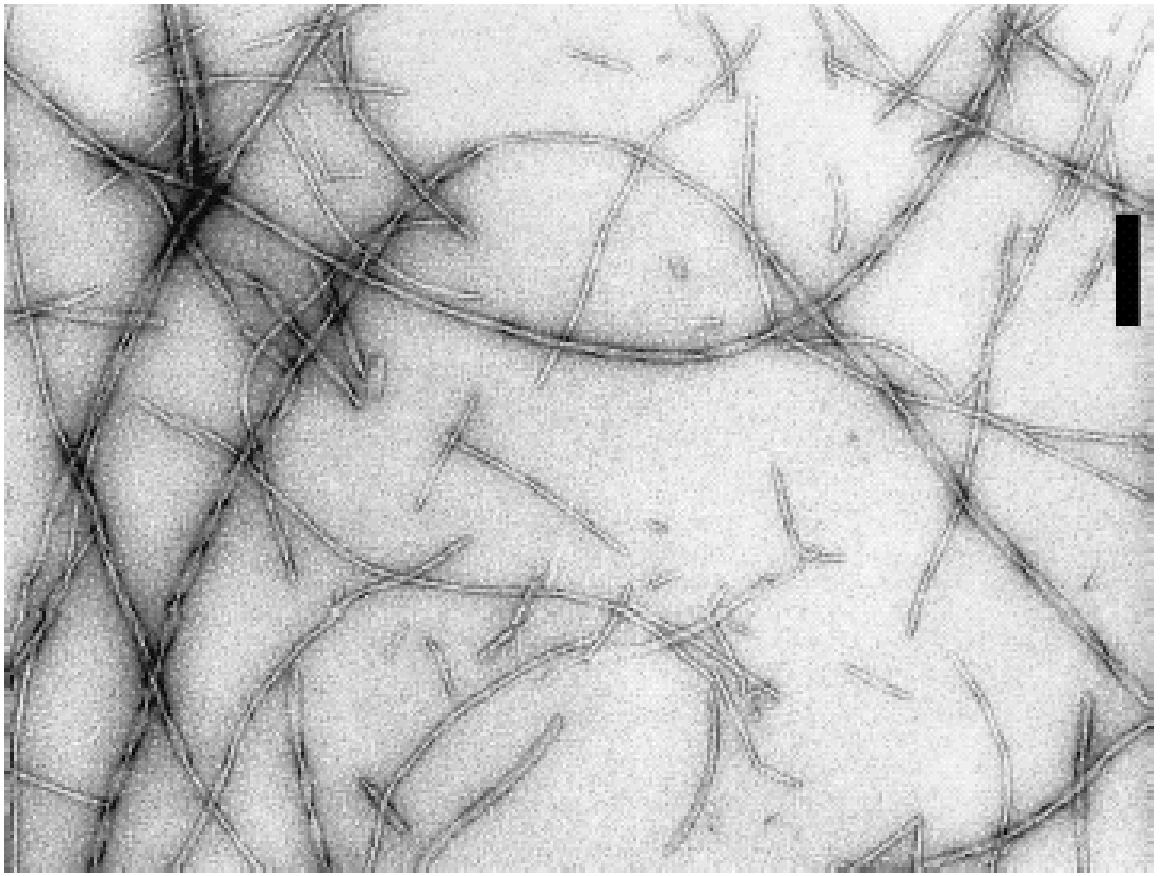


Figure 3. Amyloid fibrils from the Ile56Thr variant of human lysozyme produced by transmission electron microscopy. Scale bar, 200 nm. (From Morozova-Roche *et al.* 2000.)

Table 2. *Fibril protein components and precursors in amyloid diseases*

(Data from Sunde *et al.* (1997).)

clinical syndrome	fibril component
Alzheimer's disease	Aβ peptide, 1-42, 1-43
spongiform encephalopathies	full length prion or fragments
primary systemic amyloidosis	intact light chain or fragments
secondary systemic amyloidosis	76-residue fragment of amyloid A protein
familial amyloidotic poly neuropathy I	transthyretin variants and fragments
senile systemic amyloidosis	wild-type transthyretin and fragments
hereditary cerebral amyloid angiopathy	fragment of cystatin-C
haemodialysis-related amyloidosis	β2-microglobulin
familial amyloidotic polyneuropathy II	fragments of apolipoprotein A-I
Finnish hereditary amyloidosis	71-residue fragment of gelsolin
type II diabetes	fragment of islet-associated polypeptide
medullary carcinoma of the thyroid	fragments of calcitonin
atrial amyloidosis	atrial natriuretic factor
lysozyme amyloidosis	full length lysozyme variants
insulin-related amyloid	full length insulin
fibrinogen α-chain amyloidosis	fibrinogen α-chain variants

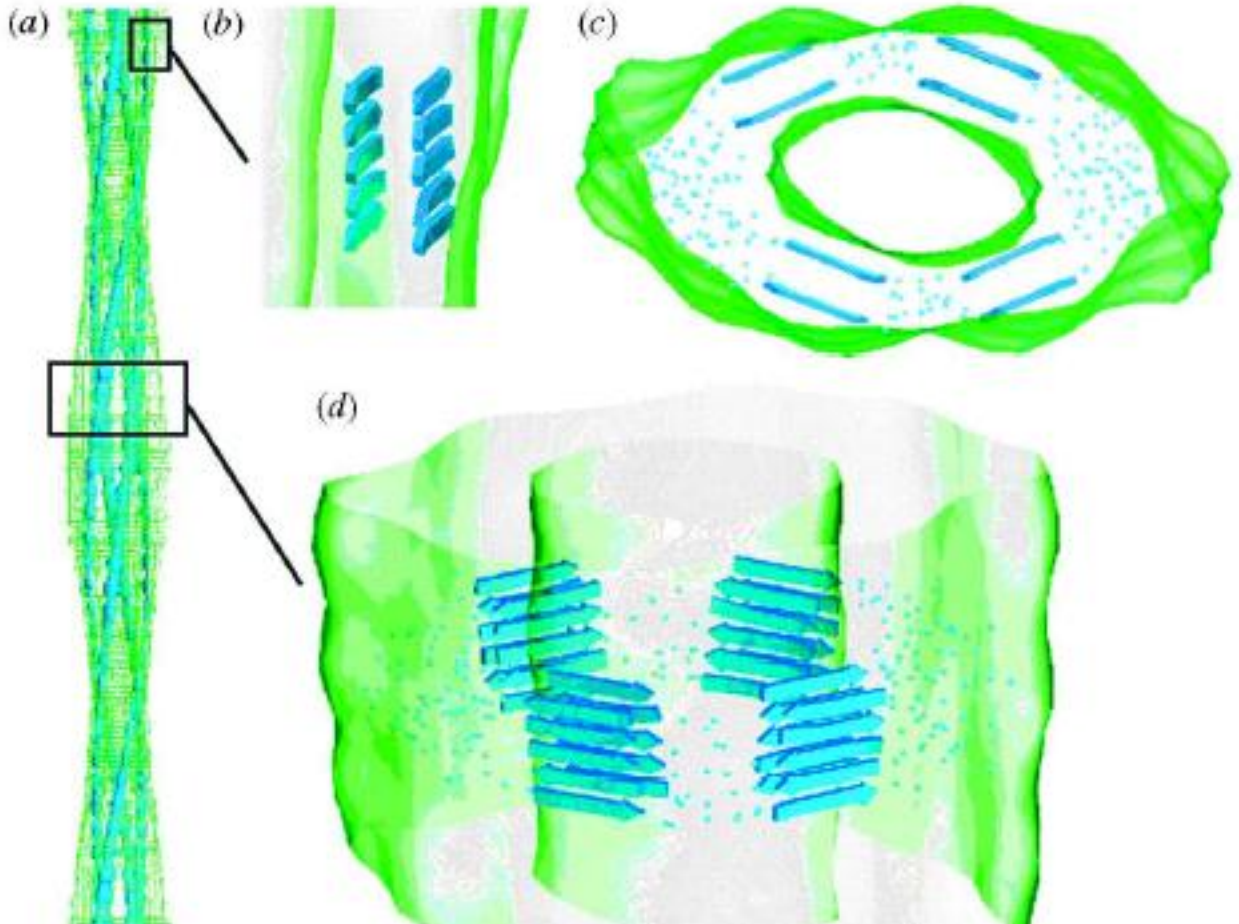


Figure 7. Model of the polypeptide fold in SH3 fibrils. (*a*) Overview of the fibril structure, showing the outer surface as
mesh and the protofilaments as solid blue surfaces. The ribbon-like protofilaments form the skeleton of the fibril structu
model for the molecular packing is shown in (*b*–*d*), with the electron micrograph map as a transparent rendered surface
view of a single protofilament. (*c*) Cross-section of the fibril and (*d*) slightly tilted side view of the fibril. β-sheets derived
SH3 structure have been fitted into the map, after opening the β-sandwich fold and reorientating and straightening the
The remaining regions of polypeptide sequence are shown as disconnected dots, to indicate the number of residues prese
not the conformation. The β-sheets contain a mixture of parallel and antiparallel strands. This particular arrangement
arbitrary and was chosen because it required the least rearrangement of the native β-sheet structure, although there is n

Structure of the cross-β spine of amyloid-like fibrils