Protein Folding In Vivo

Biochemistry 412

March 7th, 2006

But first, before we talk about
***in vivo* phenomena…**

some more theory!

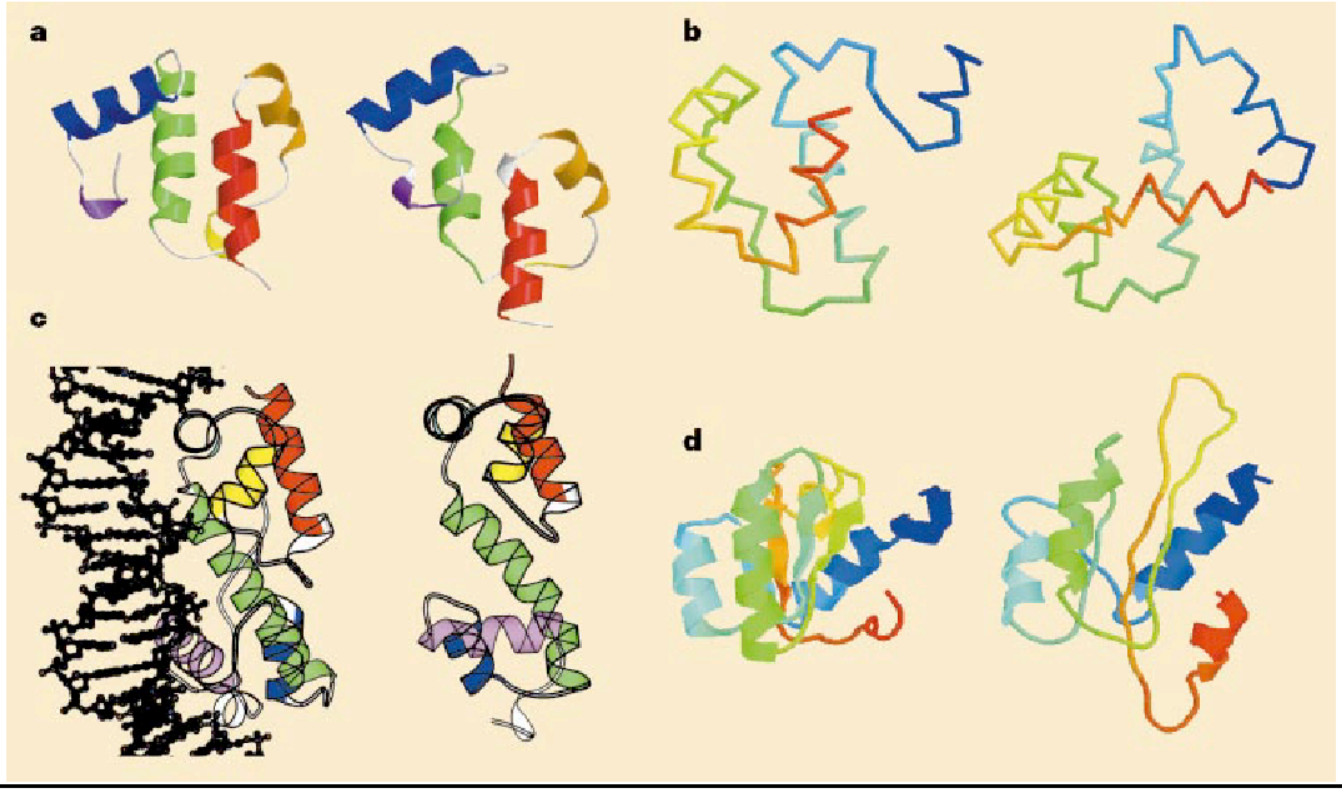
Computational Protein Folding

How are the theorists doing lately?

Box 3

Ab initio structure predictions

Blind *ab initio* structure predictions for the CASP3 protein structure prediction experiment. For each target, the native structure is shown on the left with a good prediction on the right (predictions by Baker[39](**a**, **c**), Levitt[40](**b**) and Skolnick[41](**d**) and colleagues; for more information see http://predictioncentre.llnl.gov/ and *Proteins* Suppl. **3,** 1999). Segments are colour coded according to their position in the sequence (from blue (amino terminus) to red (carboxy terminus)). **a**, DNA B helicase[41]. This protein had a novel fold and thus could not be predicted using standard fold-recognition methods. Not shown are N- and C-terminal helices which were positioned incorrectly in the predicted structure. **b**, Ets-1 (ref. 43). **c**, MarA[44]. This prediction had potential for functional insights; the predicted two-lobed structure suggests the mechanism of DNA binding (left, X-ray structure of the protein–DNA complex). **d**, L30. A large portion of this structure was similar to a protein in the protein databank but the best *ab initio* predictions were competitive with those using fold-recognition methods. The three approaches that produced these predictions used reduced-complexity models for all or almost all of the conformational search process.



Baker (2000) *Nature* **405**, 39.

Progress in *de novo* protein structure prediction & design

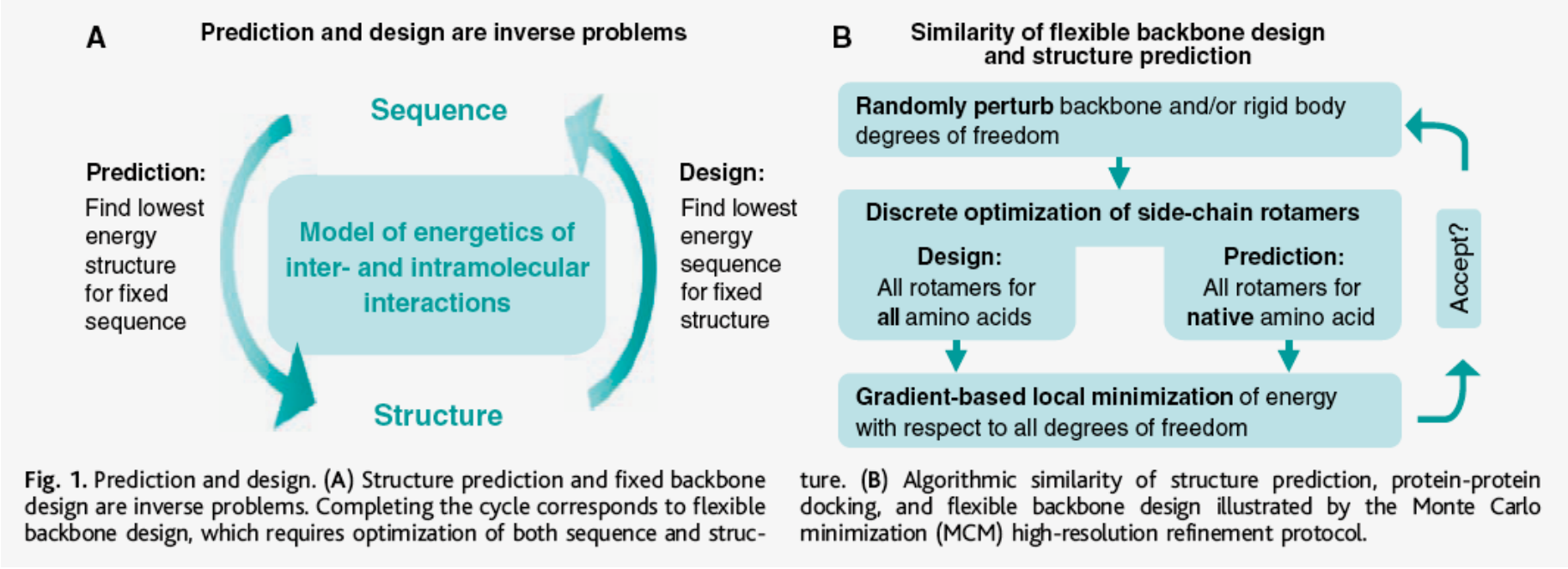


Fig. 1. Prediction and design. (A) Structure prediction and fixed backbone design are inverse problems. Completing the cycle corresponds to flexible backbone design, which requires optimization of both sequence and structure. (B) Algorithmic similarity of structure prediction, protein-protein docking, and flexible backbone design illustrated by the Monte Carlo minimization (MCM) high-resolution refinement protocol.

Schueler-Furman et al (2005) *Science* **310**, 638.

Present protein folding theory suggests that proteins must find the right polypeptide chain topology ("topomer") first, then they can form 2° structure and snap into the correct 3D conformation relatively quickly.

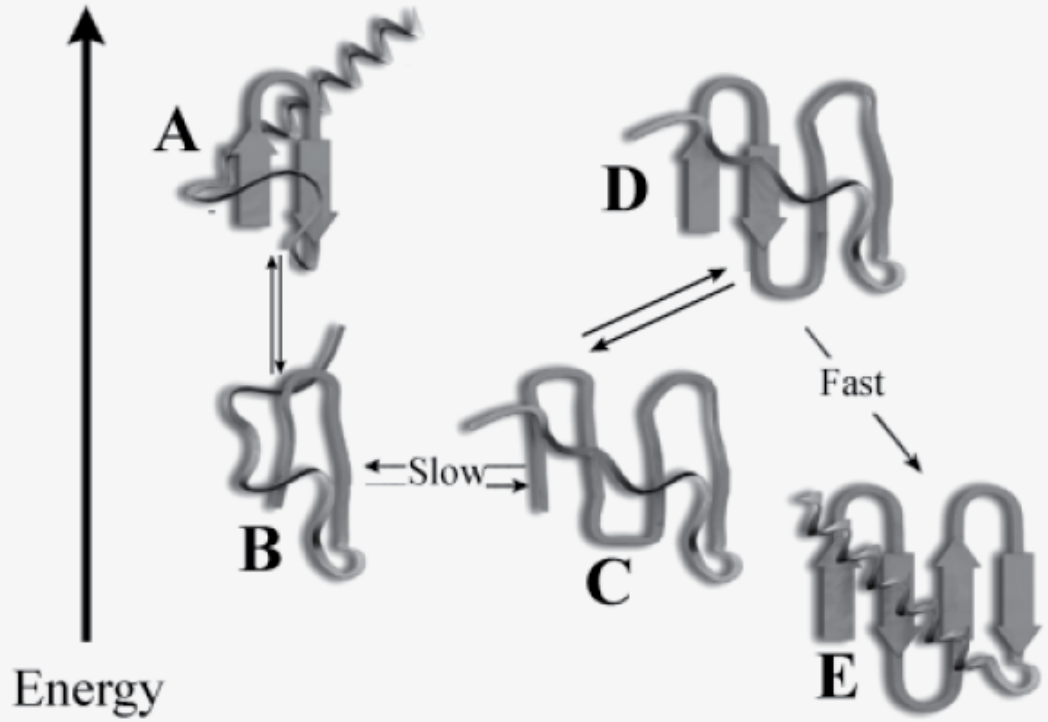


Figure 10 The essence of the topomer search model is that the rate with which an unfolded polymer diffuses between distinct topologies is much slower than the rate with which local structural elements zipper (and, critically, unzip) [reviewed in (59)].

Gillespie & Plaxco (2004) *Ann. Rev. Biochem.* **73**, 837.

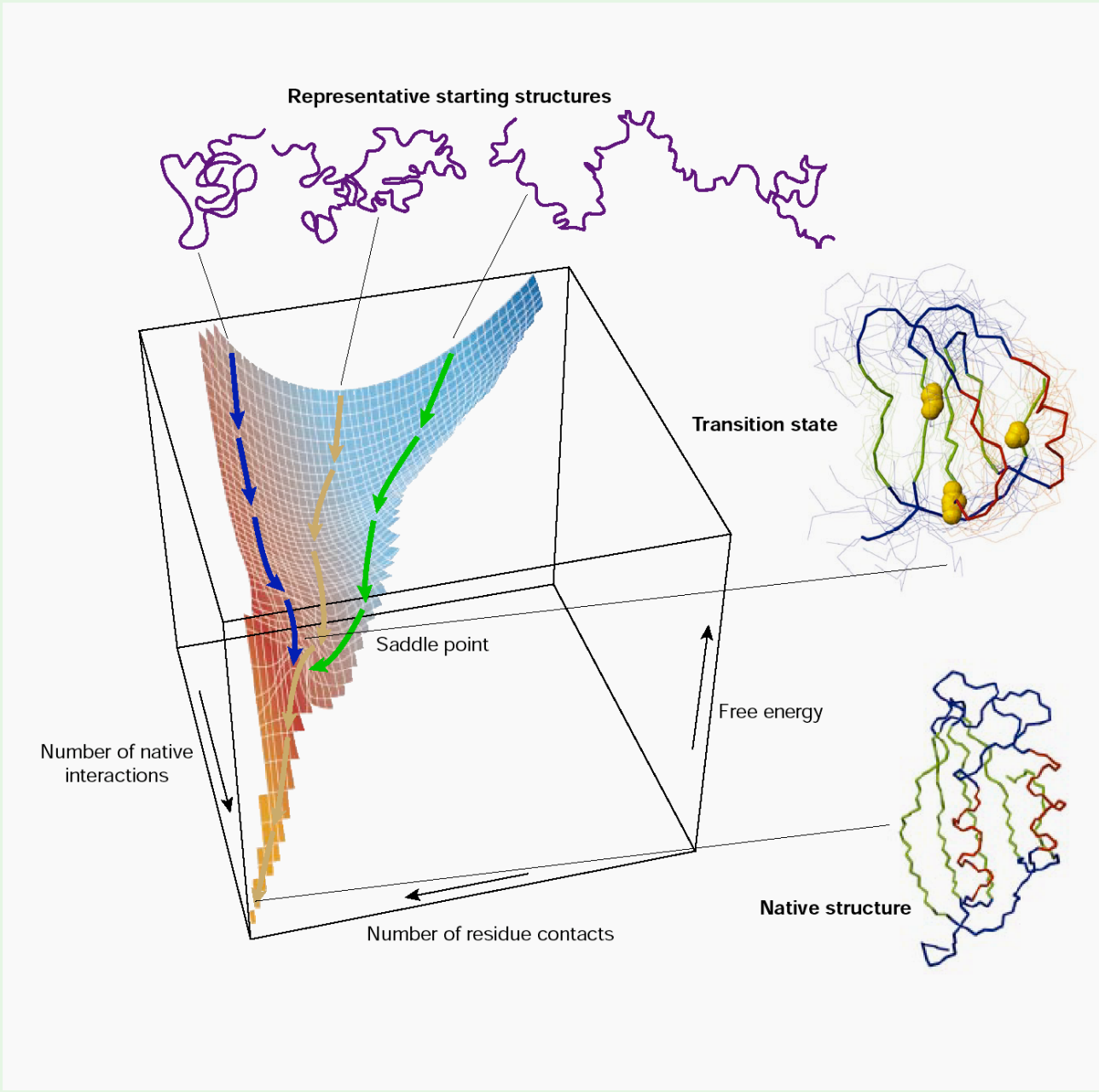


Figure 1 A schematic energy landscape for protein folding. The surface is derived from a computer simulation of the folding of a highly simplified model of a small protein. The surface 'funnels' the multitude of denatured conformations to the unique native structure. The critical region on a simple surface such as this one is the saddle point corresponding to the transition state, the barrier that all molecules must cross if they are to fold to the native state. Superimposed on this schematic surface are ensembles of structures corresponding to different stages of the folding process. The transition state ensemble was calculated by using computer simulations constrained by experimental data from mutational studies of acylphosphatase[18]. The yellow spheres in this ensemble represent the three 'key residues' in the structure; when these residues have formed their native-like contacts the overall topology of the native fold is established. The structure of the native state is shown at the bottom of the surface; at the top are indicated schematically some contributors to the distribution of unfolded species that represent the starting point for folding. Also indicated on the surface are highly simplified trajectories for the folding of individual molecules. Adapted from ref. 6.

Dobson (2003) *Nature* **426**, 884.

Calculating structures with their associated energies: Deep & narrow energy wells are a hallmark of near-correct structures

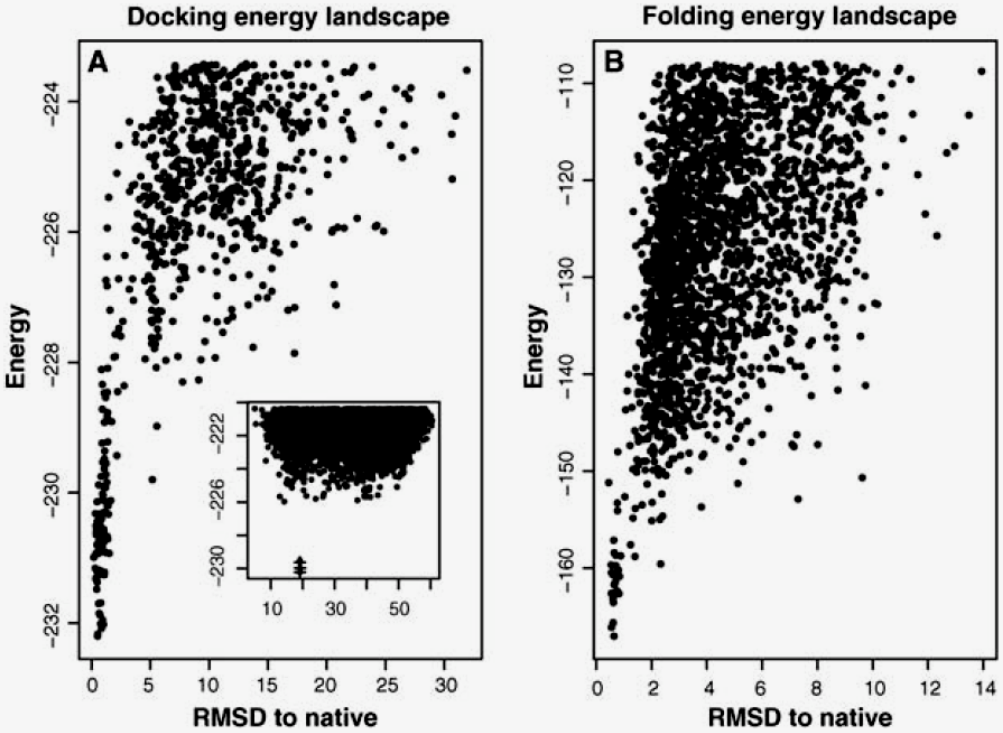


Fig. 2. Energy landscapes. Each point represents the lowest energy structure sampled in a single MCM trajectory. (**A**) Docking energy landscape for Capri Target 12 [cohesin-dockerin complex; PDB (Protein Data Bank) ID 1ohz (*66*)]. (**Inset**) In a large collection of trajectories starting from different random orientations that were carried out for the CAPRI experiment, a small number of structures (+) are distinguished from the background population by a significant energy gap. The *x* axis is the RMSD to an arbitrary reference orientation. (Main panel) Trajectories starting from these low-energy structures map out a narrow energy funnel. The *x* axis is the RMSD to the native structure. A deep energy funnel, as in this example, is a strong indicator that a prediction is correct. (**B**) Folding landscape for double-stranded RNA binding protein [PDB ID 1di2 (*67*)]. The backbone RMSD is to the native structure. The energy function (units are in kcal/mol) includes entropic contributions from solvation effects, but not the configurational entropy associated with protein vibrational and side-chain degrees of freedom, and hence is not the true free energy.

Schueler-Furman et al (2005) *Science* **310**, 638.

For small, single domain proteins, some of the predictions are getting very good!

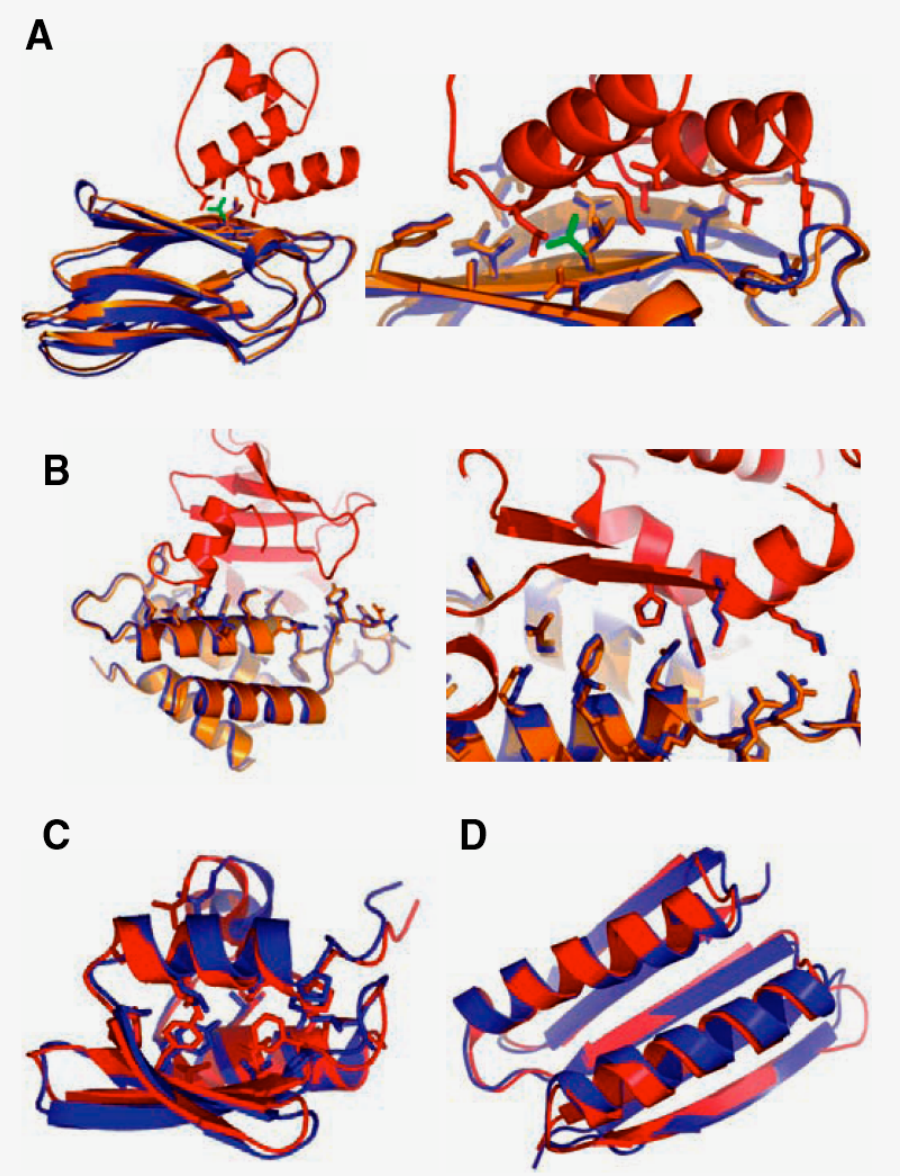


Fig. 3. Examples of high-resolution prediction and design. (**A**) CAPRI Target 12 [dockerin–cohesin (*66*); interface residue backbone RMSD = 0.27 Å]. The lowest energy structure in Fig. 2A, main panel, is shown here. The side chain of Leu-83 (green in the free monomer) changes conformation upon binding. Side-chain conformations in red were provided; those in blue were predicted. (**B**) CAPRI Target 15 [ColicinD–Immunity protein D (*68*); interface residue backbone RMSD = 0.23 Å]. No side-chain information was provided for either partner. (**C**) CASP6 de novo structure prediction Target 0281 [hypothetical protein from *Thermus thermophilus* Hb8, PDB ID 1whz (*69*); backbone RMSD = 1.59 Å]. (**D**) TOP7 (RMSD = 1.2 Å) (*42*). (A) and (B) are adapted from figure 1 of (*70*). Blue: models; red and orange: x-ray structures.

Schueler-Furman et al (2005) *Science* **310**, 638.

You, too, can do theoretical protein folding with David Baker!!

See http://boinc.bakerlab.org/rosetta

But make sure your computer doesn't overheat!

Protein Folding In Vivo

Molecular Chaperones

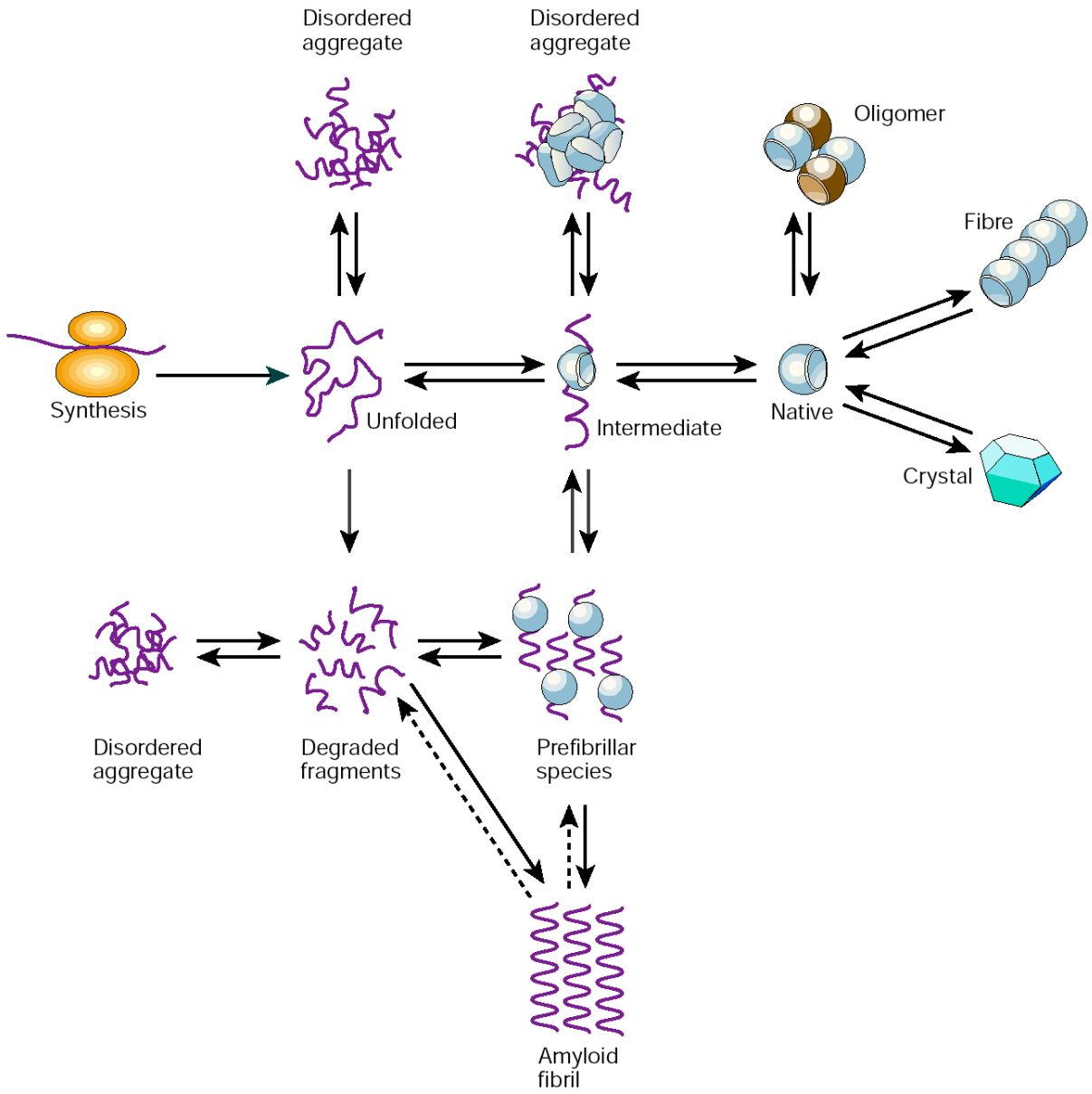


Figure 4 A unified view of some of the types of structure that can be formed by polypeptide chains. An unstructured chain, for example newly synthesized on a ribosome, can fold to a monomeric native structure, often through one or more partly folded intermediates. It can, however, experience other fates such as degradation or aggregation. An amyloid fibril is just one form of aggregate, but it is unique in having a highly organized 'misfolded' structure, as shown in Fig. 3. Other assemblies, including functional oligomers, macromolecular complexes and natural protein fibres, contain natively folded molecules, as do the protein crystals produced *in vitro* for X-ray diffraction studies of their structures. The populations and interconversions of the various states are determined by their relative thermodynamic and kinetic stabilities under any given conditions. In living systems, however, transitions between the different states are highly regulated by the environment and by the presence of molecular chaperones, proteolytic enzymes and other factors. Failure of such regulatory mechanisms is likely to be a major factor in the onset and development of misfolding diseases. Adapted from ref. 54.

Dobson (2003) *Nature* **426**, 884.

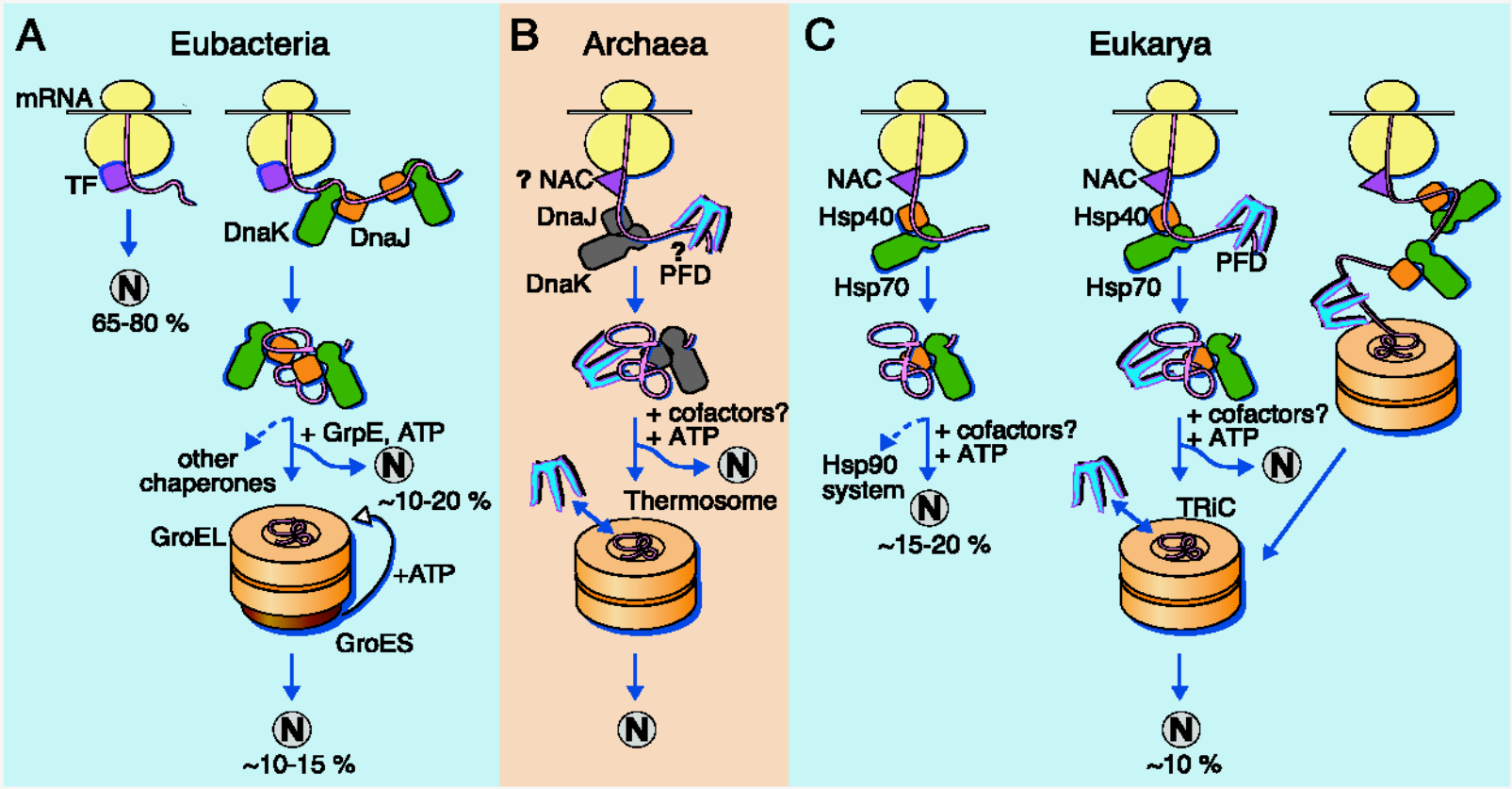
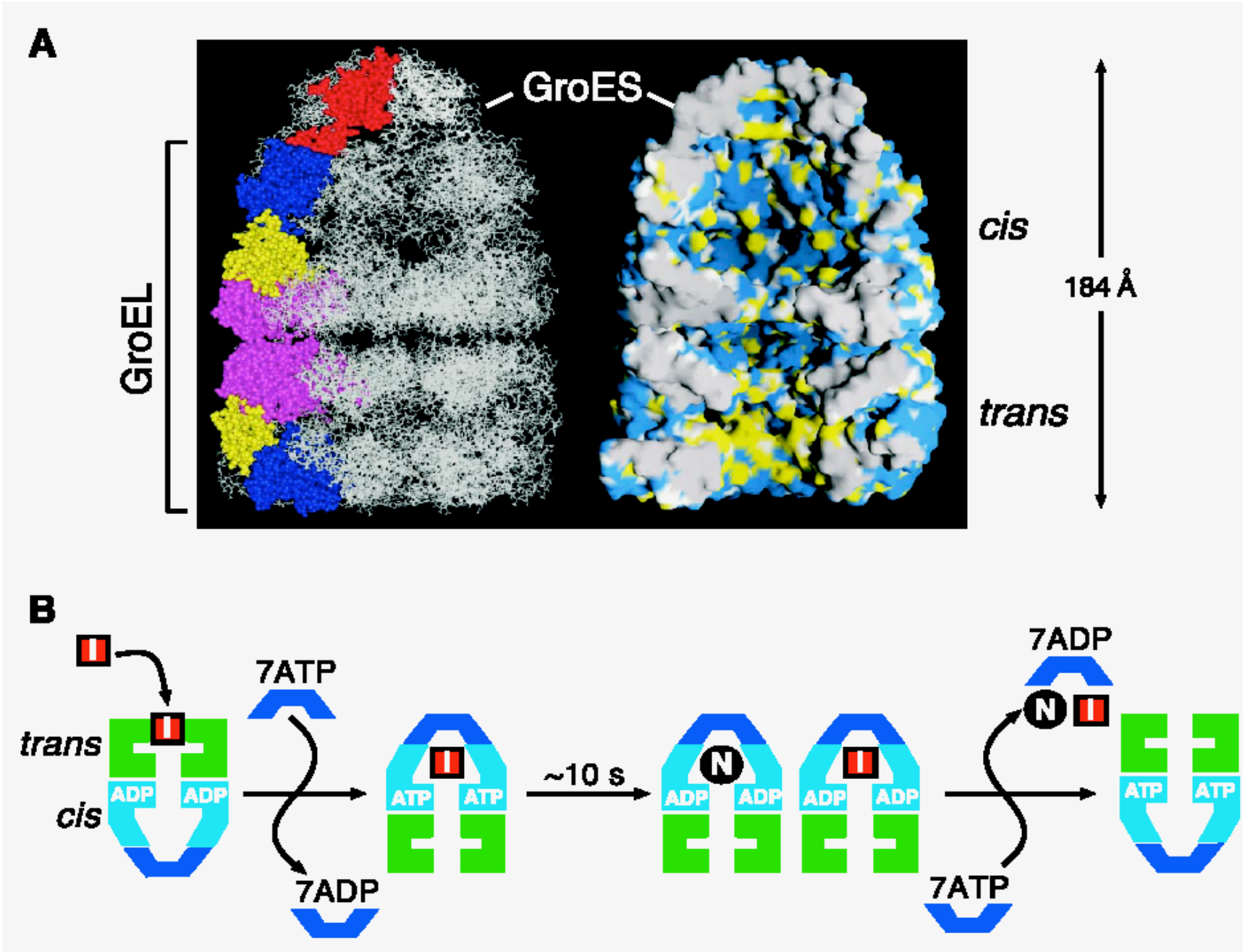


Fig. 2. Models for the chaperone-assisted folding of newly synthesized polypeptides in the cytosol. (**A**) Eubacteria. TF, trigger factor; N, native protein. Nascent chains probably interact generally with TF, and most small proteins (~65 to 80% of total) fold rapidly upon synthesis without further assistance. Longer chains (10 to 20% of total) interact subsequently with DnaK and DnaJ and fold upon one or several cycles of ATP-dependent binding and release. About 10 to 15% of chains transit the chaperonin system—GroEL and GroES—for folding. GroEL does not bind to nascent chains and is thus likely to receive an appreciable fraction of its substrates after their interaction with DnaK. (**B**) Archaea. PFD, prefoldin; NAC, nascent chain–associated complex. Only some archaeal species contain DnaK/DnaJ. The existence of a ribosome-bound NAC homolog, as well as the interaction of PFD with nascent chains, has not yet been confirmed experimentally. (**C**) Eukarya—the example of the mammalian cytosol. Like TF, NAC probably interacts generally with nascent chains. The majority of small chains may fold upon ribosome release without further assistance. About 15 to 20% of chains reach their native states in a reaction assisted by Hsp70 and Hsp40, and a fraction of these must be transferred to Hsp90 for folding. About 10% of chains are co- or posttranslationally passed on to the chaperonin TRiC in a reaction mediated by PFD.

Hartl & Hayer-Hartl (2002) *Science* **295**, 1852.



Hartl & Hayer-Hartl (2002) *Science* **295**, 1852.

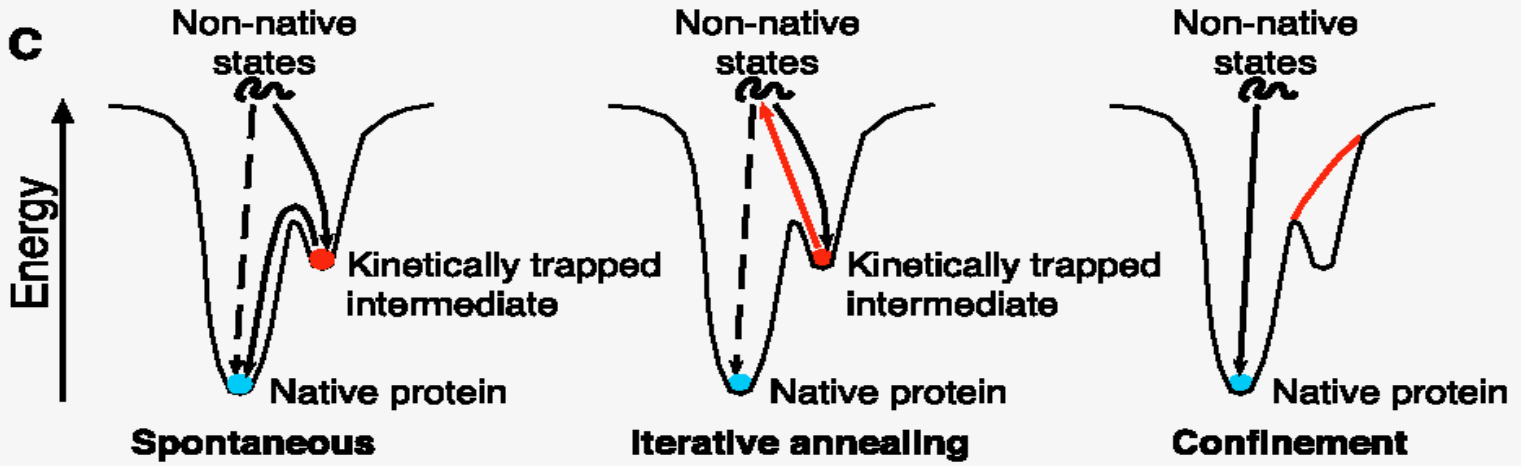


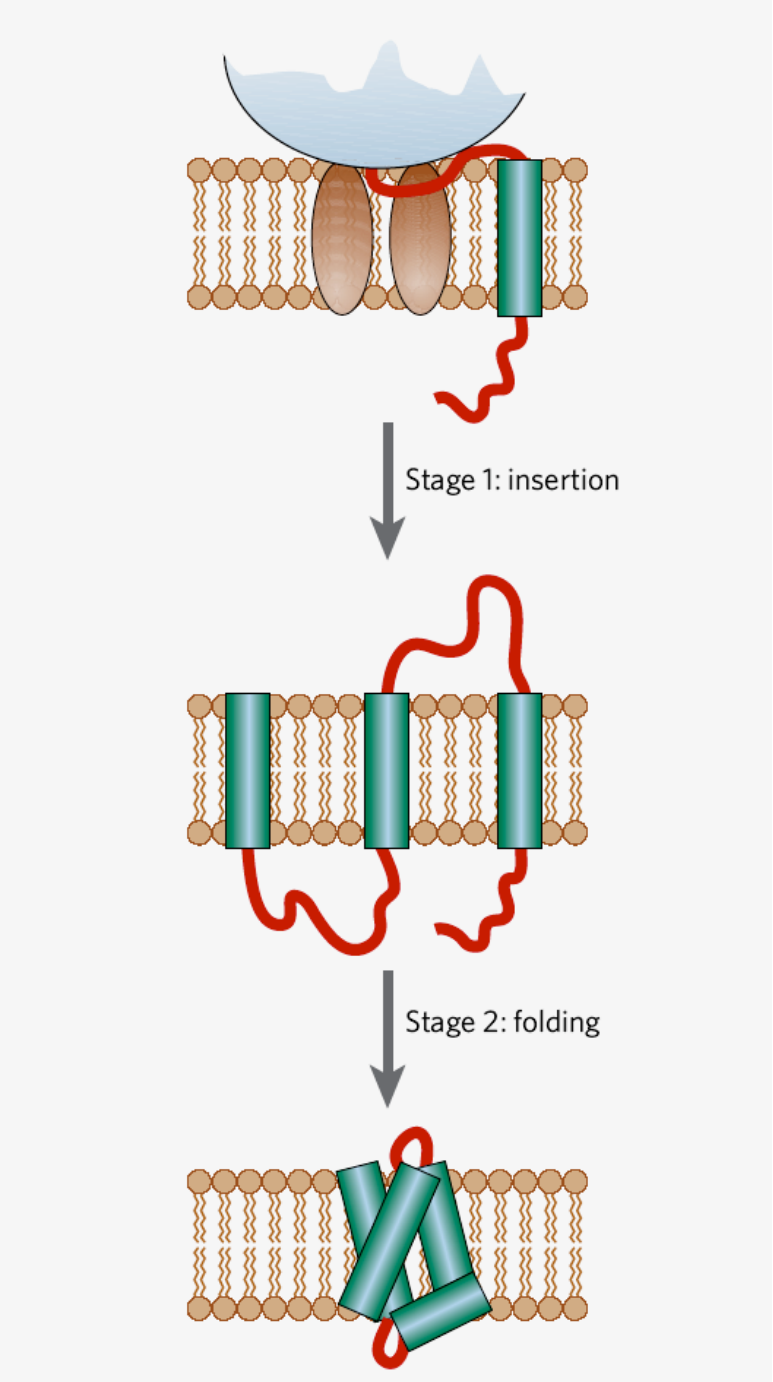
Fig. 4. The GroEL-GroES chaperonin system. (**A**) (Left) View of the asymmetric GroEL-GroES-$(ADP)_7$ complex generated with the coordinates 1AON (*61*) and program Weblab ViewLite 4.0 (Molecular Simulations). The equatorial, intermediate, and apical domains of one subunit each in the cis and trans ring of GroEL are colored in pink, yellow, and dark blue, respectively, and one subunit of GroES is colored red. (Right) The accessible surface of the central cavity of the GroEL-GroES complex. Polar and charged side-chain atoms, blue; hydrophobic side-chain atoms, yellow; backbone atoms, white; and solvent-excluded surfaces at subunit interfaces, gray. [Reprinted from (*61*) with permission] (**B**) Simplified reaction of protein folding in the GroEL-GroES cage. I, folding intermediate bound by the apical domains of GroEL; N, native protein folded inside the cage. For a typical GroEL substrate, multiple rounds of chaperonin action are required for folding; both I and N accumulate after a single reaction cycle and exit the cage upon GroES dissociation. I is then rapidly re-bound by GroEL. (**C**) Mechanisms of accelerated folding. Simple energy diagrams are shown for a protein that forms a kinetically trapped intermediate during spontaneous folding (left). In the iterative annealing model, this intermediate is thought to be actively unfolded by GroEL/GroES (*69*) and allowed to repartition (middle), whereas confinement of nonnative protein in the narrow, hydrophilic environment of the GroEL-GroES cage is suggested to result in a smoothing of the energy landscape (right), such that formation of certain trapped intermediates is avoided (*67*). Both proposed mechanisms would result in accelerated folding.

Hartl & Hayer-Hartl (2002) *Science* **295**, 1852.

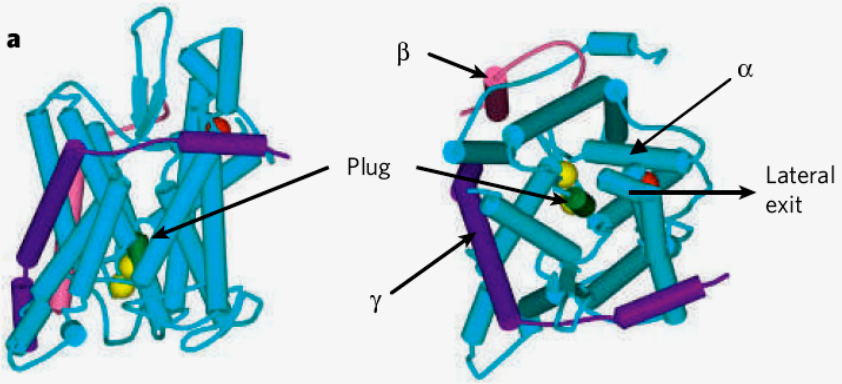
Folding of membrane proteins in the cell

Two stages of membrane protein folding

Bowie (2005) *Nature* **438**, 581.



Mechanistic model for membrane protein insertion



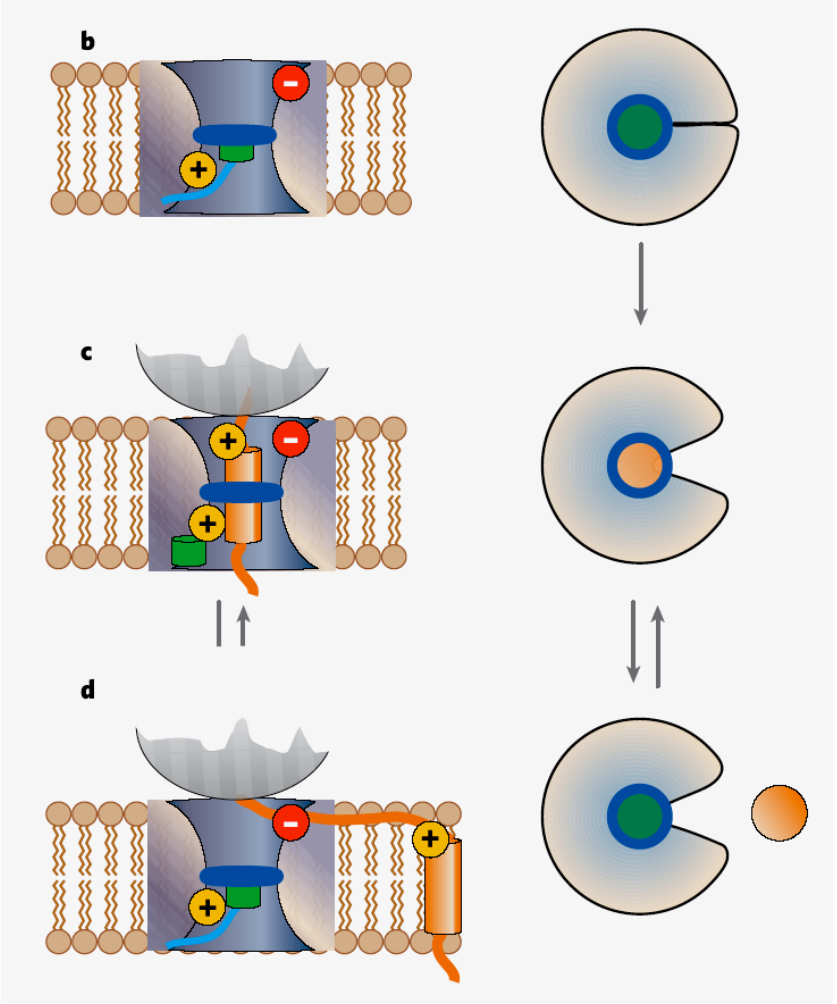


Figure 6 | The Sec61 translocon structure and mechanism model[23]. **a**, The α subunit is shown in light blue, the β subunit in purple and the γ subunit in pink. The plug helix that blocks the pore is shown in green. Key charged residues that help define the topology of translocating polypetides in *S. cerevisiae* are shown in yellow (positive charge) and red (negative charge). **b**, Schematic side and top views of the Sec61 translocon showing the locations of the charges, the plug helix in green and the hydrophobic collar shown by the dark blue ring. **c**, A schematic view of a the translocon with a nascent polypeptide (orange) emerging from the ribosome (grey). A positively charged residue helps define the topology of this segment as N-terminal first. **d**, A schematic view of the translocon opening the lateral gate so the helix can exit to the membrane if the partitioning is favourable.

Bowie (2005) *Nature* **438**, 581.

Protein degradation in the cell

Natural turn-over as well as removal of defective proteins

Protein Degradation *In Vivo*

Goldberg (2003) *Nature* **426**, 895.

Table 1 **Abnormal proteins rapidly degraded in cells**

Type of abnormality	Cause
Incomplete proteins	Nonsense mutations, incorporation of puromycin, premature termination, proteolytic cleavage
Missense proteins	Mutations, incorporation of amino-acid analogues, biosynthetic errors
Free subunits of multimeric complexes	Excess synthesized subunits
Postsynthetic damage	Oxygen radicals, intracellular denaturation
Genetic engineering	Gene fusions, frame-shifts, incorrect localization
Protein misfolding	

Box 1

Intracellular conditions that damage cell proteins

Temperature of 37 °C or higher (denaturing conditions).
Many reactive small molecules — these cause oxidation, deamidation, glycation or nitrosylation.
Many enzymes that modify proteins, for example proteases or kinases.
High salt concentrations (which favour dissociation of multimers)
Many fatty acids, which act like detergents.
Other unfolded proteins — nascent polypeptides, damaged or mutant polypeptides and insoluble inclusions are sticky.

Conclusion: to maintain a protein's function, avoid the intracellular milieu.

Nature's meat grinder:

Misfolded proteins are degraded (and ultimately recycled to their constituent amino acids) by the *proteasome*

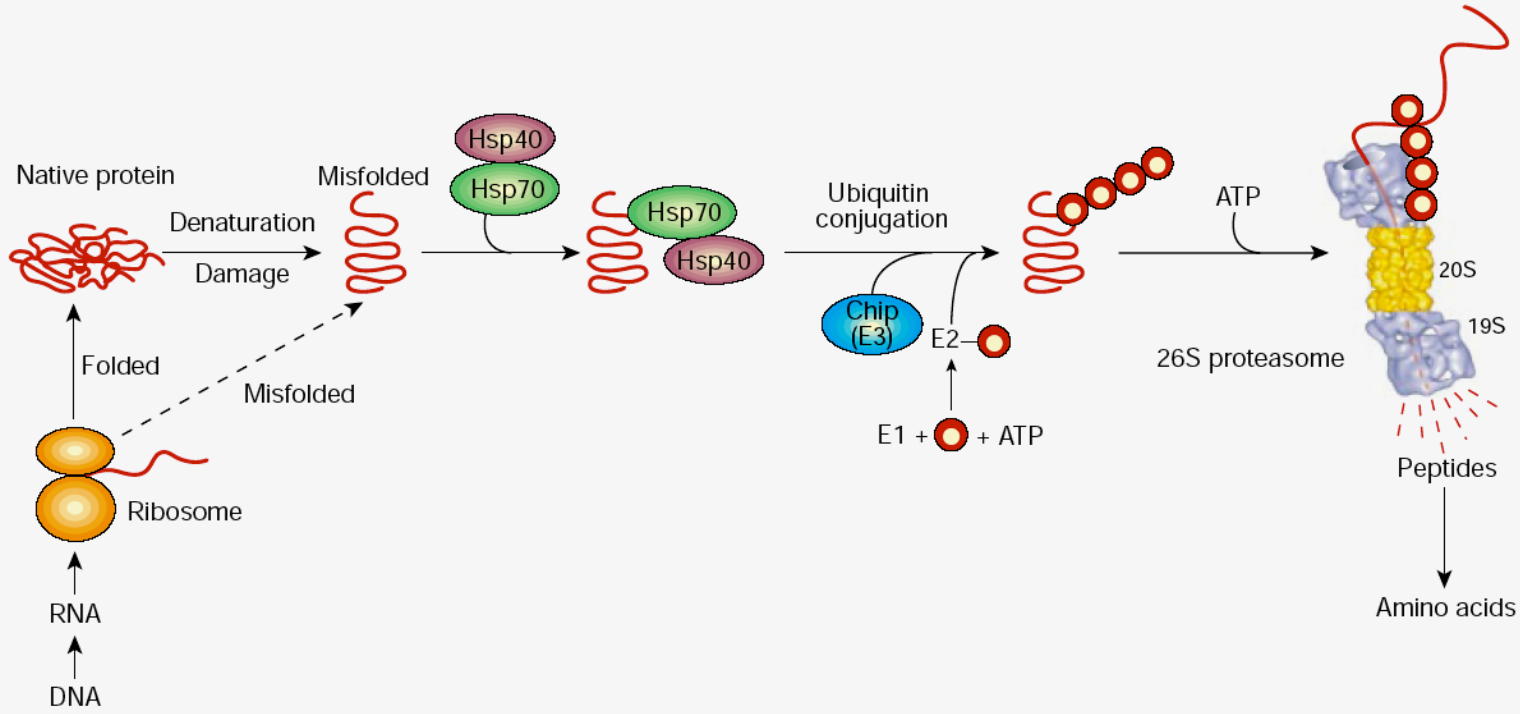


Figure 1 The ubiquitin–proteasome pathway. Molecular chaperones may function in protein folding and in the degradation of misfolded species. By associating with exposed hydrophobic domains, chaperones Hsp70/40 promote the folding of newly synthesized proteins and favours their refolding. Alternatively, they can facilitate the recognition of abnormal proteins, leading to their ubiquitylation by CHIP, the E3, and their degradation by the 26S proteasome. The red circles represent ubiquitin.

Goldberg (2003) *Nature* **426**, 895.

Protein Folding and Disease

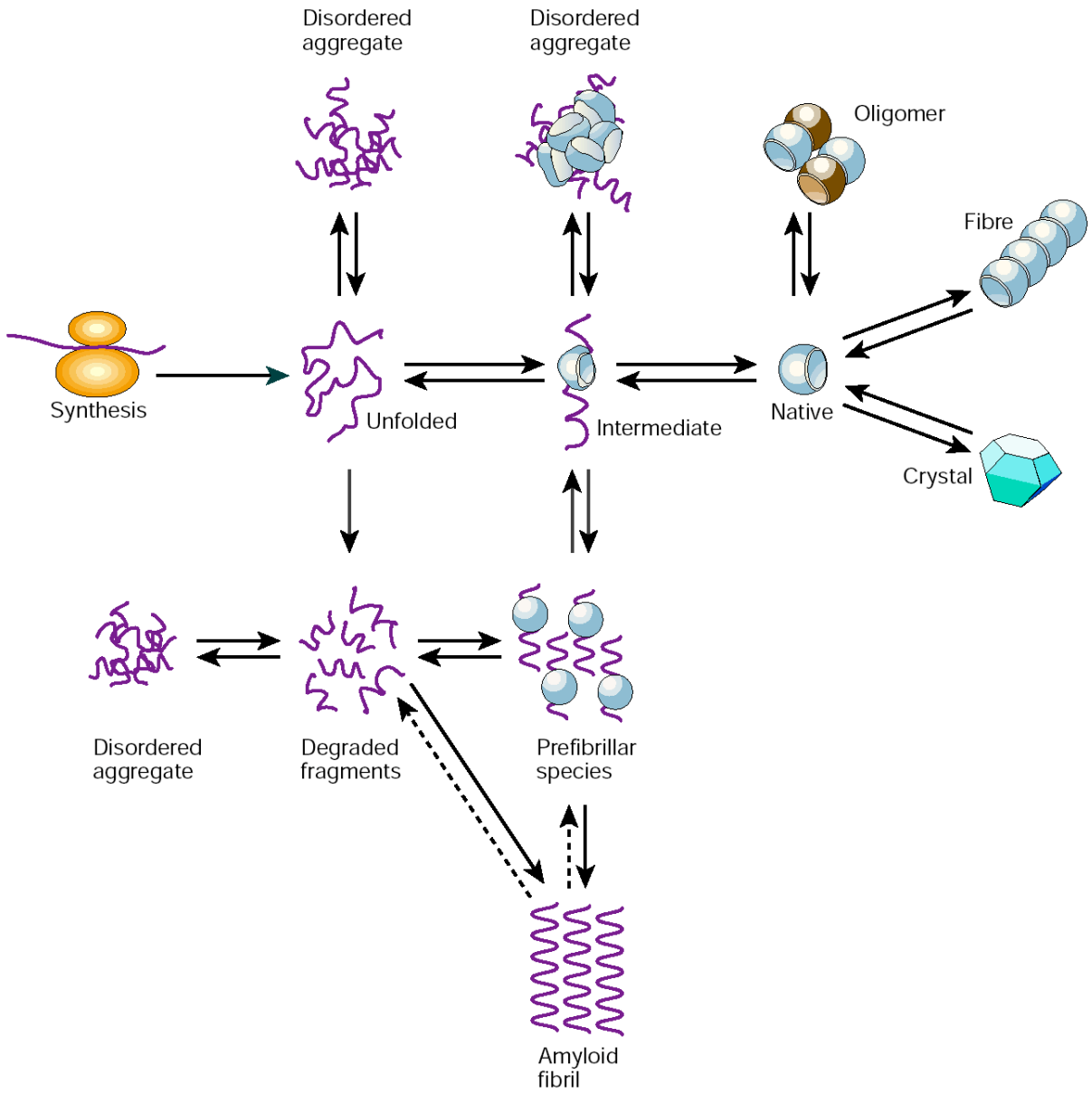


Figure 4 A unified view of some of the types of structure that can be formed by polypeptide chains. An unstructured chain, for example newly synthesized on a ribosome, can fold to a monomeric native structure, often through one or more partly folded intermediates. It can, however, experience other fates such as degradation or aggregation. An amyloid fibril is just one form of aggregate, but it is unique in having a highly organized 'misfolded' structure, as shown in Fig. 3. Other assemblies, including functional oligomers, macromolecular complexes and natural protein fibres, contain natively folded molecules, as do the protein crystals produced *in vitro* for X-ray diffraction studies of their structures. The populations and interconversions of the various states are determined by their relative thermodynamic and kinetic stabilities under any given conditions. In living systems, however, transitions between the different states are highly regulated by the environment and by the presence of molecular chaperones, proteolytic enzymes and other factors. Failure of such regulatory mechanisms is likely to be a major factor in the onset and development of misfolding diseases. Adapted from ref. 54.

Dobson (2003) *Nature* **426**, 884.

Some examples (of many!) of human diseases where amino acid substitutions (mutations) cause pathological protein misfolding

• **Sickle cell anemia** - The Glu6 --> Val mutation in the beta subunit changes the solubility properties of hemoglobin, causing aggregation and shape changes in the red blood cells (see http://sickle.bwh.harvard.edu/scd_background.html)

• **Emphysema** - Alpha-1 antitrypsin regulates the activity of elastase in the lungs; too much elastase activity can cause destruction of lung tissue. Mutations in the gene for alpha-1 antitrypsin cause misfolding during its synthesis in the liver, which then leads to defective export from liver cells and a deficiency in the lungs (see http://health.enotes.com/genetic-disorders-encyclopedia/alpha-1-antitrypsin).

• **Alzheimer's disease** - Mutations in a gene product called "APP", or in APP processing enzymes, can cause build-up of a peptide breakdown product. This peptide (known as beta peptide) can aggregate into so-called *amyloid fibrils*, which can form deposits (known as *plaques*) that build up over time in the brain, killing neurons (see Selkoe [2001] *Neuron* **32**, 177).

Model structures for amyloid peptide fibrils

(note: these are not Alzheimer's amyloid peptide fibrils, but the structures are thought to be similar)

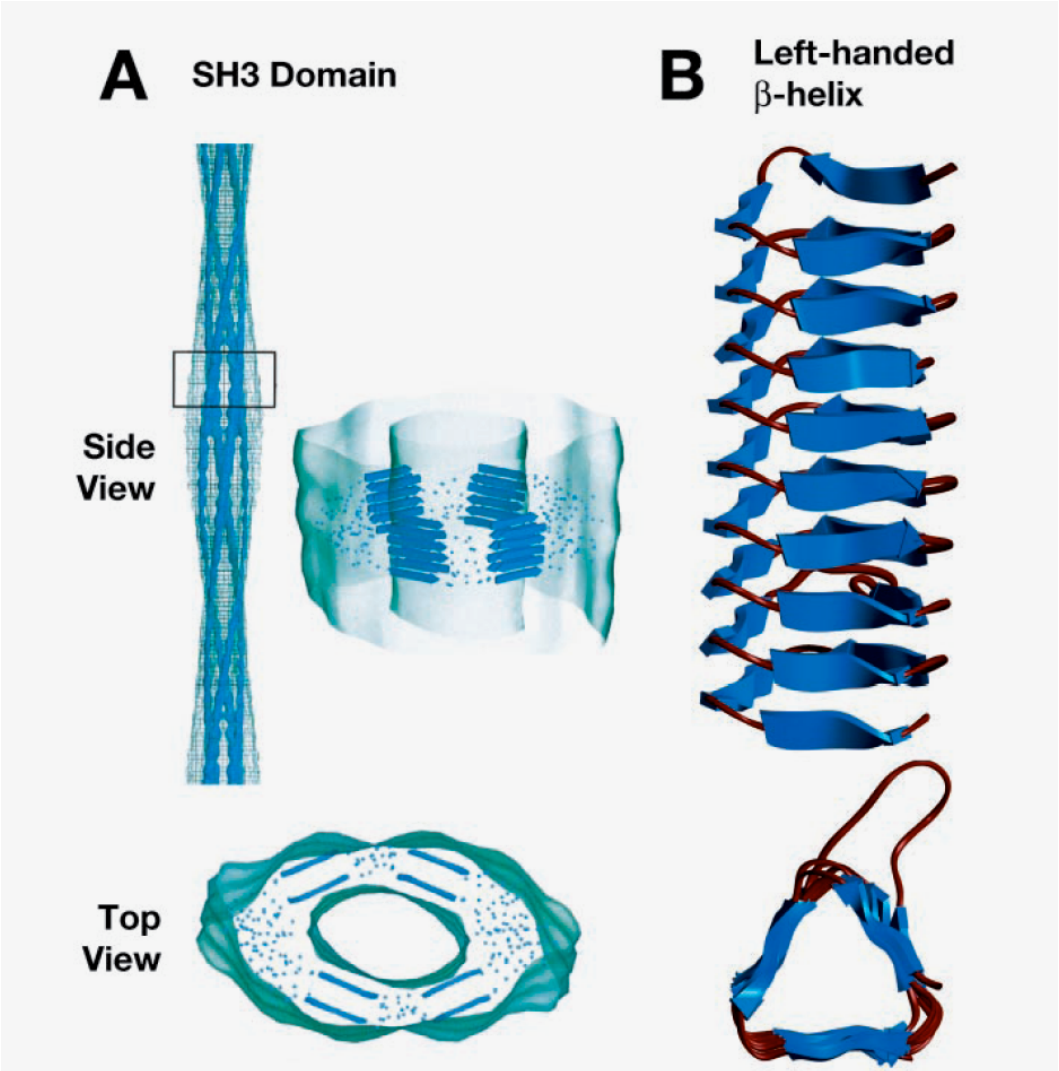


Figure 4 Two models for amyloid structure. Both fulfill the requirements of the cross-β fold in which individual β-strands are oriented perpendicular to the fiber axis, whereas β-sheets are oriented parallel to it. (*A*) Model from cryo-EM studies of amyloid formed by the SH3 domain from PIP_3 kinase (83). (*B*) An example of a left-handed β-helix (from UDP-N-acetylglucosamine pyrophosphorylase of *Streptococcus pneumoniae,* PDB ID 1G97), which has been proposed to resemble PrP^{SC} (86). Image is courtesy of Dr. Cedric Govaerts (unpublished material).

Chien et al (2004) *Ann. Rev. Biochem.* **73**, 617.

Electron micrographs of various amyloid fibril preparations

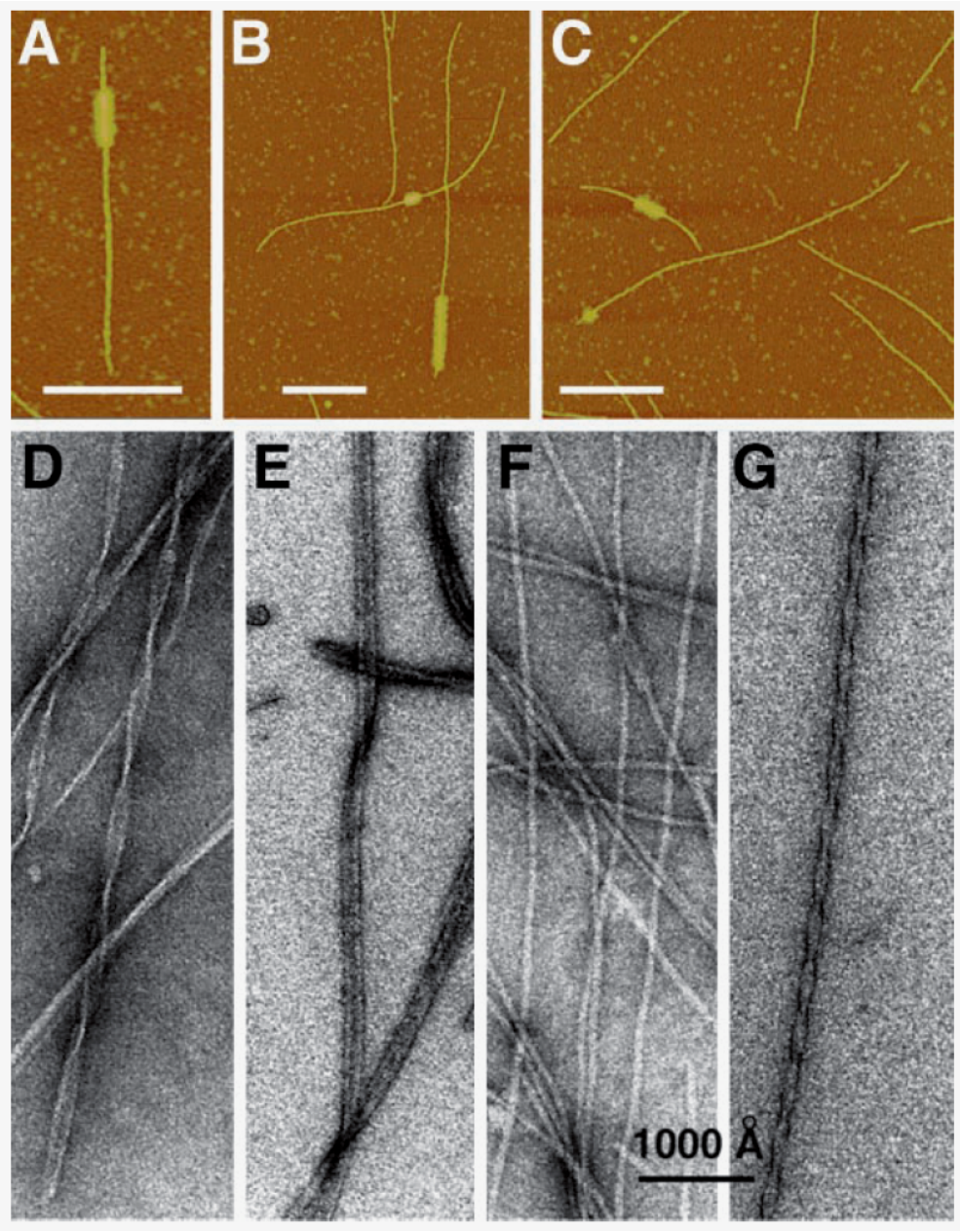
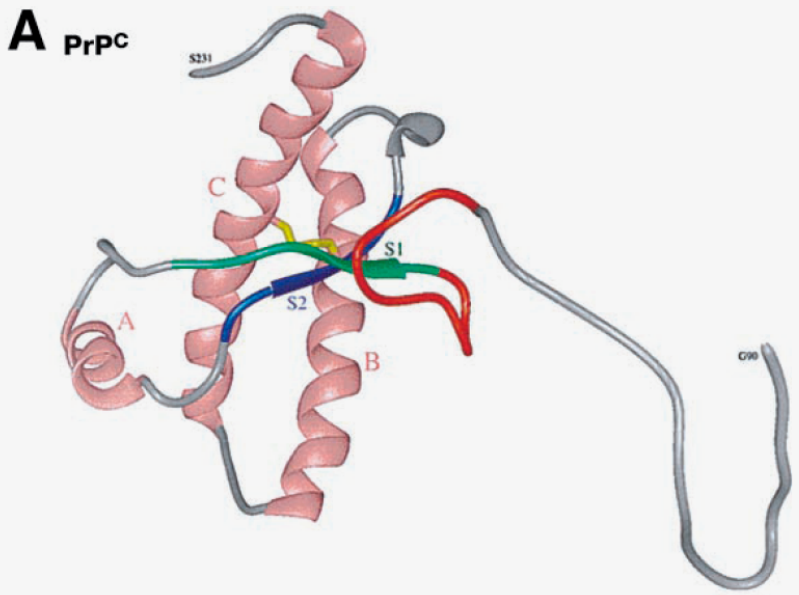


Figure 6 Amyloid fibers adopt multiple distinguishable structures. (*A, B, C*) Amyloid fibers formed spontaneously by Sup35NM vary in their growth patterns, including overall rate and polarity of growth (138). Four kinetic fiber types visualized by an AFM single fiber growth assay are shown. The original seed is labeled with antibody and is therefore wider than the new growth extending from its ends. Note the presence of long and short symmetric and asymmetric fibers. Scale bar is 500 nm. (*D, E, F, G*) Negative stain EM of amyloid fibers formed spontaneously by the SH3 domain from PIP_3 kinase illustrates that they vary in the number of protofilaments and helical pitch (83). Scale bar is 100 nm.

Chien et al (2004) *Ann. Rev. Biochem.* **73**, 617.



B PrP27-30

C **Conformational Differences Between PrPC and PrPSC**

PrPC	PrPSC
monomeric	multimeric
soluble	insoluble
protease sensitive	protease resistant
predominantly α-helical	predominantly β-sheet

Figure 1 PrP^{C} and PrP^{SC} are conformationally distinct. (*A*) Solution NMR structure of Syrian hamster PrP^{C}, residues 90–231. The structure is predominantly alpha helical with an unstructured amino terminus (199). (*B*) Negative stain EM of Syrian hamster $PrP^{27–30}$ (Sc237 strain), stained with uranyl acetate. The material is in insoluble protease-resistant high molecular weight aggregates that are predominantly β-sheet. Scale bar is 100 nm. Image courtesy of Dr. Holger Wille (unpublished material). (*C*) Summary of differences between PrP^{C} and PrP^{SC}.

Prions also form amyloid-like deposits and fibrils

(see next slide)

Chien et al (2004) *Ann. Rev. Biochem.* **73**, 617.

Electron micrographs of amyloid-like fibers formed by prions

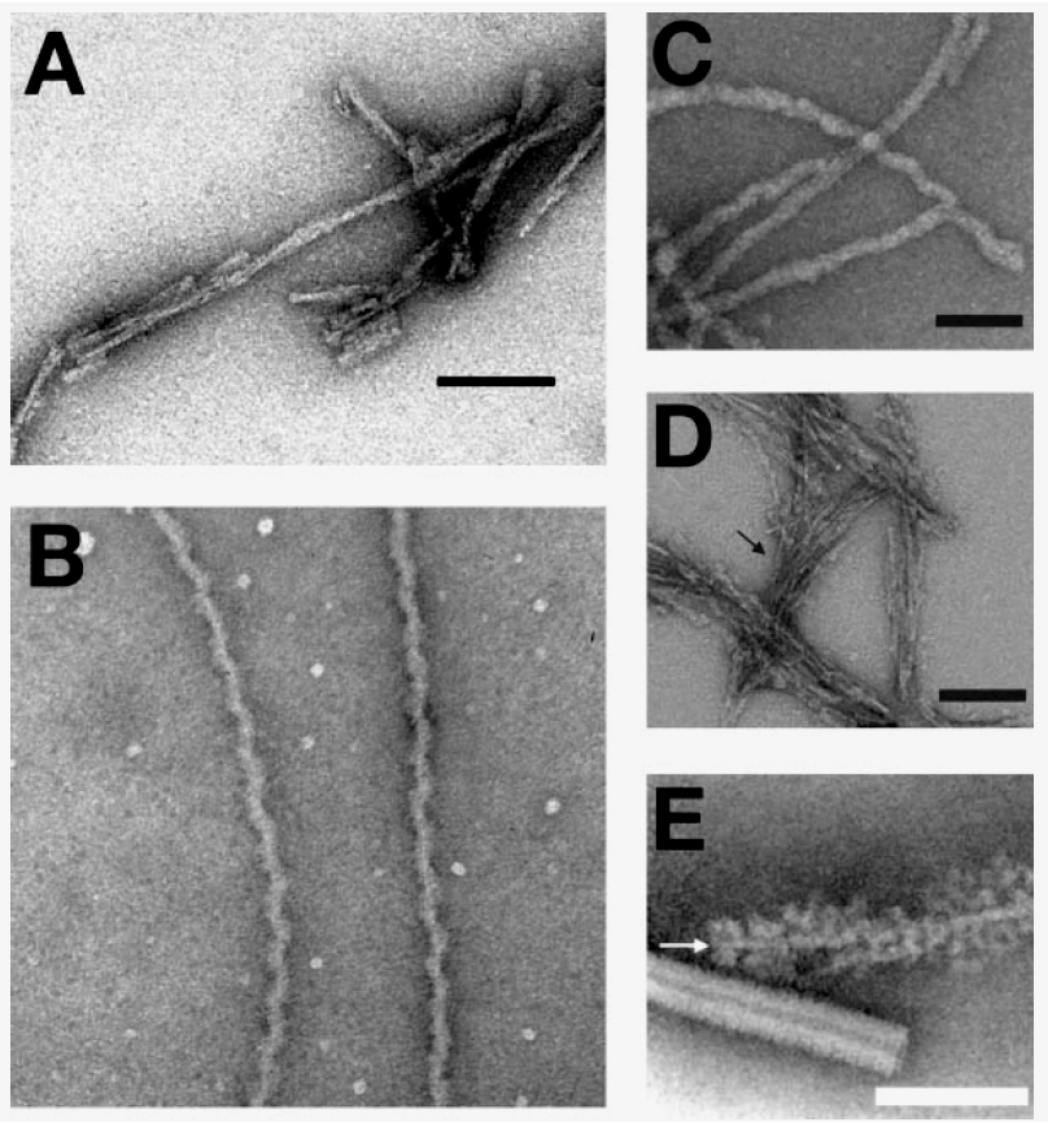


Figure 3 Amyloid-like fibers are formed by a variety of prion proteins. (*A*) EM of Syrian hamster PrP^{27-30} (Sc237 strain), stained with uranyl acetate. Bar = 100 nm. Image courtesy of Dr. Holger Wille (unpublished material). (*B*) Amyloid fibers formed by Sup35NM, stained with uranyl acetate. Sup35NM fibers are on average 5–10 nm in diameter. (*C*, *D*) EM of full-length Ure2p fibers stained with uranyl acetate before (*C*) and after (*D*) digestion with proteinase K. Arrow in *D* indicates position of a single fiber. Bar = 100 nm (89). (*E*) Amyloid fibers formed by full-length Ure2p, stained with vanadate and visualized by dark-field scanning transmission electron microscopy (STEM). Arrow indicates the core of the fiber. Bar = 50 nm (89).

Chien et al (2004) *Ann. Rev. Biochem.* **73**, 617.

What about therapeutic interventions in "protein folding-opathies"?

For example, can pharmacological rescue of misfolded and/or misrouted proteins be achieved *in vivo*?

Guess what, yes! *It works!!*

GnRHR has many known mutations that disrupt its folding and sorting within the cell.

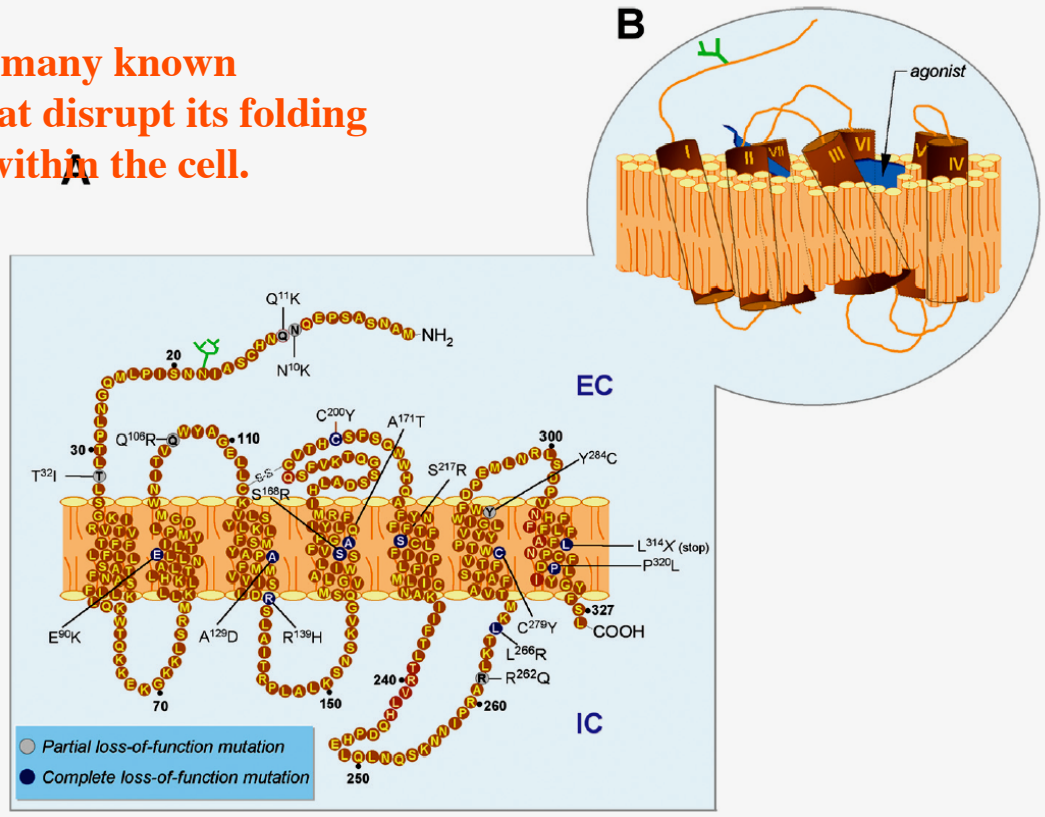


Figure 1: A) Sequence of the human gonadotropin-releasing hormone (GnRH) receptor (GnRHR) and location of the inactivating (loss-of-function) mutations identified to date (31). A member of the G protein-coupled receptor superfamily, the GnRHR consists of a single polypeptide chain that traverses the cell surface membrane seven times, forming characteristic transmembrane helices, interconnected by alternating extracellular (EC) and intracellular (IC) loops. The GnRHR is a ligand-activated switch that activates the heterotrimeric $G_{q/11}$ guanosine nucleotide-binding protein. This receptor exhibits several unique features compared to other members of the superfamily, including the reciprocal exchange of the conserved D and N residues in the transmembrane domains II and VII (from left to right), the replacement of Y with S in the highly conserved DRY motif located in the junction of the transmembrane domain III and the second IC loop (residues 138–140), and the lack of the cytosolic carboxyl-terminal extension into the cytosol, which plays an important role in receptor cell surface expression, internalization and desensitization (32–34). Following GnRH binding, the GnRHR activates phospholipase Cβ as a result of coupling to the membrane associated trimeric $G_{q/11}$ protein associated with the intracellular domains of the receptor, leading to phosphatidylinositol 4,5-biphosphate hydrolysis and formation of inositol 1,4,5-triphosphate (IP) (34,35). B) Counterclockwise orientation of a prototypic G protein-coupled receptor from transmembrane domains I to VII (with the cytoplasmic side at bottom). The closed loop structure is representative of receptors for peptide ligands such as GnRH. In this arrangement, the core is composed mainly of transmembrane domains II, III, VI and VII, whereas domains I, IV and V are peripherally sequestered. GnRH interacts with the second EC loop (at N^{102}), transmembrane domain III (at K^{121}), and the third EC loop-transmembrane domain VII junction (D^{302}) of the GnRHR (34).

Ulloa-Aguirre et al (2004) *Traffic* **5**, 821.

Note: the exogenously added drug

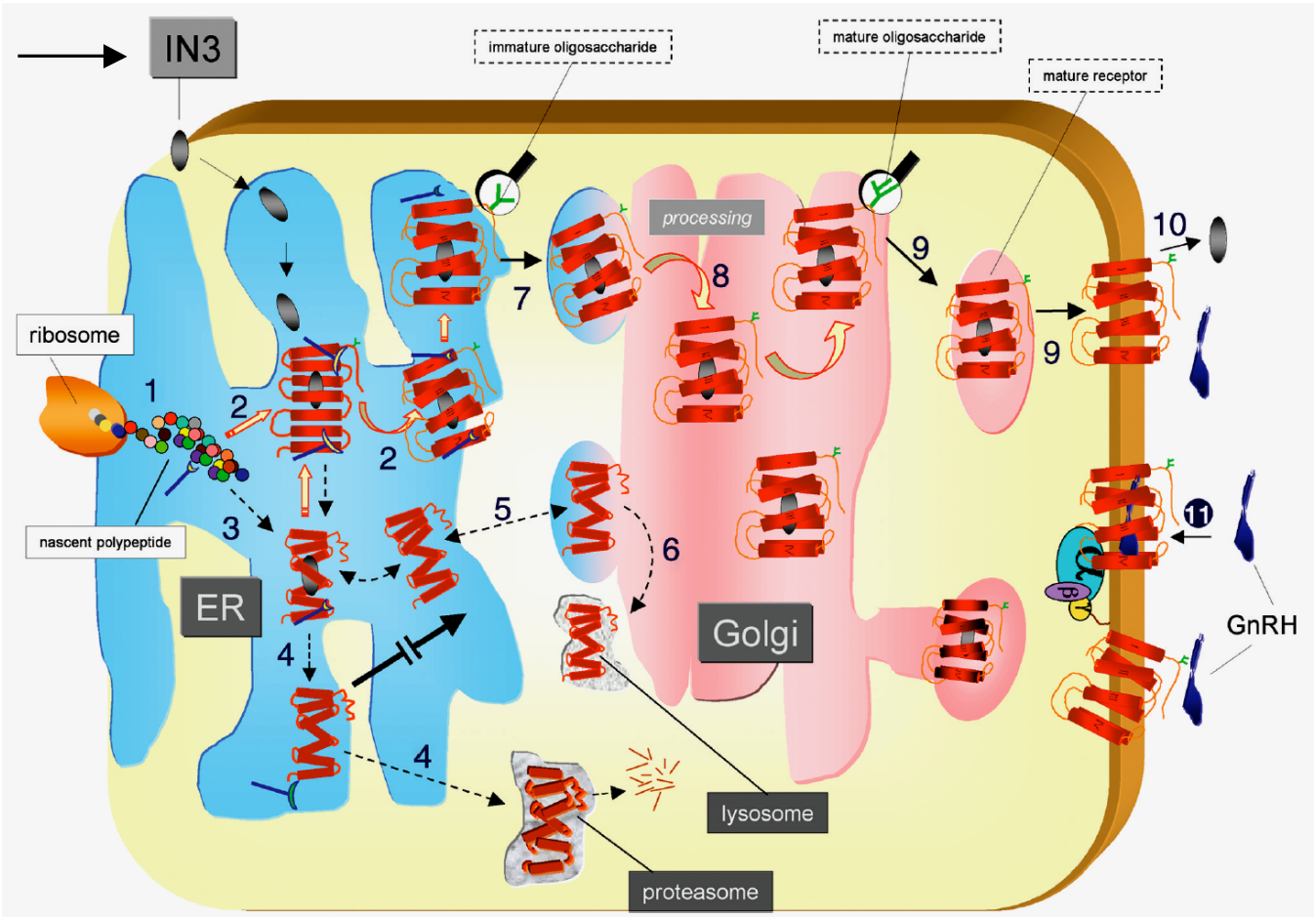


Figure 2: Quality control in the endoplasmic reticulum (ER). Newly synthesized polypeptides are translocated to the lumen of the endoplasmic reticulum (step 1). Folding is facilitated by interaction of the nascent polypeptide with molecular chaperones (rod-like structures) (step 2). Misfolded and misassembled products are retained in the ER and exposed to resident chaperones to attempt folding (step 3). Eventually, misfolded proteins may be dislocated into the cytoplasm for proteosomal degradation after dissociation of the molecular chaperones (step 4). Alternatively, defective proteins may be exported to and retained by the Golgi, and retrotranslocated to the ER, where correct folding is again attempted (step 5), or diverted to lysosomes for degradation (step 6). Proteins properly folded are translocated to the Golgi (step 7) where processing (e.g. glycosylation) and maturation of the protein molecule is completed (step 8). Mature products are then exported to their final destination (e.g. the cell surface membrane) (step 9). In the presence of pharmacological chaperones (e.g. IN3), folding and assembly of a defective (or mutated) protein is facilitated early during its biosynthesis, so that it can escape degradation, exit the ER and be properly localized (27–30). Molecular chaperones dissociate in the ER, whereas pharmacological chaperones may remain associated (37). Once the newly synthesized protein (e.g. the GnRHR) reaches its destination, the pharmacoperone can dissociate from the rescued molecule (step 10) to allow interaction of the ligand with its binding site at the receptor protein (step 11).

Ulloa-Aguirre et al (2004) *Traffic* **5**, 821.

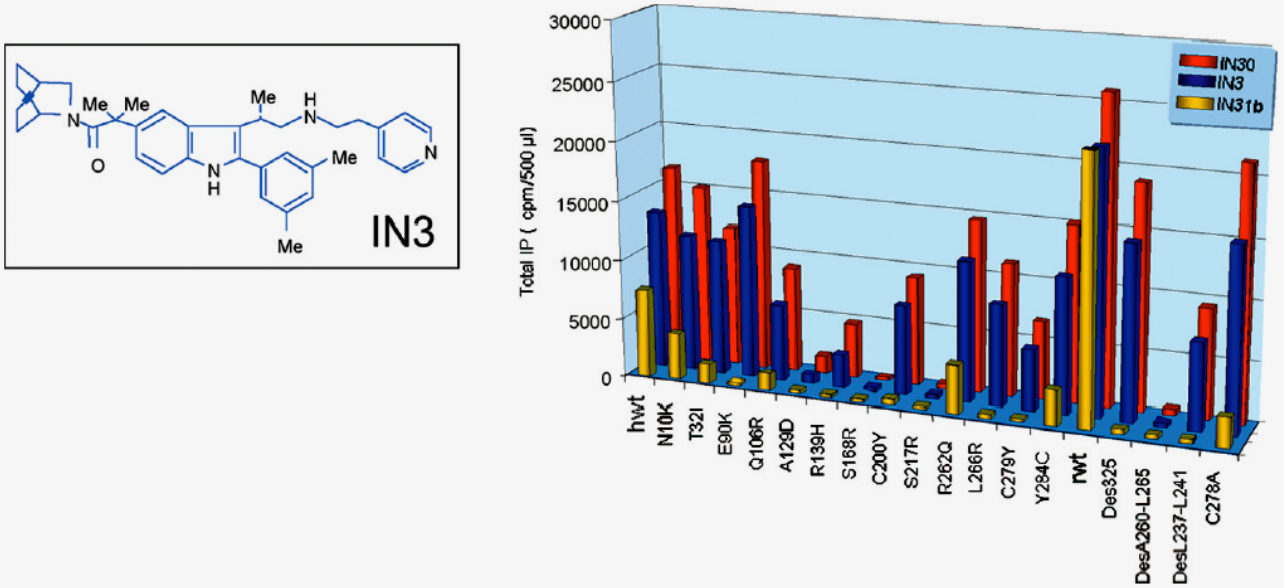


Figure 5: Pharmacological rescue of WT and mutant human GnRH receptors by three cell membrane-permeant peptidomimetics. The figure on the right shows inositol phosphate production (IP) by COS-7 cells transiently expressing each receptor in the presence of 10^{-7} M Buserelin. Hydrophobic peptidomimetics are able to penetrate cells and interact specifically with protein targets; one such antagonist of GnRH is the indole, IN3 (132–134). In the experiment shown, cells were exposed to a 1.78 µM concentration of each pharmacoperone at the time of transfection. The pharmacological chaperones tested were the indoles IN30, IN31b, and IN3 (left figure). For 12 of 13 naturally occurring human GnRHR mutants tested in these studies, exposure to IN3 has resulted in both $G_{q/11}$ coupling and specific ligand binding (27–29,31), demonstrating that the effect of the mutational error leading to intracellular retention is either completely or partially corrected by this approach. Two of the mutants, $S^{168}R$ and $S^{217}R$ (Figure 1) that could not be rescued by this approach are only about 20 Å apart in the WT molecule. Conceivably, some mutants exhibiting either low responsiveness or refractoriness to pharmacoperone treatment may bear structural defects involving microdomains critical for ligand binding, receptor activation and/or effector coupling (34), but not those essential for proper protein folding or intracellular routing. All peptidomimetics studied with an IC_{50} value for the human GnRHR of 2.3 nM [including quinolones and erythromycin-derived macrolides (not shown in the figure)] displayed a measurable efficacy in rescuing GnRHR mutants, and within a single chemical class, this ability correlated to these IC_{50} values (29). The results from four laboratory manufactured loss-of-function rat GnRHRs [$des^{325–327}$, $des^{237–241}$ and $des^{260–265}$ rat GnRHRs (three nonfunctional deletion mutants), and $C^{278}A$ rat GnRHR (a nonfunctional Cys mutant)] are also shown. Reproduced from [29], with permission from The American Society for Pharmacology and Experimental Therapeutics.

Ulloa-Aguirre et al (2004) *Traffic* **5**, 821.