How to fold proteins

Is there hope to predict the native fold of proteins ?

Remember the results from Anfinsen (1959, 1961, 1963):

proteins are able to fold by themselves (denature them in vitro, and let them fold without any assistence : success !)

⇒ 1) The recipe to fold is written in the sequence

2) The native state is stable if it is the minimum of the free-energy

→ COMPATIBLE WITH WHAT WE LEARN FROM SIMPLE MODELS

Idea (dates back 50 years, early '70s):

integrate the equations of motion for the system, and let it relax to the most probable state (= minimum of free-energy ⇒ native state)

$$m \ddot{\vec{x}}_i = \sum_j \vec{F}_{ij}$$

$\vec{x}_i$: position of each atom (here i)

$\vec{F}_{ij}$: forces on atom i due to particles j : includes all other atoms of the protein and of the solvent

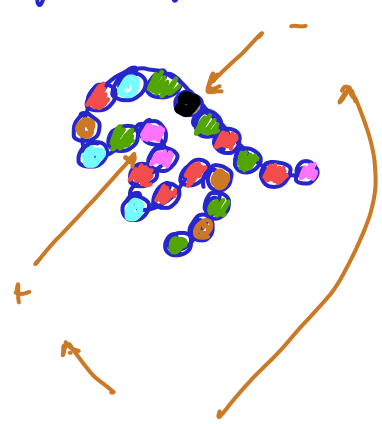
Not here the case of describing the technical methods.

Result of the approach : ZIP

Essentially no major success, apart for very small proteins that fold rapidly (energy landscape rather smooth).

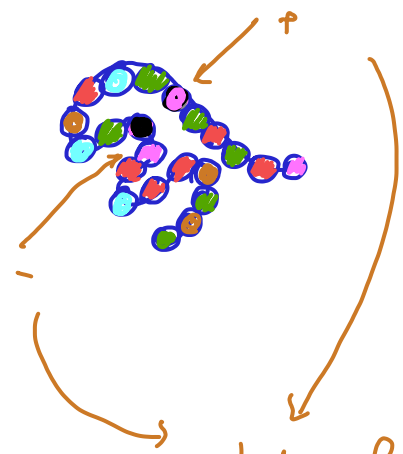
And yet, the solution is coming from physics by an unexpected route : co-evolution

A simple primer on co-evolution



(+ −) attract each other

⇒ they stabilise the structure

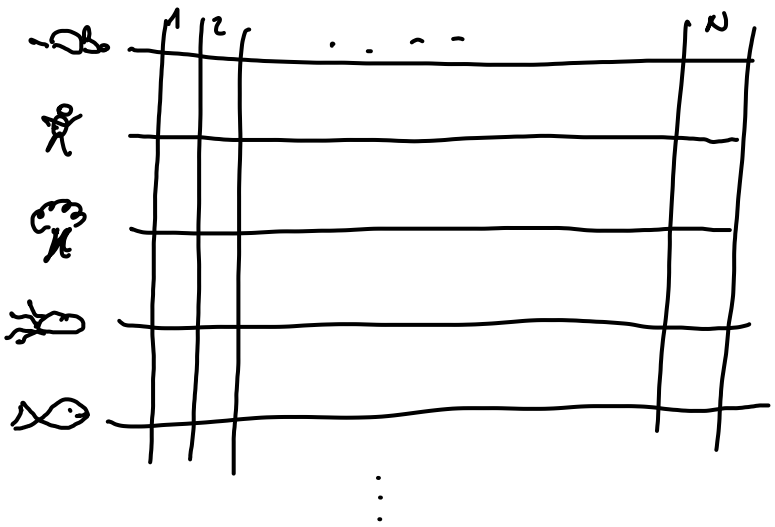


mutations have changed the amino-acids, but only concerted mutations that have maintained the stability can be selected and observed

the plan is thus

- Collect very many sequences for the same protein from different organisms. Typically need for $10^3 - 10^4$ sequences (mostly available).

- Align them



- for each position count the fraction of times that each amino-acid appears: $f_i^\alpha \quad \alpha = 1, \dots, 20 \quad \sum_\alpha f_i^\alpha = 1$
- for each pair of positions, count the number of times that each pair of amino-acids appear:

$$f_{ij}^{\alpha\beta} \quad \alpha = 1, \dots, 20 \quad \beta = 1, \dots, 20 \quad \sum_{\alpha\beta} f_{ij}^{\alpha\beta} = 1$$

If $f_{ij}^{\alpha\beta} \neq f_i^{\alpha} f_j^{\beta}$ (significantly)

this means that the amino-acids in positions i and j are not selected independently, but are correlated!

This correlation, as proposed in the picture, is born out of the need of the two positions to coordinate their physical properties because they are close in space ⇒ can be identified and used as constraints to improve the computational methods to fold proteins.

This is what Alpha-fold (from Google's Deepmind) does.

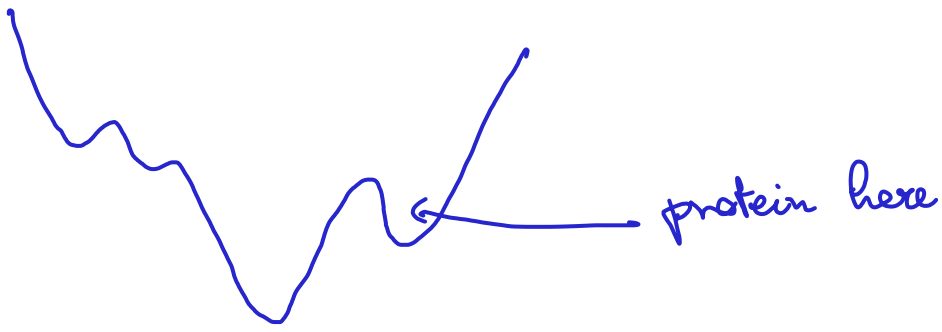
(http://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery)

This idea has been proposed by biologists about 30 years ago, and turned in an effective algorithm about 10 years ago by physicists. Now is mixed with machine-learning (see Alphafold)

There are a few details yet that will be explored in an exercise.

The idea of adding information on top of physical principles is called INTEGRATIVE MODELING.
In principle, if enough experimental or co-evolutionary information is known, then the physical principles might even be side-stepped. This is what AlphaFold does.

What happens when the folding is not successful?



amino-acids

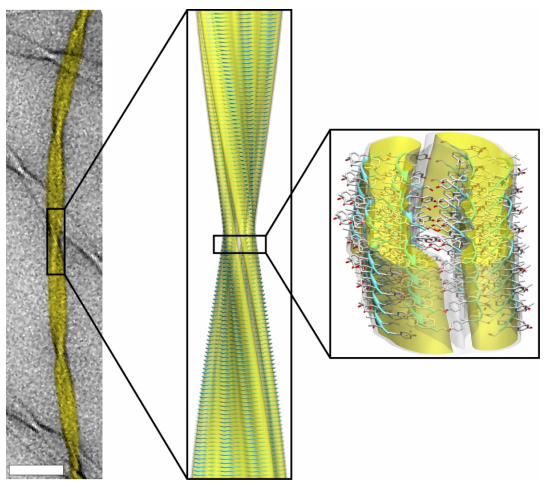
Typically, the misfolded protein exposes some hydrophobic residues on the surface. They are unhappy, and look around for other hydrophobic regions to stick to: other misfolded proteins!

Misfolded proteins can aggregate, precisely to satisfy their desire to hide hydrophobic regions.

Protein aggregation is extremely dangerous for the cell, and for organisms: Alzheimer, Parkinson, ALS, Huntington, ... are all "misfolding diseases".

fibrils

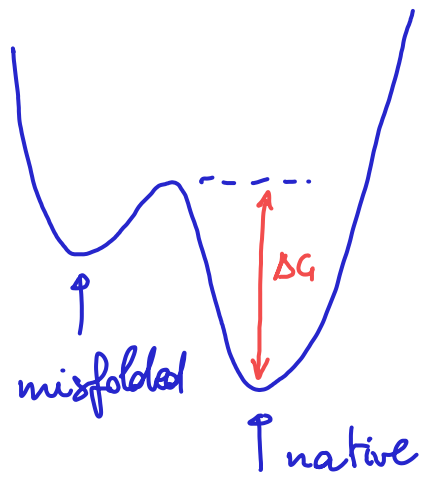
Some aggregates take the form of "fibers", long, mostly 1D assemblies also known as "amyloids"



they are long, semi-flexible polymers!

The big question is: what happens really in the cell?

Let's look at the energy landscape



if $\Delta G \gg k_B T$ ⇒ native state very stable

but $\Delta G \simeq 10\text{-}15\ k_B T$

But different factors can affect the landscape enough to make $\Delta G > k_B T$ (but not $\gg k_B T$)

1) Mutations
2) Stresses (e.g. higher temperatures)

⇒ the misfolded population increases ⇒ aggregation

How do cells cope with it?

A set of proteins, collectively called "chaperones" can, by means of a disparate set of molecular mechanisms, favor the native state of proteins by transiently binding to misfolded conformations.

Problem: how can they increase the concentration of native proteins while not being constantly bound to them?

They are ATPases: they consume ATP (an external source of energy) and in doing so they move away from thermodynamic equilibrium.

Think about "pumps" taking proteins in their misfolded state and moving them in their native state.

What are chaperones, and how do they work?

Chaperones are ancient! We know that because they are present in each and every organism on Earth. There are 5 families, meaning 5 types of chaperones structurally and genetically unrelated.

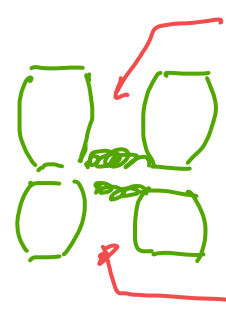
Hsp 60 : Heat - Shock Protein 60 kDa → molecular mass

↳ 1Da = 1 proton

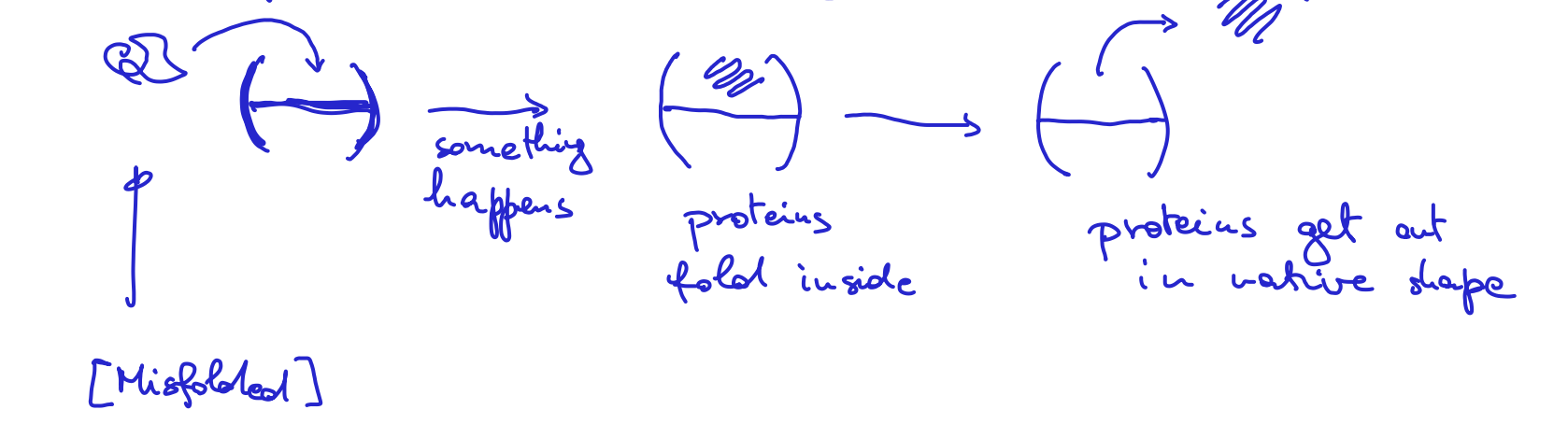
They are overexpressed upon heat shock (and other stresses) when misfolding and aggregation are more likely.

Two rings with an internal cavity each made of 7 units (1 unit is 60 kDa)

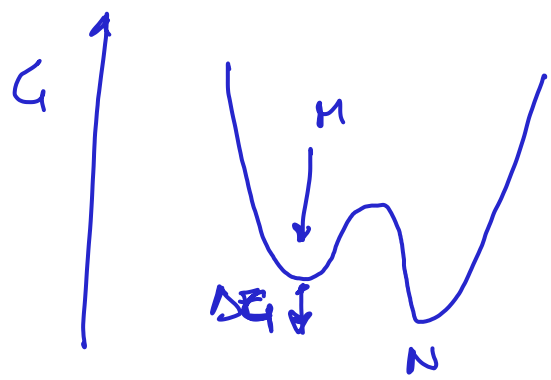
Section



Misfolded proteins enter the cavity



Let's draw the energy landscape again:



$$\Rightarrow \quad \frac{[M]}{[N]} = e^{-\beta \Delta G}$$

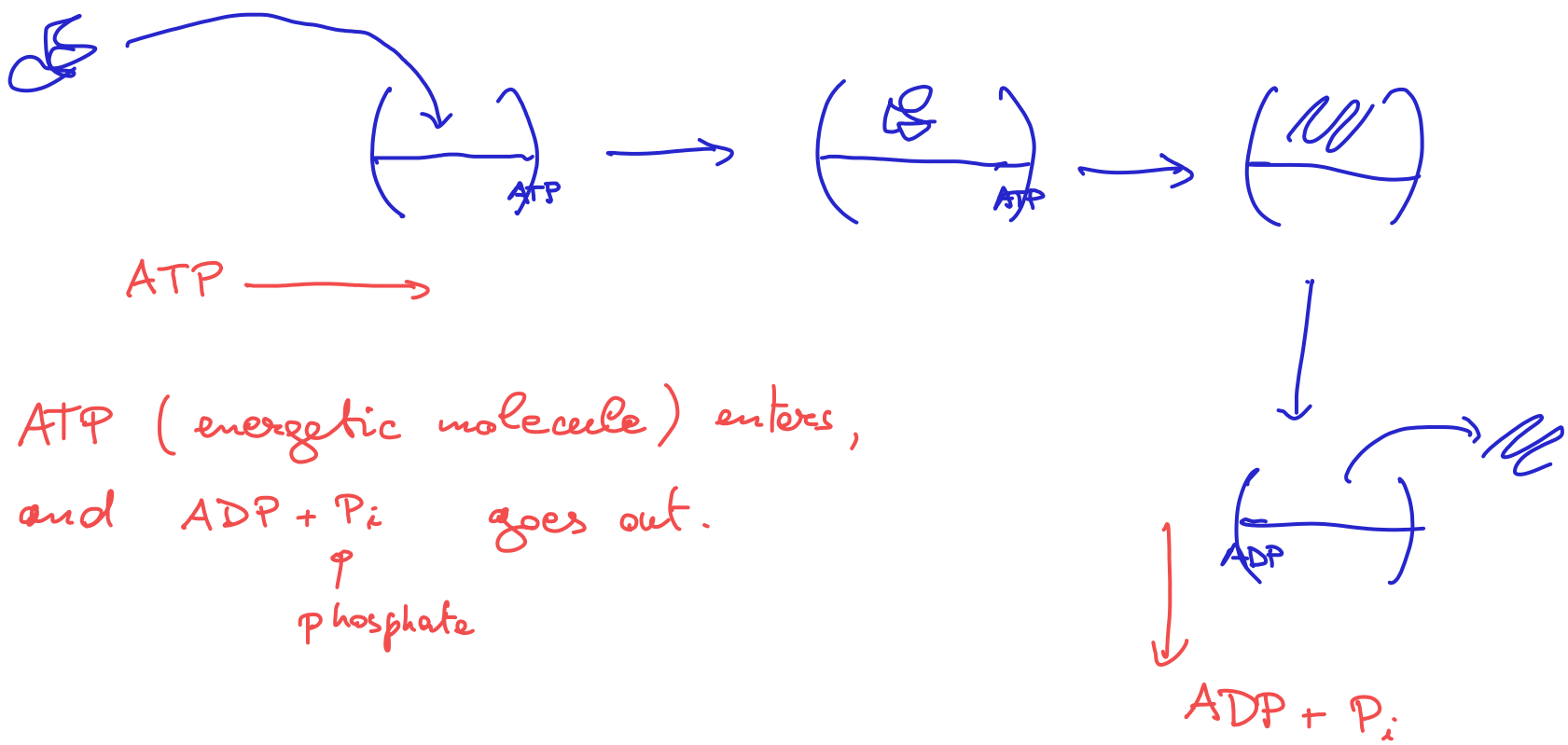
The action of the chaperone would be useless at thermodynamic equilibrium:

$$\frac{[M]}{[N]} = e^{-\beta \Delta G}$$ always!

If misfolding is dictated by thermodynamics, it's useless.

[M] → free molecules, not bound to the chaperone

But in reality

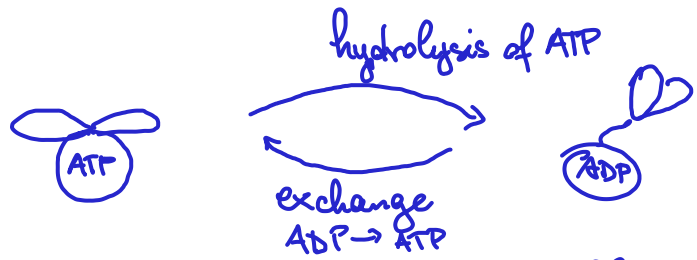


ATP (energetic molecule) enters, and ADP + P_i goes out.

P_i = phosphate

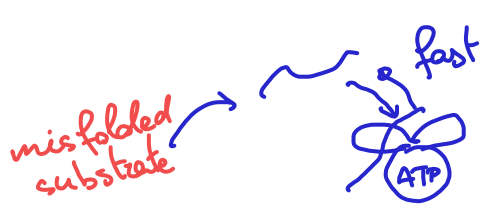
The system is consuming energy: it's an open dissipative system and, when looked at overall, it does not need to obey Boltzmann (although there are other rules that must be obeyed).

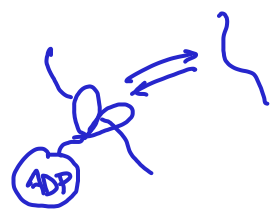
Hsp70



two completely different conformations depending on the bound nucleotide.

Why are the two conformations useful?





slow binding / unbinding

The switch between the 2 conformations occurs upon ATP hydrolysis, which is stimulated by the substrate, and by exchange (ADP exits and ATP enters).

Recall that the affinity between 2 molecules is defined by the dissociation constant, $K_d = \frac{k_{off}}{k_{on}}$ $\left(\frac{\text{unbinding rate}}{\text{binding rate}}\right)$

$A + B \rightleftharpoons AB$ (for simplicity, $[A]_{tot} \gg [B]_{tot}$)

rate equation: (free molecules)

$$\frac{d}{dt}[AB] = k_{on}[A][B] - k_{off}[AB]$$

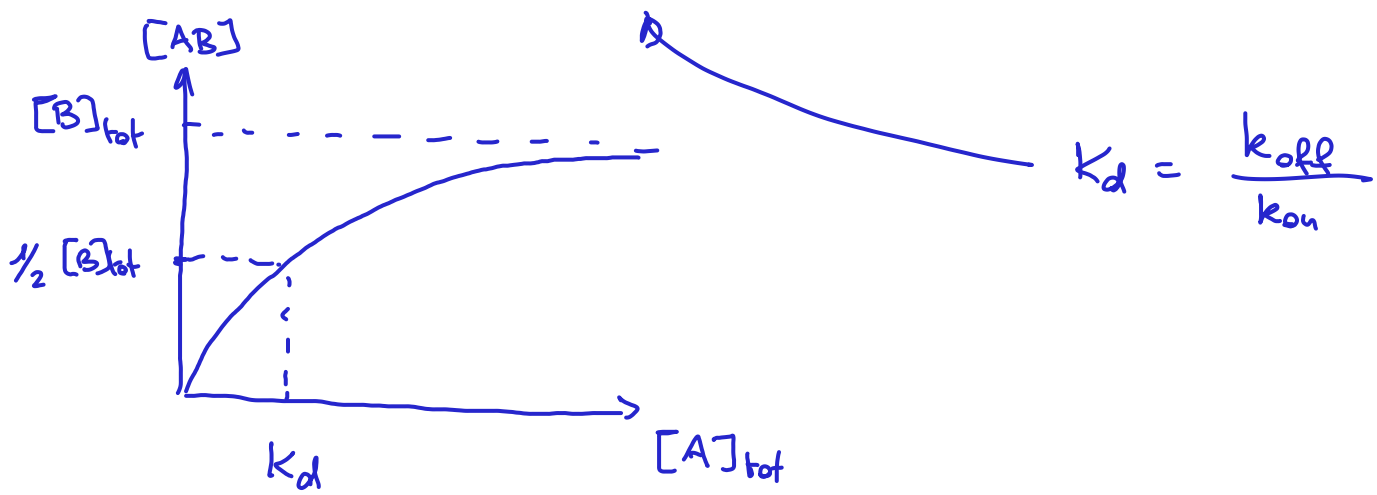
Then $[B] = [B]_{tot} - [AB]$ by conservation

$[A] = [A]_{tot} - [AB] \simeq [A]_{tot}$

At steady state ($\dot{[AB]} = 0$) :

$$k_{on}\,[A]_{tot}\left([B]_{tot} - [AB]\right) - k_{off}\,[AB] = 0$$

$$\Rightarrow \quad [AB] = \frac{[A]_{tot}}{K_d + [A]_{tot}}\,[B]_{tot}$$



K_d is the concentration $[A]_{tot}$ necessary for binding half of B

So, the smaller K_d, the higher the affinity.

Hsp70 uses a trick:

$$K_{d,ATP} = \frac{k_{off,ATP}}{k_{on,ATP}} \qquad K_{d,ADP} = \frac{k_{off,ADP}}{k_{on,ADP}}$$

both fast — can be roughly equal — both slow

But: if there is hydrolysis of ATP upon substrate binding,

then

$$K_d^{obs} = \frac{k_{off,ADP}}{k_{on,ATP}}$$

that is: smallest numerator
largest denominator

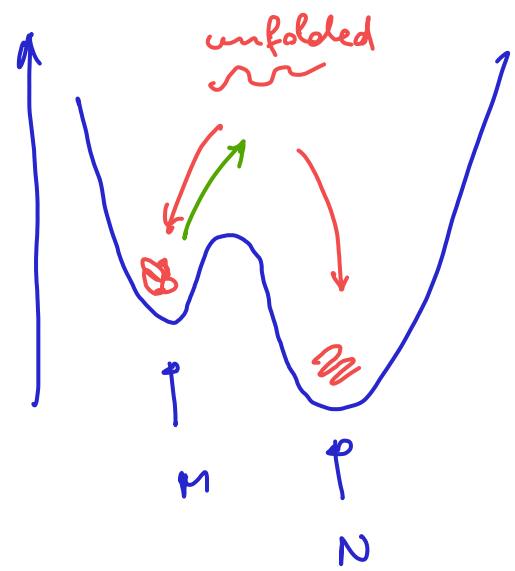
The energy of ATP is used to have a large affinity (a strong grip) on the substrate.

What is the large affinity useful for?

When Hsp70 binds to a substrate, its large volume (it is really large!) disturbs the substrate conformation. How? Imagine that it corresponds to a monomer that suddenly becomes much larger. By excluded volume it leads to an increase of the end-to-end distance. But increasing the end-to-end distance against the energy (it's like expanding a collapsed polymer) increases the free-energy of the polymer.

The polymer tries thus not to do it, by getting rid of the chaperone. If the binding energy of the chaperone is small enough, the substrate expansion is the "weak link".

Why is expanding the substrate useful?



↗ = Hsp70 action

upon unfolding, the protein can natively fold, or misfold again, going through another chaperone cycle.

How does Hsp70 know which substrates to act on?

Only mildly hydrophobic substrates activate hydrolysis.

Thus, native proteins without accessible hydrophobic surfaces do not activate Hsp70, whereas misfolded proteins do.