Ion PETRE

Academy of Finland
Åbo Akademi University & TUCS, Turku, Finland

# Computational Models in Systems Biology
# - Heat shock response in Eukaryotes -

Joint work with

- R.Back
- J.Eriksson (BTK)
- L.Sistonen (BTK)
- A.Mikhailov (BTK)
- A.Pada (ÅA-CS)
- S.Saxen (ÅA-CS)
- K.Nylund (ÅA-CS)

# Main research topics

- Gene assembly in ciliates
- The heat-shock response in eukaryotes
- Self-assembly

# Agenda

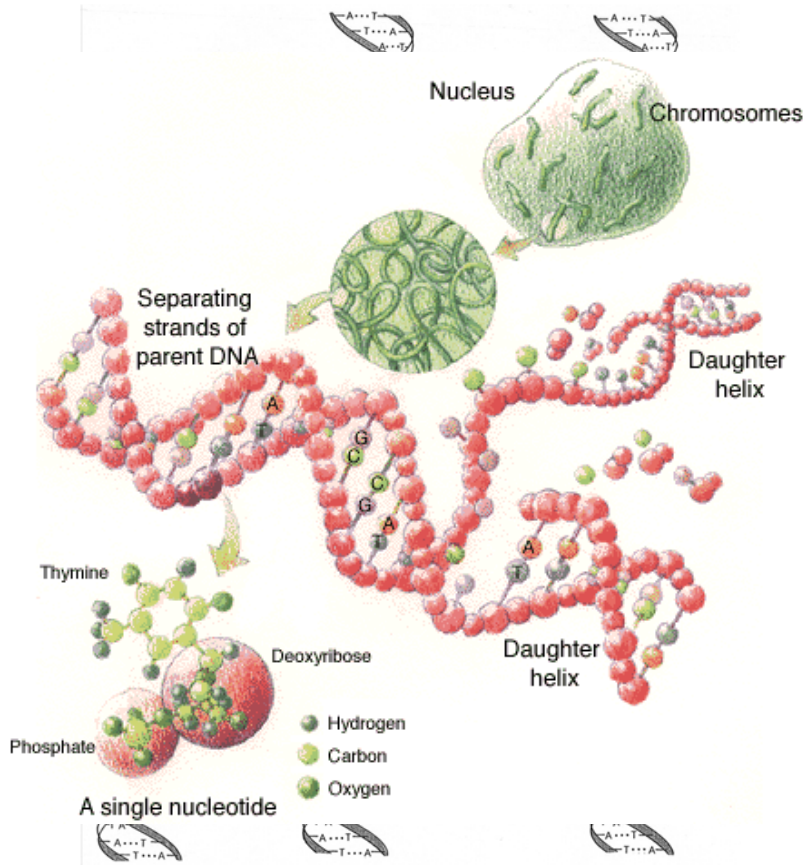
- A crash course on Molecular Biology
- The problem: heat shock response
- A continuous model
- A discrete model
- Perspectives

# A (5 min?) crash course on Molecular Biology

- DNA
- Proteins
- Enzymes
- Genes
- Chaperons

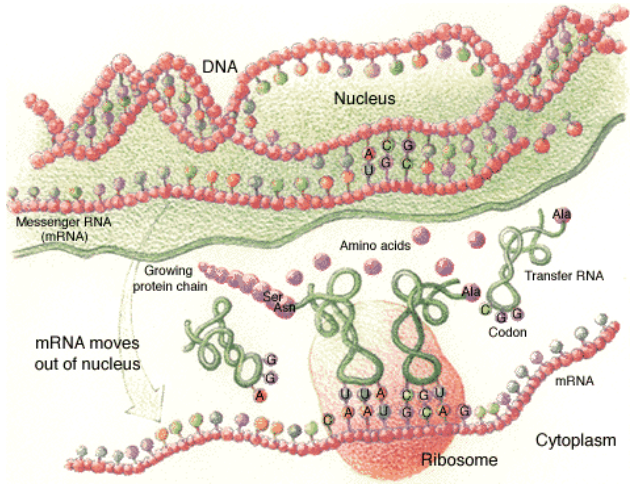
# A crash course on Molecular Biology

- DNA
  - Single strands: sequences of nucleotides (A, C, G, T)
  - Watson-Crick complementarity: A-T, C-G
  - Double strands: two single strands with complementary nucleotides bind together forming a double helix
  - Contains the blue print of the organism, each cell has a complete copy
  - Humans: some 3 billion base pairs in every single cell
- Proteins
- Enzymes
- Genes
- Chaperons



# A crash course on Molecular Biology

- DNA
- Proteins
  - Sequences of amino-acids (20 possible)
  - Translated from RNA based on a universal code
  - 3 nucleotides (codon) code for an amino acid, some amino acids correspond to several codons
    - Only one start codon, 3 stop codons
  - Form a 3D fold – determines the function of the protein
  - The fold is determined by the sequence (and the outside conditions)
  - "Holy grail" of Bioinformatics: the protein folding problem – predict the 3D fold based on the (linear) amino acid sequence
- Enzymes
- Genes
- Chaperons




Second base of codon

| First base of codon | U | C | A | G | Third base of codon |
|---|---|---|---|---|---|
| U | UUU, UUC Phe; UUA, UUG Leu | UCU, UCC, UCA, UCG SER | UAU, UAC Tyr; UAA, UAG | UGU, UGC Cys; UGA; UGG Trp | U, C, A, G |
| C | CUU, CUC, CUA, CUG Leu | CCU, CCC, CCA, CCG Pro | CAU, CAC His; CAA, CAG Gln | CGU, CGC, CGA, CGG Arg | U, C, A, G |
| A | AUU, AUC, AUA Ile; AUG Met | ACU, ACC, ACA, ACG Thy | AAU, AAC Asn; AAA, AAG Lys | AGU, AGC Ser; AGA, AGG Arg | U, C, A, G |
| G | GUU, GUC, GUA, GUG Val | GCU, GCC, GCA, GCG Ala | GAU, GAC Asp; GAA, CAG Glu | GGU, GGC, GGA, GGG Gly | U, C, A, G |

The genetic code, written by convention in the form in which the Codons appear in mRNA. The three terminator codons, UAA, UAG, and UGA, are boxed in red; the AUG initiator codon is shown in green.

# A crash course on Molecular Biology

- DNA
- Proteins
- Enzymes
  - Special type of proteins, specialize in recognizing very specific blocks of DNA and binding to it
  - Some of them may then cut the DNA in a precise way, others may copy or repair DNA, etc.
- Genes
- Chaperons

# A crash course on Molecular Biology

- DNA
- Proteins
- Enzymes
- Genes
  - DNA has coding blocks (genes) and non-coding (junk?) blocks
  - Humans: some 20 000 – 30 000 genes (**in every cell!**)
  - Genes are transcribed into RNA that is then translated into proteins
  - RNA: similar structure as DNA, T replaced with U, mostly single stranded
  - Not all genes transcribed in all cells
  - Controllers: some non-coding blocks upstream of the gene – promoter regions
  - The RNA polymerase enzyme cannot bind to DNA on itself – helped by other enzymes that bind to the promoter region
  - Promoter region may be inhibited by other regions
  - **A robust computer science-like system: "if-then-else"**
- Chaperons

# A crash course on Molecular Biology

- DNA
- Proteins
- Enzymes
- Genes
- Chaperons
  - Proteins assisting other proteins in achieving proper folding.
  - Many chaperones are heat shock proteins: proteins expressed in response to elevated temperatures.
    - Protein folding is severely affected by heat, and therefore chaperones act to counteract the potential damage.
  - Chaperones recognize unfolded proteins by the hydrophobic residues these expose to the solvent.
  - Incompletely folded proteins or misfolded proteins with exposed hydrophobic groups have a tendency to aggregate.
    - This aggregation is extremely detrimental to the cell: see Alzheimer's and Creutzfeld-Jacob's (human version of mad cow disease)
    - Chaperones help to prevent this by providing encapsulated hydrophobic environments that allow the protein to properly fold.

# Agenda

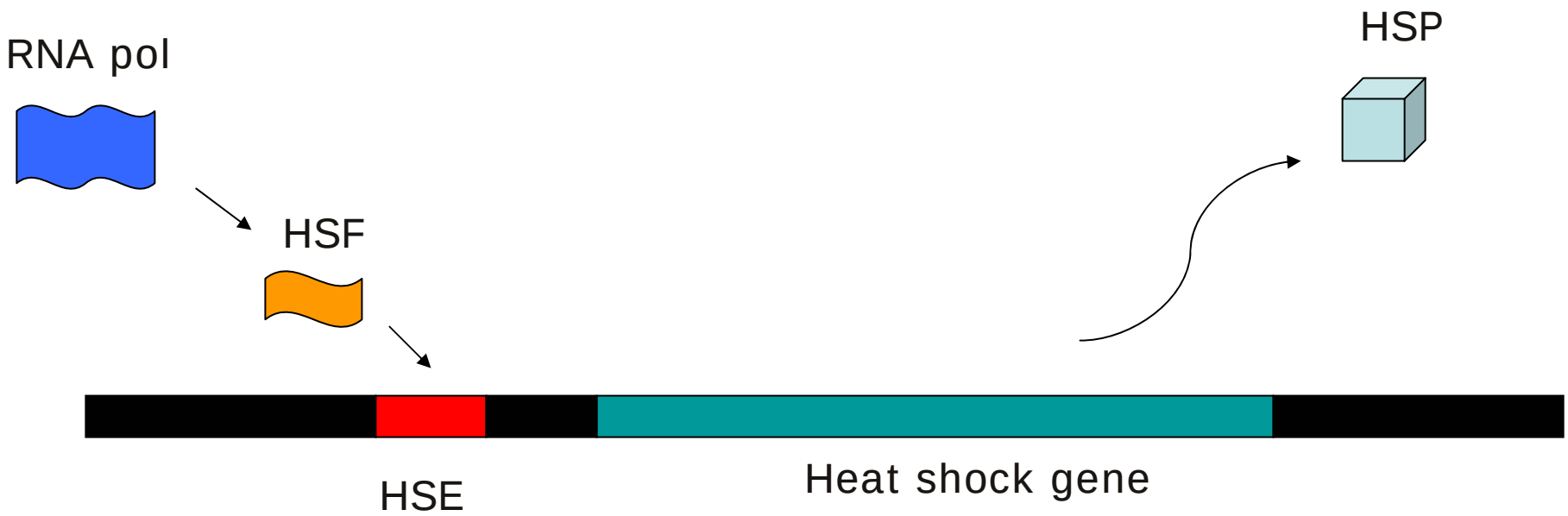
- A crash course on Molecular Biology
- The problem: heat shock response
- A continuous model
- A discrete model
- Perspectives

# Heat shock response

- Proteins tend to misfold under heat and exhibit their hidrophobic cores
  - Misfolded proteins may accumulate with disastrous effects for the cell
- The eukaryotic cell reacts to this by producing “heat shock proteins” (HSP) – several types of them (HSP 27, HSP70, HSP90,…)
  - HSP acts as a chaperone: assists in refolding of other proteins or directs their degradation
  - Under shock the cell produces massive amounts of HSP to handle the staggering amount of misfolded proteins
- A certain limited amount of HSPs exists also without stress
  - Not convenient (energy expensive) to keep up a huge amount of HSPs at all times
- Challenge (for the cell):
  - *React promptly to an elevated temperature*
  - *Find the optimum amount of HSPs for survival in those conditions*
  - *Return to the original level of HSPs once heat shock is remove*

# Heat shock response

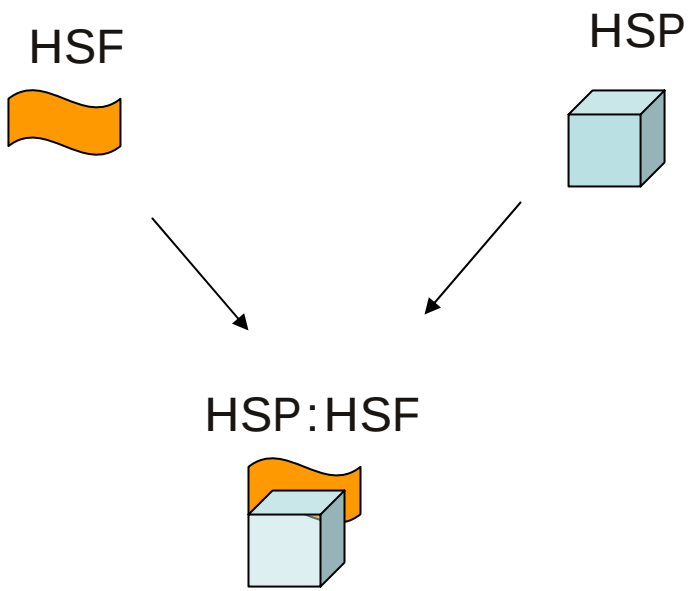
- Take a look at how production of HSP is being controlled
  - The transcription of the heat shock genes is regulated through the promoter HSE (heat shock element)
  - The heat shock gene can only be transcribed if a certain protein HSF (heat shock factor) binds to the promoter HSE



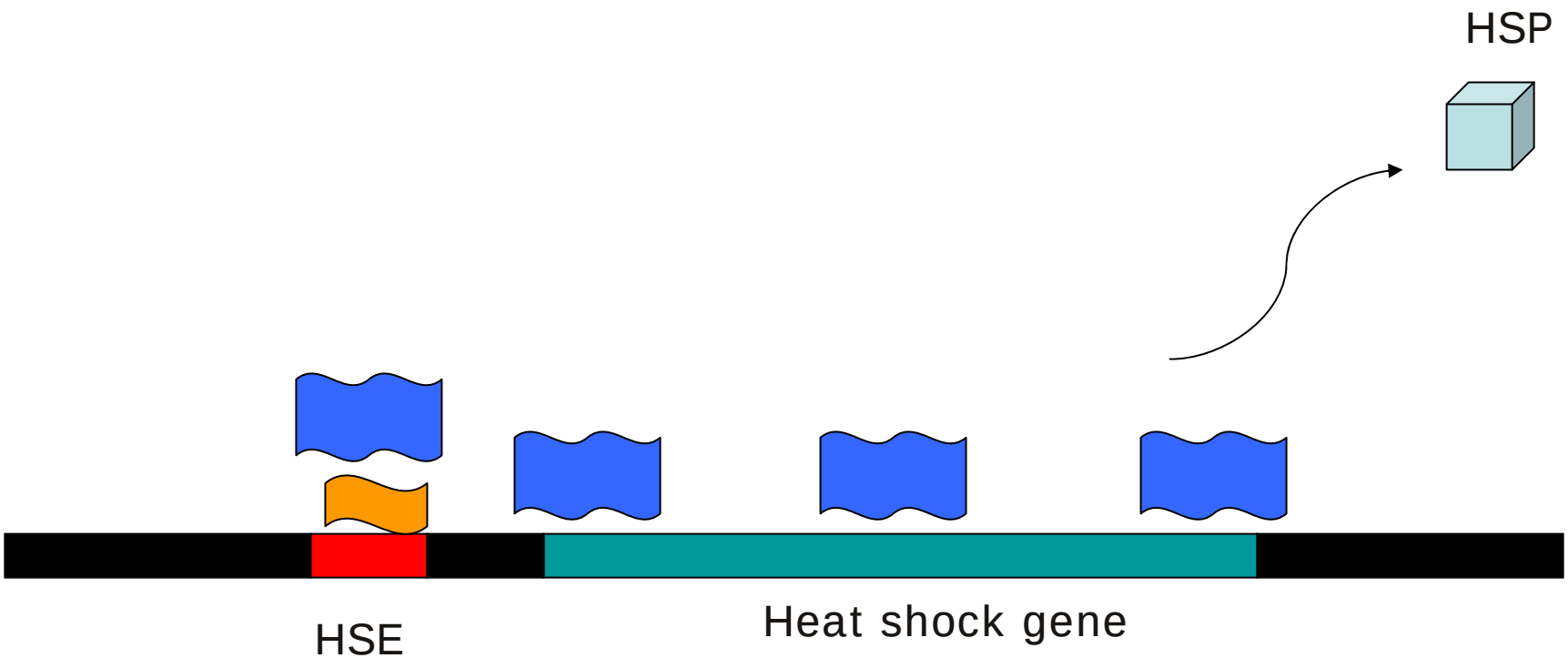
# Heat shock response

- Production of HSP needs to be stopped once shock is removed
  - If heat is reintroduced the cell is reacting much quicker to the challenge than in the beginning
  - How is this “implemented”?
- HSP has an affinity for HSF
  - HSP binds to HSF rendering HSF inactive
  - HSP removes HSF from DNA (HSE) thus making RNA pol unable to bind to the heat shock gene and produce more HSPs

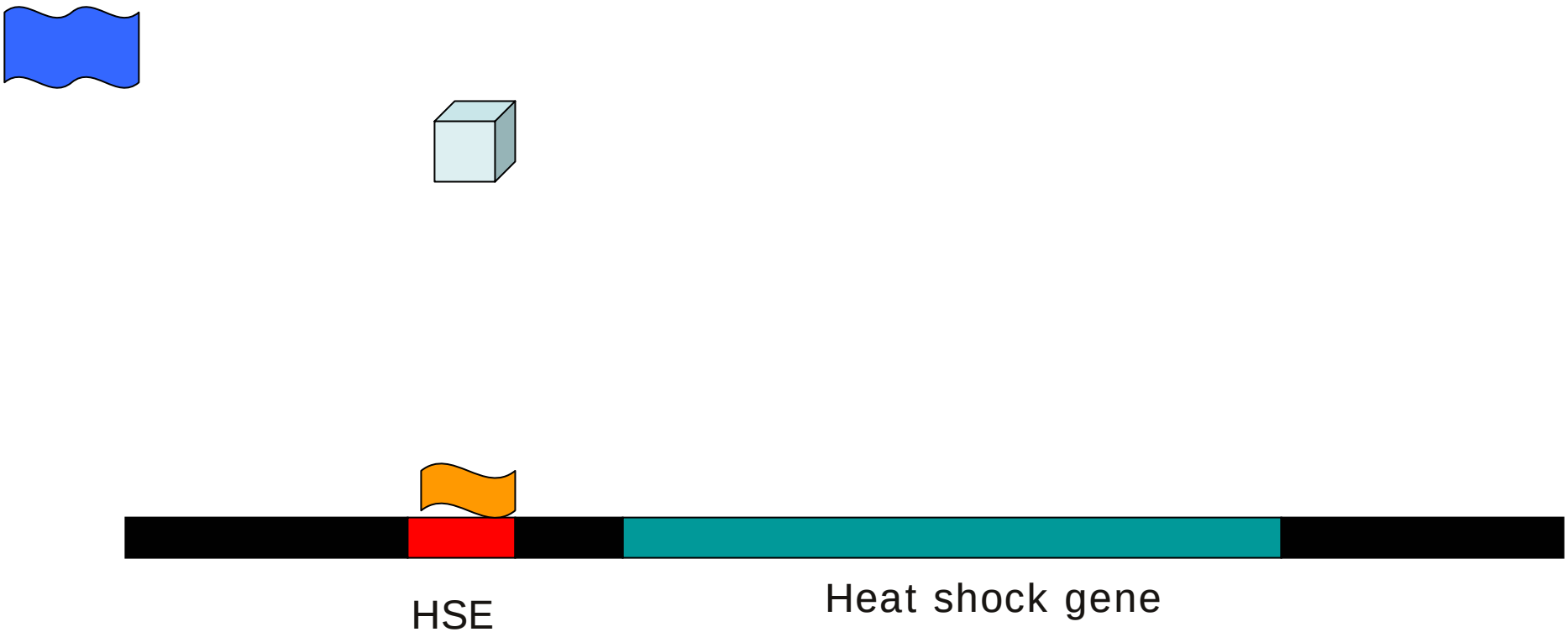
# HSP-HSF back regulation



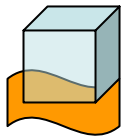
# HSP-HSF back regulation



# HSP-HSF back regulation



# HSP-HSF back regulation

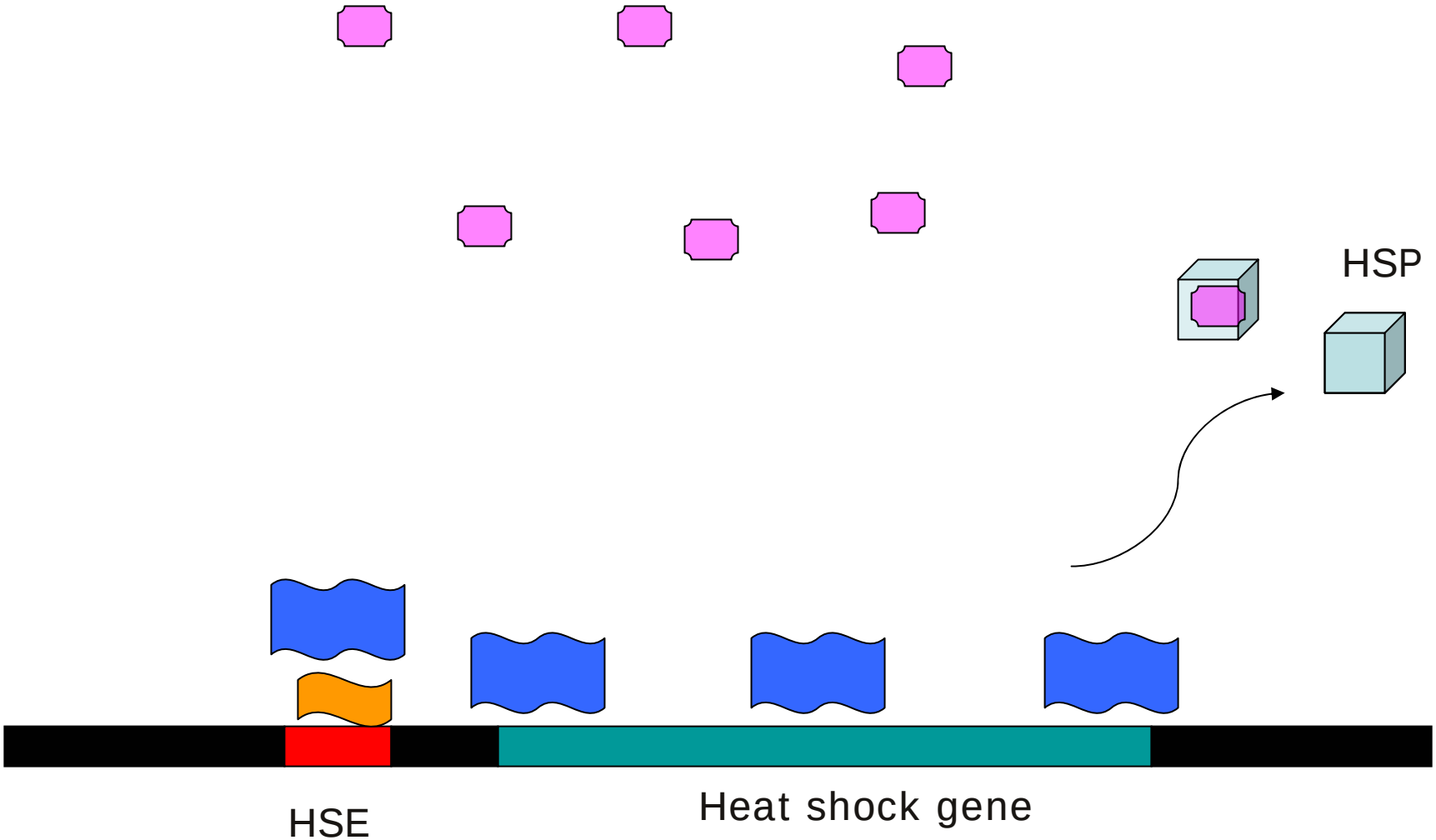


MFP

HSE

Heat shock gene

# HSP-HSF back regulation



# Heat shock response: summary

- Proteins get misfolded under heat: MFPs produced massively
- HSP binds to MFP and either helps them to refold or directs their degradation
  - HSP+MFP→HSP:MFP
- HSF binds to HSE (DNA) promoting translation of more HSP
  - HSF+HSE→HSF:HSE
  - HSF:HSE→HSP
- Once there is an excess of HSPs in the cell, HSP binds to HSF rendering it inactive; it also breaks bonds HSF:HSE
  - HSP+HSF→HSP:HSF
  - HSP+HSF:HSE→HSP:HSF+HSE

# More details

- Our molecular model has several other details that I ignore in this presentation for sake of clarity
  - HSF is anyway inactive by itself; it must form trimers HSF:HSF:HSF before it is able to bind to HSE
    - How are trimers being formed?
  - A trimer HSF3:HSE can only promote translation of the heat shock gene if it is hyper-phosphorylated
    - How is DNA translation controlled by HSF phosphorylation?
  - Two types of HSP, only one is heat-induced. HSP gets degraded after a while.
  - The level of household HSP is constant
  - The level of HSF is constant

# Objectives (for biologists)

- Give a quantitative model (at least in relative terms)
  - How strong is each reaction in itself or at least compared with other reactions?
  - Framework for “virtual experiments” (either too expensive to perform the real experiments, or too time consuming, or simply impossible)
    - Prolonged heat shock
    - Recovery experiments
    - Measurements
  - Framework for reasoning about heat shock
    - Is an equilibrium reached at a certain temperature, how quickly, what are its characteristics?
    - Scenarios: what if some reaction is slower/faster – outcome, possible “treatments”, etc.

# Objectives for computer scientists

- Defining the formal framework required by the biologists, reasoning about it
- How is it best to reason about such problems?
  - Continuous methods (differential equations) as in physics and chemistry?
  - Discrete methods (agent-based, automata, Petri nets, etc) as in computer science?
- Heat shock response as a computing paradigm: given an input/trigger, a response is computed
  - How much can it compute, how reliably
  - For a well defined set of requirements is Nature's solution the optimal implementation
  - Refinement of a different (and better!) implementation?
  - Plant/sensor/actuator model

# Two computational models

- The continuous model
  - Andreas Pada, Stefan Saxen
  - Based on differential equations
  - Implemented in Kitano's CellDesigner and simulated in Jarnac (freely available tools), based on SBML
  - Fundamentally deterministic model
- The discrete model
  - Kristian Nylund
  - Based on stochastic distributions
  - Implemented through Markov chains and simulated in our own Java-based simulator
  - Fundamentally non-deterministic model

# Perspectives: phosphorylation

- Phosphorylation only sketched so far: we consider that only 5% of all trimers are hyper-phosphorylated at any time – computed this assuming independent, all-activating sites
  - Model the phosphorylation
    - Inhibitor vs. activator sites
    - Dependency between the (de)phosphorylation of various sites
    - Increase/decrease in phosphorylation with heat stress
  - Plug the phosphorylation into both models
    - As a heat-dependent average percentage of activated HSFs, or
    - As individual characteristic of each HSF
  - Estimate the distribution of HSF in various phosphorylation states depending on the temperature

# Relevance for Biology

- Trimer formation
- Clarifying the point where cell detects stress
- Long simulations
- No contamination
- Costs
- Clean drop in temperature
- Scenarios

# Relevance for Computer Science

- Modeling and computational aspects
  - **Model forming** for both the discrete and the continuous model, starting from the molecular model
  - General discussion on **continuous vs. discrete modeling**
- Mathematics of both models
  - **Convergence** to steady or exploding state
  - **Time to reach steady state**
  - **Characterize steady state** depending on temperature and initial conditions

# Relevance for Computer Science

- Heat shock as a computing paradigm
  - Given an input (temperature), the cell reacts by “computing” the optimal level of HSP
  - The challenge here is to understand the principles of this computation: reverse engineer Nature’s design
  - How much can one compute in this way, how fast, reliably, efficiently, etc.
    - Give a computational model mimicking the heat shock module
  - Define the **specifications** of a heat shock response module and **refine** it to an implementation– verify if Nature’s own design is in some way unique or optimal
    - Conditions to be fulfilled: robustness, energy efficiency, transient dynamics
  - **The “plant/sensor/actuator”-based model**

# Achievements, novelty

- Our molecular model is more detailed than the others
  - Heat shock response mostly modeled for bacteria, ours for eukaryotes
  - Our molecular model is refined to more details than most others
    - Misfolded proteins rather than “stress kinase” as triggers of the response
    - HSP of two types: household and “dynamic”
    - Degradation of HSP included in the model
    - HSF forms dimers, trimers
    - Working under the natural hypothesis that trimers are formed from dimers and monomers rather than from monomers directly
    - Proportion of dimers vs. trimers according to observation
    - Phosphorylation taken into account (still to a limited degree)
    - Time to refold considered
    - Transcription delay and time from transcription to translation

# Goals of this project

- How to model cellular reactions
  - Discrete vs. continuous
  - Modeling the particle motion
  - Modeling the bond formation and bond breaking
  - Deducing parameters
  - Experimental data
- How to build a robust (computing) system out of unreliable units
- Based on a certain model, deduce (reason) how to overcome failures of various kinds, identify sensitive points for various reactions, etc.
- What kind of formalism is best for describing cellular systems and reasoning about them
- Can the heat-shock response be used as a (theoretical, Turing-universal) computing device?