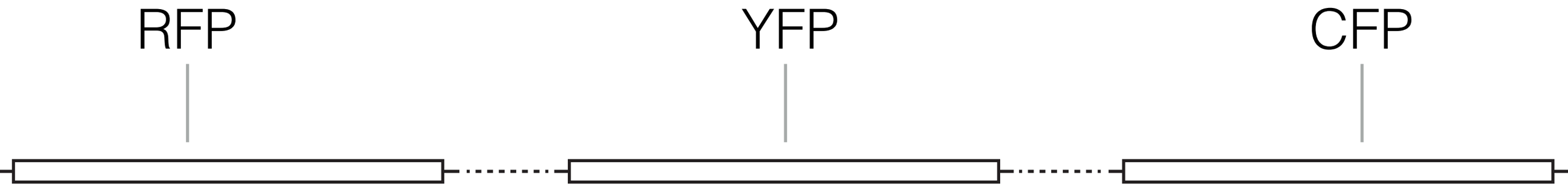
Protein Design Language

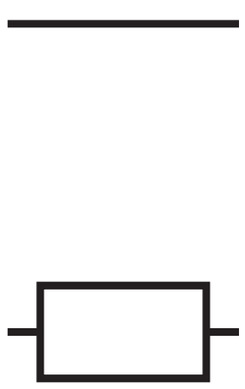
Robert Sidney Cox
Kobe University

Unspecified protein


RFP
YFP
CFP

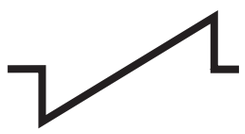
Membrane region

region types describe long protein sequences with variable lengths



Unspecified region

Named region, specified function
(protein domain)



Membrane region

Site glyphs decorate protein regions

DNA binding domain

Inducer binding pocket

Repressor

d

b

DNA binding domain

Inducer binding pocket

Repressor

d

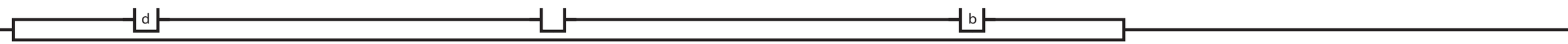
b

Dimerization domain
(convention: blank label)

DNA binding domain

Inducer binding pocket

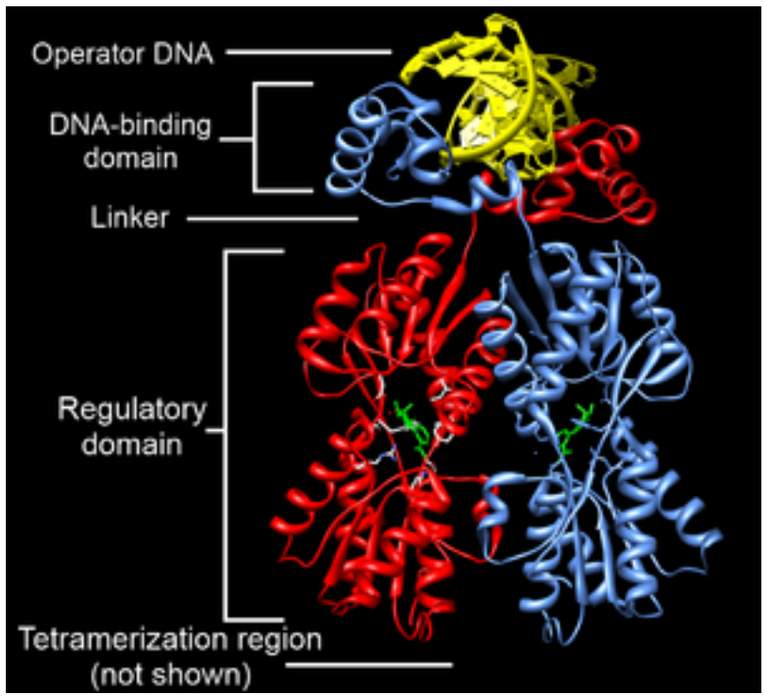
Repressor

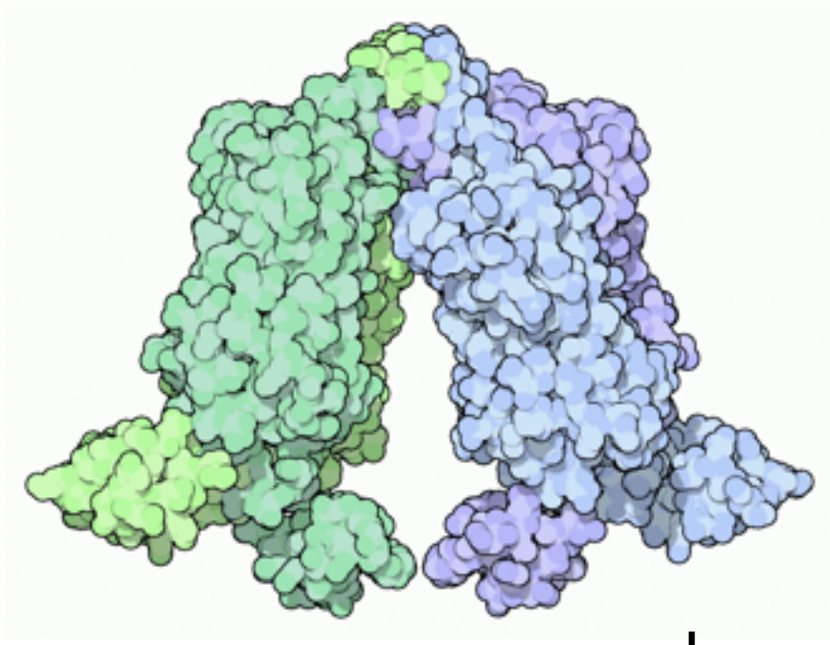


DNA binding domain

Inducer binding pocket

Repressor





Lac Repressor

Drawing Constraints

- One line per protein, of any shape
- Glyphs can be omitted
- Three classes of glyphs
  - Regions: variable size & display
  - Sites & Tags: single residues or small patches
- Glyphs must lay centered on the line
- Regions (only) can be scaled horizontally
- Overlap
  - Allowed: sites and tags onto specific or nonspecific regions
- Not allowed
  - overlapping sites or tags
  - overlapping region types

Modification glyphs describe protein modifications

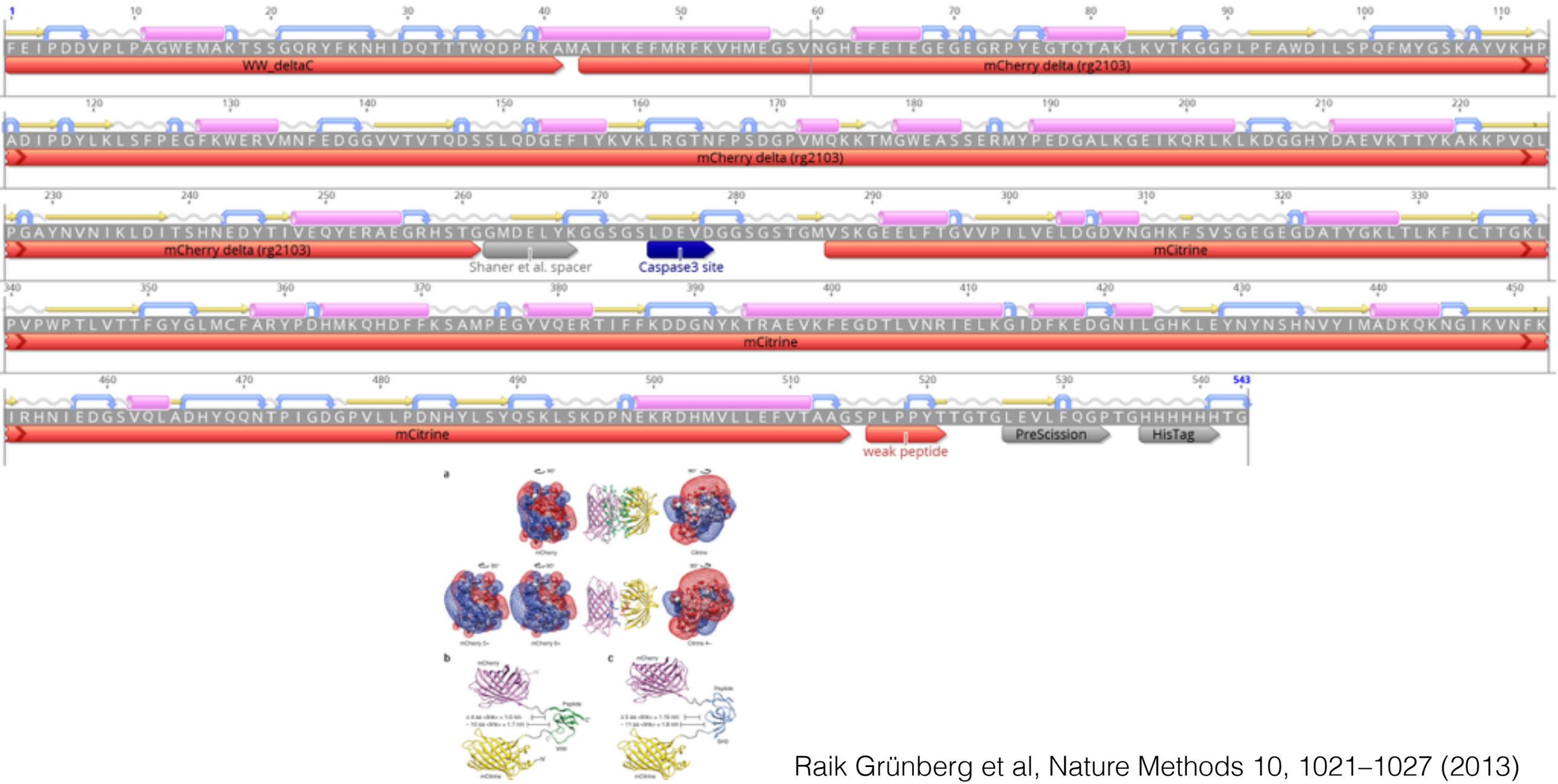
Cleavage site, protease

Covalent binding, post-translational

Protein stability element

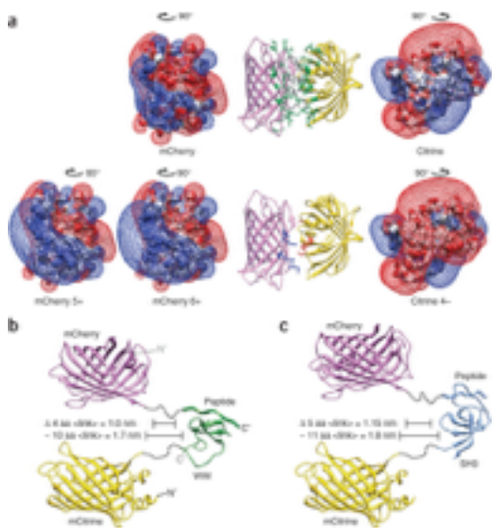
Protein degradation signal

Biochemical tag



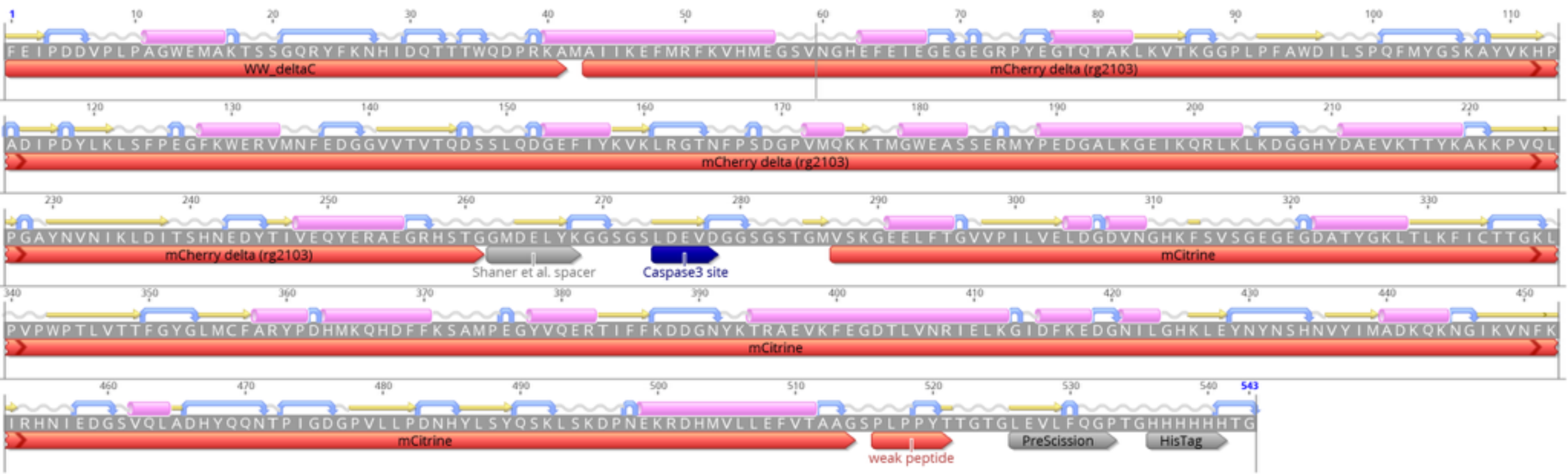
Raik Grünberg et al, Nature Methods 10, 1021–1027 (2013)

Separate structural information from design



Raik Grünberg et al, Nature Methods 10, 1021–1027 (2013)

Separate structural information from design



ww_deltaC stability

mCherry (rfp) domain

caspase 3 site protein cleavage

citrine (yfp) domain

prescission site protein cleavage

His tag

2 types of binding sites, catalytic and non-covalent

Active site, enzyme catalytic

Binding site, non-covalent

Substrate

Cofactor

Catalytic enzyme

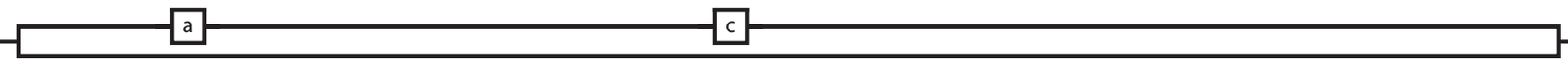
a

c

Substrate

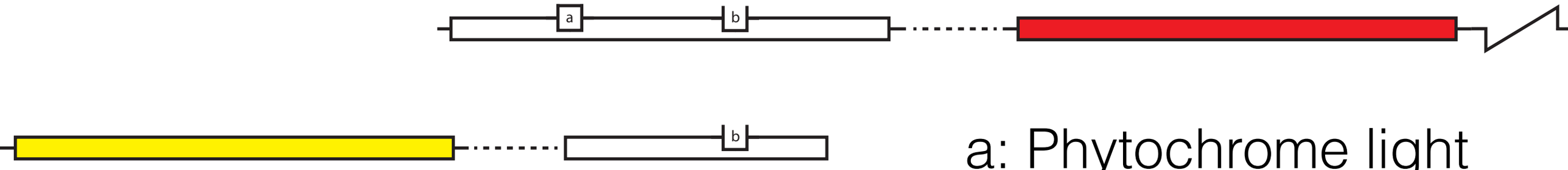
Cofactor

Catalytic enzyme



can put sites and signals into specified protein regions

Special protein interactions can be shown with binding sites and catalytic sites



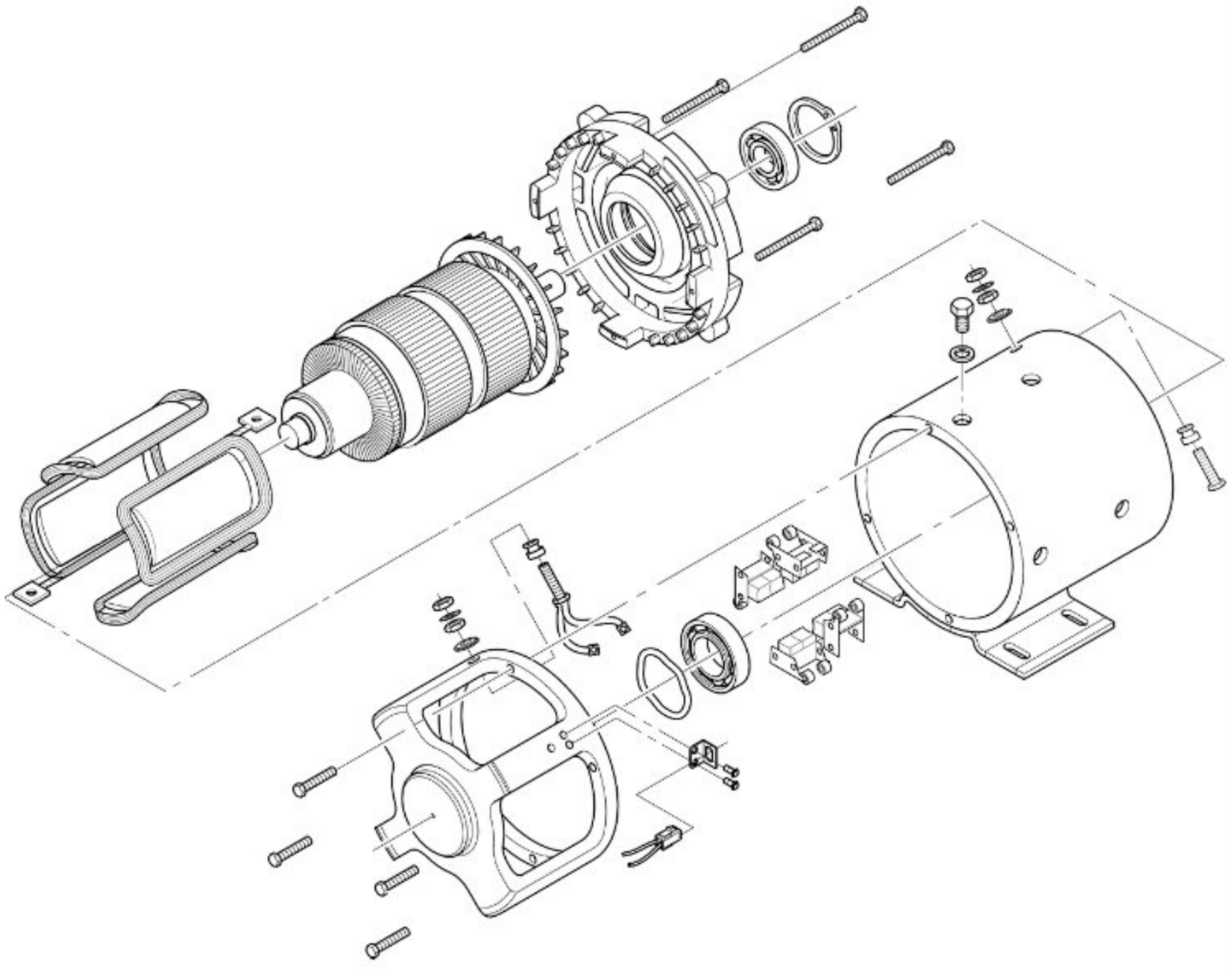
a: Phytochrome light activated binding site
b: PIF binding site

Patterned membrane recruitment of YFP

Conway's glider from the game-of-life cellular automata is projected onto membrane as Red/IR pattern (inset).

Imaging membrane with TIRF

Oscillating Membrane Recruitment and Dissociation of YFP by Alternating Red/Infra-red Light




c
e


b
d

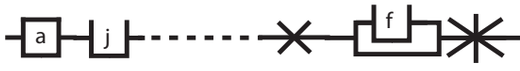

a
j
f

three repressors

LacI

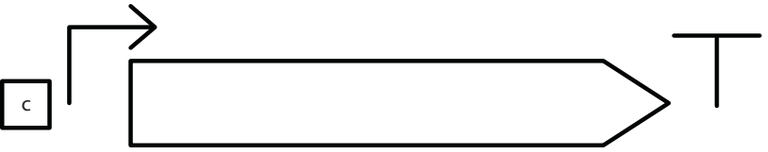
TetR

CI

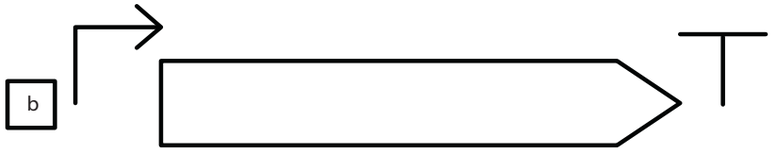


three repressors

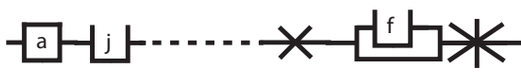
LacI

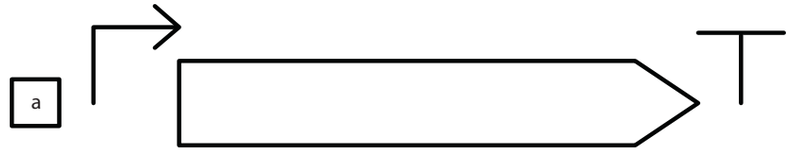


TetR



CI



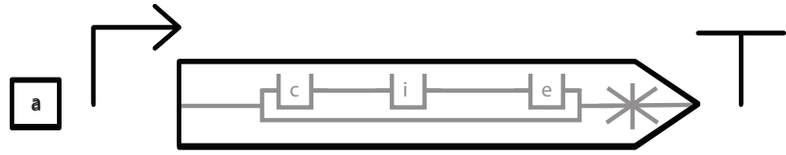


the repressilator can be described without arrows

LacI

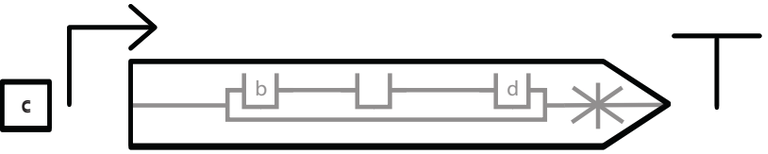
TetR

CI

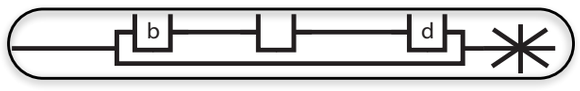


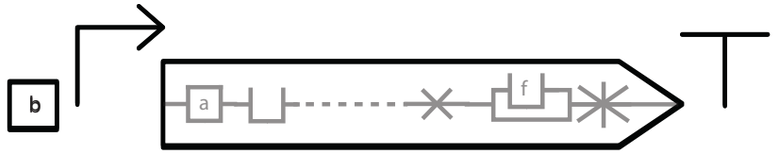
the repressilator can be described without arrows

LacI

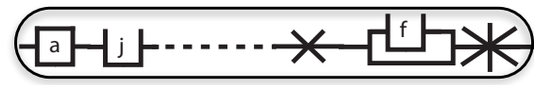


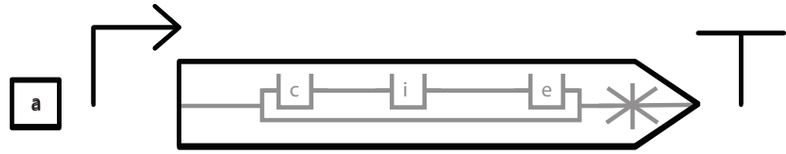
TetR





CI





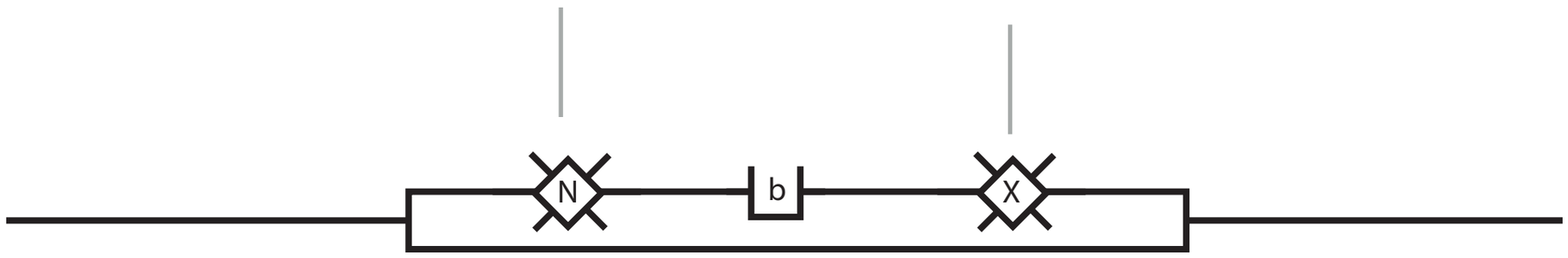
Localization signals determine where proteins are expressed

Localization signal, irreversible

Localization signal, reversible

Nuclear
Localization
Sequence

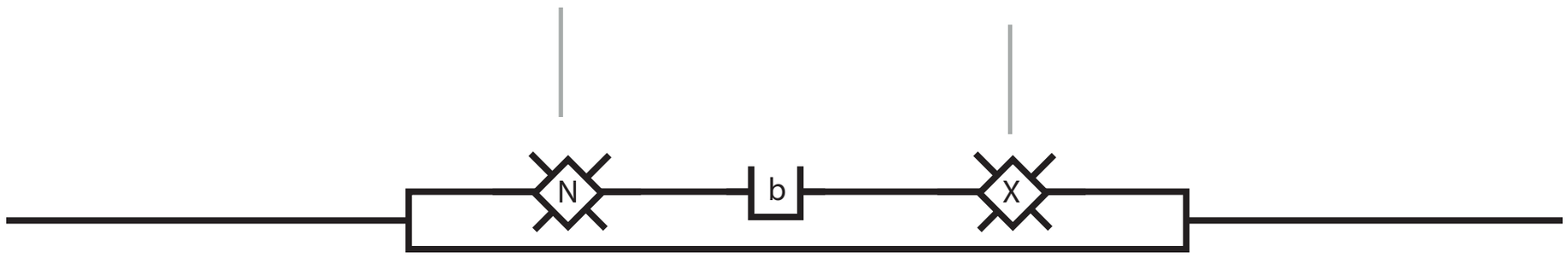
Nuclear
Export
Sequence



Estrogen Receptor

Nuclear Localization Sequence

Nuclear Export Sequence



Estrogen Receptor

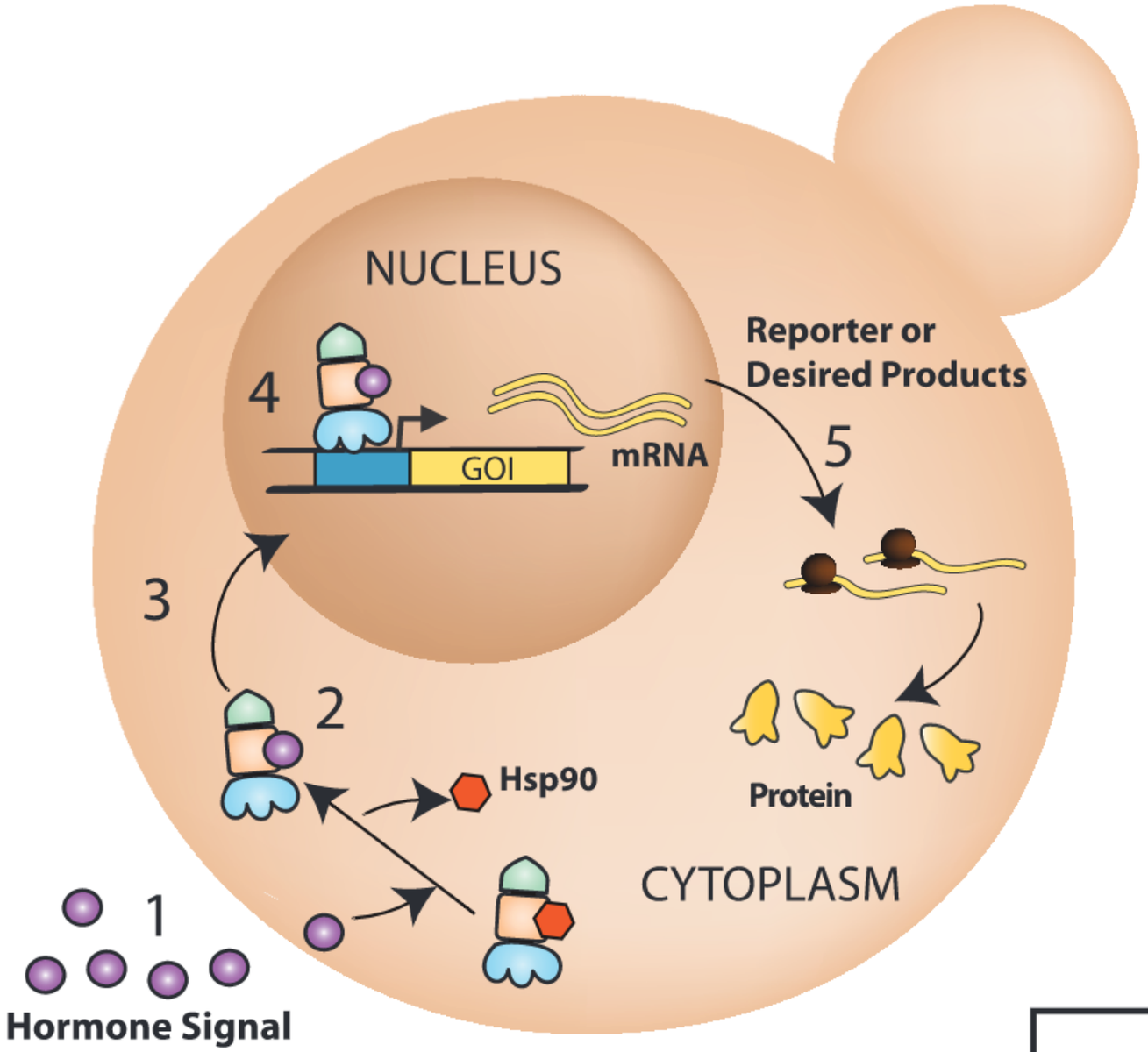
Nuclear Localization Sequence

Inducer binding

Nuclear Export Sequence

DNA binding domain

Activation domain



McIsaac, R. S. et al. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* 41, e57 (2013).

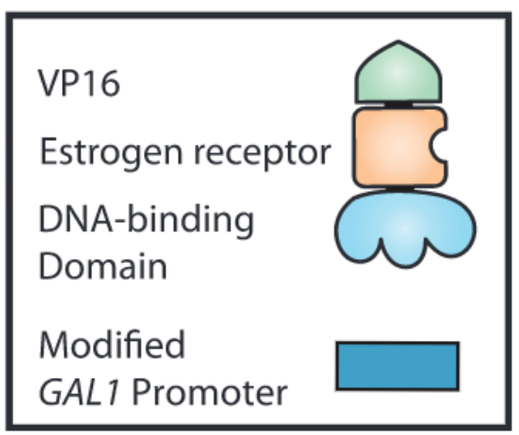
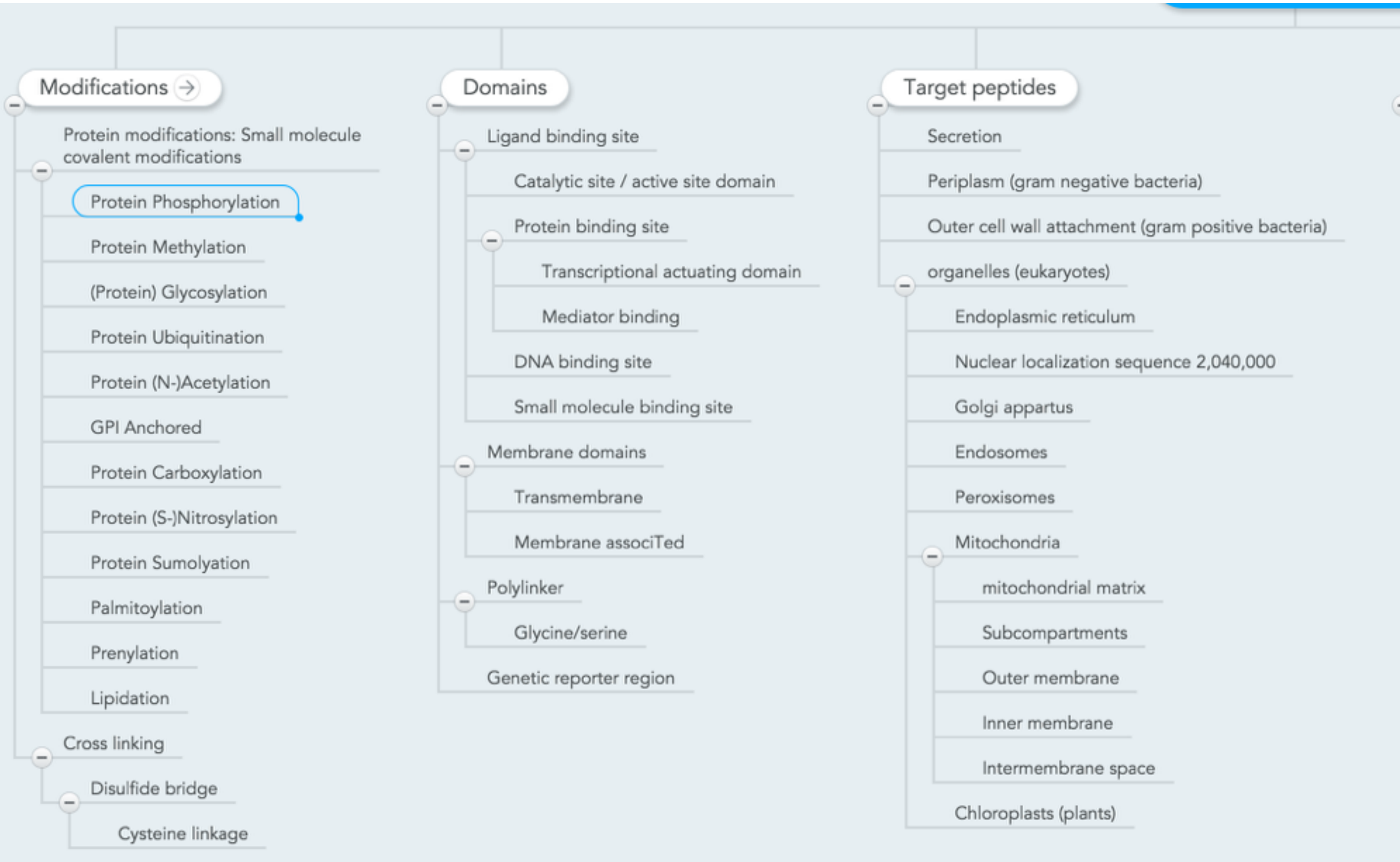
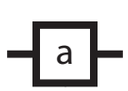


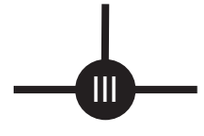
Figure 1. Schematic of hormone-based gene expression system. ATFs contain a DNA-binding zinc-finger array, the ligand binding domain of the human estrogen receptor and the VP16 activation domain. In the presence of β-estradiol (1), ATFs dissociate from Hsp90 (2), translocate to the nucleus (3) and activate transcription of a gene of interest (GOI) (4). Once produced (5) the gene products can be detected using a variety of methods.

Sites and Tags



Glyphs for describing protein designs





Visual Protein Design Language

- Shorthand for communicating engineered protein features
- Highlight specific design features / protein function
- Separate structural information from design elements
- Allow user customization
- Reduce dependency on arrows
- Can be embedded into other diagramming languages (SBOL Visual, etc.)

glyphs

unspecified

linker

membrane

specified, short
(24 pt height)

specified, tall
(36 pt height)

active site, catalytic
24 pt height box

binding site, noncovalent
24 pt height box

protein cleavage site
24 pt height box

protein modification, covalent
24 pt diameter circle

protein stability element
24 pt diameter circle

localization tag, irreversible
36pt height diamond

localization tag, reversible
36pt height diamond

degradation tag
36pt height

biochemical tag
36pt height

covalent crosslink
variable length height

Character and label constraints

- Site and tags may have 1 character type only: each character defines a different glyph
  - CAPS -> tags
  - lower -> binding
  - Roman -> modifications
- Any line not 3 pt is considered a label
- Any text besides single letters inside a glyph is considered a label
- Labels can be associated with a glyph with coordinates from glyph center label(glyph, x, y)