

## System-biologi – introducerende biologi og matematik<sup>#</sup>

Andreas Vedel, Aalborg City Gymnasium

### Cellers ekspression af gener

Et gen kan defineres som DNA-stykker der indeholder den nødvendige information for at cellen kan producere et bestemt protein.

Man taler om genes ekspression eller at gener udtrykkes når DNA transkriberes til RNA og, for de fleste genes vedkommende, translateres til protein.

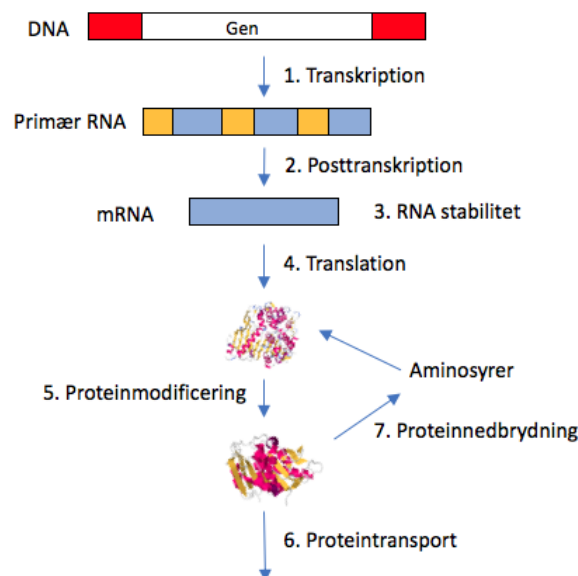
For at en celle kan udnytte den tilgængelige energi optimalt, er det vigtigt at cellen er i stand til at regulere genekspressionen, så koncentrationen af genprodukter svarer til cellens behov.

Nogle gener, de såkaldte "husholdningsgener", udtrykkes hele tiden (kaldes konstitutiv ekspression) fordi cellen har brug for at disse gener er "tændt" konstant (fx enzymet RNA-polymerase).

Langt de fleste gener udtrykkes under nogle bestemte omstændigheder og så taler man om reguleret ekspression. Ekspression af gener i bakterier kan variere mere end en faktor tusind som følge af ændringer i næringsstoffer eller andre miljøændringer.

### Regulering af cellers proteinkoncentration

Den cellulære koncentrationen af et protein er kontrolleret vha en fint afstemt balance mellem mindst 7 forskellige reguleringsmekanismer (figur 1). I dette dokument fokuseres primært på reguleringsmekanisme nr. 1 og (i mindre omfang) nr. 4.



**Figur 1**

Den cellulære koncentrationen af et protein er kontrolleret vha en fint afstemt balance mellem mindst 7 forskellige reguleringsmekanismer.

<sup>#</sup> Dette dokument bygger hovedsagligt på kap. 1 i bogen "An introduction to Systems Biology" (CRC Press 2020) af Uri Alon. Dokumentet er ikke færdigt endnu, men kan (i færdig form) forhåbentlig fungere som basisviden til at forstå anvendelsen af transkriptionsnetværk i gymnasiets biologi- og matematikundervisning.

### Cellers 'kognitive' udfordring

Levende celler eksisterer i et komplekst miljø og er i stand til at sanse mange forskellige eksterne signaler eller stimuli, fx temperatur, pH eller koncentrationen af næringsstoffer og skadelige stoffer samt biologiske signalmolekyler fra andre celler. Information om cellens indre tilstand, såsom metaboliske nedbrydningsprodukter og skader på DNA eller membranproteiner, er også vigtige for cellen. Celler er i stand til at respondere på sådanne eksterne og interne signaler ved producere forskellige proteiner.

Hvis fx sukker sanses i cellens omgivelser, så begynder cellen at producere proteiner, der dels kan transportere sukker ind i cellen og dels kan udnytte sukker som energikilde. Hvis der fx sker indre eller ydre skader, så producerer cellen nogle reparationsproteiner. Celler måler altså konstant eksterne og interne miljøændringer for at kunne beregne en passende respons, som i praksis er en passende mængde af nødvendige proteiner.

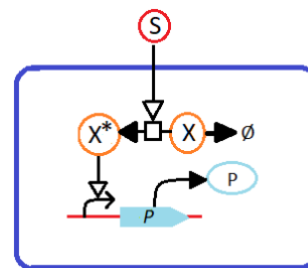
### Transskriptions-faktorer

En speciel gruppe af proteiner, der kaldes transskriptions-faktorer, er designet til at cellen kan skifte mellem aktive og inaktive molekulære tilstande med en hastighed som er moduleret af interne eller eksterne stimuli,  $S_1, S_2, S_3, \dots, S_n$ . Her anvendes  $X_1, X_2, X_3, \dots$  som symbol for en inaktiveret transskriptionsfaktor og  $X_1^*, X_2^*, X_3^*, \dots$  som symbol for en aktiveret transskriptionsfaktor.

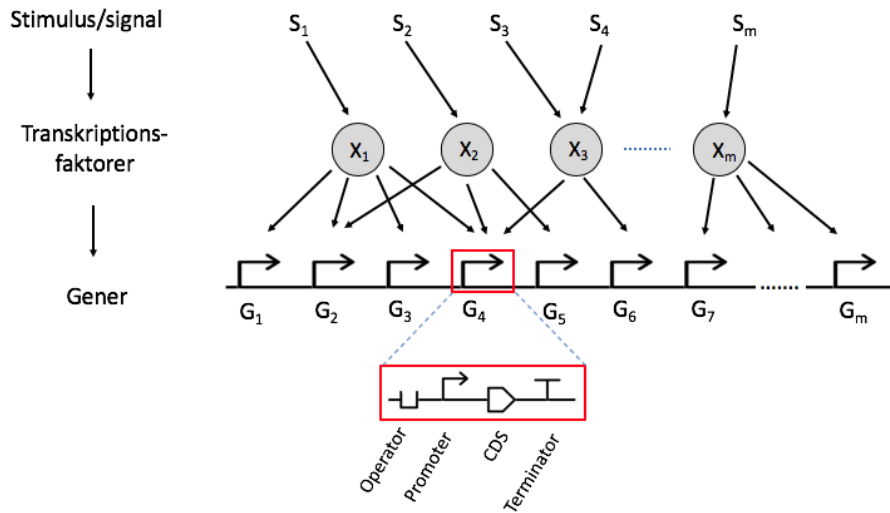
En aktiveret transskriptionsfaktor kan bindes til cellens DNA (ved promotoren) og derved regulere hastigheden af et eller flere specifikke geners ekspresion.

**Figur 2**

Et simpelt transskriptions- og translations-feedbacksystem, hvor genet  $P$  aktiveres af stimulus  $S$  fra cellens omgivelser til at danne proteinet  $P$ .  $X$  er en transskriptionsfaktor, som efter kort tid nedbrydes ( $\emptyset$ ).  $X^*$  er den aktiverede transskriptionsfaktor.



Aktivitetsniveauet af transskriptionsfaktorer kan betragtes som en celledes interne repræsentation eller 'opfattelse' af cellens virkelighed. En ganske almindelig tarmbakterie som *Escherichia coli* kan producere ca. 4500 forskellige proteiner hvoraf ca. 300 er forskellige transskriptionsfaktorer. Den interne repræsentation af et sæt transskriptionsfaktorer er en kompakt beskrivelse af et stort antal stimuli fra cellens interne og eksterne miljø. Mange forskellige situationer kan summeres vha et bestemt aktivitetsniveau af forskellige transskriptionsfaktorer til at betyde "Jeg sultner". Andre situationer kan summeres vha et andet aktivitetsniveau af transskriptionsfaktorer til at betyde "Mit DNA er beskadiget". En skematisk illustration af en sådan "transkriptionsfaktor-repræsentation" er vist i figur 3.



**Figur 3**

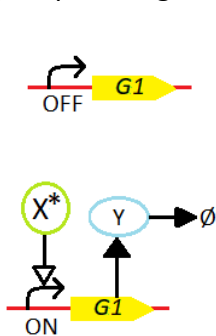
**Transkriptionsnetværk består af interaktioner**

Interaktioner mellem transkriptionsfaktorer og gener kan beskrives vha såkaldte transkriptionsnetværk, der er opbygget af nogle bestemte interaktionsmønstre. Figur 1 viser et simpelt eksempel på et transkriptionsnetværk hvor genet P aktiveres af stimulus S fra cellens omgivelser til at danne proteinet P.

Hastigheden hvormed et gen transskriberes, dvs antallet af mRNA-molekyler pr. tidsenhed, reguleres vha en regulatorisk enhed på DNA, en promoter. En promoter, der i figur 2 symboliseres vha en "knækket pil", styrer hvor RNA-polymerase skal begynde transskriptionen af et gen. RNA-polymerase bindes til promotoren, inden den begynder transskription, og derfor kan transskriptionen reguleres vha RNA-polymerases binding til promotoren.

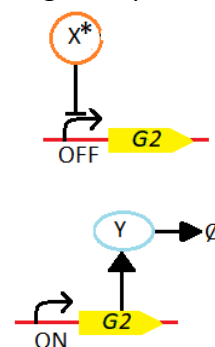
**Aktivator og repressor**

Transkriptionsfaktorer påvirker altså hastigheden hvormed RNA-polymerase transskriberer DNA til RNA. Denne påvirkning kan medføre hhv en forøgelse eller reduktion af transkriptionshastigheden hvor de ansvarlige transkriptionsfaktorer kaldes for hhv aktivatorer eller repressorer. En aktivators effekt symboliseres vha en udfyldt pilespids og en repressors effekt symboliseres vha en stump pilespids. I figur 2 og 3 er vist eksempler på en aktivator og en repressor.



**Figur 4**

Den aktiverede transkriptionsfaktor, X\*, er en aktivator, der forøger transkriptionshastigheden. Lidt forenklet forårsager en aktivator at ekspressionen af genet "tændes".



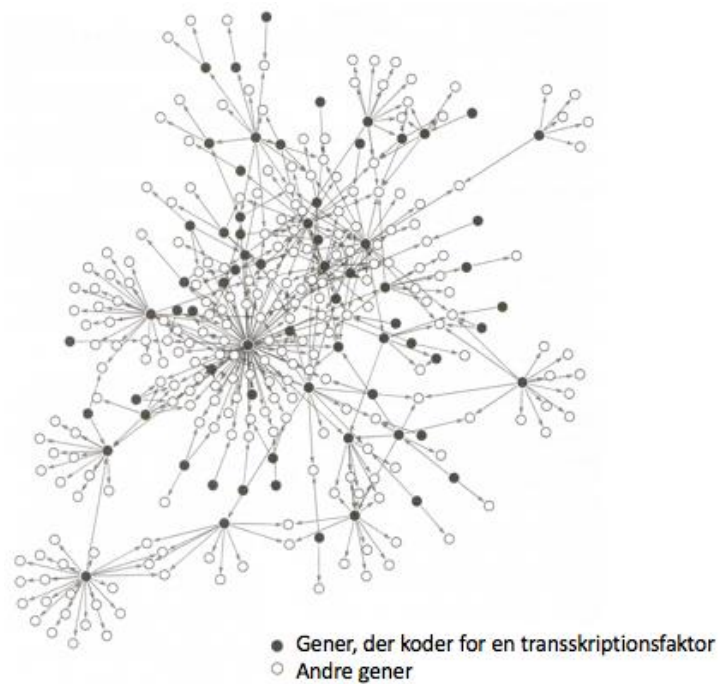
**Figur 5**

Den aktiverede transkriptionsfaktor, X\*, er en repressor, der sænker transkriptionshastigheden. Lidt forenklet forårsager en repressor at genets ekspression "slukkes".

**Figur 6**

Et udsnit af transkriptionsnetværk med interaktioner for ca. 20 % af generne i bakterien *E. coli*.

Hver cirkel (knudepunkt) i figuren symboliserer et gen og pile symboliserer transkriptionsregulering af et gen vha protein-produktet fra et andet gen. En pil  $\rightarrow$  fra et knudepunkt X til et andet knudepunkt Y betyder altså at genproduktet fra genet X er en transkriptionsfaktor, der bindes til promotoren af genet Y for at regulere transkriptionshastigheden af genet Y.



Hillfunktion – matematisk model for simpel regulering vha aktivator og repressor

Transkriptionsfaktorerers effekt kan i nogle simple tilfælde beskrives matematisk vha en såkaldt Hill-funktion:

Vi betragter produktionen af et protein Y, der er reguleret vha transkriptionsfaktoren X. Når X regulerer Y, symboliseret i transkriptionsnetværket som  $X \rightarrow Y$ , så er antallet af Y-molekyler pr tidsenhed en funktion af X-koncentrationen i den aktiverede form,  $X^*$ :

$$Y = f(X^*)$$

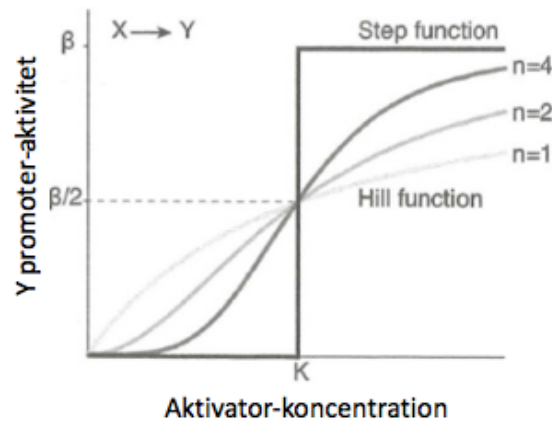
Funktionen  $f(X^*)$  er en voksende funktion når X er en aktivator og en aftagende funktion når X er en repressor. Denne type funktion kan beskrives vha en såkaldt Hill-funktion. En Hill-funktion kan udledes ved en ligevægtsbetragtning af bindingen mellem en transkriptionsfaktor og en promoter<sup>1</sup>.

Hill-funktionen for en aktivator er en kurve der stiger fra nul og (asymptotisk) nærmer sig et maksimalt mættet niveau (figur 7),  $\beta$ . Funktionsudtrykket er givet ved:

$$f(X^*) = \beta \frac{X^{*n}}{K^n + X^{*n}}$$

Hill-funktionen beskriver ofte empiriske data med relativ god præcision.

<sup>1</sup> Den matematiske udledning vil ikke blive gennemgået her. I stedet henvises til Appendix A i Uri Alons bog "An introduction to Systems Biology" (CRC Press 2020).



**Figur 7**  
 Forskellige kurveforløb af en Hill-funktion for en aktivator.

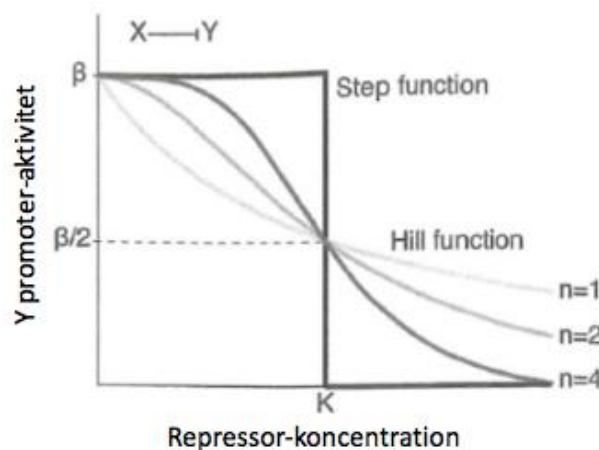
Der er tre parametre,  $K$ ,  $\beta$ , og  $n$  hvor

- $K$  kaldes aktiveringskoefficienten (med enheden koncentration) og definerer den nødvendige koncentration af  $X^*$  der skal til for at aktivere gen-ekspressionen (her målt som promoteraktivitet).
- $\beta$  er den maksimale promoter-aktivitet, som opnås ved høje aktivator-koncentrationer, dvs. når  $X^* \gg K$ .
- Den sidste parameter, Hill-koefficienten  $n$ , beskriver stigningen af kurven. Jo større  $n$ , jo mere får responsen karakter af en off-on-respons hvor genet enten er "slukket" eller "tændt".

Fra funktionsudtrykket og i figur 7 ses at 50 % af den maksimale promoteraktivitet (der er et udtryk for genekspressionen) opnås når  $X^* = K$ .

Funktionsudtrykket for en repressor er vist nedenfor og grafen er vist i figur 6.

$$f(X^*) = \beta \frac{X^{*n}}{K^n + X^{*n}}$$



**Figur 8**  
 Forskellige kurveforløb af en Hill-funktion for en repressor.

I figur 8 ses at der er tale om en aftagende kurve hvor kurveforløbet igen afhænger af tre parametre:

- En repressor forårsager størst transskription (maksimal promoteraktivitet,  $\beta$ ) når repressoren ikke er bundet til promoteren. Maksimal promoteraktivitet opnås derfor når koncentrationen af repressoren,  $X^* = 0$ .
- 50 % af den maksimale promoteraktivitet opnås når  $X^* = K$ .
- Hill-koefficienten,  $n$ , bestemmer hvor brat kurven falder. Jo større  $n$ , jo mere får responsen karakter af en on-off-respons hvor genet enten er "tændt" eller "slukket".

En interaktionspil i et transskriptionsnetværk kan, uanset om der er tale om en aktivator eller repressor, altså modelleres vha en Hill-funktion. En interaktionspil i et transkriptionsnetværk kan således tildeles tre værdier,  $K$ ,  $\beta$ , og  $n$ .

### Basal genekspression

I de to beskrevne Hill-funktioner ovenfor antages at genekspressionen går fra nul til maksimal promoteraktivitet  $\beta$ , men mange gener i celler er aldrig helt "slukket". Disse gener har en såkaldt **basal genekspression** eller en minimums-promoteraktivitet,  $\beta_0$ . Denne situation kan beskrives matematisk ved at addere  $\beta_0$  til ovenstående to Hill-funktioner:

$$Y = \beta_0 + f(X^*) = \beta_0 + \beta \frac{X^{*n}}{K^n + X^{*n}}$$

### Logisk approksimation: Transskription er "On" eller "off"

Hill-funktioner er anvendelige i detaljerede modeller af gentransskription, men ofte kan en glidende overgang mellem lave og høje værdier af promoteraktivitet,  $\beta$  ( $\approx$  genekspression) erstattes af en enten-eller-forsimpling. Her antages at genet enten er "slukket",  $f(X^*) = 0$ , eller "tændt",  $f(X^*) = \beta$ .

Denne forsimpling kaldes en **logisk approksimation** og svarer i figur 7 og figur 8 til situationer med høje værdier af Hill-koefficienten,  $n$ . I figur 7 ses at en logisk approksimation for en aktivator svarer til den anførte "Step function") med tærskelværdien  $K$ . Når  $X^*$  overstiger  $K$ , er genet "tændt".

Tilsvarende svarer en logisk approksimation af en repressor til den i figur 8 anførte "Step function" med tærskelværdien  $K$ . Når koncentrationen af repressoren  $X^*$  overstiger  $K$ , er genet "slukket".

Matematisk kan en logisk approksimation beskrives vha en såkaldt step-funktion,  $\theta$ :

- Logisk approksimation for en aktivator:  $f(X^*) = \beta\theta(X^* > K)$ ,
- Logisk approksimation for en repressor:  $f(X^*) = \beta\theta(X^* < K)$ ,
- hvor  $\theta$  enten har værdien 0 eller 1.

### Steady-state og simpel genregulering

I figur 1 ses at cellen skal modtage en stimulus før en transskriptionsfaktor aktiveres:  $X \rightarrow X^*$ . Når cellen modtager et signal,  $S$ , bliver  $X$  omdannet til  $X^*$  og bindes til promoteren for genet, der transskriberes og translateres til et protein,  $Y$  (jf. figur 2 og 3). Cellen producerer  $Y$  med en hastighed  $\beta$  (enhed: koncentration pr. tidsenhed).

Produktionen af Y er balanceret af to processer der reducerer koncentration af Y i cellen: Protein-nedbrydning,  $\alpha_{ned}$  (pga enzymer i cellen) og fortynding  $\alpha_{fort}$  (koncentrationen falder fordi cellens volumen forøges ifm cellevækst).

Den totale koncentrationsreduktion er summen af nedbrydningshastigheden og fortyndingshastigheden (begge med enheden pr. tidsenhed, 1/tid):

$$\alpha = \alpha_{ned} + \alpha_{fort}$$

Koncentrationsændringen af proteinet Y skyldes forskellen mellem produktionshastigheden  $\beta$  og koncentrationsreduktionen  $\alpha$ :

$$dY/dt = \beta - \alpha Y$$

hvor koncentrationsreduktionen beregnes som koncentrationen af Y gange koncentrationsreduktionen  $\alpha$ .

Ved steady-state har Y en konstant koncentration,  $Y_{st}$  der kan findes når  $dY/dt = 0$ .  $Y_{st}$  er ikke overraskende forholdet mellem produktionshastigheden  $\beta$  og koncentrationsreduktionen  $\alpha$ :

$$Y_{st} = \beta/\alpha$$

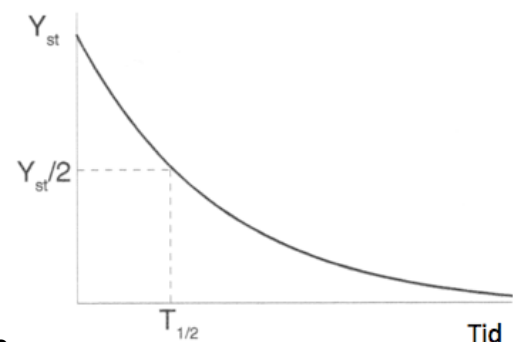
Jo højere produktionshastighed  $\beta$ , jo højere  $Y_{st}$ . Jo højere reduktionshastighed  $\alpha$ , jo lavere  $Y_{st}$ .

### Responstid

Hvis produktionen af Y stopper, dvs  $\beta = 0$ , så falder Y eksponentielt (figur 9):

$$Y(t) = Y_{st}e^{-\alpha t}$$

Hvor hurtigt falder Y? For at vurdere dette indføres den såkaldte responstid, der defineres som  $T_{1/2}$ . Det er den tid der går før Y er halveret i forhold til  $Y_{st}$ .



Figur 9

Responstiden kan findes når  $Y(t) = Y_{st}/2$ :

$$T_{1/2} = \log(2) / \alpha$$

Bemærk at reduktionshastigheden  $\alpha$  bestemmer responstiden: Hurtig reduktion medfører en hurtig koncentrationsændring. Produktionshastigheden  $\beta$  påvirker steady-state, men ikke responstiden.

Hvis proteiner har hurtige reduktionshastigheder (stor  $\alpha$ ), så er det nødvendigt med en stor produktionshastighed for at opretholde steady-state,  $Y_{st} = \beta/\alpha$ .

Fordelen ved hurtige reduktionshastigheder er at cellen dermed har en hurtig responstid, når en koncentrationsændring er nødvendig. Altså en effektiv reguleringsmekanisme.

Hvis en celle ikke har modtaget stimuli om proteinproduktion i et stykke tid, så kan det antages at proteinkoncentrationen  $Y = 0$ . Hvis cellen pludselig modtager et kraftigt signal S, så vil Y begynde at stige gradvis mod steady-state. Denne situation kan beskrives matematisk vha udtrykket

$$Y(t) = Y_{st}(1 - e^{-\alpha t})$$

Koncentrationen af Y stiger gradvis fra nul mod steady-state hvor  $Y_{st} = \beta/\alpha$ .

Responstiden kan igen findes når  $Y(t) = Y_{st}/2$  og (uden at begrunde det matematisk) viser det sig at

$$T_{1/2} = \log(2) / \alpha$$

Responstiden er altså den samme for begge situationer: Når proteinkoncentration stiger og når den falder.  $T_{1/2}$  er altså udelukkende bestemt af nedbrydningshastigheden. Jo større  $\alpha$ , jo hurtigere responstid.

### Stabile proteiner: Responstid og cellers generationstid

Mange proteiner bliver ikke aktivt nedbrudt af celler i vækst ( $\alpha_{ned} \approx 0$ ), hvorfor disse proteiner kaldes stabile proteiner. Produktionen af sådanne stabile proteiner påvirkes derfor kun af fortynding pga volumenforøgelsen,  $\alpha = \alpha_{fort}$ .

For stabile proteiner er responstiden lig med generationstiden. Argumentet er at en celle, der vokser fordobler sit volumen ved at dele sig til to celler, hvorfor koncentrationen er halveret, hvilket netop svarer til definitionen på responstid. Efter en cellegenerationstid,  $\tau$ , er proteinkoncentrationen halveret:

$$T_{1/2} = \log(2) / \alpha_{fort} = \tau$$

Bakteriers celle-generationstid,  $\tau$ , er i størrelsesordenen 30 min til nogle få timer. I plante- og dyreceller er  $\tau$  typisk på ca. en dag eller længere. Umiddelbart er det overraskende at  $T_{1/2} = \tau$ , da transkriptionsnetværk jo er ansvarlige for cellens evne til at reagere på stimuli. Med så lang en responstid for cellens stabile proteiner, kommer responsen på ydre stimuli først rigtigt til udtryk i næste cellegeneration.

Således kan responstiden være en begrænsende faktor, der udgør en hindring for udformning af effektiv genregulering. I dokumentet "Netværksmotiver" beskrives forskellige typer af genreguleringer som kan øge cellers responstid.

### Vigtige parametre

Parameter	Dansk navn	Engelsk navn	Betydning	Faktorer, der påvirker dens størrelse
X, Y, Z...	Transkriptionsfaktor	Transcription factor	Aktivator eller repressor	
X*, Y*, Z* ...	Aktiveret transkriptionsfaktor X, Y, Z...			
S	Stimulus (S <sub>x</sub> , S <sub>y</sub> , S <sub>z</sub> ...)		Ydre /Indre Range: ?	Affinitet Koncentration
f(X*)			rate af produktion af Y, aktiveret af X* Range: ?	
K	Aktiveringskoefficient	activation coefficient	Den koncentration af X*, som er nødvendig for signifikant aktivering /hæmning af Y	Den kemiske affinitet mellem X* og promotor samt andre faktorer



			Range: koncentration Enhed: Koncentration	
$\beta$	maximal promotoraktivitet	maximal promotor activity	(vandret asymptote for Y-promotoraktivitet) Range: [0,1]	Høj $X^*$ -koncentration ( $X^* \gg K$ )
$n$	Hill-koefficient	Hill coefficient	Stigning af Hill-funktions kurve Range: typisk 1-4, for stepfunktion $\infty$	Høj $n$ giver stepvis aktivering
$\alpha$	Fjernelsesrate af transkriptionsfaktor	Removal rate	$\alpha_{\text{fort}}$ Fortyndingsrate (eller hastighed?) $\alpha_{\text{ned}}$ Nedbrydningsrate (eller hastighed?) Range: [0,1] (?)	$\alpha_{\text{fort}}$ afhænger af celledeling $\alpha_{\text{ned}}$ afhænger af enzymatisk nedbrydning eller deaktivering. Kan være lav for stabile proteiner
$X \rightarrow Y$	Aktivering		X aktiverer genet Y, så protein Y stiger	
$X \dashv Y$	Hæmning		X hæmmer genet Y, så protein Y falder	
$\tau$	Cellegenerationstid		Tid mellem to celledelinger	
$Y_{\text{st}}$	Steady-state koncentration af translateret protein		Når	
$T_{1/2}$	Responstid		Den tid der går før translateret protein Y er halveret ift $Y_{\text{st}}$ .	