

Genomet

- systembiologiens rygrad

Her præsenteres en del af den moderne genomforskning og systembiologi, og det beskrives, hvordan man kan bruge denne viden til at optimere kemikalieproduktion i cellefabrikker, resulterende i en bæredygtig produktion af kemikalier

Af Mikael Rordam Andersen, Michael Lyng Nielsen og Jens Nielsen, Center for Mikrobiel Bioteknologi, BioCentrum-DTU, DTU

De sidste års udvikling af DNA-sekventeringsteknologi har muliggjort sekventering af genomet af et hastigt stigende antal organismer med industriel, medicinsk og/eller biologisk interesse. Med denne nye viden er der sket en udvidelse af den bioteknologiske forskning med mange nye teknikker og fagområder til følge. Et af disse nye fagområder er systembiologi, hvor man gennem en holistisk og ofte matematisk tilgang til biologien forsøger at forstå komplekse systemer.

Genomet - livets komponentliste

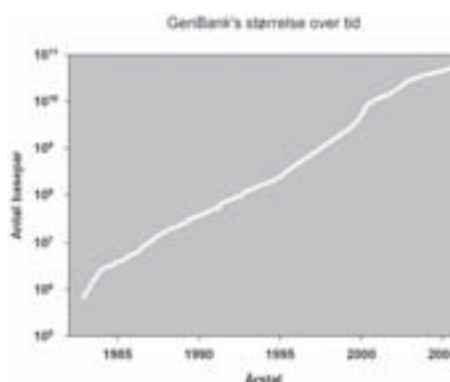
Ordet genom er en sammentrækning af ordet »gen« og den græske endelse -om, der betyder »alt« eller »komplet«, og bruges til at betegne en organismes samlede arvemasse, der er repræsenteret kemisk i DNA. Da ordet blev introduceret i 1920 af Hans Winkler, professor i botanik [1], var tanken om, at man kunne have adgang til denne information, stadig abstrakt. Det ændrede sig i 1995, hvor den første fulde genomsekvens af organismen *Haemophilus influenzae* blev offentliggjort [6]. Siden da er genomet af over 300 organismer blevet sekventeret [17], spændende fra tarmbakterier over bagegær til planter og pattedyr så som hund, kat og mennesket. Faktisk er stigningen i tilgængelige genomiske data så stor, at man har kunnet iagttage eksponentiel vækst i den centrale database GenBank [14] med en fordobling af genomisk information ca. hver 18. måned siden første optælling (figur 1).

Drivkraften bag denne indsats har været den indsigt i en organismes biologiske potentiale, som gensekvensen kan give. Det centrale dogme i biologien er, at et gens DNA transskriberes til et mRNA-molekyle, som oversættes af cellen til et protein. Kender man derfor en celledes sum af gener (gennem genomet), kan man forudsige alle potentielle mRNA-molekyler (kaldet transkriptomet) og dermed også alle proteiner, som kan være til stede i cellen (proteomet) (figur 2). Man har derfor adgang til organismens komponentliste, når man kender genomet.

Genombaseret forskning - reverse engineering af livet

Som mange forskere og investorer sidst i 90'erne måtte sande, så er en genomsekvens dog ikke en liste af svar og patenter, som blot skal renskrives, men snarere en liste af spørgsmål. Det er ikke umiddelbart muligt ud fra en DNA-sekvens at forudsige funktionen af det tilsvarende protein, ligesom man ud af et bibliotekskatalog ikke kan se, hvad man kan finde i bøgerne. Det har igangsat en række initiativer, som har til formål at kortlægge funktionen af DNA-sekvenser og proteiner.

Den klassiske metode til at klarlægge enzymfunktioner er en oprensning med en efterfølgende undersøgelse af, hvilke bio-



Figur 1. Graf over GenBanks vækst som funktion af årstallet. Bemærk at ordinaten har logaritmisk skala.

kemiske reaktioner enzymet kan katalysere. Hvis man kombinerer dette med en bestemmelse af proteinets sekvens, er det muligt herfra at bestemme gensekvensen. På denne måde er der skabt sammenhæng mellem mange enzymfunktioner og deres tilsvarende gen.

Et meget ambitiøst projekt for at udfylde den manglende viden, er *Saccharomyces Genome Deletion*-projektet [18]. Projektet har som mål, vha. molekylærbiologiske metoder, metodisk at fremstille et antal gærstammer som hver mangler ét af de ca. 6000 gener, der findes i *Saccharomyces cerevisiae* (bagegær). Studiet af disse stammer har ført til stor indsigt i funktionen af de enkelte gener, og bagegær har af denne grund udviklet sig til en meget vigtig modelorganisme for genombaseret forskning (foruden dens store industrielle betydning).

Når man har knyttet funktion til et gen i en organisme, kan man vha. computeralgoritmer finde mulige kandidater for den samme funktion i andre organismer. Det forholder sig oftest sådan, at gener med ens funktioner har lignende sekvenser, også på tværs af organismer. Man kan derved benytte computers

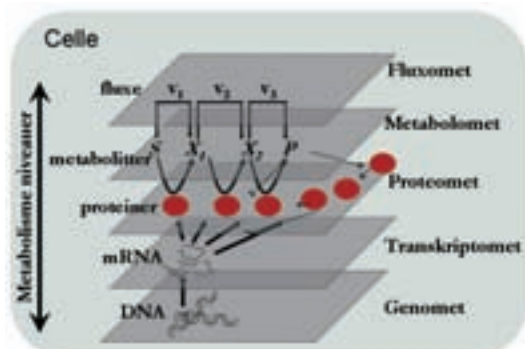
Kemisk Arbejdsmiljø

13 "rigtige" kurser indenfor kemisk arbejdsmiljø – hvilket kursus supplerer din hverdag?

Find vores kursusprogram vedlagt Dansk Kemi nr. 1/06, på vores hjemmeside eller bestil pr. telefon.

** også firma-tilpassede kurser **

ALTtox a/s • www.alttox.dk • e-mail: alttox@alttox.dk • tlf: 38 34 77 98



Figur 2. Beskrivelse af de forskellige systemniveauer i cellens metabolisme. DNA transskriberes til mRNA som oversættes til proteiner. Kemikalier i stofskiftet (metabolitter) omdannes af proteiner (enzym). Strømmene af masse gennem systemet kaldes for fluxe.

evne til hurtigt at sammenligne store mængder data til at genomsøge et genom[3]. Adskillige internetsider gør dette muligt for et stort antal genomer, og der er dermed en stor genomisk vidensbase tilgængelig for de nye genomsekvenser, som løbende kommer til. Derved identificeres komponenter i ny-sekventerede organismer lettere.

Systembiologien - genominformation sat i sammenhæng

Biokemi og biologi har indtil for nylig været baseret på den ovenstående reduktionistiske tilgang til bestemmelsen af livets byggesten. Selv i storskalaprojekter som deletions-studier bestemmer man stadig proteins funktion ét ad gangen og prøver på den måde at forstå komplekse systemer ud fra deres enkelte dele.

Argumentet for det nye fagområde systembiologi har været, at komponenter i komplekse systemer ofte ikke opfører sig på samme måde, når de er isoleret fra resten. Det gælder i biologi- en såvel som i kemiske reaktorer eller i trafikken (en bilist opfører sig forskelligt i myldretiden, ift. når personen er alene på vejen). I stedet forsøger man at forstå cellen som et netværk eller et komplekst sæt af interaktioner, og derigennem opnå forståelse af de enkelte komponents rolle, når de fungerer i deres naturlige sammenhæng.

Nøglen til observationer på systemniveau har været opfindelsen af en række nye højeffektivitets teknikker, der gør forskeren i stand til at se på forskellige niveauer af cellens metabolisme (jf. figur 2). Metoder på genomniveau kaldes genomics og har givet den sproglige stamme til en lang række -omics. Herunder findes bl.a.:

- Transcriptomics: Studier af hele cellens mRNA-niveauer under givne forhold.
- Proteomics: Bestemmelse af alle proteiner i cellen.
- Metabolomics: Måling af koncentrationen af alle stofskifteprodukter (metabolitter) i cellen.

Alle niveauer baserer sig dog i større eller mindre grad på den information der findes i genomet.

Et gennemgående træk i systembiologien til forskel fra den klassiske biologi er tilstedeværelsen af matematiske modeller. Ofte søger man at opstille en matematisk model for sit system (organismen), og vha. beregninger og praktiske forsøg afprøve og forbedre den. Disse modeller kan have mange forskellige former. Således anvendes der både traditionelle kinetiske modeller i systembiologi til at beskrive dynamikken af enkelte biosynteseveje og modeller der er afledt direkte fra data.

Systembiologien er blevet en tværfaglig videnskab, og man finder derfor både ingeniører, kemikere, biologer, fysikere, statistikere og computerprogrammører inden for feltet.

Den systembiologiske tilgang til metabolismens lag

Genomet: Ud over at se genomet som en liste af gener arbejder forskere også med genomet set som et system, hvor eksempel-

vis den fysiske placering af gener, den samlede DNA-sekvens inkl. stykker mellem gener (ikke-kodende DNA) og DNA-strengens fysiske/kemiske egenskaber tages i betragtning. Disciplinen kaldes »comparative genomics« og bruges ofte til at finde ligheder og forskelle på tværs af arter. Nogle af de førende i verden til at udvikle nye metoder til dette og grafisk repræsentation af genoms enorme datamængder (figur 3) er Center for Biologisk Sekvensanalyse (CBS) på DTU. På deres meget besøgte hjemmeside stilles et antal værktøjer og sammenligninger til rådighed [16].

Transkriptomet: Måling af niveauet af alle forskellige mRNA-molekyler i cellen kan give et indblik i den komplekse regulering af stofskiftet. På trods af, at mRNA oversættes til proteiner, er denne oversættelse ikke udelukkende kvantitativt afhængig, men snarere afhængig af RNA-strengens kemiske sammensætning og fysiske egenskaber. Man kan derfor ikke umiddelbart bruge målinger af RNA-koncentrationer til at bestemme koncentrationen og aktiviteten af enzymer og andre proteiner. Derimod kan man se på relative ændringer mellem cellens transkriptom i to tilstande. Til dette formål bruges DNA-chips eller DNA-mikroarrays, som kan måle på over 30.000 forskellige mRNA-molekyler fra samme prøve samtidigt. Eksperimenter udført med disse metoder giver ved matematikken og statistikens hjælp en indsigt i reguleringen af hele cellens stofskifte, og dermed hvordan cellen tilpasser sig forskellige forhold. Disse metoder er afhængige af, at genomsekvensen for den interessante organisme er tilgængelig.

Proteomet: Den samlede biologiske funktion af en celle er i høj grad defineret af dens niveau og sammensætning af proteiner. Af den grund er det særdeles interessant at analysere proteomet [10]. Den oftest benyttede måde til at evaluere en kompleks blanding af proteiner er en kombination af separation og massespektrometri (MS). Proteinernes teoretiske masse kan beregnes ud fra genomsekvensen, og dermed er det muligt at identificere dem. Processen er meget arbejdskrævende, men giver nogle interessante muligheder for studier af eksempelvis protein-protein-interaktion i dynamiske komplekser [8], som ellers ikke ville være mulige med en standardtilgang. Endvidere er der udviklet metoder, der muliggør analyse af det såkaldte phospho-proteom [13], dvs. måling af alle proteiner der har en fosforylgruppe påsat (for mange proteiner spiller fosforylering en stor rolle i deres funktion i cellen).

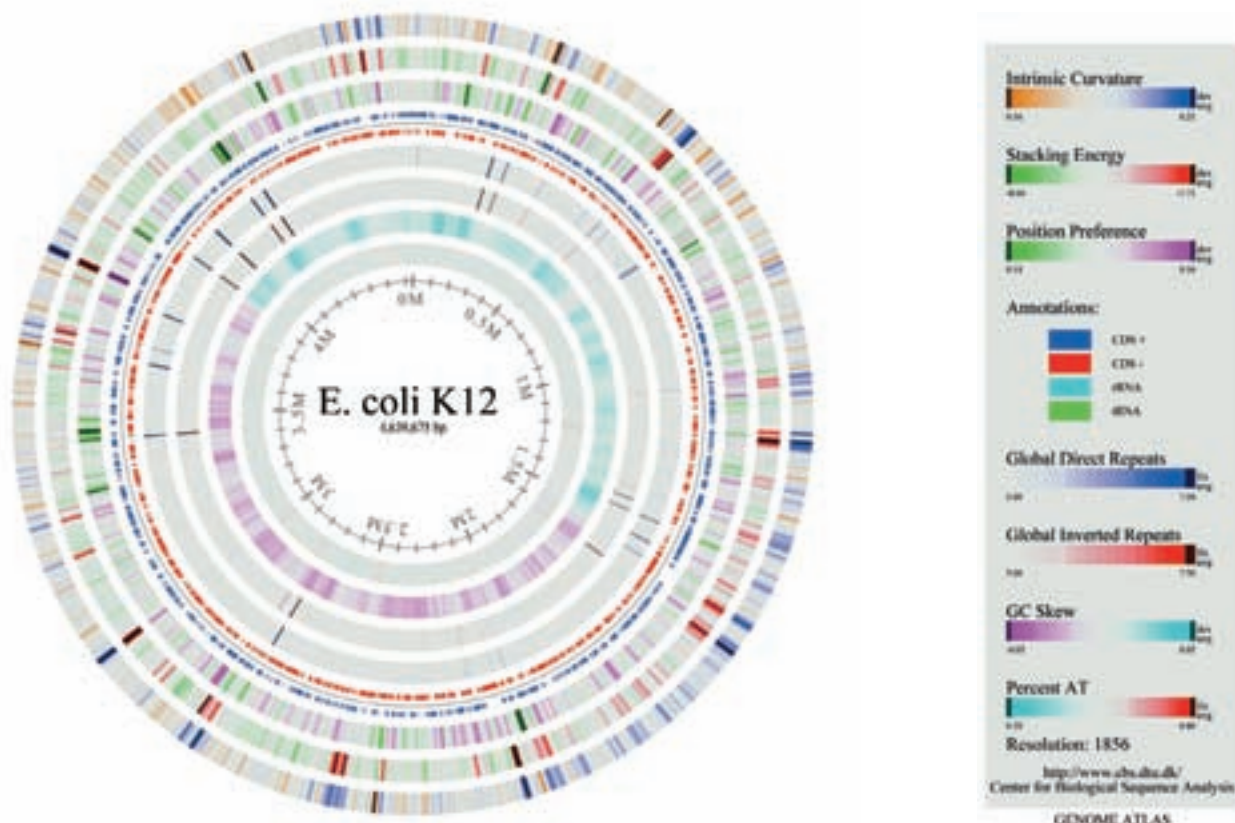
Der findes også metoder til at måle på metabolit- og flux-niveau (jf. figur 2). Nogle af disse beskrives nærmere i andre artikler i denne udgave af Dansk kemi.

Metabolisme-modeller - liv i computeren

På Center for Mikrobiel Bioteknologi (CMB) på DTU forsøger vi bl.a. at koble information fra genomet med målinger og undersøgelser fra de andre -omer i matematiske metabolisme-modeller. I disse modeller beskrives hver bioreaktion - katalyseret af enzymer - vha. en ligning, hvor de ubekendte er specierne i den biokemiske reaktion. En simuleret celle består derved af et stort antal ligninger (typisk større end 750) og et endnu større antal ubekendte. Ligningssystemet løses vha. et stort antal simple massebalancer og optimering efter størst mulig væksthastighed. Resultatet er værdier for strømme (fluxe) af metabolitter gennem alle de enkelte enzymer i cellens metabolisme.

På CMB er der blevet konstrueret modeller af bakterier [2], bagegær[5] og skimmelsvampe[4], alle brugt i industriel produktion af kemikalier. Ved at slette og tilføje gener (dvs. ligninger) i computermodellen kan man i flere tilfælde med sikkerhed forudsige måder at optimere kemikalieproduktion i cellen på. På CMB arbejdes især med ethanol-, antibiotika- og enzymproduktion.

En anden styrke ved disse modeller er, at de kan betragtes som et netværk af enzymer, reaktioner og kemiske stoffer og



Figur 3. Grafisk repræsentation af tarmbakterien *E. coli*'s kromosom vha. et såkaldt genomatlas. De koncentriske ringe viser forskellige fysiske egenskaber for DNA-strengen. De blå og røde felter i fjerde ring er placeringen af gener [7,9,12]. Lavet af Dave Ussery, CBS, DTU.

dermed en slags kort over cellens stofskifte. Det er derfor muligt at integrere viden fra DNA-chips, proteom- og metabolom-målinger i netværket, og på denne måde vise på systemniveau hvordan cellen reguleres [11].

Den nyere tids genomforskning har altså muliggjort systembiologien, som gennem sine metoder er et værdifuldt værktøj til at lære om cellen som biologisk reaktor. Især den matematiske modellering har vist sig at kunne give meget information om organismers potentiale som celle-fabrikker.

E-mail-adresse

Mikael Rørdam Andersen: mr@biocentrum.dtu.dk

Referencer:

1. Oxford Dictionary of English. Oxford University Press. 2005.
2. Borodina I, Krabben P, Nielsen J. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res.* 2005;15(6):820-829.
3. Cummings L, Riley L, Black L, Souvorov A, Resenchuk S, Dondoshansky I, Tatusova T. Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiol Lett.* 2002;216(2):133-138.
4. David H, Akesson M, Nielsen J. Reconstruction of the central carbon metabolism of *Aspergillus niger*. *Eur J Biochem.* 2003;270(21):4243-4253.
5. Forster J, Famili I, Fu P, Palsson BO, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 2003;13(2):244-253.
6. Fleischmann RD, et al. Whole-Genome Random Sequencing and Assembly of *Haemophilus Influenzae* Rd. *Science.* 1995;269(5223):496-512.
7. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature.* 2006;Jan.
8. Griffin TJ, Goodlett DR, Aebersold R. Advances in proteome analysis by mass spectrometry. *Current Opinion in Biotechnology.* 2001; 12(6):607-612.
9. Jensen LJ, Friis C, Ussery DW. Three views of microbial genomes. *Res. Microbiol.* 1999; 150:773-777.
10. Keilberg V, Rasmussen L. *Proteiner - oprensning og karakterisering.* Gads Forlag. 2001.
11. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA.* 2005;22;102(8):2685-2689.
12. Pedersen AG, Jensen LJ, Brunak S, Staerfeldt HH, Ussery DW. A DNA structural atlas for *Escherichia coli*. *J. Mol. Biol.* 2000; 299:907-930.
13. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, Guo H, et al. Global analysis of protein phosphorylation in yeast. *Nature.* 2005;438(1):679-684.
14. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
15. <http://www.cbs.dtu.dk/services/GenomeAtlas/>
16. <http://www.cbs.dtu.dk/>
17. <http://www.cbs.dtu.dk/databases/DOGS/index.php>
18. http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html

Eksisterer din produktion efter REACH?

Det ved vi på www.trotters.dk

trotters
- sikkerhed for mennesker og miljø!