

Eksperimentel identifikation af gener

***Bacillus subtilis*-forsøg viser vejen til eksperimentelle undersøgelser af hele genomer fra andre bakterier**

Af ph.d.-studerende Simon Rasmussen og lektor Hanne Jarmer, CBS, DTU

Hovedparten af de gener, som vi kender i dag, er fundet ud fra DNA-sekvenser ved brug af computerbaserede metoder. Om disse gener er virkelige og rent faktisk har en biologisk funktion, er kun fastslået i laboratoriet for en forsvindende lille del. Det er dog indenfor de senere år blevet muligt at aflæse hvor på en organismes DNA, der ligger aktive gener - og det endda i mange tilfælde i et enkelt laboratorieforsøg. Det er netop et sådan forsøg, vi har gennemført med bakterien *Bacillus subtilis*.

Hvad er *Bacillus subtilis* for en størrelse?

B. subtilis er en harmløs gram-positiv bakterie, som findes naturligt stort set alle steder. Den lever hovedsageligt som en epifyt, dvs. på planter - uden at gøre skade på disse. Det har vist sig, at *B. subtilis* er en vigtig bestanddel af den bakterieflora, som vokser på planter og er med til at forhindre invasion af bakterier, som kan være skadelige - helt på linje med funktionen af den menneskelige huds normale bakterieflora [1]. Den blev fundet og beskrevet første gang allerede i 1872 af den franske forsker, Ferdinand Julius Cohn (1828-1898), som arbejdede med klassifikation af bakterier. *B. subtilis* er kendt for at være enormt omstillingsberedt og kan endda danne sporer, hvorfor den også ofte findes steder, hvor den kun lejlighedsvis kan vokse, som f.eks. i udtørret jord. *B. subtilis*, anvendes som produktionsstamme - en levende fabrik - til industriel fremstilling af forskellige enzymer til f.eks. vaskepulver. Der har med tiden udviklet sig en del interesse for *B. subtilis* og andre af dens familiemedlemmer i forbindelse med udviklingen af probiotika. Det er et kosttilskud bestående af bakterier, der har vist en gavnlig effekt på både dyr og menneskers almene helbreds-tilstand [2]. Tilsætter man en blanding af sporer fra *Bacillus licheniformis* og *B. subtilis* i foderet til svin øges udbyttet og raten af infektioner nedsættes i samme størrelsesorden, som hvis dyrene modtog vækstfremmer (antibiotika), hvilket pga. den stigende udvikling af resistens bør holdes på et absolut minimum [3].

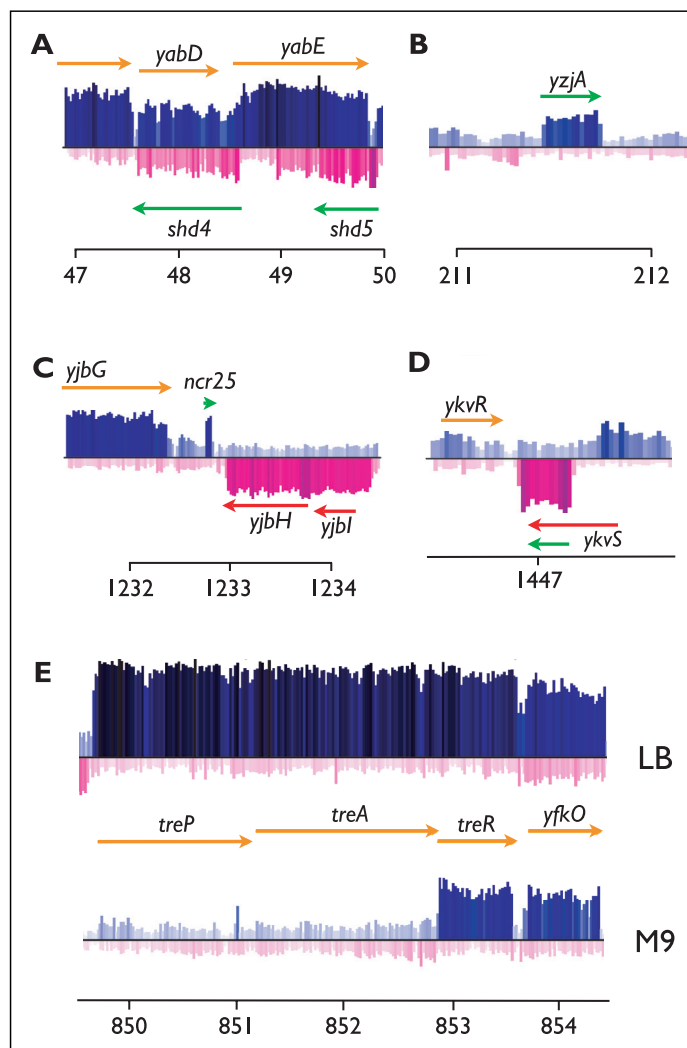
Hvad ved vi indtil nu?

Det bakterielle genom (DNA) er en særdeles kompakt struktur. Begge DNA-strengene er tæt pakkede med gener, hvoraf mange er organiseret tæt sammen i mindre enheder. Disse enheder kaldes operoner og er defineret ved, at generne aflæses samtidig. Med aflæsning forstås syntesen af RNA ud fra DNA. RNA er det primære produkt af et gen eller en operon (med flere gener), og det oversættes i de fleste tilfælde til protein ud fra det enkelte gens åbne læseramme. Den åbne læseramme er en speciel række af koder (tri-nukleotid kodons), som cellen oversætter til en bestemt række af aminosyrer, der er byggestenene i proteiner. Proteiner har utallige funktioner inde i cellen. De er enzymer, de transporterer stof ind og ud af cellen, de fremstiller energi, de giver cellen og dens organeller den rigtige struktur, osv. Der er dog en stigende erkendelse af, at RNA selv spiller en større rolle end hidtil antaget. Man har længe kendt til funktionen af de ribosomale RNA'er (rRNA) og transfer RNA'er (tRNA), der er hhv. del af ribosomet og nøglerne til oversættelsen af den

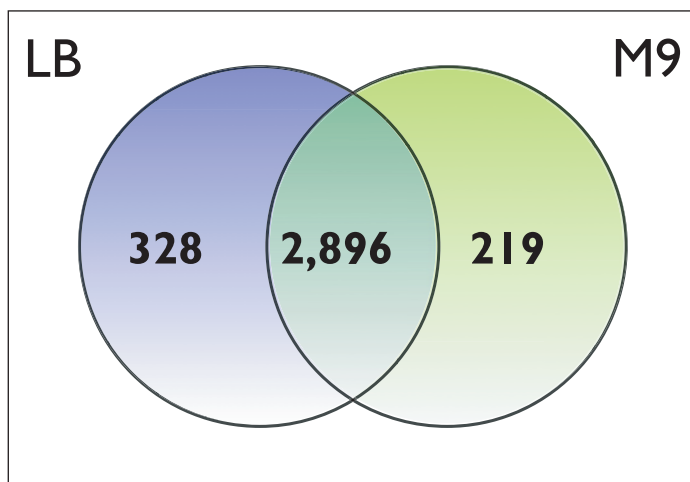
åbne læseramme. Ud over disse er der begyndt at vise sig en mangfoldighed af RNA-funktioner. Hovedårsagen, til at RNA-området har haltet bagefter ift. proteinerne, er, at disse ikke-kodende gener - RNA-gener - ikke er lette at finde i en sekvens, da de bl.a. ikke har en åben læseramme.

Nyopdagelser

Når man i et forsøg som vores aflæser, hvor på genomet, der



Figur 1. Aktiviteten af aflæsningen af DNA til RNA i forskellige regioner af *B. subtilis*' genom. Watson-strengen er blå, Crick-strengen magenta. Farvens intensitet og søjlens højde indikerer, hvor kraftigt det givne signal er. Gener vises med pile. Røde og orange er allerede annoterede gener, grønne er nyopdagede. (A) Udtrykket i området: 47-50 Kb, hvor der ses to eksempler på antisense transskription (*shd4* og *shd5*) af *yabE* og *yabD*. (B) 211-212 Kb-regionen, hvor det nyopdagede proteinkodende gen, *yzjA*, er fundet. (C) Området: 1231-1234 Kb, hvor det nye ikke-kodende gen, *ncr25*, er fundet. (D) Området omkring 1.447 Kb, hvoraf det fremgår, at *ykvS*-genet i virkeligheden ikke er så langt som antaget. (E) Aktiviteten i området: 850-854 KiloBaser (Kb) i hhv. rigt (Luria Broth, LB) og fattigt (Minimal medium 9, M9) medium.



Figur 2. Venn-diagram over antallet af gener, som findes udtrykt i LB- og M9-medier.

dannes RNA, så ser man både proteinkodende og ikke-kodende transskripter (*non-coding RNA*, ncRNA). I figur 1 ses en håndfuld udvalgte eksempler af det 4,2 Mega baser (Mb) store *B. subtilis*-genom. De proteinkodende, kendte gener er vist med røde og orange pile, alt efter hvilken af de to strenge de ligger på, og de nyopdagede gener er vist med grønne pile. I figuren ses også to eksempler på *skyggegener* (*shd4* og *shd5*), dvs. signalgivende områder, der ligger på den modsatte streng ift. eksisterende proteinkodende gener. Funktionen af disse er for størstedelens vedkommende stadig ukendt, men for nogle få er det kendt, at de ved at binde sig til genets RNA regulerer enten stabiliteten eller tilgængeligheden heraf. Det kan dog også

tænkes, at en del af disse skyggegener slet ingen funktion har, men bare er et produkt af et aktivt aflæsningskompleks, som indimellem springer til den modsatte streng ved en tilfældighed, og måske kan det simpelthen ikke betale sig for cellen at blokere denne fejlaflæsning.

I vores projekt har vi undersøgt, hvor der udtrykkes gener i både rigt og i fattigt medium. Som det fremgår af den nederste del af figur 1 og figur 2, er der forskel på, hvilke gener der udtrykkes i disse meget forskellige vækstvilkår. I hvert af de to medier ser vi lige omkring 70% af de kendte gener udtrykt, og hvis begge forsøg inddrages, ser vi næsten 82% af de 4.225 til dato beskrevne *B. subtilis*-gener. Vi identificerer 44 nye protein-kodende gener og 85 nye ncRNA-gener. Mange af de nye protein-kodende gener er meget korte - sammenlignet med de kendte gener. Deres funktion er selvfølgelig ukendt, men ved at sammenligne deres aminosyre-rækkefølge med kendte proteiner kan vi komme med kvalificerede gæt på, hvad deres funktion er. Der er gennemgående ikke overvældende høj lighed med allerede kendte proteiner, hvilket i kombination med den relative korte længde er hovedårsagen til at de ikke tidligere er blevet identificeret ved brug af *in silico*-metoder. Vi ser dog mulige funktioner indenfor bl.a.: Regulering af transskription, transport og bakteriofag-relaterede funktioner, men hovedsageligt ligner disse nye proteiner andre ukendte proteiner.

Hvorfor gør vi det her?

Dette projekt er kun en lille del af et større europæisk samarbejde, BaSysBio, der involverer 15 andre partnere fordelt ud over Europa og med en enkelt i Australien. I BaSysBio-projektet stræber vi efter eksperimentelt at finde alle gener, deres regulering og aktivitet under forskellige forhold, samtidig med

Patenter | Varemærker | Kommerciel rådgivning

„Ingen jubler hvis markedsandelen skrumper“

Patenter er en investering, der sikrer dig eneretten til din opfindelse og værner om din markedsposition. Beskyt din viden og optimér din konkurrenceevne. Den rette patentstrategi forvandler dine gode idéer til forretningsmæssige aktiver.

Chas. Hude A/S er blandt Danmarks førende virksomheder inden for rådgivning om Intellectual Property Rights (IPR). Vi beskytter, overvåger, håndhæver og forsvare immaterielle rettigheder for danske og udenlandske kunder. Chas. Hude blev grundlagt i 1896 og har i dag 90 medarbejdere. Mød vores kompetente profiler på www.chashude.dk

Steven Richard Kitchen,
European Patent Attorney

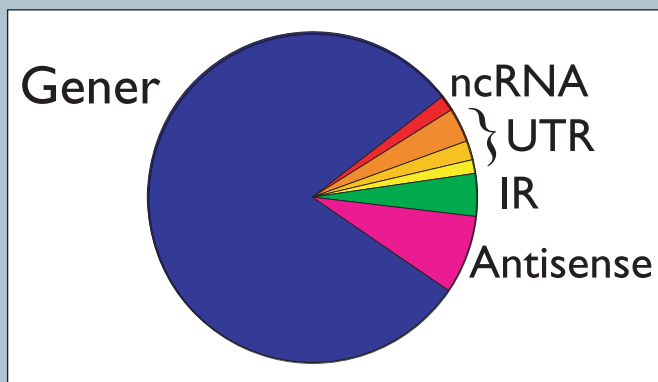
CHAS. HUDE

København | H. C. Andersens Boulevard 33 | 1780 København V | Telefon 33 19 34 00
Århus | Marselisborg Havnevej 36 | 8000 Århus C | Telefon 33 19 35 60 | www.chashude.dk

af koncentrationen af proteiner og metabolitter måles. Det overordnede mål er at gøre det muligt at fremstille modeller og dermed mulighed for at simulere cellens funktioner ved brug af en systembiologisk indgangsvinkel. Herved vil vi lære en masse nyt om bakterien - viden, som gradvist vil kunne overføres til sygdomsfremkaldende bakterier forholdsvis nært beslægtede med *B. subtilis*, såsom: *Bacillus anthracis* og *Staphylococcus aureus*. Vi planlægger at udføre forsøg, som det her beskrives med disse patogener.

Metoden

Den metode vi har anvendt i dette studium kaldes *tiling array*-analyse og bygger på kemi, som allerede er kendt fra DNA-chips/-microarrays. En glasplade er fyldt med bitte små områder, der hver især indeholder en stump enkelt-strengt DNA (en probe). Disse prober indeholder alle hver deres unikke sekvens, som hver er i stand til at binde en komplementær DNA-streng, *target*, med den modsatte sekvens. Så dannes den dobbeltstrengede helix-struktur, som DNA er kendt for. Hvis target er mærket med en fluorofor, synliggøres de små områder på glaspladen, der har "fanget" noget. I vores forsøg har vi fået firmaet *NimbleGen* til at fremstille en *tiling chip*, som vi kalder *BaSysBio-Bsub T1*. Proberne har vi designet, således at hele genomet er dækket - og så de er mellem 45 og 65 nukleotider (nt) lange, alt afhængigt af deres basesammensætning. Ved at lade længden variere opnås en mere ensartet smeltetemperatur, som resulterer i mindre variation indenfor et gen, der forventes at give nogenlunde samme signal fra start til slut. Probernes target-områder på genomet er placeret, så de starter 22 nt forskudt ift. hinanden - på begge strenge, hvilket vil sige, at hver probe overlapper med naboproben med mellem 23 og 43 nt og med naboens nabo 1-21 nt. Proberne er desuden

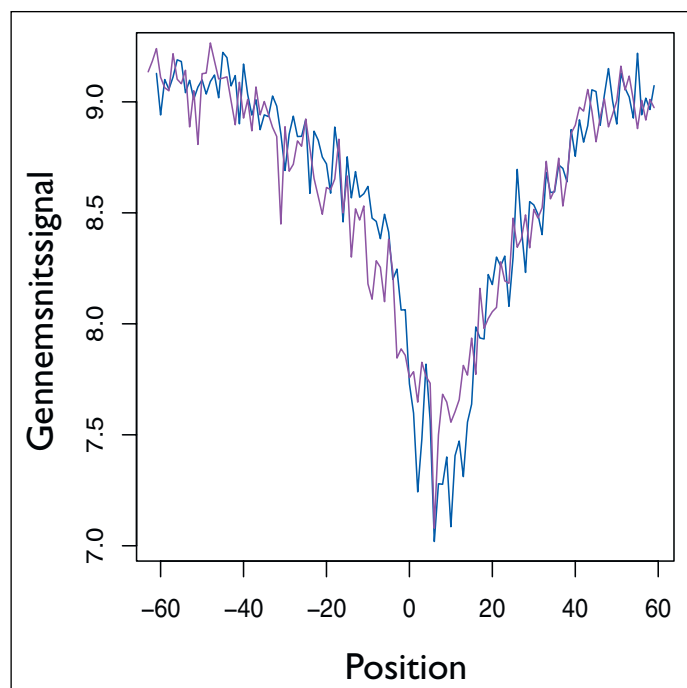


Figur 3. Lagkageplot over hvor de signalgivende prober er placeret. Kategorien "Gener" indeholder prober, som ligger indenfor de 4.105 proteinkodede gener, som på nuværende tidspunkt er beskrevet i *B. subtilis*. "ncRNA" indeholder de prober, som ligger i kendte ncRNA'er. "UTR" kategorien er de områder foran, imellem (i operoner) og bagved de proteinkodende gener, som ikke oversættes til protein, som de er fundet i vores studium. "IR" er kort for Intergenic Region og indeholder de prober med signal over baggrund, som ligger i områderne mellem generne. "Antisense" er skyggegenerne, som beskrevet i teksten.

forskudt strengene imellem, således at der er blot 11 nt mellem hver probestart. Dette udnyttes kun, hvis chippen benyttes til forsøg, som ikke er strengspecifikke. Vi har målt (hybridiseret) 3 uafhængige oprensninger fra vækst i både det rige og fattige medium, hvor ca. 40% af proberne har givet et signal højere end baggrunden. Langt størstedelen ligger i områder, hvor vi forventer et signal (figur 3). Derudover har vi 4 gange hybridiseret genomisk DNA til den samme slags chip.

Hvad ser vi, når vi undersøger det signal, som DNA'et giver?

Vi forventede, at resultatet ville hjælpe os med at vurdere, hvordan vores prober fungerede og herved normalisere vores data. Men vi fandt, at variationen i vores DNA-data kun øgede støjen i vores RNA-data, hvis vi normaliserede med disse. Dette blev konkluderet ud fra to test. I den ene målte vi, hvor stor en del af resultaterne, der var allerede kendte gener. I den anden undersøgtes hvor præcist, vi kunne forudsige starten af de transskripter, hvis start allerede er kendt. Resultaterne af det genomiske DNA viste til gengæld nogle ting, som vi ikke havde tænkt over. Eftersom alle områderne i genomet vil være til stede i omtrent samme mængde, ville man forvente et signal, som gennemgående var det samme over det hele. Men vi ser et par tusinde regioner af varierende længde, som giver et signal, der ligger langt under gennemsnittet. Mange af disse skyldes formentlig fejl i sekvensen, men påfaldende mange (~40%) falder sammen med både kendte og forudsagte Rho-uafhængige terminatorer. De er signal til det enzym, der oversætter DNA til RNA om at stoppe. Terminatorerne består af en meget stabil struktur - en stem-loop, som dannes i RNA'et - efterfulgt af en række Thymin (T - en af de 4 baser, som DNA er opbygget af), som tilsammen forårsager, at enzymet frigøres og dermed stopper transskriptionen. I figur 4 ses gennemsnitssignalet henover de Rho-uafhængige terminatorer. Fænomenet er ikke overraskende. Når der er grundlag for en stabil struktur, vil den givne probe folde med sig selv og derved forhindre videre hybridisering. Vi kan altså ud fra vores DNA-data finde regioner, hvor der enten er fejl i sekvensen eller basis for foldning. Vi har fundet 7 regioner, som ud fra vores data tilsyneladende er "haler" på gener (3' UnTranslated Regions, 3' UTRs), hvor der er sekvensbevaring og en stabil sekundær struktur. Funktionen af de gener, som har denne stabile struktur, er forbundet med membranen: Nogle er deciderede membranproteiner, andre eksporteres og atter andre er medvirkende til opbygning af membranen. Funktionen af denne 3' UTR kan derfor meget vel være at forhindre oversættelsen fra RNA til protein, indtil et bestemt signal er modtaget. Dette signal kan være at halen enten bliver klippet af eller bundet af et protein, som herved ophæver blokeringen.

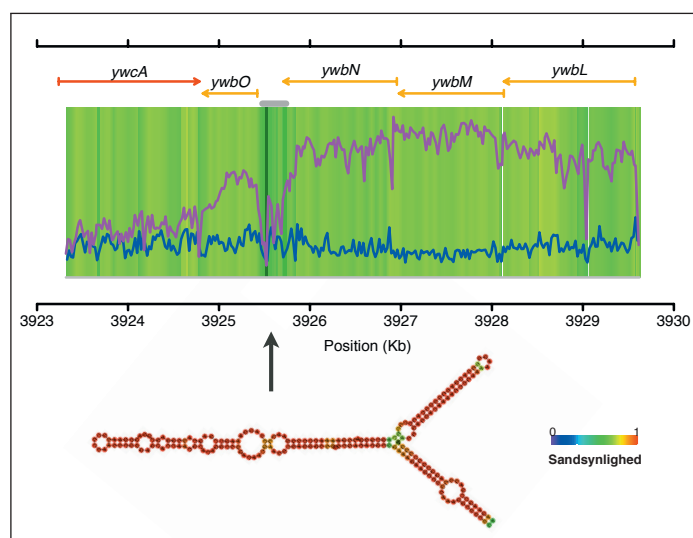


Figur 4. Signalet for genomisk DNA henover forudsagte Rho-uafhængige terminatorer. Den blå er gennemsnitssignalet fra Watson-strengen og magenta for Crick-strengen.

Mest sandsynlig er en mekanisme, udløst tæt på membranen. Herved kan det sikres, at proteinet ikke folder eller på anden vis modnes, før det er i den rigtige sammenhæng. Denne hypotese mangler dog stadig eksperimentel evidens.

Fremtidsperspektiver

Forsøgene er de første, der eksperimentelt viser, hvor på *B. subtilis*' genom, der udtrykkes gener. Vi har undersøgt dette i både et rigt og et fattigt medium, hvilket verificerer ca. 82% af de eksisterende gener og tilføjer næsten 130 nye. Vi forventer, at en lille del af de kendte gener ikke er rigtige gener, men også, at der vil være gener, som ikke kan findes under de her undersøgte omstændigheder. Der bør derfor, som nævnt tidligere, udføres flere tilsvarende studier, hvor alle mulige vækstbetingelser benyttes. I sidste ende vil det være muligt at bruge denne facitliste til at rette annoteringen af bakterien *Bacillus subtilis*.



Figur 5. Et eksempel på den velbevarede og strukturelle "gen-hale": Området omkring YwbN. Den øverste del af figuren viser udtrykket på Watson-strengen (blå) og Crick-strengen (magenta). Baggrunden viser signalet fra det genomiske DNA i samme område og varierer fra gul til grøn til sort, jo mørkere desto mere struktur (eller sekvensfejl) er der. Generne er vist som tidligere angivet. Den nederste del er den forudsagte sekundære struktur af "gen-halen" efter ywbN, som er bestemt ved brug af programmet RNAfold [4].

Projektet er støttet af EU, som et EU-IP-FP6-projekt (LSHG-CT2006-037469). Fremstilling og klargøring af RNA- og DNA-prøverne blev finansieret af Augustinus Fonden. Vi ønsker her at takke Pia Friis for hendes håndtering af det eksperimentelle arbejde.

E-mail-adresser:

Simon Rasmussen: simon@cbs.dtu.dk

Hanne Jarmer: hanne@cbs.dtu.dk

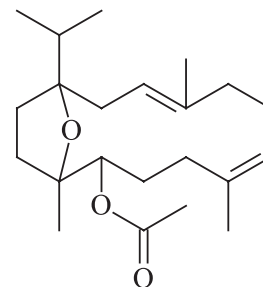
Referencer

1. Choudhary DK, Johri BN. Interactions of *Bacillus* spp. and plants - With special reference to induced systemic resistance (ISR). Microbiol Res. 2008 Oct 7.
2. Huang JM, La Ragione RM, Nunez A, Cutting SM. Immunostimulatory activity of *Bacillus* spores. FEMS Immunol Med Microbiol. 2008 Jul;53(2):195-203. Epub 2008 Apr 21.
3. Alexopoulos C, Georgoulakis IE, Tzivara A, Kritas SK, Siochu A, Kyriakis SC. Field evaluation of the efficacy of a probiotic containing *Bacillus licheniformis* and *Bacillus subtilis* spores, on the health status and performance of sows and their litters. J Anim Physiol Anim Nutr (Berl). 2004 Dec; 88(11-12): 381-92.
4. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA Websuite. Nucleic Acids Res. 2008

Nyt om...

.... Røgelse er godt for andet end sjælen

Røgelse er et centralt element i gudstjenesten i en række religioner, man træffer det såvel i katolske kirker som buddhistiske templer, hvor det begge steder er med til at skabe en speciel stemning. Man brænder forskellige harpikser fra træer, olibanum fra *Boswellia* arter og myrra fra *Commiphora* arter. Man har fundet tabletter til dette brug i Tutankhamons grav. Alkoholopløsning af sådanne harpikser er brugt til at præservere admiral Nelsons lig med.



incensolacetat

Forskere fra det hebræiske universitet i Jerusalem har undersøgt olibanum og fundet, at det indeholder et stof, incensolacetat, der har antidepressiv og angstdæmpende effekt på mus. Selvom dette ikke er bekræftet hos mennesker, mener forfatterne, at det er sandsynligt, at incensolacetat forøger den euforiske stemning, der opstår under religiøse handlinger.

Carl Th.

Culture and Chemistry meet in fragrant plant-based incense *Chemical and Engineering News* December 22, 2008 side 38; Incensol acetate, an incense component, elicits psychoactivity, *The Journal of the Federation of American Societies for Experimental Biology*, 2008, 22 side 3024.



**Få mailbesked,
når der er nyheder på
dit fagområde**

Tilmeld dig på www.techmedia.dk

TechMedia