



I SPALTEN

Skicka ditt genom som epostbilaga

Idag, när nya och snabbare DNA-sekvenseringstekniker står som spön i backen, börjar man allt oftare stöta på problem av ett mer prosaiskt slag, som att datafilerna man arbetar med blir för stora att hantera på ett effektivt sätt. Ett mänskligt genom består ju som bekant av cirka 3 miljarder baspar, egentligen 6 miljarder i sin diploida form, vilket ger mycket stora filstorlekar även efter komprimering. Forskare från University of California Irvine har nu utvecklat en uppsättning kompressionsmetoder som gör att man kan få ner storleken på ett mänskligt genom till den grad att det kan skickas som en e-postbilaga. Grundidén är egentligen självklar: man jämför genom som ska komprimeras med ett referensgenom och registrerar endast avvikelserna, som SNP:er och kopietalsvariationer. Med sin metod kan forskarna till exempel komprimera ner James Watsons genom till fyra megabyte. Programmet, DNazip, är skrivet i C++ och kan gratis laddas ned från <http://silver.ics.uci.edu/~dnazip/> tillsammans med exempelindata, som just Watsons genom. **MH**

Dyk längre ner i den djupa webben

Enligt vissa beräkningar är det mindre än 10 procent av allt material på webben som kan hittas med en Google-sökning. Resten ligger i det som kallas "the deep web", vilket består av till exempel sidor som är mindre populära och därför rankas lågt av Googles länkfrekvensbaserade metod, eller av material som ligger bakom "betalväggar" och alltså är tillgängligt för betalande kunder. Obskyra och specialiserade sajter kan dock innehålla information av värde för forskare. Företaget Deepdive lanserade nyligen en betaversion av sin webbtjänst som trålar igenom den djupa webben med hjälp av mönstermatchningsalgoritmer som ursprungligen utvecklades för DNA-sekvensanalys. Deepdive genomsköter exempelvis databaser över patent och kliniska prövningar. Life science har hittills varit företagets huvudfokus, men nyligen lades bland annat informations- och miljöteknologi till specialområdena. En skillnad jämfört med Google är att man kan klistra in hela stycken av text att söka på snarare än att använda enstaka termer. Google är begränsad till sökfraser på högst 32 ord. **MH**

iPhone-version av biologisk sökmotor

Nextbio har släppt en gratis iPhone-version av sin biologiska sökmotor för gener, sjukdomar och kemiska föreningar. Med hjälp av denna mobila mjukvara kan man alltså, precis som med webbversionen, söka efter till exempel PubMed-artiklar och experimentella genuttrycksdata samt korrelationer mellan olika geners uttryck. NextBio meddelar att intresset för applikationen varit stort sedan den släpptes under hösten, med upp till 2000 nedladdningar per vecka. **MH**

Simuleringsverktyg för biologer som inte kan stava

Cellucide är kanske ett tidigt exempel på hur nästa generation av biokemiska simuleringsverktyg kommer att se ut. Mycket eftertanke verkar ha lagts ner på utformning, användarvänlighet och att gömma komplexiteten från användaren.

Cellucide är den första produkten från Plectix, ett Bostonföretag som startades 2004 av bland annat systembiologer från Harvard. Det är ett verktyg för att bygga upp, lagra och sprida kunskap om cellulära signaleringsvägar samt simulera dem. Ett grafiskt symbolspråk för just signalering används för att rita upp signaleringsvägar för hand. Uppbyggandet av simuleringsmodellen sker bakom kulisserna utifrån det grafiska schema användaren ritat upp.

De idéer som ligger till grund för Plectix kommer från forskare som har tänkt mycket djupt runt biologisk modellering och hur modelleringsprocessen kan effektiviseras, säger Plectix vd Paul Edwards.

Istället för komplicerade system av differentialekvationer använder sig Cellucide av regler för interaktioner mellan biomolekyler. Dessa regler är uttryckta i ett formellt språk som kallas Kappa. Man behöver inte förstå Kappa för att kunna använda Cellucide.

Förutom att Cellucide erbjuder nya spännande möjligheter för erfarna modellerare kan det även användas av biologer som aldrig kunnat drömma om att skriva ner en differentialekvation, säger Paul Edwards.

Vi brukar skoja och säga att det kan användas av folk som inte ens kan stava till ordet modell.

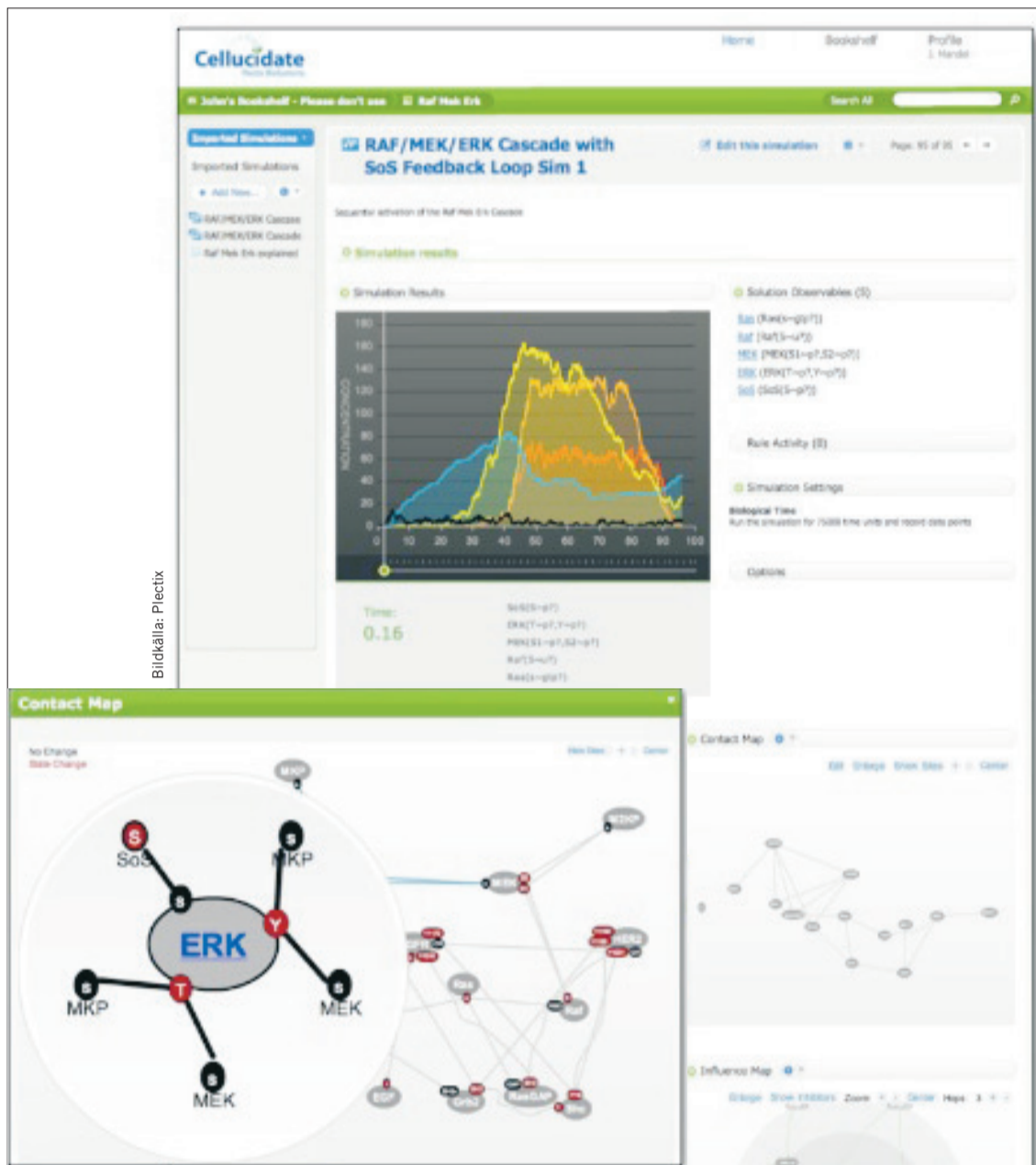
Cellucide är webbaserat och användarna kan arbeta med olika nivåer av publik läsbarhet i sina modeller. Dessa kan vara helt öppna, helt privata eller öppna för utvalda användare inom till exempel ett samarbetsprojekt mellan olika institutioner.

Aktiverar samlingar av fakta

Plectix använder begreppet "aktiva data" för att beskriva en aspekt av vad som är nytt i Cellucide. Eftersom systemet använder sig av en konsekvent formell representation, Kap-

Kappa

Kappa är ett formellt språk som beskriver interaktioner mellan biomolekyler i form av regler. I Cellucide handlar det om interaktioner mellan proteiner, men Kappa är en allmän formalism. Genom att använda en regelbaserad representation undviker man en rad problem med modeller baserade på differentialekvationer, som svårigheten att hitta rätt parametervärden och risken för kombinatoriska explosioner i antalet ekvationer i stora system.



Kamouflerad komplexitet. Användarvänliga Cellucide kan användas även av biologer som inte ens kan stava till äckvasjon.

pa-språket, är det lätt att lägga till nya interaktionsregler till de redan existerande i en modell, varefter intressanta effekter kan uppstå.

Regler är fakta om på vilka sätt molekylslag kan interagera med varandra. Dessa regler representeras med Kappa-notation och återfinns i grafisk form i Cellucide, säger Paul Edwards.

Den aktiva egenskapen vi talar om kommer från att man kan samla ihop en hög sådana här regler och bara trycka på "play": forskare kan snabbt och enkelt börja utforska de emergenta konsekvenserna av sina fakta och hypoteser. Du kan genast se effekterna av en ny regel genom en enkel "drag-and-drop"-manöver.

Att modellerna automatiskt blir körbara har också intressanta implikationer för publikationsmekanismer, påpekar Paul Edwards. Ett stort problem med dagens matematiska biologiska ämnen är att reproducera publicerade modeller.

Ett verktyg som Cellucide skulle kunna fungera som en plattform för exekverbara matematiska modeller tillhörande publicerade artiklar.

Facebook för proteiner

Vad är då Plectixs vision för Cellucide? Paul Edwards förklarar:

När mina barn frågar mig vad jag jobbar med, brukar jag säga att jag gör "Facebook för proteiner".

Analgin handlar inte bara om nätverk av proteininteraktioner: vi ser framför oss att Cellucide ska bli en miljö som användare kan köra sina egna verktyg i, precis som det finns en massa olika applikationer inuti Facebook som skapats av användare.

Baserat på Kappa kan användare skapa sina egna kraftfulla verktyg, kanske en simulationsmotor eller ett 3D-visualiseringsverktyg.

Cellucide är nu i betatestfasen och affärsmodellen är fortfarande under utveckling. Plectix tar just nu emot

feedback från testare och Cellucide är gratis under denna period.

Det finns också konkreta prov på plattformens effektivitet. En av Plectixs industriella samarbetspartners använder antikroppsteknik för att karakterisera skillnader i signalering mellan friska och transformerade onkogen cellinjer. Plectix lyckades i en simuleringsmodell reproducera dessa celler med tillräcklig precision för att möjliggöra en kausal tolkning av experimentella observationer. Därmed pekade man ut möjligheter att använda paneler av observationer som skulle kunna resultera i mer tillförlitliga diagnoser av cancer.

MIKAEL HUSS

Mer information Plectixs blogg (<http://blog.plectix.com/>) innehåller intressanta reflektioner om att simulera biologiska system i allmänhet och Cellucide i synnerhet. Mer om Kappa: <http://www.plectix.com/intro-kappa.htm>



Devex utvecklar ditt nya instrument

Vi hjälper dig genom hela, eller delar av, din produktutvecklingsfas; elektronik, mekanik, mjukvara, projektledning, från idé till serieproduktion. Vår kompetens gör skillnad för din framgång! Ring Lars Bredal-Hansen på tfn +46-735 195 236. www.devexmekatronik.se



English Translation

Simulation Tool For Biologists Who Cannot Spell

Cellucidate is perhaps one early example of how the next generation of biochemical simulation tools will look. Much reflection seems to have been done on the design and usability in order to hide complexity from the user.

Cellucidate is the first product from Plectix, a Boston business launched in 2004 by among others system biologists from Harvard. It is a tool to build, store and disseminate knowledge about cellular signaling pathways in order to simulate them. A graphical symbol language designed for signaling is used to draw the signaling pathways by hand. The construction of the simulation model happens behind the scenes from the graphical schematic that the user has drawn.

“The ideas behind Plectix come from researchers who have thought very deeply about biological modeling and how the modeling process can be made more effective,” says Plectix CEO, Paul Edwards.

Instead of complicated systems of differential equations, Cellucidate uses rules for interactions between biomolecules. These rules are expressed in a formal language called Kappa. You do not need to understand Kappa to use Cellucidate.

In addition, Cellucidate offers exciting new opportunities for the non-experienced modeler. “It can also be of use for biologists who never dream of writing down a differential equation,” says Paul Edwards. “We jokingly say that it can be used by people who cannot even spell the word model.”

Cellucidate is web-based and users can work with different levels of public accessibility in their models. They can be completely open, completely private or open to selected users, for example, in a collaboration project between different institutions.

Activates collections of facts

Plectix uses the concept "active data" to describe one aspect of what is novel about Cellucidate. As the system uses a standardized formal representation, Kappa language, it is easy to add new interaction rules to existing ones in a model, which can lead to interesting effects.

"Rules" are the facts about the ways in which biomolecules may interact with each other. “These "rules" are represented in Kappa notation and in graphical form in Cellucidate,” says Paul Edwards. “The "active" property we talk about comes from the possibility to gather a collection of rules and just press "play". Researchers can quickly and easily begin to explore the emergent consequences of their facts and hypotheses. You can instantly see the effects of a new rule by simply dragging and dropping”

“The fact that the models are executable also has interesting implications for publication mechanisms,” as Paul Edwards points out. A big problem with today’s mathematical biology is the difficulty of reproducing published models. A tool like Cellucidate could serve as a platform for executable mathematical models associated to published articles.

Facebook for proteins

What is Plectix’s vision for Cellucidate? Paul Edwards explains, “when my children ask me what kind of work I do, I say I am making a "Facebook for proteins". This analogy is not just about networks of protein interactions. We are envisioning Cellucidate as an environment where users can run their own tools, similar to the user-created applications in Facebook. Based on Kappa, users can create their own power tools, maybe a simulation engine or a 3D visualization tool.”

Cellucidate is currently in the beta testing phase and the business model is still under development. Plectix is receiving feedback from testers and Cellucidate is free during this period.

There is also concrete proof of the platform's effectiveness. One of Plectix's industrial partners is using antibody technology to characterize differences in signaling between healthy and transformed oncogenic cell lines. Plectix managed to reproduce these cells in a simulation model with sufficient precision to enable a causal interpretation of experimental observations. This pointed the way toward possibilities of using panels of observations, which could result in more reliable diagnosis of cancer.

Kappa

Kappa is a formal language that describes the interactions between biomolecules in the form of rules. The rules in Cellucidate describe interactions between proteins, but Kappa is a general formalism. By using a rule-based representation one avoids a number of problems with differential equation based models, such as the difficulty of finding the correct parameter values and the risk of a combinatorial explosion in the number of equations in large systems.

By Mikael Huss

For more information:

Plectix's blog (<http://blog.plectix.com/>) contains interesting thoughts on the simulating biological systems in general and Cellucidate in particular.

More about Kappa at <http://www.plectix.com/introkappa.htm>