

**EN DANSK STRATEGI
FOR FORSKNING OG FORSKERUDDANNELSE
INDEN FOR BIOINFORMATIK**

Udarbejdet på opfordring af Danmarks Grundforskningsfond

Trykt april 2002

Indholdsfortegnelse	side
1. Sammenfatning	3
2. Introduktion.....	4
3. Baggrund.....	4
4. Den post-genome fase: Behov for øget integration af bioinformatik i forskningen	6
5. Bioinformatik i videnssamfundet og som konkurrenceparameter.....	7
6. Behov for øget satsning på bioinformatikbaseret forskning i Danmark.....	8
7. Bioinformatik strategi.....	11
Detaljeret budget.....	14

En dansk strategi for forskning og forskeruddannelse inden for bioinformatik

Udarbejdet på opfordring af Danmarks Grundforskningsfond ved møde på Carlsberg Laboratorium den 30/10-2001 med oplæg af professor Søren Brunak, DTU, professor Matthias Mann, SDU og professor Jotun Hein, AU, vedrørende behovet for styrkelse af bioinformatik i Danmark

af

Ole Nørregaard Jensen, Syddansk Universitet <jenseno@bmb.sdu.dk>
Peter Roepstorff, Syddansk Universitet <roe@bmb.sdu.dk>

I samarbejde med en *ad hoc* nedsat bioinformatikarbejdsgruppe bestående af:

Søren Brunak, DTU	<brunak@genome.cbs.dtu.dk>
Freddy B. Christiansen, Århus universitet	<freddy@biology.au.dk>
Merete Fredholm, KVL	<Merete.Fredholm@ihh.kvl.dk>
Just Jensen, Danmarks JordbrugsForskning	<just.jensen@agrsci.dk>
Olaf Nielsen, Københavns Universitet	<onigen@biobase.dk>
Ole Skovgaard, Roskilde Universitetscenter	<olesk@ruc.dk>
Karen Welinder, Aalborg Universitet	<welinder@bio.auc.dk>
Børge Diderichsen, Novo-Nordisk	<bqd@novonordisk.com>

1. Sammenfatning

Det foreslås, at der i Danmark tages initiativ til en indsats indenfor bioinformatik for at styrke forskningen og det danske erhvervsliv gennem forbedret adgang til de bioinformatiske værktøjer og informationer.

Dette tænkes opnået gennem oprettelse af en forskerskole i bioinformatik, gennem styrkelse af bioinformatik miljøer på de enkelte forskningsinstitutioner, og gennem oprettelse af et nationalt center for bioinformatik.

Formålet er:

- at styrke forskningen i bioinformatik.
- at skabe adgang til bioinformatiske data for alle danske forskere gennem koordineret anskaffelse af bioinformatiske data og værktøjer.
- at forbedre muligheden for systematisk organisation af bioinformatiske data gennem definition af datastandarder samt udvikling, integration og vedligeholdelse af databaser og bioinformatik værktøjer.

Det foreslåede initiativ forudsætter tilførsel af 168,5 mio. kr. over fem år samt medfinansiering fra institutionerne på 35,5 mio. kr. i samme periode

2. Introduktion

Cellebiologisk, biomedicinsk og bioteknologisk forskning genererer i disse år store mængder information om levende organismer, inklusive mennesket. Denne detaljerede information om biologiske systemer deponeres i offentlige eller private databaser og er til stor nytte for akademiske og industrielle forskningsgrupper, nationalt såvel som internationalt.

Adgang til, koordinering af, og udnyttelse af denne biologiske information (bioinformatik) er af afgørende betydning for den natur- og lægevidenskabelige grundforskning og for den anvendelsesorienterede forsknings konkurrenceevne.

Formålet med dette oplæg er:

- (i) at gøre opmærksom på bioinformatikkens grundlæggende betydning for moderne biomedicinsk og bioteknologisk forskning i Danmark, samt
- (ii) at foreslå et initiativ der vil styrke Danmarks position og konkurrenceevne indenfor bioinformatik og dens anvendelser indenfor bioteknologisk og biomedicinsk forskning.

Dette kan opnås ved etablering af en national bioinformatikinfrastuktur, styrkelse af bioinformatikforskningen på de involverede institutioner, samt oprettelsen af en forskerskole i bioinformatik.

3. Baggrund

Netop i disse år befinder den biomedicinske og bioteknologiske forskning sig i en unik situation. Komplette genomer – hele den molekylære arvmasse – for en lang række levende organismer afdækkes ved hjælp af avancerede molekylære biologiske analyseteknikker.

Bedst kendt er de humane genomprojekter, der i 2001 udmundede i offentliggørelse af DNA sekvensen for størstedelen af den menneskelige arvmasse, om end i forskellige versioner. Genomer for mange bakterier, gærtyper, planter, fisk, og gnavere og andre pattedyr er enten allerede færdige, eller bliver det om ganske få år.

Indenfor videnskaben findes der næppe nogen anden form for data, der omfattes med så stor offentlig interesse. Det skyldes bl.a., at nye højeffektive teknikker til bestemmelse af DNA sekvenser har ført til store landvindinger indenfor molekylærbiologien og bioteknologien, og har fået stor betydning for behandling af sygdomme, herunder individualiserede sygdomsprognoser og diagnostik, samt for opklaringen af forbrydelser, fastlæggelse af familieforhold osv.

De mange genomprojekter medfører at nye DNA- og proteinsekvenser tårner sig op i de internationale databaser. Det samme gør information om sekvensernes funktion og afstamning, og ofte også om deres molekylære struktur.

Data, der kvantitativt beskriver hele cellers genekspression, protein-protein vekselvirkning og intercellulære kommunikation, er også i meget kraftig vækst. Den voldsomme vækst i datamængdernes omfang har radikalt ændret mulighederne for at udvikle seriøse computermetoder til klassifikation, forudsigelse og modellering af funktionen, strukturen og dynamikken af cellulære komponenter - fra det enkelte molekyle til den komplette organisme. Hermed kan man ikke bare udvikle nyttige metoder til anvendelse indenfor bioteknologi og farmakologi men også uddrage ny grundvidenskabelig viden om komplicerede biologiske mekanismer.

Det er interessant at notere sig, at væksten af i de biologiske datamængder svarer til en fordobling af databasernes størrelse ca. hver 10. måned. Denne vækst overstiger udviklingen af hurtigere computer-processorer, hvor hastigheden fordobles ca. hver 18. måned. En af fremtidens udfordringer og krav til bioinformatikken er derfor udvikling af nye meget effektive programmer og algoritmer til udforskning af disse store mængder biologisk information.

En af de største udfordringer for biologisk forskning er således at omsætte disse massive mængder af biologiske data til ny viden og erkendelse samt at udnytte denne indsigt i biomedicinske og bioteknologiske sammenhænge, f.eks. til udvikling af nye lægemidler og forbedring af miljø og fødevarer.

I denne sammenhæng spiller udvikling af nye bioinformatiske værktøjer en helt central rolle, ligesom det er meget væsentligt at sikre at de allerede eksisterende anvendes og udnyttes optimalt. Hertil er det ofte en fordel at opbygge multidisciplinære forskningsgrupper, der er i stand til at løse opgaverne kompetent både hvad angår de biomedicinske og bioteknologiske mål samt de datalogiske og systemvidenskabelige aspekter.

Indenfor bioinformatikken udvikles og anvendes avancerede computermetoder til DNA og protein sekvensanalyse og til fortolkning af kvantitative data, opnået ved f.eks. gen-ekspressionsanalyse. Bioinformatiske metoder kan i vid udstrækning identificere tilstedeværelsen af gener i DNA, forudsige hvilke proteiner de giver anledning til, og hvilken struktur og funktionalitet proteinerne får, og ikke mindst hvilke gener og proteiner, der vekselvirker med hinanden i en bestemt levende organisme.

Den optimale anvendelse af disse computerbaserede metoder til systematisk undersøgelse og udforskning af biologisk relevante data er for stor en opgave for den enkelte forsker. Der er derfor et udækket behov for en koordineret indsats og et integreret koncept, som kan understøtte den enkelte forskers maksimale udnyttelse af den eksisterende viden på området. Der er ligeledes et stort behov for en koordineret uddannelsesindsats indenfor bioinformatikken, således at såvel den akademiske forsknings som erhvervslivets behov for højt kvalificerede bioinformatikere kan blive tilgodeset.

4. Den post-genome fase: Behov for øget integration af bioinformatik i forskningen

Det gennemgående tema i den egentlige bioinformatikforskning er gennem computerstudier at skabe mulighed for at udnytte den store mængde af eksperimentelt bestemte data, som er lagret i databaser, til skaffe ny viden om de komplicerede mekanismer i den levende celle.

Dette omfatter f.eks. forudsigelse af hvilke proteiner en given organisme kan fremstille samt disse proteiners rumlige strukturer, deres funktion og vekselvirkninger med andre biomolekyler. Så længe analysen udelukkende baseres på computerberegninger betegnes den ofte "in silico biologi".

Vanskeligheden ved det bioinformatiske forskningsprojekt illustreres ved de noget afvigende resultater, som er opnået ved computeranalyse af det humane genom: Antallet af gener skønnes at være fra ca. 35.000 og op til 70.000, afhængigt af hvilke computerbaserede teknikker der anvendes til identifikation af gener i det humane genom.

Et stort antal forskere anvender desuden bioinformatikken i forbindelse med den eksperimentelle forskning. Disse forskere har et stort behov for adgang til biodatabaserne og de bioinformatiske værktøjer for derigennem hurtigt at kunne vurdere og analysere de opnåede data og på basis heraf planlægge videre eksperimenter.

Forskernes anvendelse af bioinformatikken skyldes således, at samspillet mellem eksperimentel forskning og bioinformatik tillader en langt mere rationel forsøgsplanlægning og tolkning af data end hidtil muligt og derved store ressourcemæssige besparelser. Bioinformatik har derfor i det sidste 10-år udviklet sig til at være et omdrejningspunkt for biologisk, bioteknologisk og medicinsk forskning, og bioinformatik er derfor et centralt redskab i hypoteseudvikling, projektdesign og dataanalyse.

Der findes nogle få bioinformatiske forskningscentre af høj international standard i Danmark, der udvikler nye metoder, f.eks. Center for Biologisk Sekvensanalyse ved DTU, som Danmarks Grundforskningsfond startede i 1993. Herudover eksisterer der et mindre antal forskergrupper og enkelt forskere som beskæftiger sig med denne disciplin. Forskningen søger blandt andet at modellere et stort antal meget komplicerede biologiske mekanismer, som f.eks. genekspression.

På en lang række felter befinder danske forskere sig i den biologiske frontlinieforskning. Dette gælder for eksempel deltagelsen i sekvensbestemmelsen af svinegenomet, som er et samarbejde mellem danske og kinesiske forskere.

Proteomforskningen, som er en direkte konsekvens af genomprojekterne, er et andet område, hvor danske forskergrupper regnes blandt de førende i verden, hvilket også afspejles af, at tre danske forskere er medlemmer af styregruppen for HUPO (the Human Proteome Organisation, som er dannet med henblik på

at nyttiggøre den store datamængde skabt ved sekvensbestemmelsen af det humane genom).

Koblingen mellem genomet, proteomet og metabolomet (mængden af små molekyler fremstillet i den levende organisme) er ligeledes et felt, hvor danske forskere har markeret sig klart i de seneste år. Endelig er der i Danmark en markant lægevidenskabelig forskning som omsætter informationer skabt på en række af disse områder til viden om sygdomme, deres årsag og bekæmpelse. Alle disse felter er helt afhængige af en effektiv adgang til biodatabaserne og de bioinformatiske værktøjer.

Bioinformatik har et meget stort potentiale for at generere nye opdagelser, der ofte har meget store kommercielle muligheder. Investering i bioinformatik betyder i realiteten, at man vil få betydeligt mere ud af de investeringer man gør i traditionel biologisk forskning.

Dansk bioteknologisk industri og medicinalindustrien er derfor helt afhængig af samarbejde med bioinformatikforskingsmiljøerne således, at der kan skabes nye metoder med konkurrencemæssige fordele. Ligeledes har industrien et behov for højt kvalificeret arbejdskraft indenfor dette område, hvilket forudsætter en øget uddannelsesindsats fokuseret på bioinformatik.

5. Bioinformatik i vidensamfundet og som konkurrenceparameter

Som nævnt ovenfor finder bioinformatiske forskningsresultater anvendelse i mange forskellige grundvidenskabelige, bioteknologiske og farmaceutiske sammenhænge indenfor naturvidenskabelig og medicinsk forskning.

Adgang til og udnyttelse af computermetoder til biologisk dataanalyse er en afgørende faktor, både når det drejer sig om at komme først med forskningsresultater i internationale tidsskrifter og i patenteringssammenhæng. Af denne årsag investeres meget betydelige ressourcer indenfor bioinformatik i stort set alle lande i Europa, i USA og i Japan.

I Sverige er der dannet adskillige bioinformatikcentre, der støttes af statslige midler og af store private fonde (40 mio. SEK årligt). Tilsvarende i Finland (7 mio. EURO årligt). I Tyskland er der initiativer både på nationalt og på delstatsniveau (73 mio. EURO over fem år). I England er alene den del der falder ind under E-Science initiativet på 16 mio. GBP over tre år.

Det er vanskeligt at opgøre den samlede indsats, men en nedre grænse udgøres af de midler der indeholdes i nyoprettede, dedikerede centre og programmer indenfor området.

6. Behov for øget satsning på bioinformatik-baseret forskning i Danmark.

Det nuværende store behov for en øget satsning på bioinformatikforskning og uddannelse i Danmark kan anskues på fire niveauer:

- (i) Behov for styrkelse af den nationale bioinformatik infrastruktur, dvs. bedre koordination og udnyttelse af data-ressourcer samt stimulering af samarbejder på tværs af institutioner;
- (ii) Behov for styrkelse af forskningsmiljøer der anvender bioinformatik;
- (iii) Behov for ressourcer til at anskaffe og/eller generere data i stor skala;
- (iv) Behov for uddannelse af kandidater og forskere (ph.d.er) indenfor bioinformatik og relaterede områder.

Ad behovet for styrkelse af den nationale bioinformatik infrastruktur

Bioinformatikken er måske det forskningsområde, der udvikler sig allerhurtigst indenfor videnskaberne. De computerværktøjer og programpakker, der blev udviklet for bare et eller to år siden, er ofte allerede udkonkurreret af bedre løsninger. Samtidig øges informationsmængden dagligt, som nævnt ovenfor.

Optimal udnyttelse af bioinformatikkens muligheder i moderne bioteknologisk og biomedicinsk forskning forudsætter derfor, at der altid er adgang til de nyeste metoder, de senest opdaterede offentlige eller kommercielle biologiske databaser, og de bedste computer-ressourcer i landet.

Det er en nærmest umulig opgave for den enkelte forskergruppe at holde sig informeret om de seneste udviklinger, samt at vedligeholde og opdatere store mængder data.

Ukoordineret anskaffelse, udvikling, distribution og vedligeholdelse af programpakker og databaser medfører på sigt et ressourcespild, idet forskellige forskergrupper anskaffer identiske (ofte dyre) programpakker eller databaser, eller forsøger at udvikle redundante dataanalyse værktøjer. I større tværfaglige forskningsprojekter er der behov for effektiv udveksling af data imellem de involverede laboratorier via Internettet.

For at sikre bredest mulig anvendelse og størst fleksibilitet er det helt nødvendigt at følge fastlagte datastandarder for udveksling af biologisk information. I forbindelse med etablering af nye typer databaser og programmer er det derfor fra starten strengt påkrævet at gennemtænke og definere datastrukturer, der på længere sigt sikrer at data kan anvendes i et bredere udvalg af forskningsmiljøer, nationalt såvel som internationalt.

Der er derfor behov for en national satsning, der sikrer at danske forskere altid er informeret om og har direkte adgang til de nyeste bioinformatiske teknikker og data-sæt, inklusive adgang til udvalgte kommercielle databaser. Optimal udnyttelse af de foreliggende bioinformatiske teknikker og ekspertiser forudsætter også øget, effektiv udveksling af information og data imellem de forskningsmiljøer, der udvikler og anvender bioinformatikken.

Definition og fastlæggelse af retningslinier og standarder for udveksling af biologiske datasæt via Internettet vil derfor være helt centralt for et sådant tiltag. Der er også øget behov for centrale ressourcer til detaljeret analyse af meget store biologiske datasæt samt til simulering af biologiske processer.

Mulighederne for med succes at bruge computeren som værktøj til forståelse og manipulation af biologiske makromolekyleres struktur og funktion vil afhænge af, at computerkraften vokser i takt med datamængderne.

Ad behovet for styrkelse af forskningsmiljøer der anvender bioinformatik

Et stadig stigende problem er, at etablering og anvendelse af bioinformatik i de enkelte forskergrupper ofte foregår med meget specifikke problemstillinger for øje og uden overordnet koordination.

De individuelle forskningsmiljøer har derfor brug for lokale 'sparringspartnere,' der kan hjælpe med at integrere bioinformatik i forskningsmiljøerne samt sikre effektiv udbredelse af den viden og ekspertise og de bioinformatikressourcer, der akkumuleres i de enkelte grupper. Disse sparringspartnere vil være et naturligt bindeled til det nationale center og sikre en optimal infrastruktur og udnyttelse af bioinformatik ressourcer nationalt såvel som internationalt.

Et effektivt og konkurrencedygtigt bioinformatikorienteret forskningsmiljø kræver kontinuert og fokuseret indsats over længere perioder. I øjeblikket udføres megen bioinformatik forskning i de eksperimentelt orienterede forskningsgrupper på danske universiteter af ph.d.-studerende eller af andre personer med tidsbegrænsede ansættelser (post-docs), hvilket ofte kan resultere i kortsigtede og uflexible *ad hoc* løsninger både m.h.t. udvikling, anvendelse og udbredelse af biologiske databaser og software.

Fastansættelse af kvalificerede bioinformatik forskere i de enkelte miljøer vil være en helt grundlæggende forudsætning for at opnå en effektiv integration, øget kvalitet samt en bredere anvendelse af bioinformatik og derved en øget konkurrenceevne for dansk bioteknologisk og biomedicinsk forskning.

Ad behovet for ressourcer til at anskaffe og/eller generere data i stor skala

Ressourcer til danske forskningsprojekter, der genererer store mængder data, allokeres oftest i form af "instrumenter" og personale. Men stort set al biologisk forskning bliver mere og mere afhængig af "information", og det vil i fremtiden blive almindeligt, at man køber denne information hos eksterne leverandører, der både kan være offentlige forskningsinstitutioner og private firmaer, i stedet for at generere den selv.

Hvis den danske akademiske forskning skal kunne drage fordel af denne udvikling, skal det gøres lettere at søge midler til anskaffelse af information, der ikke nødvendigvis involverer anskaffelse af eget apparatur. Sådanne muligheder kan både accelerere forskningen betydeligt, ligesom den kan forøge værdien af de data forskningsmiljøerne selv har skabt, ved samkøring med andre former for data hvorved der kan opnås store synergieffekter.

Ad behovet for uddannelse af kandidater og forskere (ph.d.er)

Behovet for kvalificerede kandidater og ph.d.er med kendskab til bioinformatik er stort indenfor den akademiske verden såvel som i industrien. Behovet vil stige dramatisk indenfor de næste få år. Denne kendsgerning er imødekommet fra flere universiteters side ved etablering af nye uddannelser indenfor området i løbet af de sidste par år.

Uddannelsesforløbene inddrager aspekter fra molekylær biologi og cellebiologi med emner indenfor anvendt matematik, statistik og datalogi og integrerer koncepterne fra alle disse fag i nyetablerede bioinformatik kurser. I øjeblikket tilbyder danske institutioner enten komplette formaliserede bioinformatik kandidat-forløb eller overbygnings/efteruddannelser, samt forskeruddannelse (ph.d. grad).

Bioinformatik uddannelserne er planlagt individuelt på de enkelte institutioner, og der findes i øjeblikket ikke en overordnet koordination af de forskellige undervisningsforløb imellem institutionerne.

For at opnå den bedste uddannelse af ph.d.er er der et klart behov for en bedre planlægning og udnyttelse af alle de kompetente undervisningsressourcer og forskningseksperter indenfor bioinformatik, der ligger på de forskellige uddannelsesinstitutioner i Danmark.

Der er samtidig behov for en målrettet indsats for at styrke rekrutteringen af velkvalificerede studerende til disse programmer via øget tilførsel af midler til finansiering af forskeruddannelses forløb (ph.d. stipendier) indenfor bioinformatik. Et øget antal permanente bioinformatik-stillinger (forskere m. undervisningsforpligtelse, se ovenfor) vil også øge kvaliteten af forskerstudieforløbene.

7. Bioinformatik strategi

Formål

At forbedre dansk forsknings muligheder for optimal udnyttelse af bioinformatikken gennem koordination af den danske bioinformatikindsats på følgende områder:

- Styrket forskning i bioinformatik
- Adgang til data for alle danske forskere gennem
 - Tilvejebringelse af de nyeste data via koordinerede anskaffelser/indkøb
 - Integration af de værktøjer/software pakker, som anskaffes eller udvikles af forskningsmiljøer, i det nationale netværk
- Organisation af data ved :
 - Definition af data standarder
 - Vedligeholdelse af data
 - Udbygning/opbygning af databaser
 - Integration af databaser og søgeværktøjer
- Forbedret integration af bioinformatik i biomedicinsk og bioteknologisk forskning
- Styrkelse af bioinformatik grundforskning på universiteterne
- Uddannelse af kandidater og forskere indenfor bioinformatik området

Virkemidler

- 1) Opbygning af en national bioinformatik infrastruktur via et nationalt center for bioinformatisk forskning og koordination samt styrkelse af bioinformatikforskning lokalt på institutionerne.
- 2) Styrkelse af forskeruddannelsen gennem oprettelse af en national forskerskole i bioinformatik

Ad 1. National bioinformatik infrastruktur

Da det er vanskeligt og ressourcekrævende for de enkelte forskergrupper at udvikle, vedligeholde og løbende opdatere omfangsrige og komplekse biologiske databaser og udnytte de nyeste bioinformatik værktøjer foreslås oprettelse af et nationalt center for bioinformatik ved at tilføre ressourcer til et af de eksisterende stærke bioinformatikmiljøer.

De specifikke opgaver for det Nationale Bioinformatik center vil være:

1. Forskning på internationalt niveau.
2. Udvikling og koordinering af Danmarks bioinformatik infrastruktur

3. Integration af bioinformatik værktøjer, f.eks. ved etablering af en dansk bioinformatik internet-portal
4. Etablering og vedligeholdelse af en national biologisk database ressource
5. Licensering af kommercielle databaser og bioinformatik ressourcer
6. Kontakt til internationale bioinformatik centre (f.eks. EBI, NCBI)
7. Standardisering af dataformater og dataanalyse værktøjer
8. At indgå som aktiv samarbejdspartner med de decentrale bioinformatikgrupper
9. At formidle kontakt mellem danske bioinformatikforskere
10. Koordination af forskeruddannelse

Anvendelse af bioinformatik vil i stor udstrækning finde sted i de forskningsmiljøer hvor de eksperimentelle data genereres og fortolkes, d.v.s.: i forskningsmiljøerne på de enkelte forskningsinstitutioner og virksomheder.

Disse miljøer bør derfor udbygges til at have tilstrækkelig styrke til at kunne udnytte bioinformatik i forskningen. Det foreslås derfor at der tilføres ressourcer i samfinansiering med institutionerne til styrkelse af den lokale bioinformatikkompetance.

Ad 2. Etablering af Bioinformatik Forskerskole

På størstedelen af de danske universiteter er der allerede etableret kurser i bioinformatik baseret på de lokale ressourcer. Det foreslås at der oprettes en national forskerskole i bioinformatik koordineret af det nationale bioinformatik center.

Denne forskerskole retter sig mod studerende indskrevet på de enkelte institutioner og tilbyder et bredt ph.d. kursusprogram udbudt af de deltagende institutioner. Ph.d. projekterne vil typisk være på tværs af institutionerne, f.eks. i samarbejde mellem lokale forskergrupper og det nationale bioinformatikcenter.

Et antal ph.d. stipendier, f.eks. 10 stk. pr. år, bør indgå i forskerskolen for at kunne sikre en tilstrækkelig rekruttering. Stipendierne fordeles mellem institutionerne af forskerskolens styrelse på basis af kvaliteten af ansøgere og de foreslåede projekter. Ph.d. studerende finansieret af lokale midler eller andre eksterne midler kan ligeledes optages på forskerskolen.

Budget for det foreslåede initiativ:

Styrkelse af bioinformatikmiljøerne og infrastrukturen:

Det foreslås at der oprettes 2 til 3 bioinformatikstillinger på hvert af institutionerne med henblik på lokal udvikling og support. Oprettelsen af disse stillinger forudsætter 50% lokal medfinansiering. Det nationale bioinformatikcenter forudsættes at tildeles 8 stillinger, heraf 4 VIP, 2 TAP og 2 A-TAP stillinger.

De 4 VIP stillinger forudsættes at have 50% forskningstid. Den institution som bliver vært for dette bidrager med fuld dækning af stillinger i forbindelse med den lokale/interne understøtning af bioinformatik. I forbindelse med etablering af bioinformatikmiljøerne vil en vis investering i computere og infrastruktur blive nødvendig, igen forudsættes 1:1 samfinansiering med institutionerne.

Forskerskolen:

Det forudsættes at der bevilges ca. 15 ph.d. stipendier pr år, således at forskerskolen efter tre år når en ligevægt med ca. 45 ph.d. studerende. Omkostninger til afholdelse af kurser afholdes af de enkelte institutioner. Til administration af forskerskolen etableres en styringsgruppe som betjenes af en deltidssekretær eventuelt delt med det nationale center.

Herved bliver den nødvendige tilførsel af ressourcer til den foreslåede styrkelse af bioinformatikken i Danmark:

Detaljeret budget i mio. kr.

<i>Nationalt center for Bioinformatik</i>			
	Stillingskategori	Bioinformatik program	Institutioner
Koordination	1 VIP + ½ TAP	0,6 + 0,15	-
System administration	1 VIP + 1 A-TAP	0,5 + 0,4	-
Database administration	1 VIP + 1 TAP	0,5 + 0,4	
Programudvikling	1 A-TAP	0,4	-
Koordination af forskeruddannelse	1 VIP	0,5	-
Kontakt til deltagere	½ TAP	0,15	-
Midler til indkøb af specielle datasæt	-	2,0	-
Årlige driftsudgifter	-	2,0	-
Subtotal pr. år		7,6	
<i>Lokale bioinformatik grupper</i>			
Personale	9+9 VIP	4,5 (9 VIP)	4,5 (9 VIP)
Driftsomkostninger		2,0	2,0
Subtotal pr. år		6,5	6,5
<i>Opstartsudgifter, år 1</i>			
Nationalt center		5,0	-
Institutioner	0.3 mio. kr./VIP	2,7	2,7
Subtotal, år 1		7,7	2,7

Forskerskole (15 ph.d. stipendier/år i 5 år) mio. kr.

		Bioinformatik program	Institutioner
År 1	15 ph.d. stipendier	6,0	-
År 2	+15 ph.d. stipendier	12,0	-
År 3	+15 ph.d. stipendier	18,0	-
År 4	+15 ph.d. stipendier	18,0	-
År 5	+15 ph.d. stipendier	18,0	-
Udløb, år 6+7		18,0	
I alt		90,0	

**Total budget (mio. kr.)
(for 5 år, samt udløb af ph.d. stipendier år 6 og 7)**

		Bioinformatik program	Medfinansiering fra institutioner	
År 1	Løn	8,1	4,5	
	Drift	6,0	2,0	
	Investering/opstart	7,7	2,7	
	Stipendier	6,0	-	
	Subtotal	27,8	9,2	
År 2	Løn	8,1	4,5	
	Drift	6,0	2,0	
	Stipendier	12,0	-	
	Subtotal	26,1	6,5	
År 3	Løn	8,1	4,5	
	Drift	6,0	2,0	
	Stipendier	18,0	-	
	Subtotal	32,1	6,5	
<u>I alt for 5 år</u> Inkl. Udløb af stipendier		<u>168,2</u>	<u>35,2</u>	<u>203,4</u>