

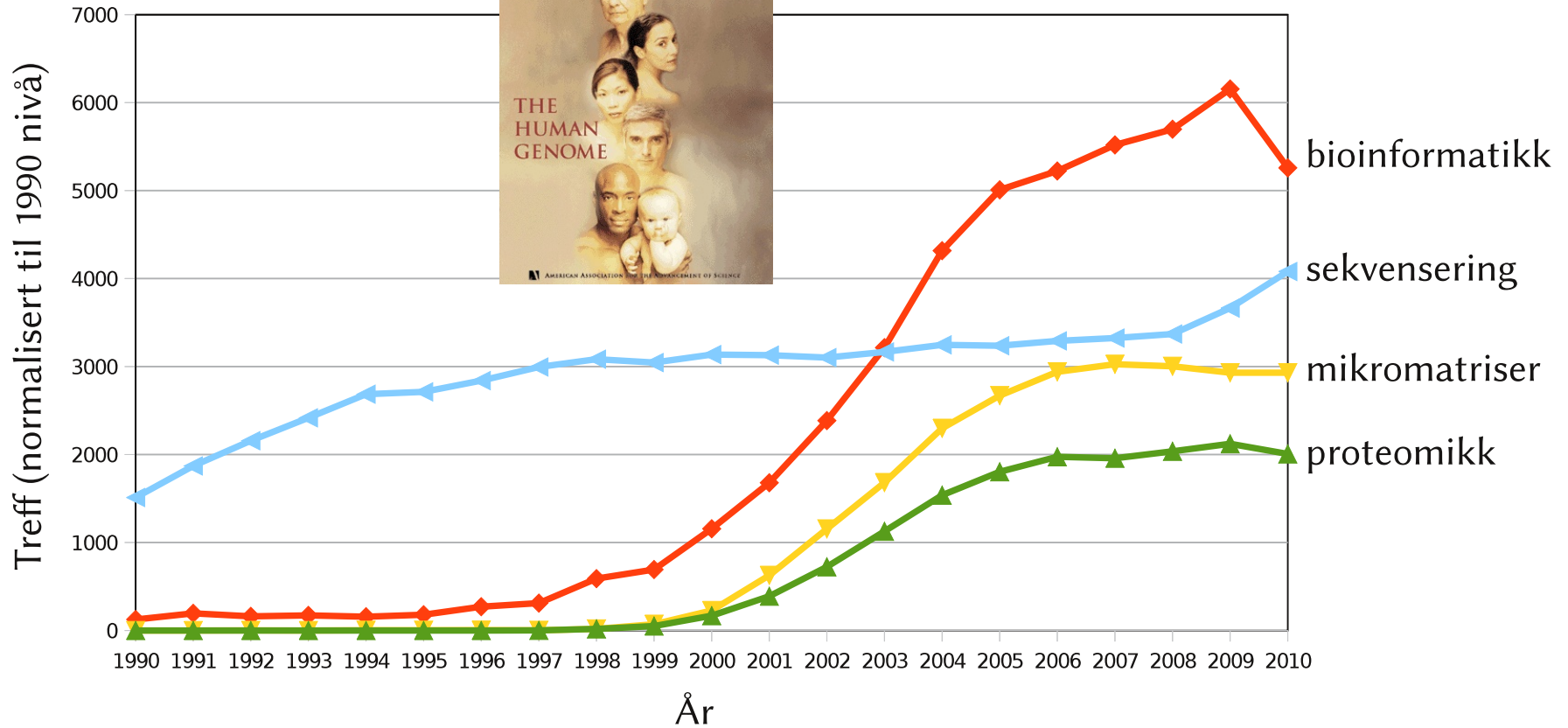
# Bioteknologi og bioinformatikk

Hvordan håndtere store mengder data og komplekse analyser?

Pål Puntervoll

“Fremveksten av høykapasitetsteknologier som produserer store mengder data har gjort at behovet for bioinformatikk og matematiske verktøy for å lagre, systematisere og modellere data fra ulike metoder og kilder har økt betydelig.”

# Teknologiutvikling (antall publikasjoner per år)



# Vekst i databaser og bioinformatiske verktøy

NAR Databaseutgave 2011:

- 96 nye
- 86 oppdaterte

NAR databasesamling:

- 1330 nøyre utvalgte databaser



# Vekst i databaser og bioinformatiske verktøy

NAR Databaseutgave 2011:

- 96 nye
- 86 oppdaterte

NAR databasesamling:

- 1330 nøyre utvalgte databaser

NAR Verktøyutgave 2010:

- 115 nye
- 7 oppdaterte

NAR verktøyoversikt:

- ca 1500 verktøy



“Fremveksten av høykapasitetsteknologier som produserer store mengder data har gjort at behovet for bioinformatikk og matematiske verktøy for å lagre, systematisere og modellere data fra ulike metoder og kilder har økt betydelig.”

“Kraftfull bioinformatikkompetanse og infrastruktur er en helt essensiell komponent for å nå målene.”

# eSysbio

## e-infrastruktur for biovitenskap

Samarbeid

Integrering av verktøy og data

Generisk

Open kildekode

Distribuert arkitektur

Standard teknologier

<http://esysbio.org>

The screenshot shows the eSysbio v0.7 web application interface. The page has a navigation bar with tabs for Dashboard, Projects, Datasets, and Tools. The main content area is titled 'Overview' and contains several sections:

- Messages:** You have 1 unread messages in your Inbox.
- Projects:** You have access to 4 eSysbio projects.
- Datasets:** You have access to 25 datasets.
- Tools:** You have access to 13 tools.

Below these sections is a link to read more on <http://www.ll.uib.no/trac/eSysBio/wiki/eSysBioWiki/FP/Documentation/eSysBioTutorial/#Datasets>.

The 'Status of your last 10 tool launches' section contains a table with the following data:

Type	Group	Name	Description	Start time	End time
●	-	fetchBatsc ...	2ec78839-8759-4a8b-b785-99 ...	17/07, 21:55	17/07, 21:55
●	maff@ebi-annotated	getResult	8c6a10e1-73f4-48fc-963f-42 ...	17/07, 16:51	17/07, 16:51
●	getResult	getResult	8c6a10e1-73f4-48fc-963f-42 ...	17/07, 16:51	17/07, 16:51
●	-	run	8c6a10e1-73f4-48fc-963f-42 ...	17/07, 16:51	17/07, 16:51
●	-	fetchBatsc ...	ffc740ef-e53d-40b2-880b-bd ...	17/07, 16:50	17/07, 16:50
●	-	Transform	fac9ecaa-1efd-4c7d-8edb-a1 ...	17/07, 16:46	17/07, 16:46
●	BlastpService	GetResult	6f3e7f0d-f9f4cc9-a804-86 ...	17/07, 16:44	17/07, 16:44
●	GetResult	GetResult	6f3e7f0d-f9f4cc9-a804-86 ...	17/07, 16:44	17/07, 16:44
●	GetResult	GetResult	6f3e7f0d-f9f4cc9-a804-86 ...	17/07, 16:44	17/07, 16:44

At the bottom of the table, there is a link: Go to [my activity](#) for complete information.

# Aktiviteter i eSysbio

## Lagre og dele data

- alle typer data
- store mengder data (StoreBioInfo)
- data tilhører prosjekter

## Verktøy

- "Web services" kan enkelt legges til og umiddelbart brukes
- Støtte for statistiske analyser: R

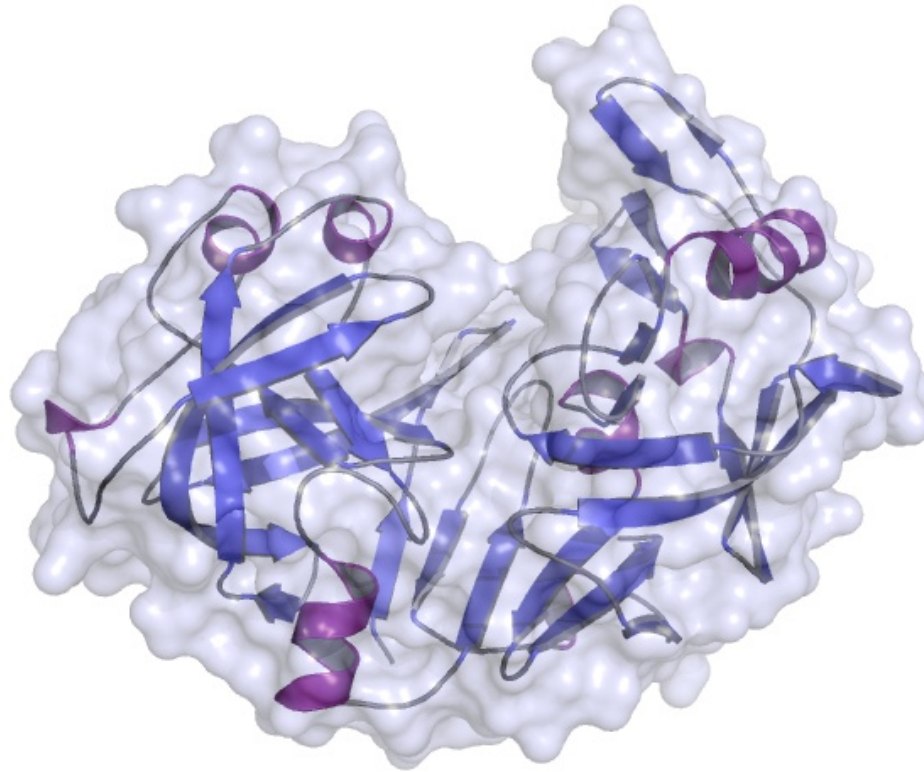
## Analyse

- Steg-for-steg analyse
- Automatiserte arbeidsflyter
- Analysehistorikk

# Bioteknologisk anvendelse

## Protein

molekylære  
maskiner



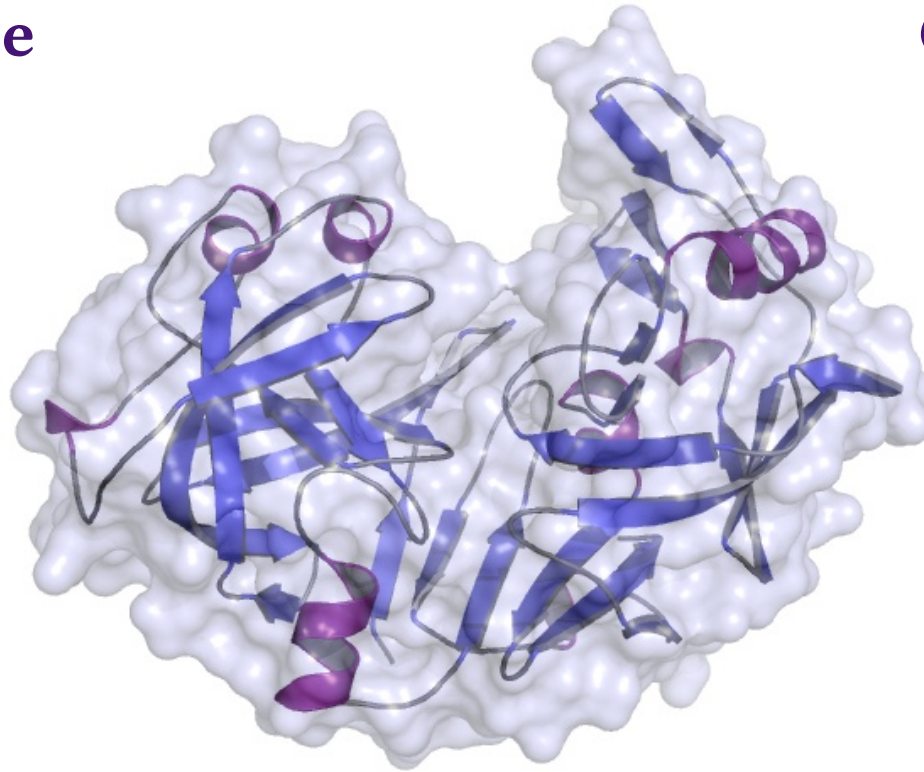
# Bioteknologisk anvendelse

## Protein

molekylære  
maskiner

Oppdage nye

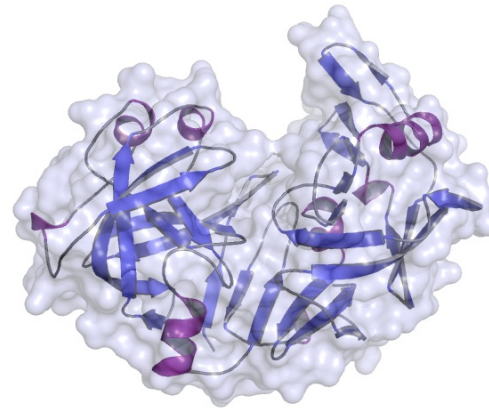
Optimalisere



# Optimalisere proteiner

sekvens → struktur → funksjon

```
MKWLLLLLGLVALSECIMYKVPLIRKKSLLR  
TLSEERGLLKDFLKKHNLNPKARYFPQWEAP  
TLVDEQPLENYLDMEYFGTIGIGTPAQDFT  
VVFDTGSSNLWVPSVYCSSLACTNHNRFNP  
EDSSTYQSTSETVSITYGTGSMTGILGYDT  
VQVGGISDTNQIFGLSETEPGSFLYYAPFD  
GILGLAYPSISSSGATPVFDNIWNQGLVSQ  
DLFSVYLSADDQSGSVVIFGGIDSSYYTGS  
LNWVPVTVEGYWQITVDSITMNGEAIACAE  
GCQAIVDGTSLLTGPTSPIANIQSDIGAS  
ENSDGDMVVSCSAISSLPDIVFTINGVQYP  
VPPSAYILQSEGSCISGFQGMNLPTEGSEL  
WILGDVFIHQYFTVFDRANNQVGLAPVA
```

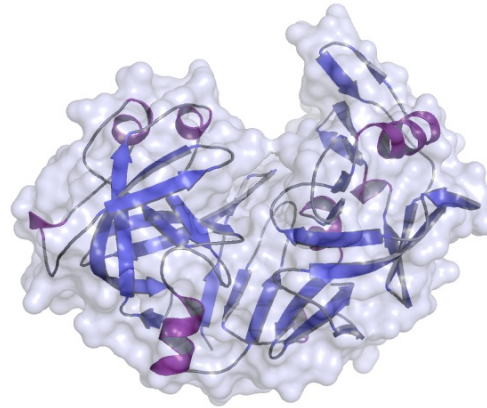


Protease

# Optimalisere proteiner

sekvens → struktur → funksjon

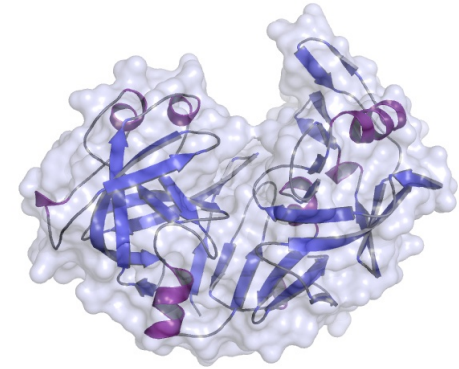
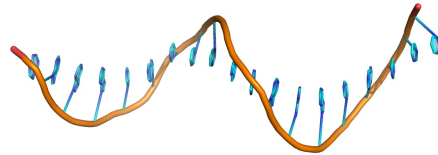
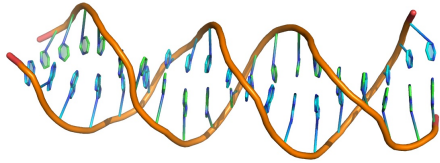
```
MKWLLLLLGLVALSECIMYKVPLIRKKSLLRR  
TLSEERGLLKDFLKKHNLNPKARYFPQWEAP  
TLVDEQPLENYLDMEYFGTIGIGTPAQDFT  
VVFDTGSSNLWVPSVYCSSLACTNHNRFP  
EDSSTYQSTSETVSITYGTGSMTGILGYDT  
VQVGGISDTNQIFGLSETEPGSFLYYAPFD  
GILGLAYPSISSSGATPVFDNIWNQGLVSQ  
DLFSVYLSADDQSGSVVIFGGIDSSYYTGS  
LNWVPVTVEGYWQITVDSITMNGEAIACAE  
GCQAIVDGTSLLTGPTSPIANIQSDIGAS  
ENSDGDMVVSCSAISSLPDIVFTINGVQYP  
VPPSAYILQSEGSCISGFQGMNLPTEGEL  
WILGDVFIHQYFTVFDRANNQVGLAPVA
```



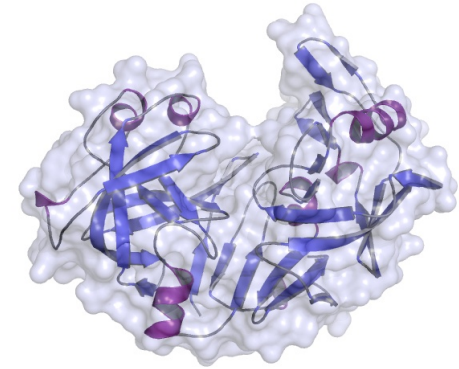
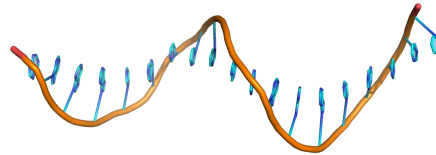
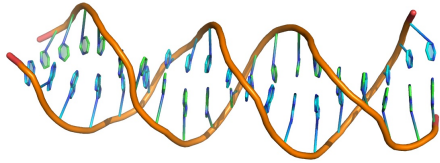
Protease

Pipeline for  
strukturprediksjon

# Oppdage nye proteiner



# Oppdage nye proteiner



**Pipeline for  
sekvenssøk**

# eSysbio gir fleksibilitet

## Tilpassede analyser:

- Nye verktøy
- Alternative verktøy

# eSysbio teamet



## **Programmerere**

Håkon Sagehaug, Prabakar Venkataraman,  
Armin Töpfer

## **PhD studenter**

Matuš Kalaš, Anne-Kristin Stavrum,  
Paweł Sztromwasser

## **Postdoc stipendiater**

Michael Dondrup, Sattanathan Subramanian,  
Siv Midtun Hollup, Francisco Roque

## **Seniorforskere**

Inge Jonassen, Kjell Petersen, Pål Puntervoll

## **Finansiering**



The Research Council  
of Norway



Universitetet i Bergen

