

# Att skilja det experimentellt verifierade från gissningarna

## FetchProt

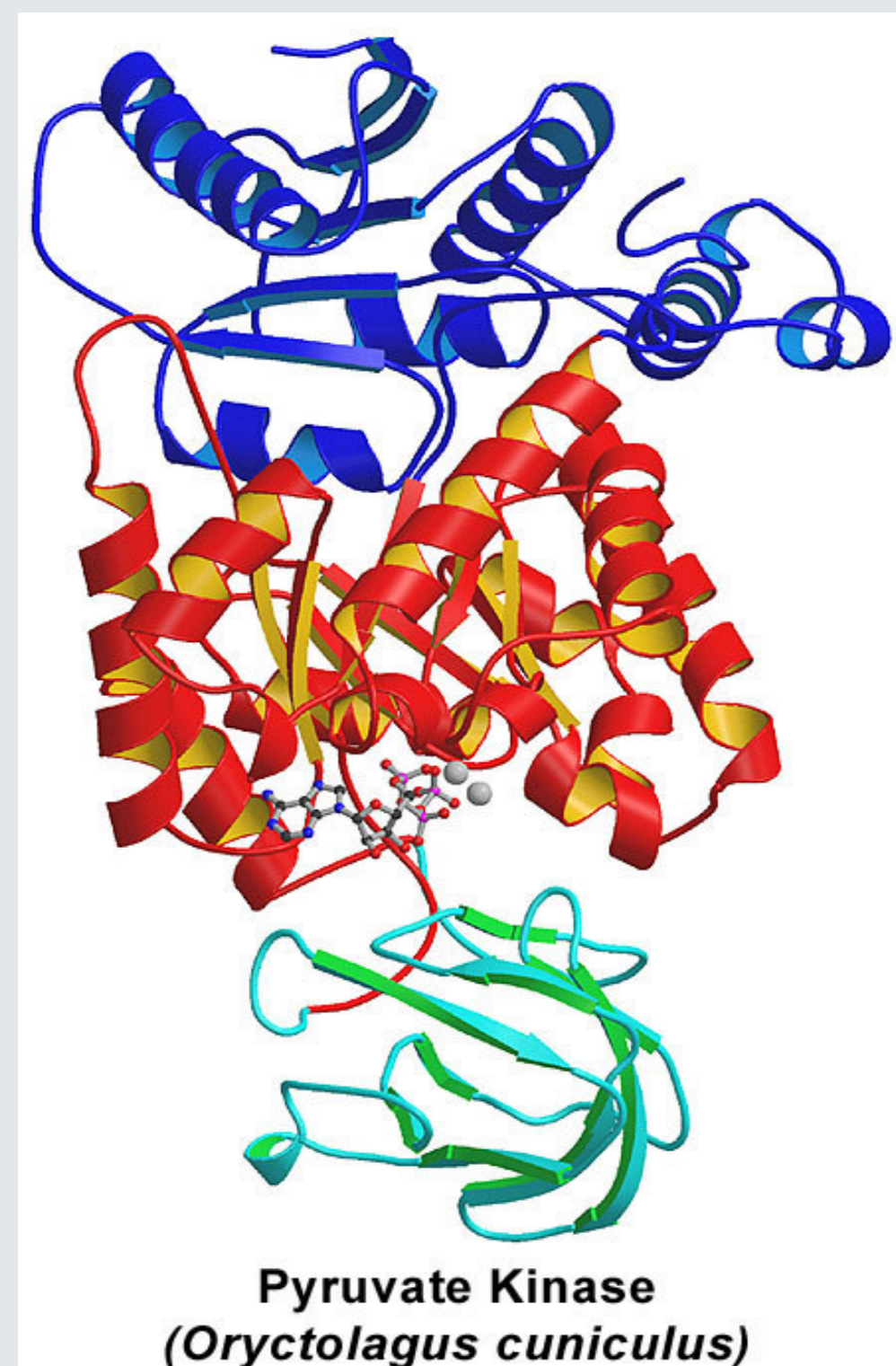
Inom projektet FetchProt samsas expertkunskap från bl. a. områdena bioinformatik, biokemi, datorlingvistik och informationsinfrastruktur för att automatiskt finna och sammanställa information om experimentellt verifierade funktioner hos proteiner, så att den blir lättillgänglig för de forskare som behöver den.

## Proteiners funktioner

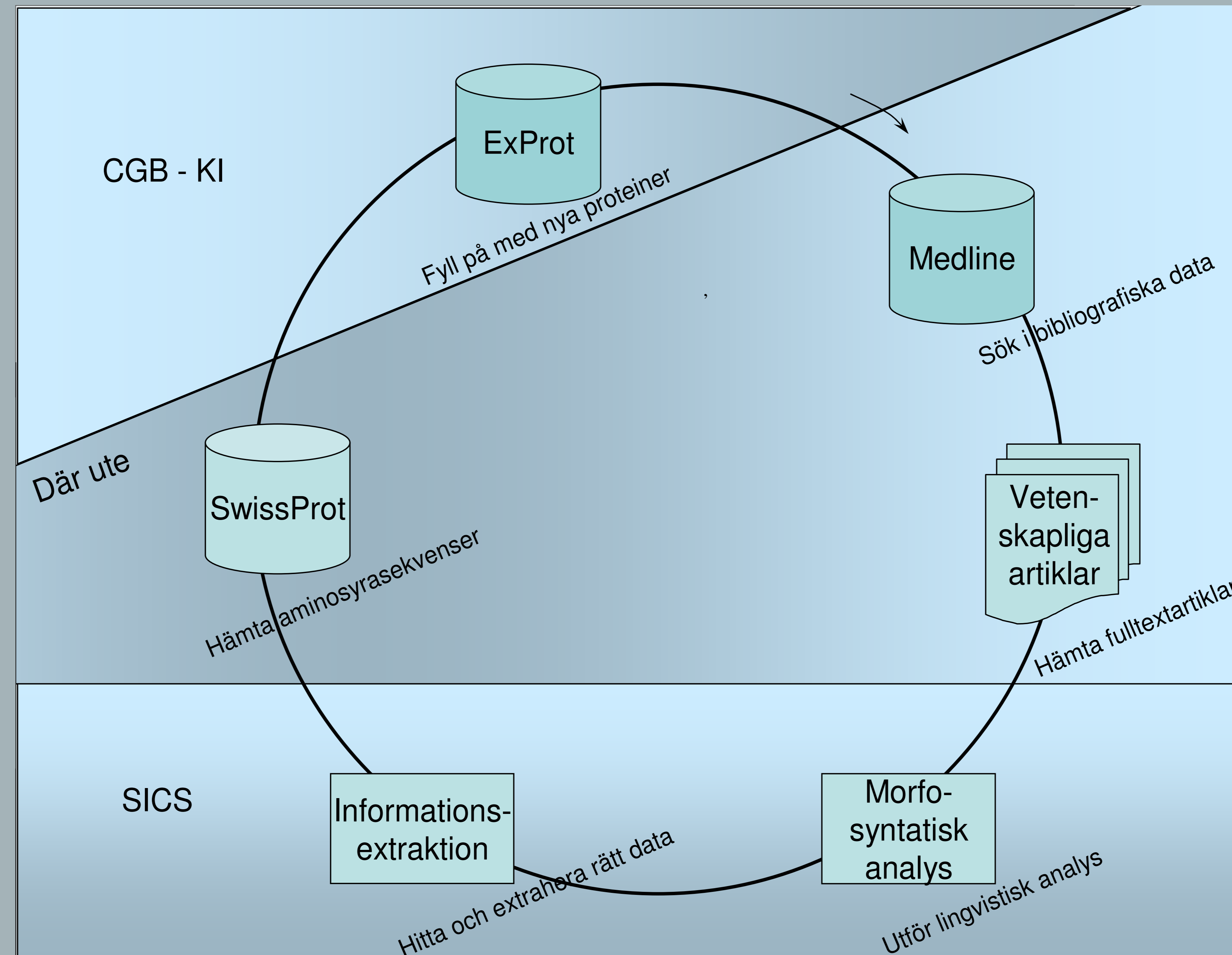
- Endast en del funktioner är experimentellt verifierade
- Övriga gissas genom analogi med strukturellt liknande proteiner
- Risken för fel ökar med avståndet

- Information om proteiners funktioner finns inte samlad och strukturerad
- Kunskapen är spridd framför allt över tusentals vetenskapliga artiklar.

[www.sics.se/humle/projects/fetchprot](http://www.sics.se/humle/projects/fetchprot)



## Systemskiss



## Metod

SOAP-baserat distribuerat system.

Informationsextraktions-system där vi prövar både handskrivna regelbaserade och maskininlärningsbaserade metoder.

Iterativ utveckling, från det specifika till det generella.

Proteiners olika typer av funktioner påvisas med olika typer av experiment.

Vi börjar med tyrosinkinaser.

## Fylla mallar

Informationsextraktionssystemet fyller mallar på dokument- och proteinnivå.

```
<Doc>
<DocID>PMID10369665</DocID>
<Protein>
<ProteinName>tyrosine-protein kinase ETK</ProteinName>
<Organism>Escherichia Coli</Organism>
<SequenceReference>EMBL;U00096</SequenceReference>
<VerifyingExpType>phosphorylation assay</VerifyingExpType>
<EvidenceSection>...</EvidenceSection>
</Protein>
</Doc>
```

## ExProt

EXProt är en databas som enbart har proteiner med en experimentellt verifierad funktion. Det är en sekundär databas med information tagen från andra databaser. Begränsningen ligger i bristen på tillgänglig strukturerad information. Ett av målen i FetchProt-projektet är att fylla på EXProt automatiskt eller semiautomatiskt.

FetchProt är ett samarbetsprojekt mellan:

Swedish Institute of Computer Science (SICS)  
*Kristofer Franzén, Jussi Karlgren*  
Språkteknologi, informationsextraktion  
Erfarenheter från "Proteinhalt i text"  
Proteinnamnsgenkännaren Yapex

Centrum för Genomik och Bioinformatik vid Karolinska Institutet (CGB)  
*Björn Ursing, Kwabena J. Sarfo*  
Biokemi, molekylärbiologi, bioinformatik.

Metamatrix AB  
*Patrik Hassel, Pär Lannerö*  
Infrastruktur, agentprogrammering, databaser

FetchProt delfinansieras av VINNOVA under programmet Infrastruktur för kunskaps-hantering.

## Annorlunda texttyp

Autophosphorylation was performed by adding 5 µl of purified Etk (~1 µg of Etk/µl) into 190 µl of reaction mix (150 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 7.4). The reaction was started by adding 5 µl of ATP (final concentration of 25 µM ATP, and 10 µCi of [ -32P]ATP). The reaction was stopped by removing 20 µl aliquots at different time points into tubes containing 5 µl of 500 mM EDTA. Then 5 µl of each aliquot was used for analysis by SDS-PAGE. Phosphorylation of poly(Glu:Tyr) (4:1, Sigma) was carried out in 40 µl containing 10 mM MgCl<sub>2</sub>, 20 mM Tris-HCl pH 7.5, 20 µM ATP, 1 µCi of [ -32P]ATP, 10 µl of suspended Talon beads with bound Etk (~1 µg of Etk/µl) and poly(Glu:Tyr) at the indicated concentration. The reaction was stopped by short centrifugation and application of 25 µl of the Etk-free supernatant onto Whatman 3MM paper which was then washed three times with 10% trichloroacetic acid and once with ethanol, dried and used to count radioactivity. In all experiments, we included control with no enzyme (Etk) to exclude non-specific association of ATP with poly(Glu:Tyr). To ensure that poly(Glu:Tyr) is the substrate, we included control without exogenous substrate [i.e. poly(Glu:Tyr)]. In this case, only low levels of labelling were detected (~400 c.p.m.) which represent low levels of Etk leakage from the Talon beads. As an additional control, we analysed the reaction by SDS-PAGE and autoradiography. The phosphorylated poly(Glu:Tyr) appears as a distinct smear 20-60 kDa in size.