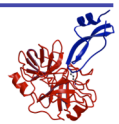
Genome annotation

Where, what and how

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

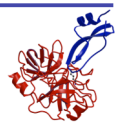


Content

- Annotation - definition and strategy
- Gene prediction (Where)
- Assigning function (What)
- Cellular processes (How)
- Annotation tools

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway



Annotation - Definition

The process of analysing a DNA sequence and adding biological information to it



DNA - Only linear sequence information

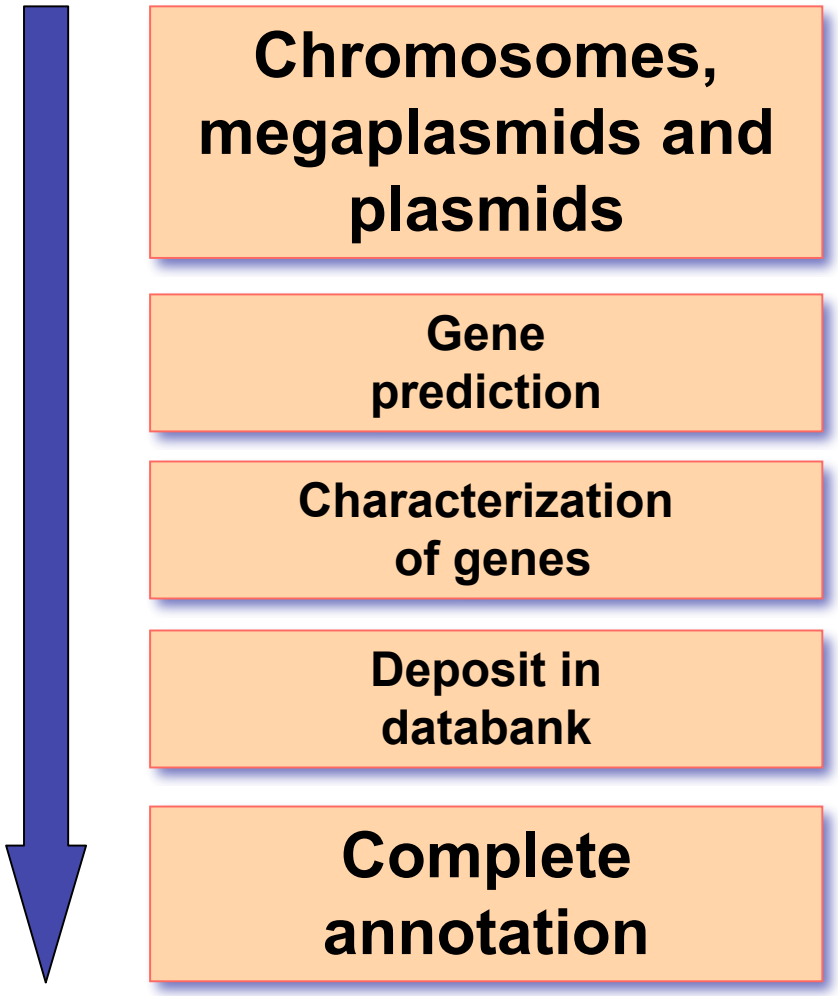
Bioinformatics Sequence analysis

Genes and encoded products with assigned functions

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Strategy for annotating a genome

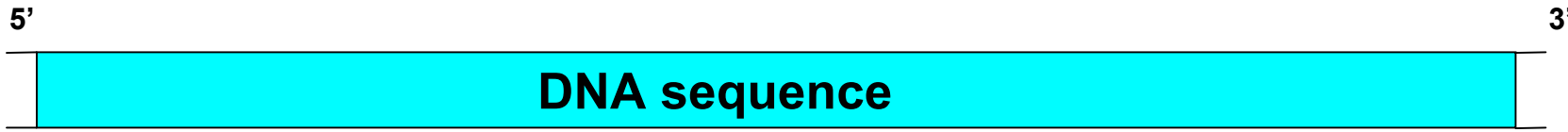


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - prokaryotic

Sequence analysis at nucleotide level

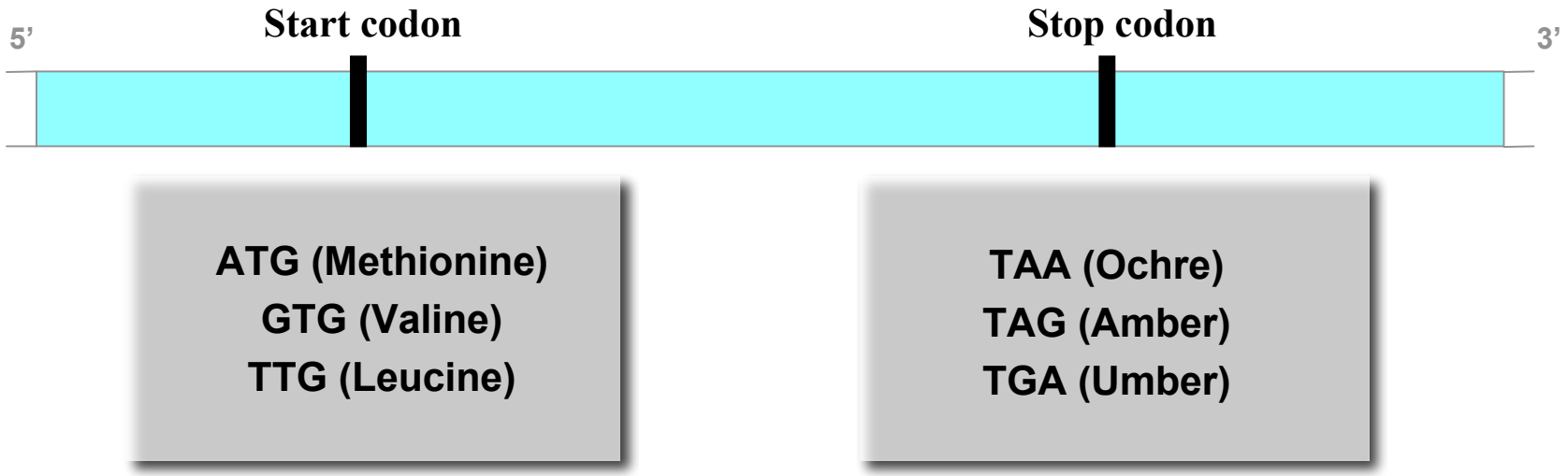


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - prokaryotic

Start and stop codon



Start codon / Stop Codon

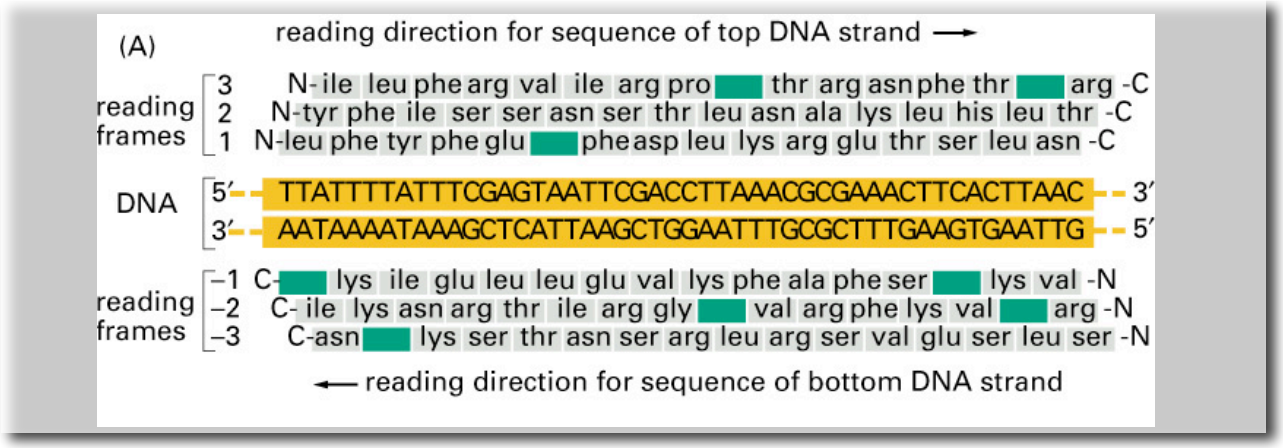
Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - prokaryotic

Open reading frame (ORF)





Start codon / Stop Codon

Open Reading Frame (protein coding)

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - prokaryotic

Regulatory elements



-10 TTGACA -35 TATAAT

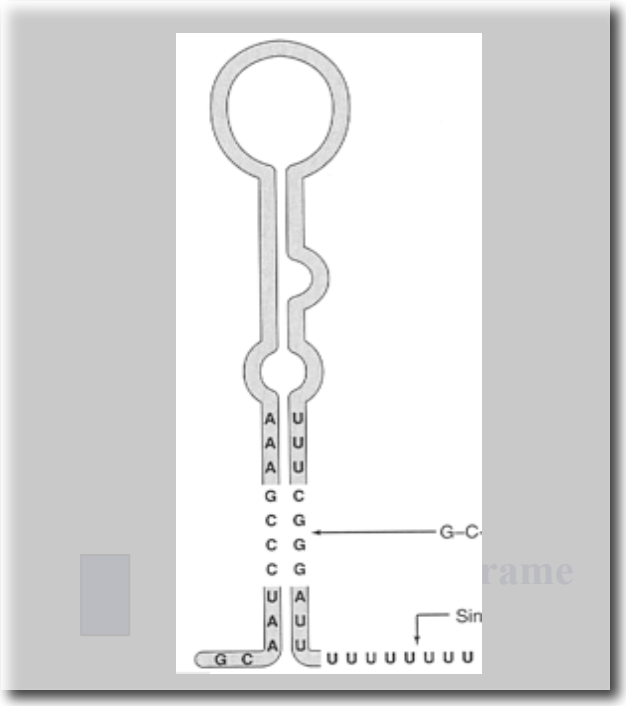
AAGGAGG

Promoter region (maybe)

Transcription start signal (maybe)

Transcription termination signal (maybe)

Start codon / Stop Codon

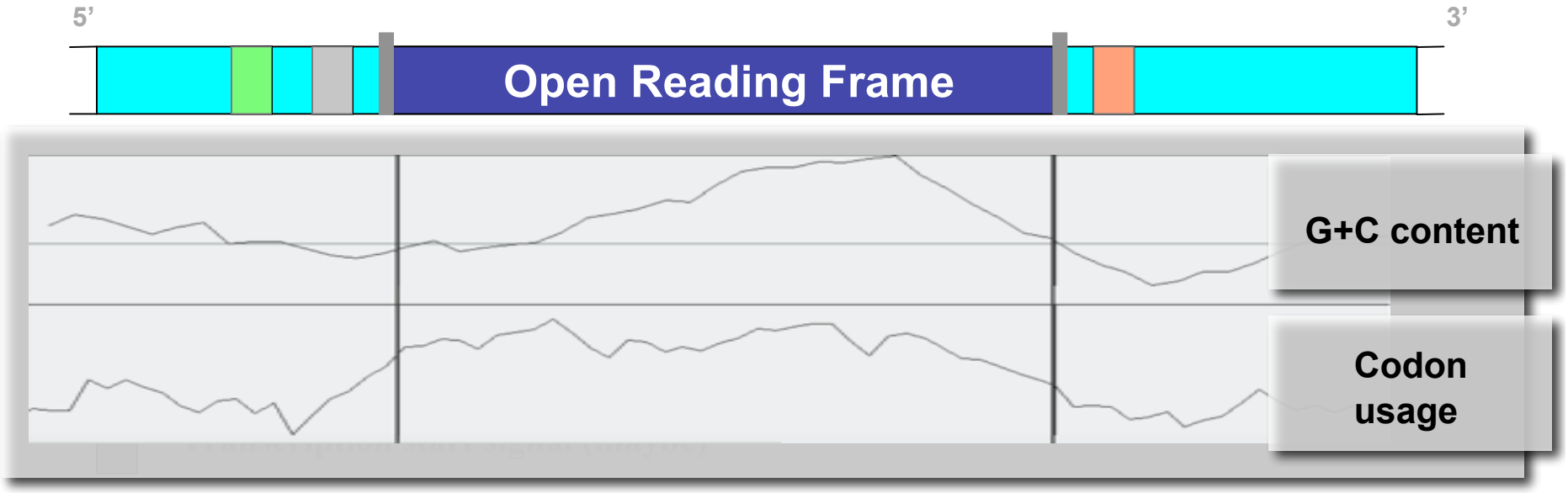


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - prokaryotic

Sequence composition - coding vs non-coding regions



Transcription termination signal (maybe)

Start codon / Stop Codon

≠ Profiles from coding and non-coding regions

Open Reading Frame (protein coding)

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - prokaryotic

Sequence composition - coding vs non-coding regions



	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr Stop Stop	Cys Cys Stop Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

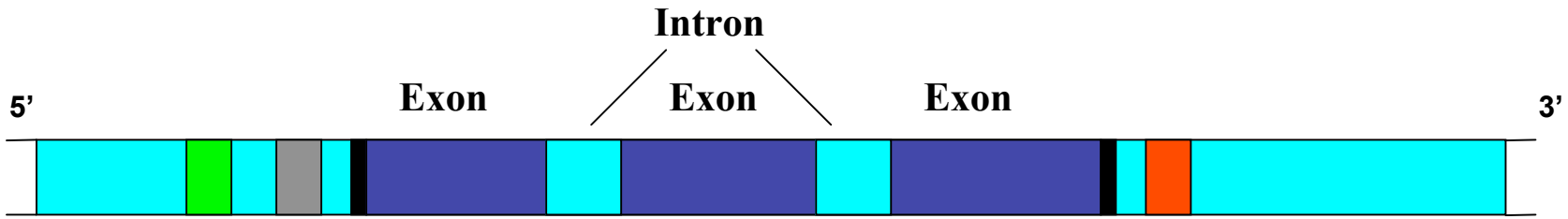
Codon	Human	Drosophila	E.coli
Arginine:			
AGA	22 %	10 %	1 %
AGG	23 %	6 %	1 %
CGA	10 %	8 %	4 %
CGC	22 %	49 %	39 %
CGG	14 %	9 %	4 %
CGU	9 %	18 %	49 %

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - Eukaryotic

Gene structure



Promoter region (maybe)

Transcription start signal (maybe)

Transcription termination signal (maybe)

Start codon / Stop Codon

≠ Profiles from coding and non-coding regions

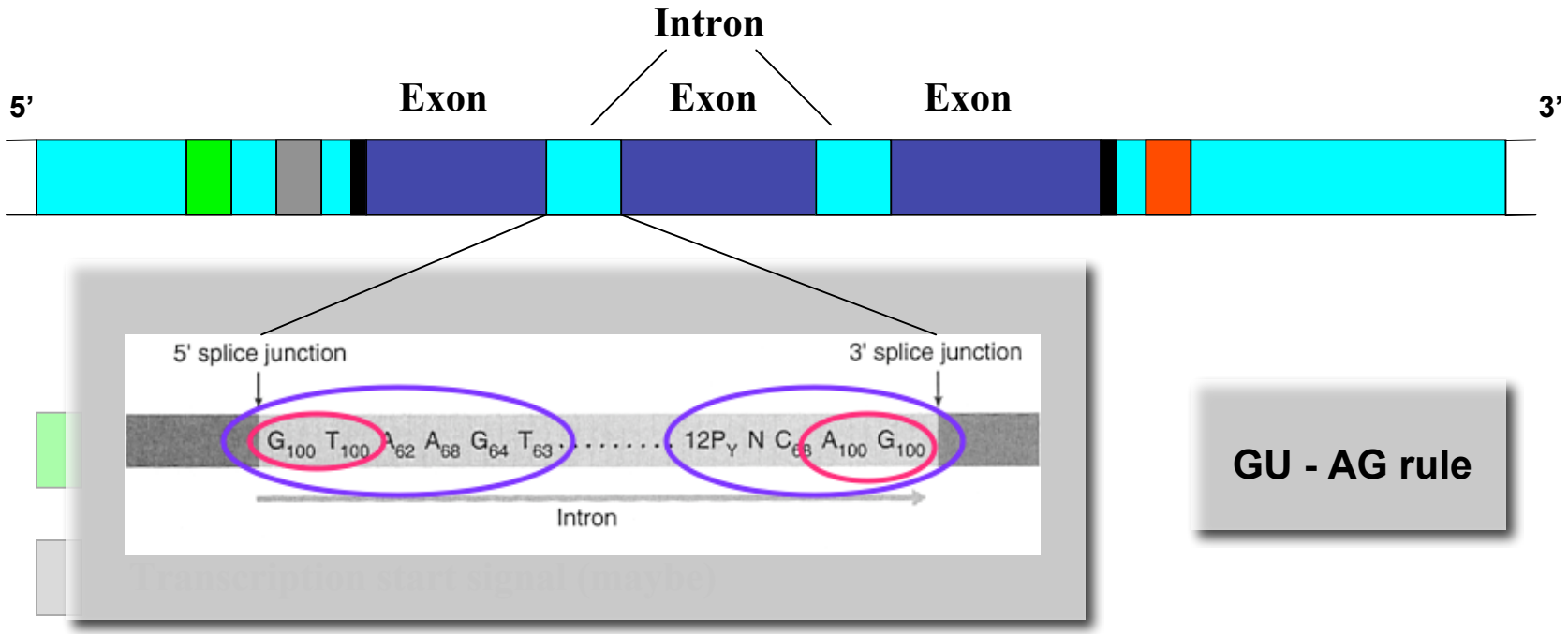
Open Reading Frame (protein coding)

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - Eukaryotic

Gene structure - Splice site



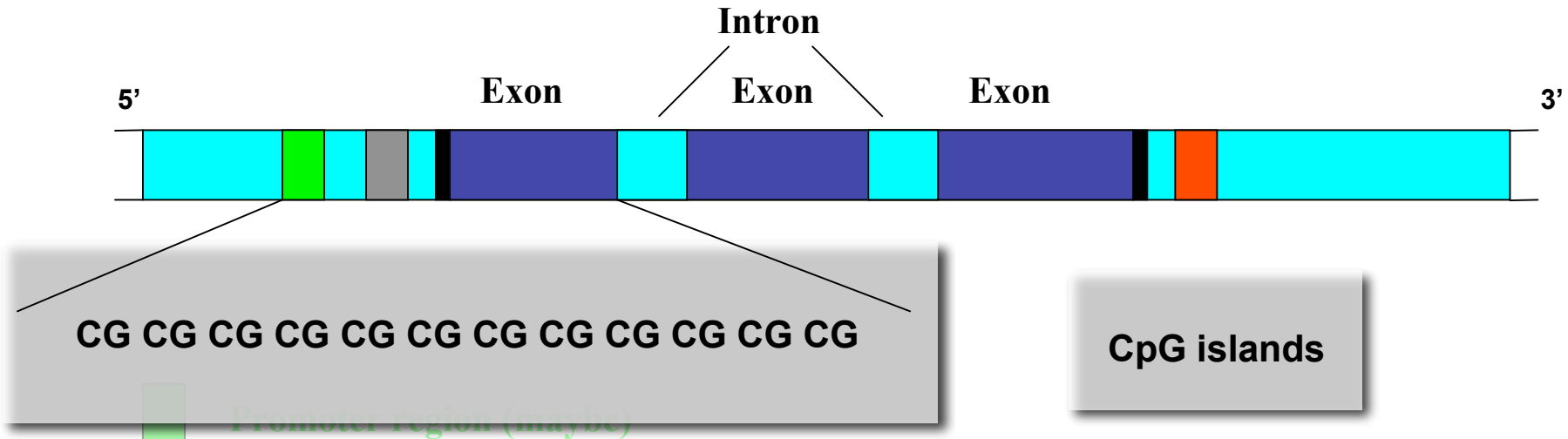
GU - AG rule

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - Eukaryotic

Gene structure - Compositional differences

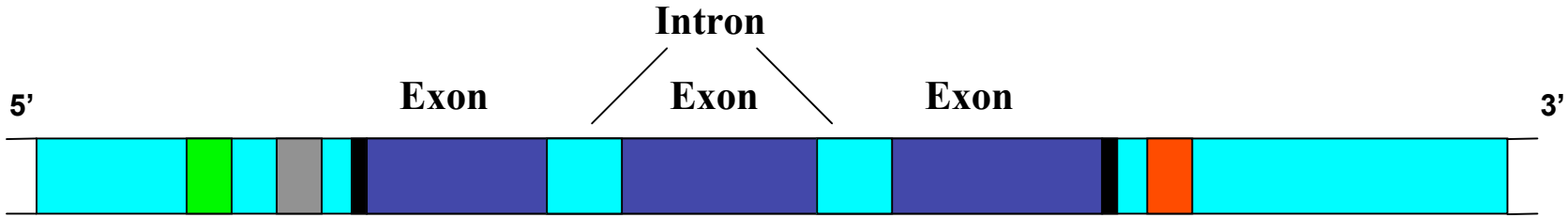


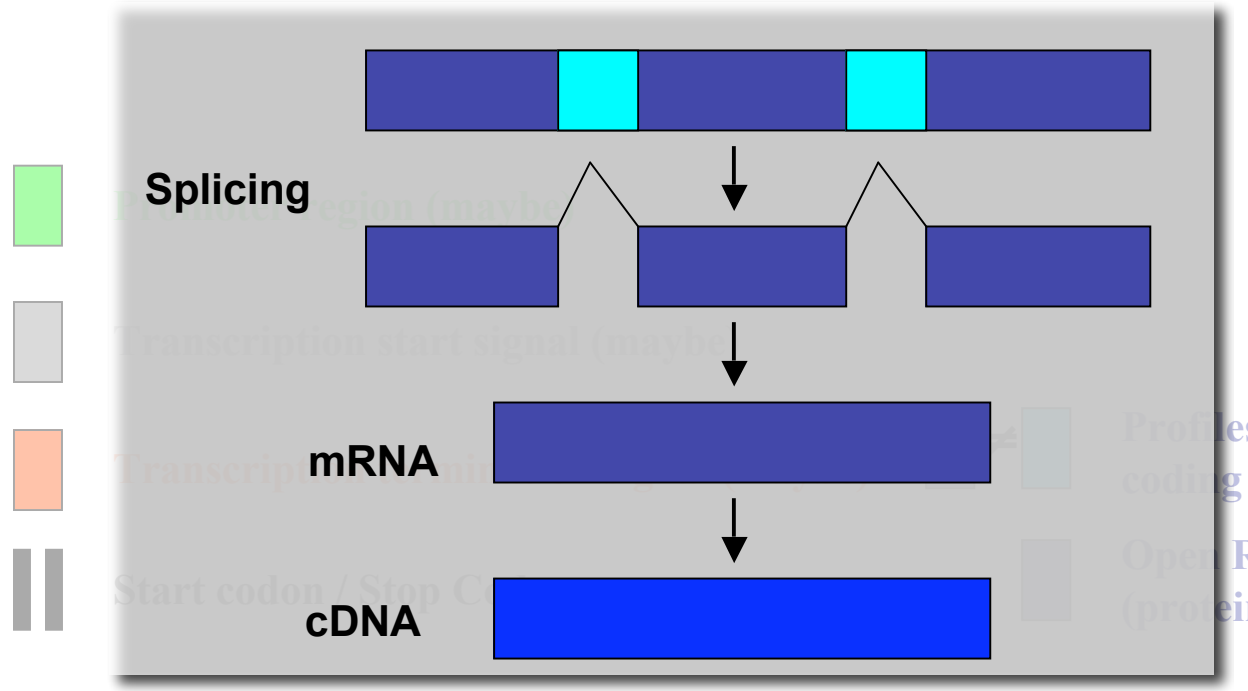
Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - Eukaryotic

Gene structure - Reverse transcribing mRNA





Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction - prokaryotic

Not only protein coding genes

Protein coding genes
- mRNA (messenger RNA)

Non-protein coding genes
- rRNA (ribosomal RNA)
- tRNA (transfer RNA)
- sRNA (small RNA)

Only in eukaryotes
- RNAi (interfering RNA)
- snRNA (nucleolar RNA)
- snoRNA (small nucleolar RNA)

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Gene prediction

Gene structure - Prokaryote



Promoter region (maybe)

Transcription start signal (maybe)

Transcription termination signal (maybe)

Start codon / Stop Codon

≠

Profiles from coding and non-coding regions

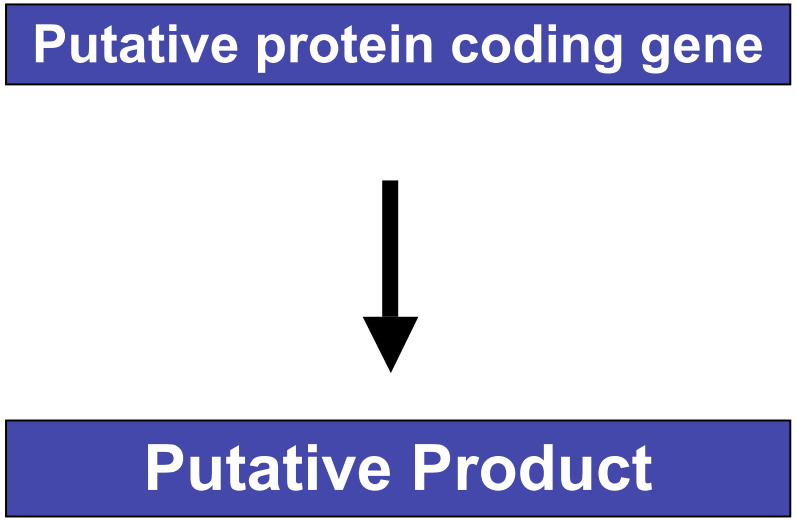
Open Reading Frame (protein coding)

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Assigning function

Sequence analysis at amino acid level

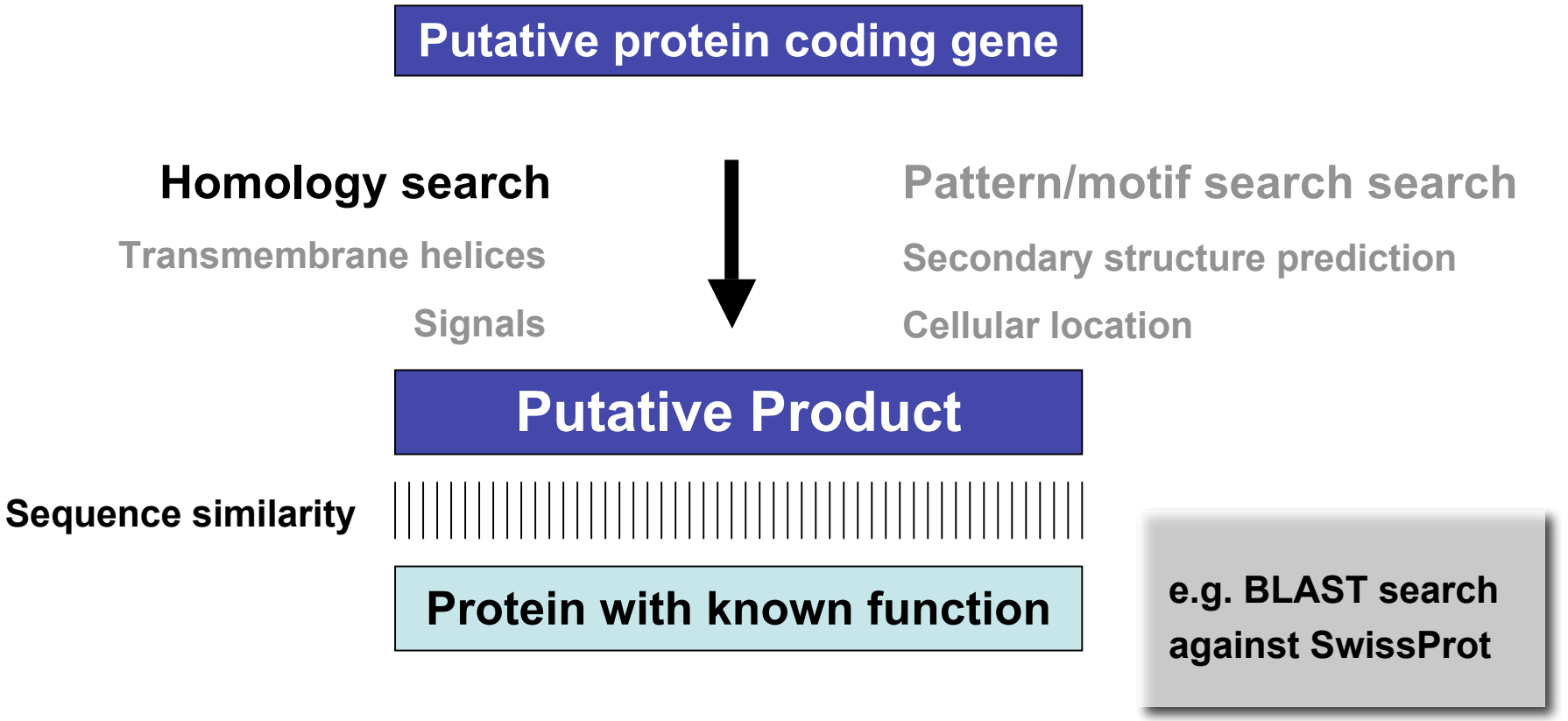


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Assigning function

Sequence similarities to characterised proteins

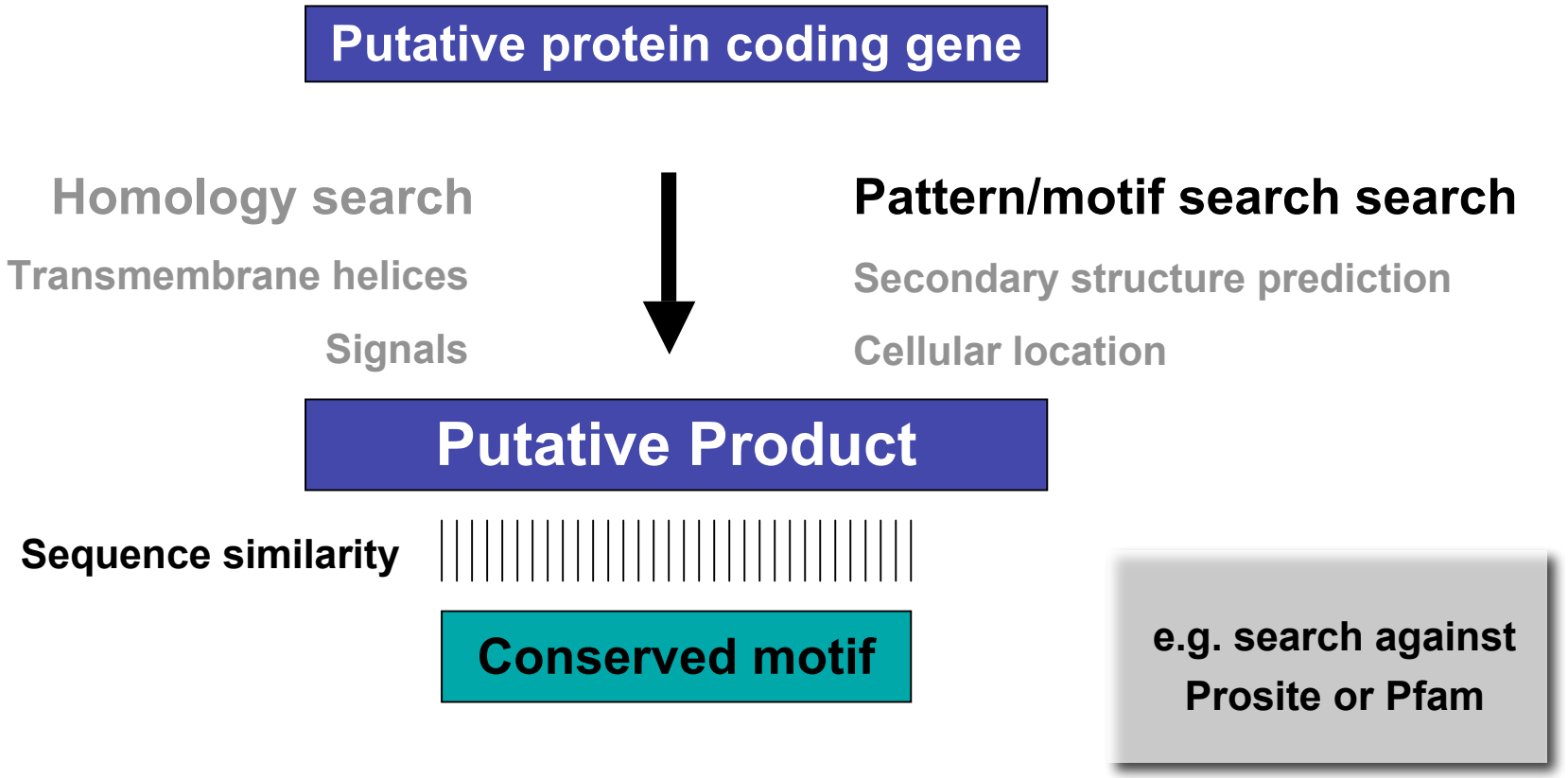


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Assigning function

Sequence analysis at amino acid level

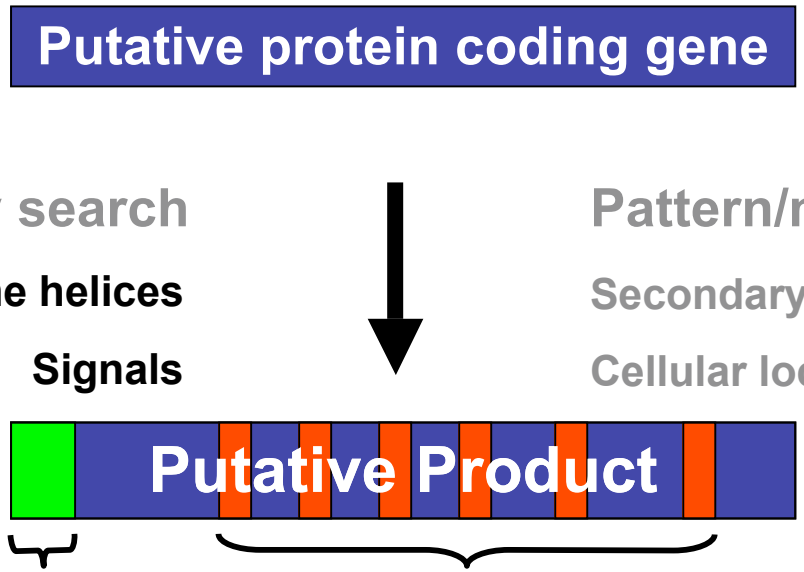


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Assigning function

Sequence analysis at amino acid level



e.g. search against SignalP and TMHMM

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Assigning function

Sequence analysis at amino acid level

Putative protein coding gene

Homology search

Transmembrane helices

Signals



Pattern/motif search search

Secondary structure prediction

Cellular location

Putative Product

Amino acid composition

e.g. Globular protein

Beta sheet

Alpha helix

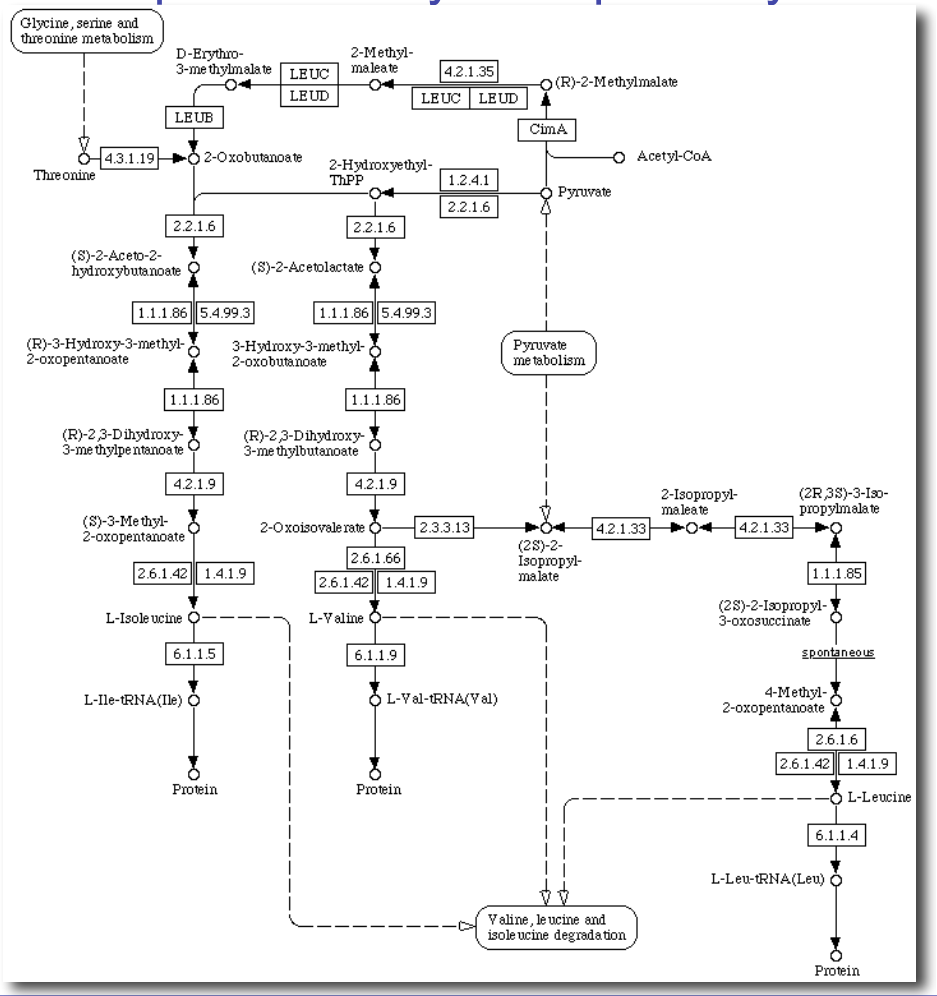
e.g. search against Cello and PredicProtein

Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Cellular processes

Sequence analysis at pathway level

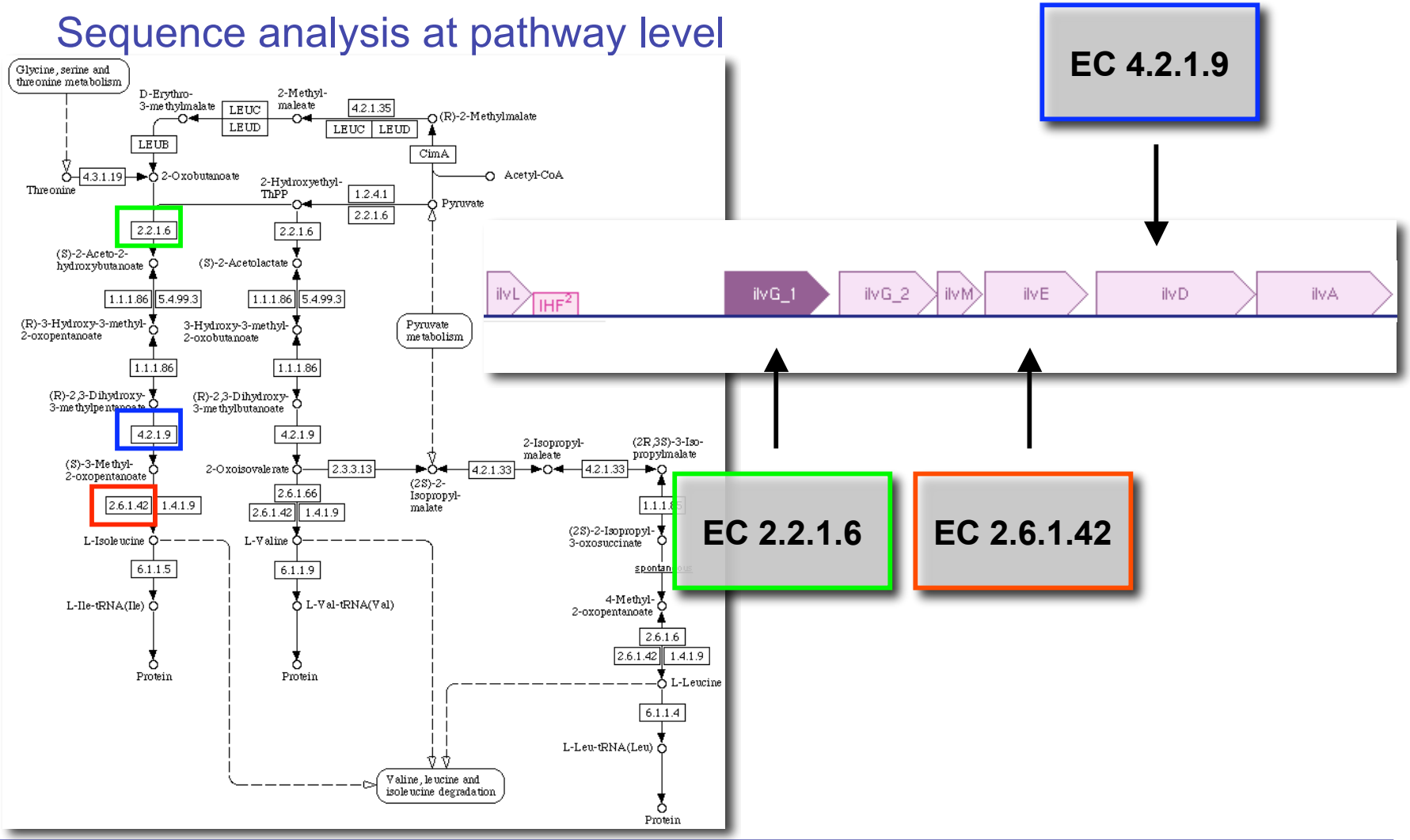


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Cellular processes

Sequence analysis at pathway level



Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Annotation tools

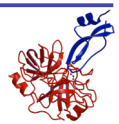
Annotation is normally performed in two stages:

1. Find all the genes (gene prediction)
2. Add layers of biological information (homology and pattern searches)

Information is collected with sequence viewer tools, e.g. Artemis

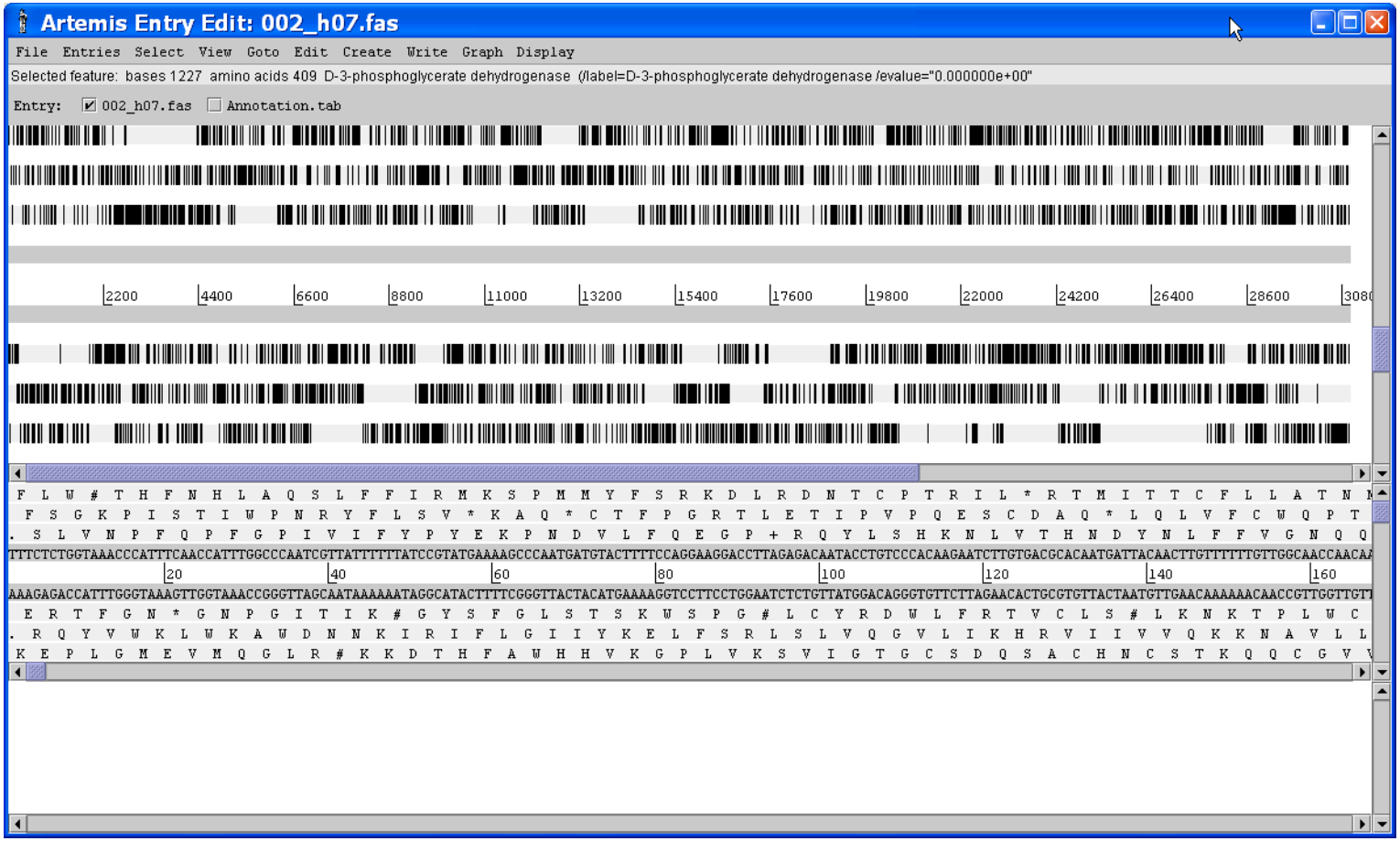
Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway



Annotation tools

Artemis

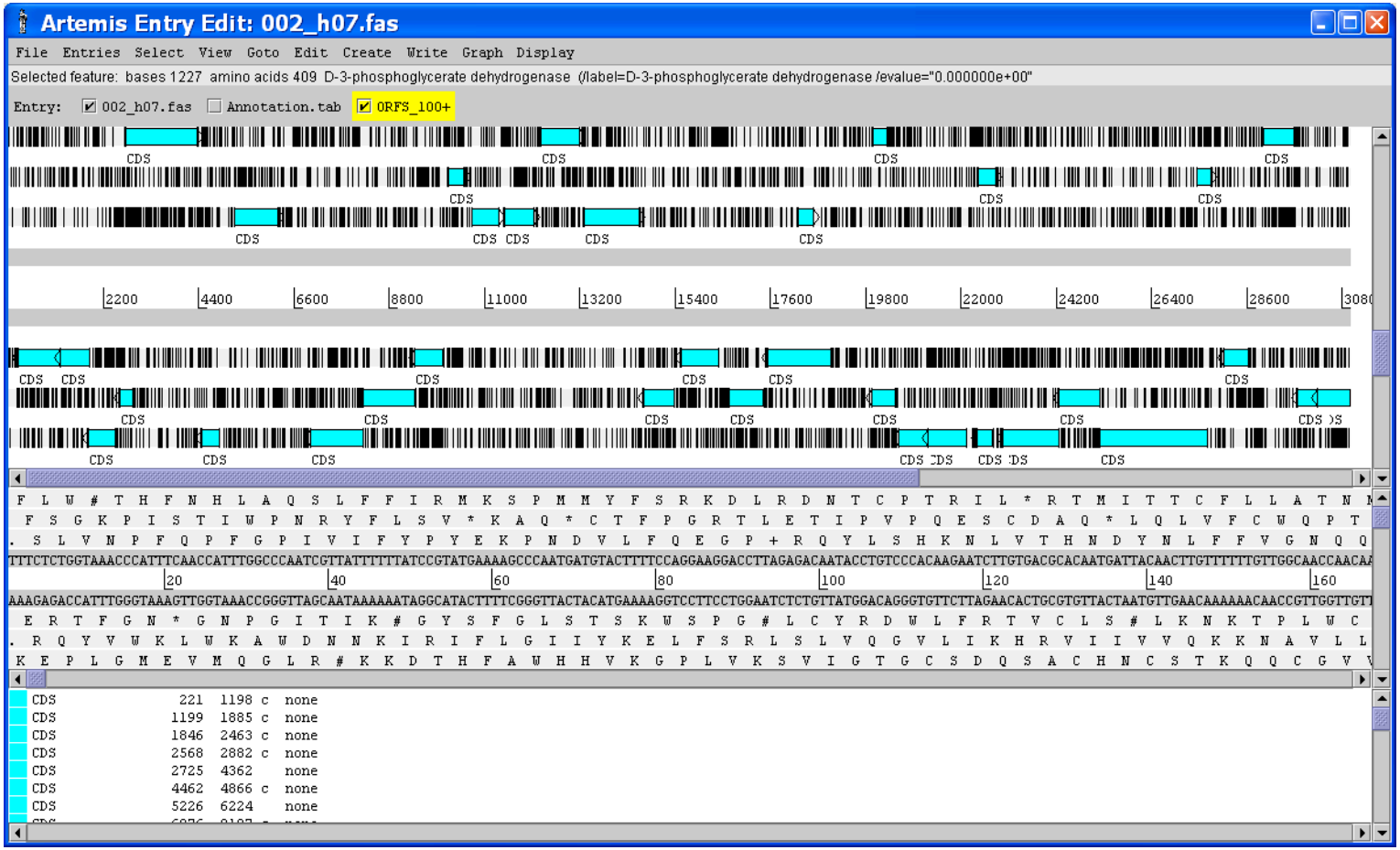


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Annotation tools

Artemis

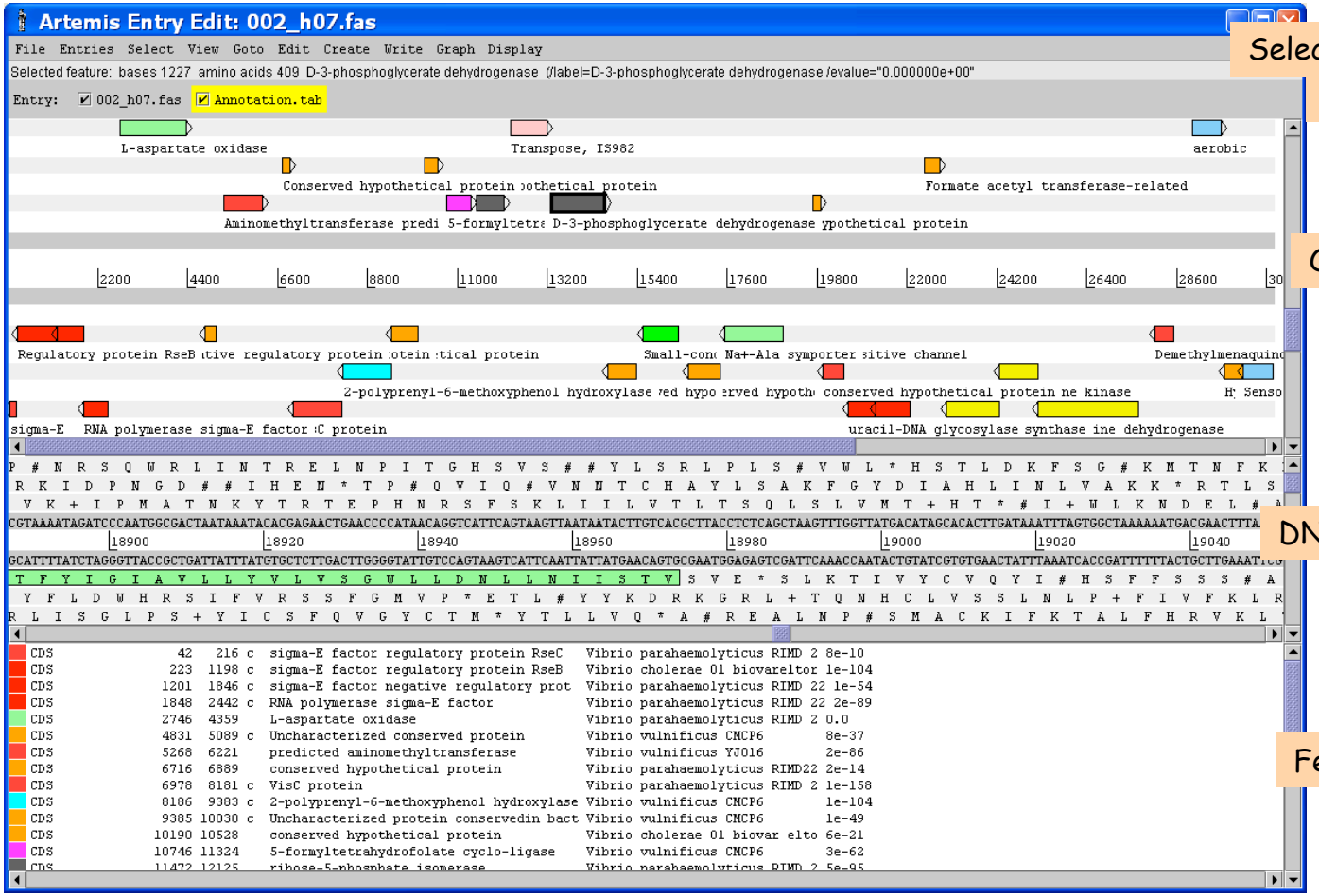


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Annotation tools

Artemis

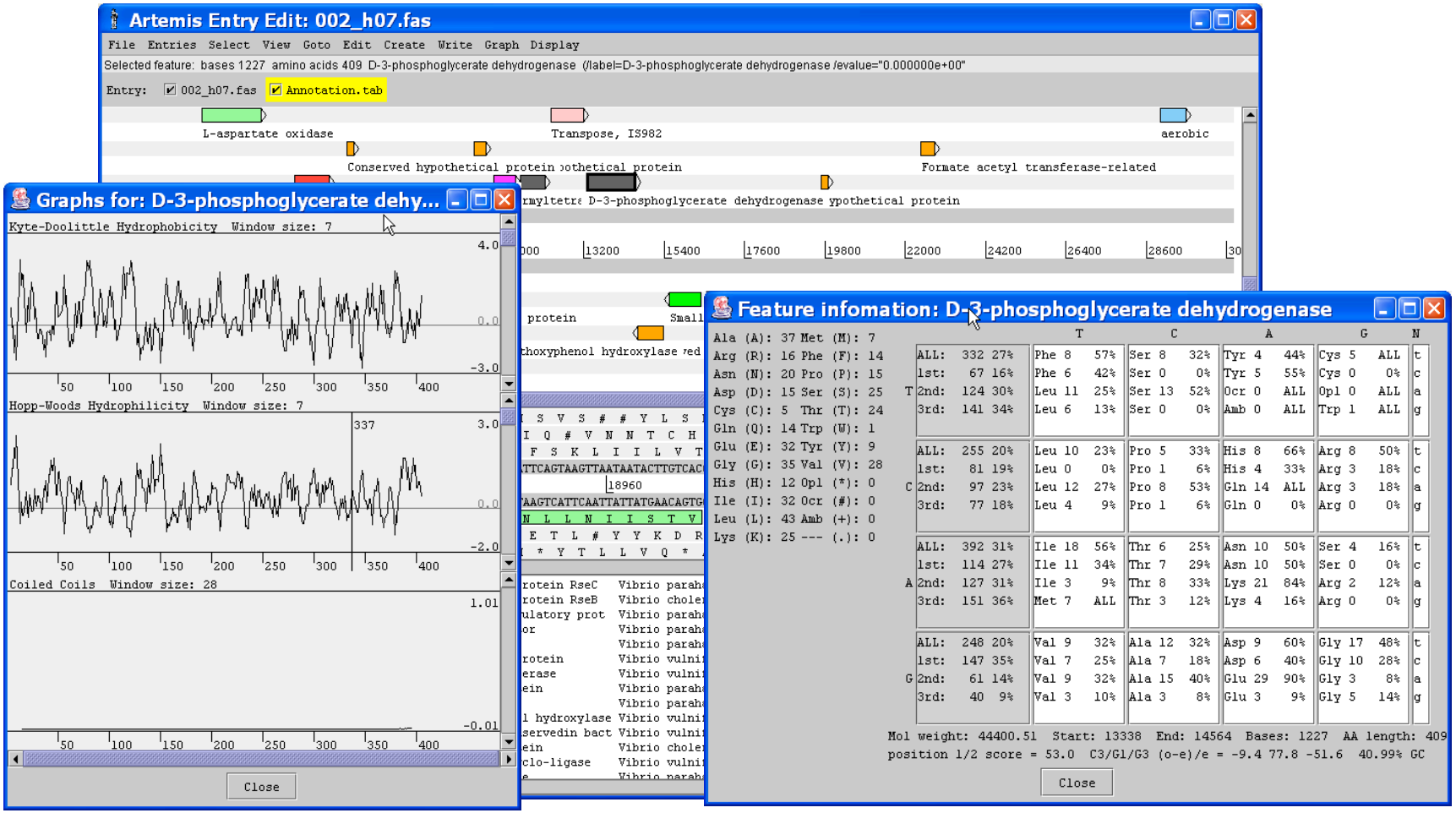


Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway

Annotation tools

Artemis



Erik Hjerde
E-mail:erikh@fagmed.uit.no
Telephone: + 47 77 64 62 63
Fax: + 47 77 64 53 50

Protein Group
Department of Molecular Biotechnology
Faculty of Medicine
University of Troms, Norway