

Functional Genomics

“Inductive reasoning, based on the integration of biological and sequence knowledge, should enable sequence and functional data to be combined in a productive way”

Antoine Danchin^o

- From Protein Sequence to Function

Current Opinion in Structural Biology 9:363-367,1999

BIOINFORMATICS , A PRACTICAL GUIDE TO THE
ANALYSIS OF GENES AND PROTEINS

Eds AD Baxevanis & FF Ouellette

Publisher Wiley-Interscience, New York, 1998

CURRENT OPINION IN STRUCTURAL BIOLOGY

VOL 9 , #3 June 1999

PROTEOME RESEARCH: NEW FRONTIERS IN FUNCTIONAL GENOMICS

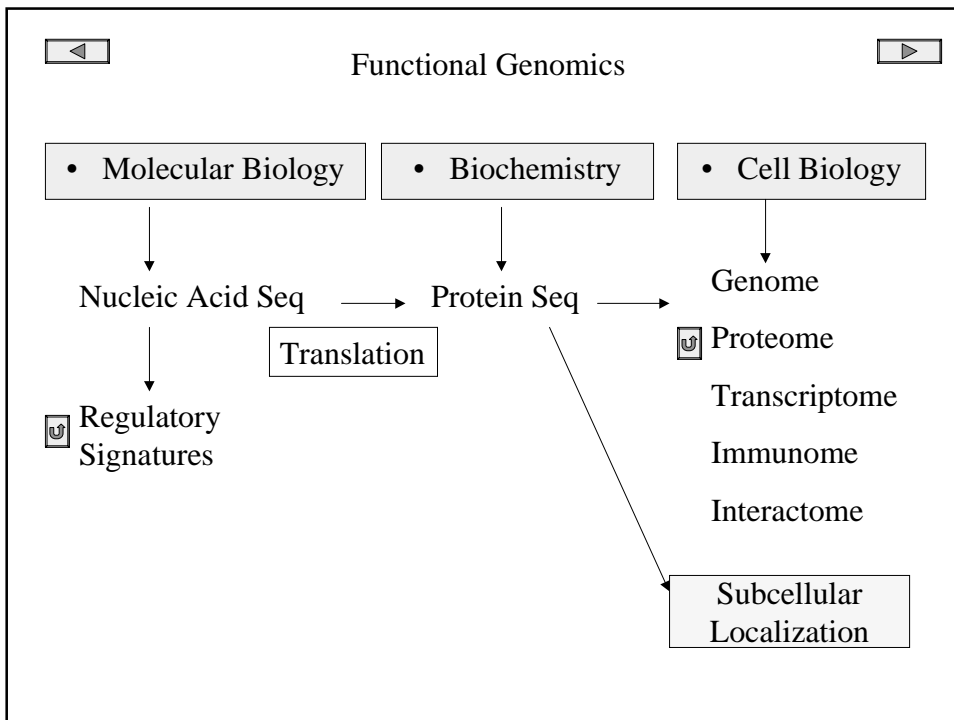
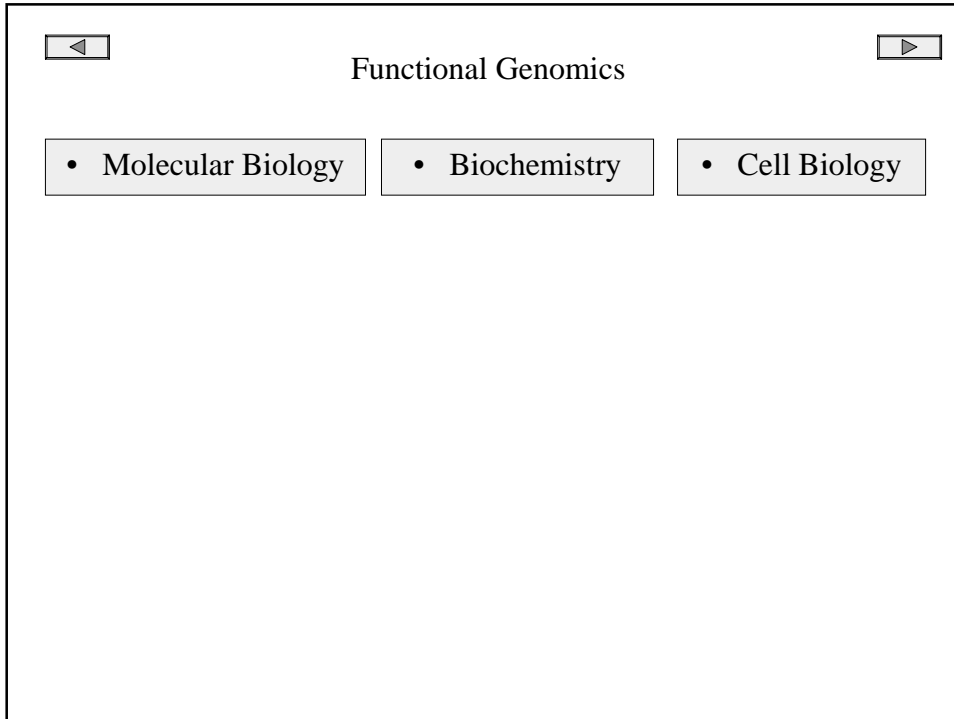
Eds MR Wilkins, KL Williams, RD Appel and DF Hochstrasser

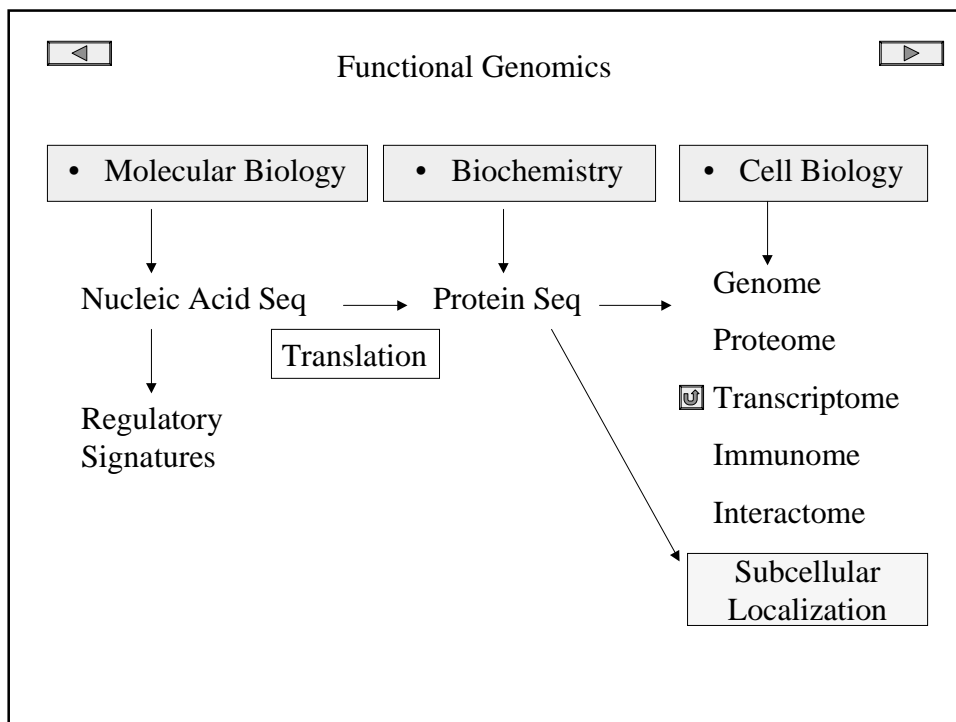
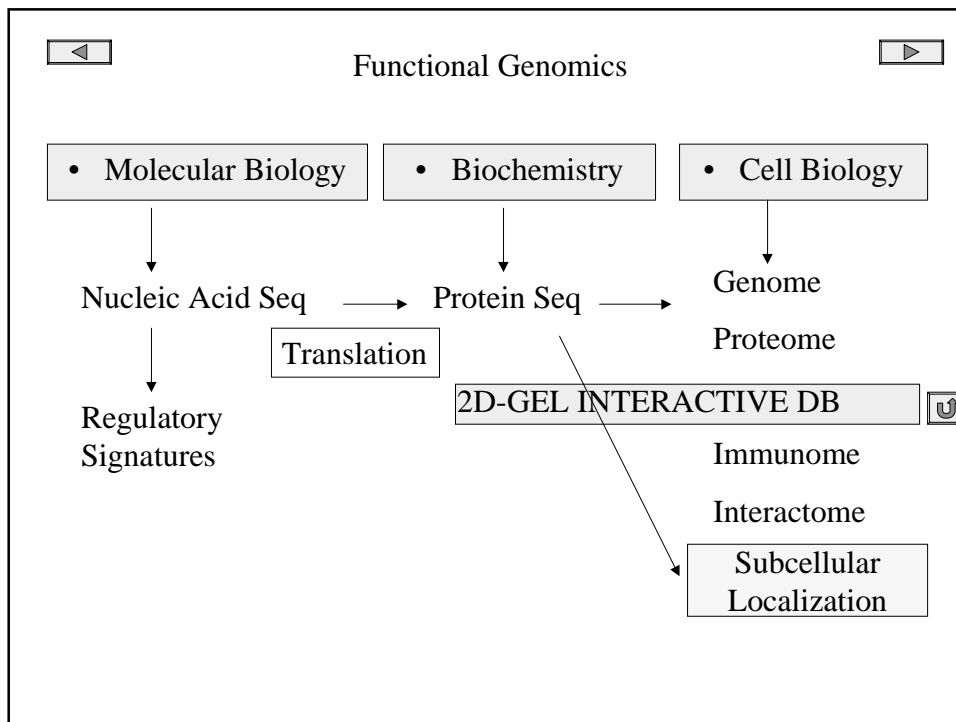
Publisher Springer, Berlin 1997


European Science Foundation Website

<http://www.functionalgenomics.org.uk/sections/programme/index.htm>





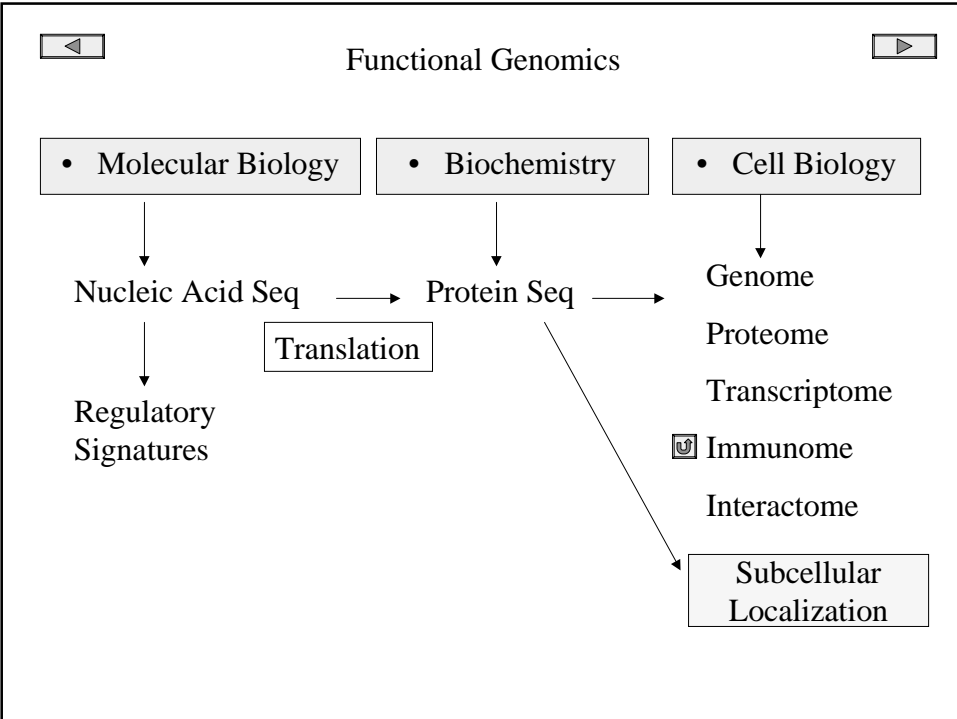




Microarray Analysis
<http://genome-www4.stanford.edu/cgi-bin/sfgf/home.pl>

The Transcriptional Program in the Response of Human Fibroblasts to Serum. Iyer et al. Science 283:83
Results for cyclin mRNA expression
<http://genome-www.stanford.edu/serum/>

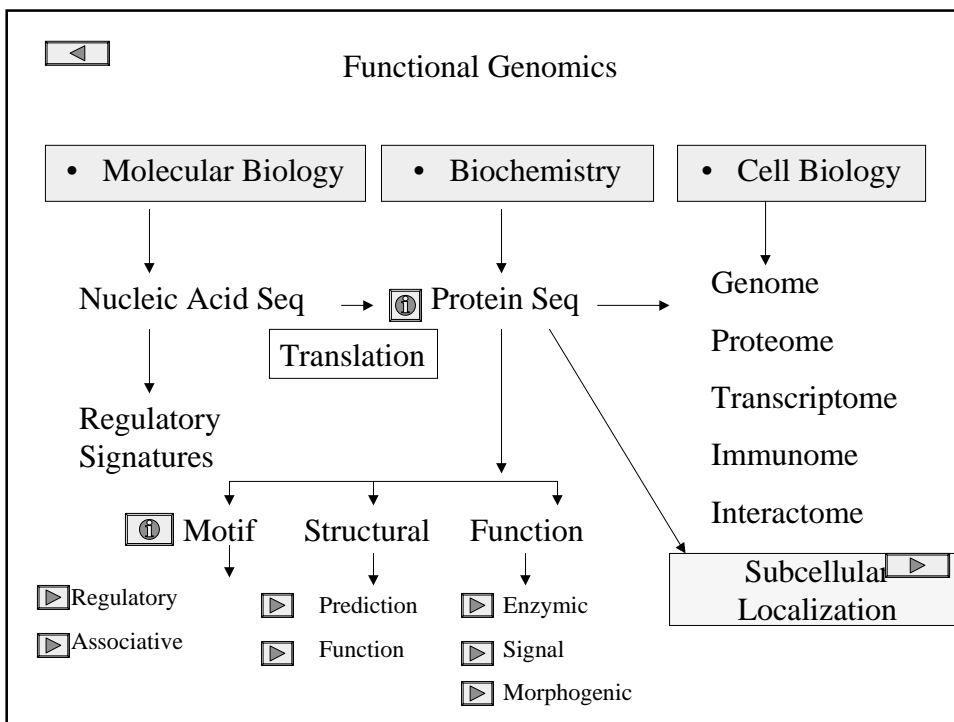
Molecular characterisation of soft tissue tumours: a gene expression study. Nielsen TO et al
Lancet. 2002 Apr 13;359(9314):1263-4.
<http://genome-www.stanford.edu/sarcoma/>



SEREX ▶

***J Exp Med* 1998 Jan 19;187(2):265-70**
Simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes. Jager E, et al

***Proc Natl Acad Sci U S A.* 2001 Dec 4;98(25):14571-6.**
Role of SEREX-defined immunogenic wild-type cellular molecules in the development of tumor-specific immunity. Nishikawa H,

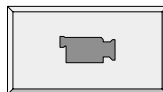




Sequence--->Structure-----> [Function]

```
MPWGRRTWL LLAFLLVFLK ISILSVTAWQ TGNCQPGPLE RSERSGTCAG PAPFLVFSQKSSISRIDPDG TNHQQLVVDA GISADMDIHY  
KKERLYWVDV ERQVLLRVFL NGTGLEKVCNVERKVSGLAI DWIDDEVLWV DQNGVITVT DMTGKNSRVL LSSLKHPNSI  
AVDPIERLMFWSSEVTGSLH RAHLKGVVDVK TLETTGGISV LTLVDLDRKRL FWVQDSGEGS HAYIHSCDYEGGSVRLIRHQ ARHSSMAF  
FGDRIFYSVL KSKAIWIANK HTGKDTVRIN LHPSFVTPGKLMVHPRAQP RTEDAARDPD PELLKQRGRP CRFGLCERDP KSHSACAEG  
YTLSDRKYCEDVNECATQN HGCTLGCECT PGSYHCTCPT GFVLLPDGKQ CHELVSCPGN VSKCSHGCVL TSDGPRCICP AGSVLGRDGK  
TCTGCSPPDN GGCSQICLPL RPSWECDCF PGYDLQSDRSCAASGPQPL LAFANSQDIR HMFHFDGTDYK VLLSRQMGMV FALDYDPVES  
KIYFAQTALKWIERANMDGS QRERLITEGV DTLEGLALDW IGRRIYWTDS GKSVMGGSDL SGKHHRIIQUERISPRGIA VHPRARRLFW  
TDVGMSPRIE SASLQGSDRV LIASSNLEP SGITIDYLDYWDYKRS VIEMANLDSG KRRRLIQNDV GHFSLAVFE DHLWVSDWAI  
PSVIRVNRKRTGQNRVRLQGS MLKPSLLVAV HPLAKPGADP CLYRNGGCEH ICQESLGTAR CLCREGFVKAWDGKMLCPQD YPILSGENAD  
LSSRCINTEGGYVCRSEGY EGDGISCFDI DEQQRGAHNC AENAACTNTE GGYNCTCAGR PSSPGRSCPDSTAPSLLED GHHLDRNSYP  
TSSRCINTEGGYVCRSEGY EGDGISCFDI DEQQRGAHNC AENAACTNTE GGYNCTCAGR PSSPGRSCPDSTAPSLLED GHHLDRNSYP  
GCPSSYDGYC LNGVCMHIE SLDSYTCNCV IGYSGDRCTRDRLRWELRH AGYQKHDIM VVAVCMVALV LLLLLGMWGT  
YYYRTRKQLS NPPKNPCDEPSGSVSSGPD SSSGAAVASC PWPWFVLEK HQDPKNGSLP ADGTNGAVVD AGLSPLQLGSVHLTSWRQK  
PHIDMGMTGQ SCWIPPSDR GPQEIEGNSH LPSYRVPGE KLHSLQSANGSCHERAPDLP RQTEPVK
```

Crystal structure of F1_ATPase



3D-Structural data on ATP Synthase reveals a miniscule engine

By Hongyun Wang and George Oster, University of California, Berkeley (see *Nature* **396**, 279-282 (1998)).

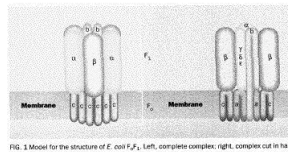
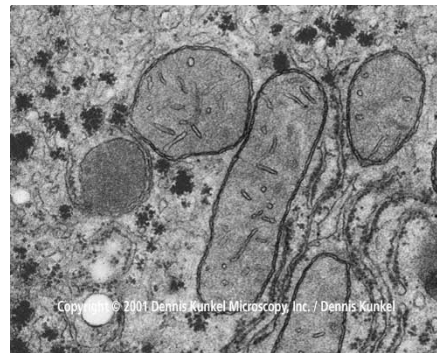


Fig. 1 Model for the structure of *E. coli* F₁F₀. Left, complete complex; right, complex cut in half.

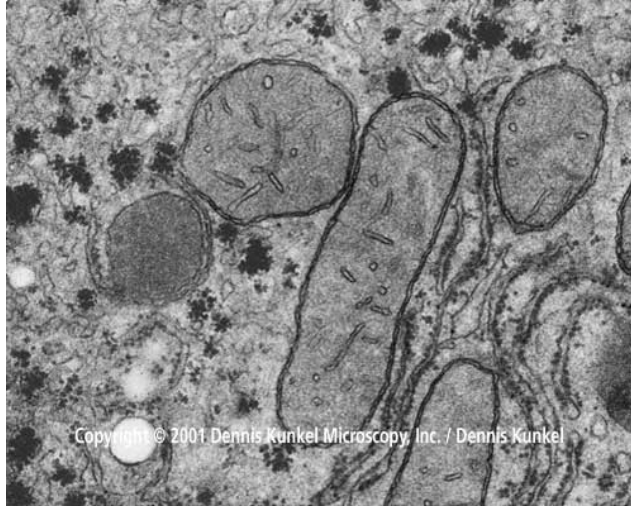
Protein pumps

FO_ATPase
F1_ATPase
F1_FO_ATPase



Cellular organelles; mitochondria, glycogen, rough endoplasmic reticulum with ribosomes on surface (liver cell).

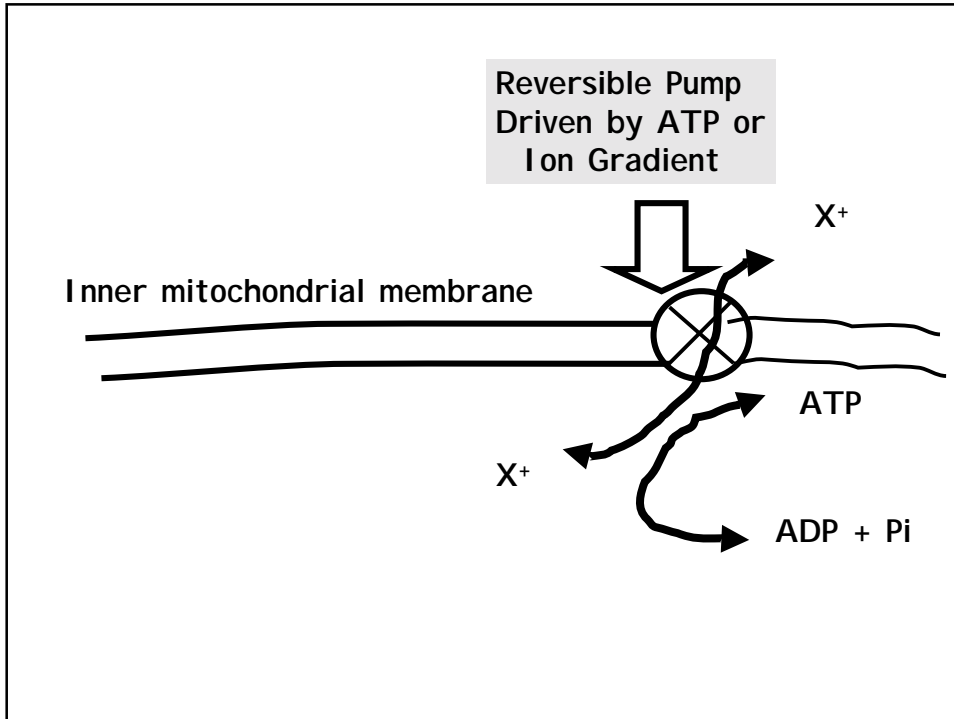
<http://www.cat.cc.md.us/courses/bio141/lecguide/unit1/eustruct/dkmito.html>



Mitochondria
Synthesize
ATP

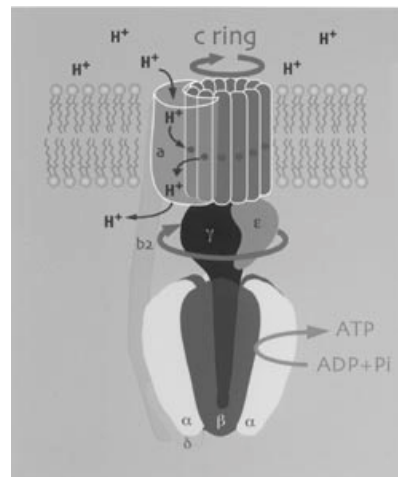


Inner
Mitochondrial
Membrane

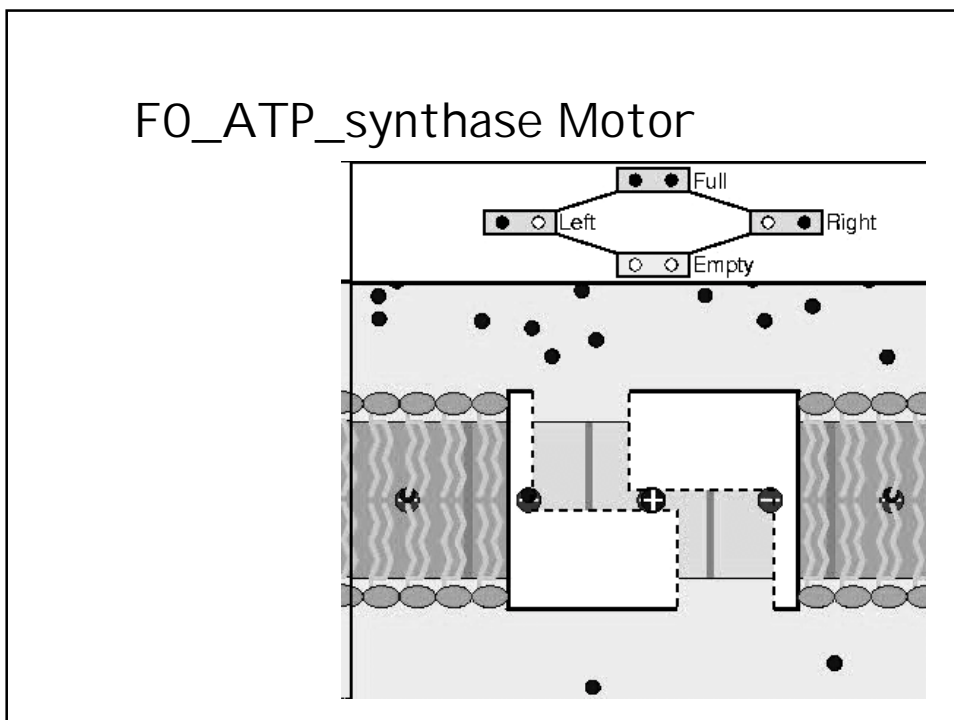
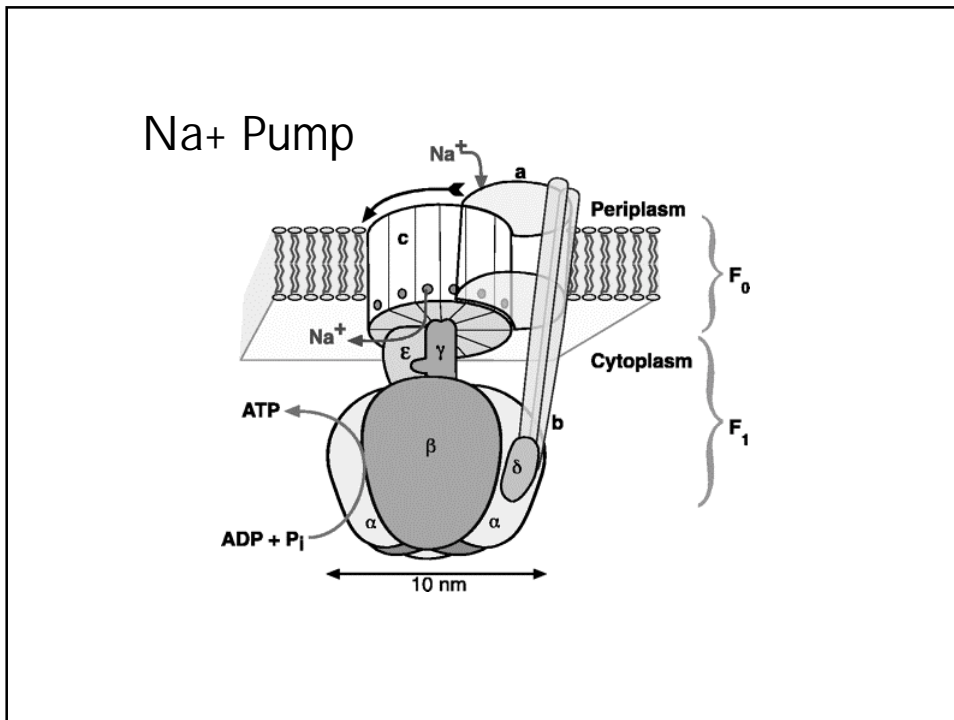


Proton Pump

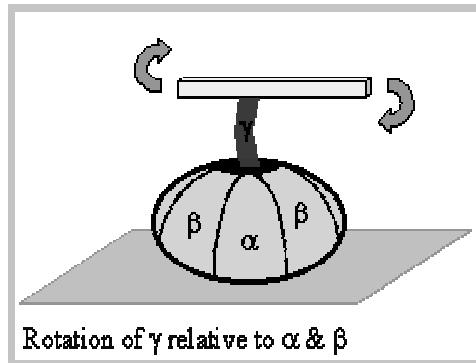
The F1F0_ATP_synthase is a small molecular motor in which catalytic site events in the F1 part are coupled to proton translocation in the F0 part by rotation of the mobile domain or crankshaft within the protein complex.



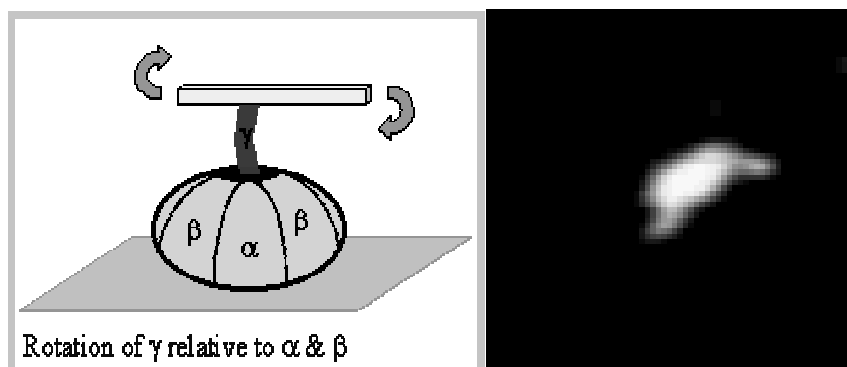
<http://www.molbio.uoregon.edu/home97/capaldi.html>

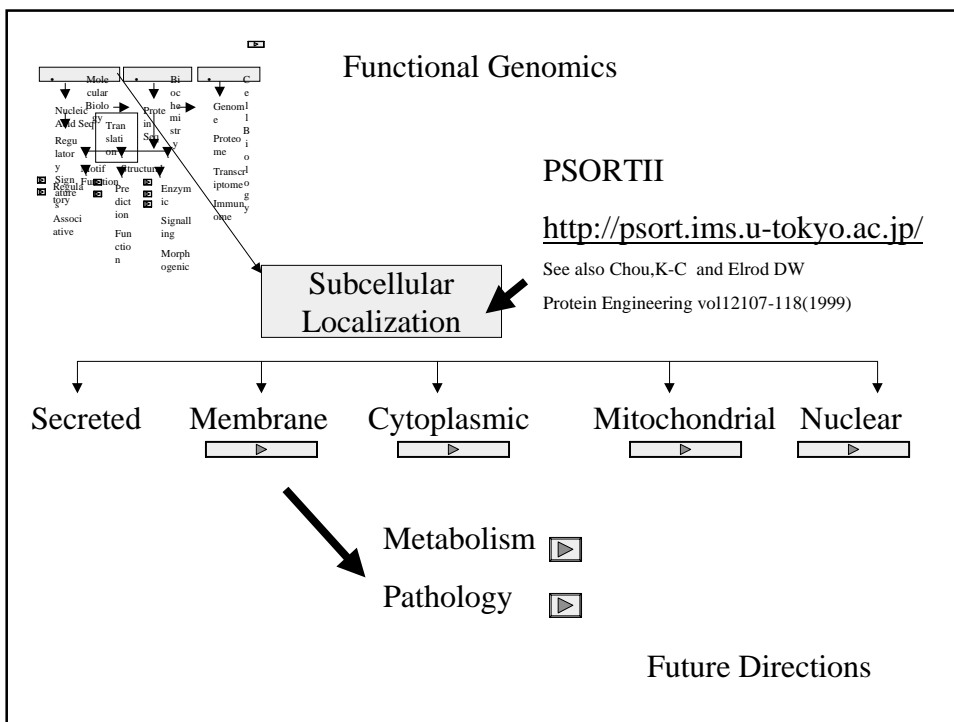
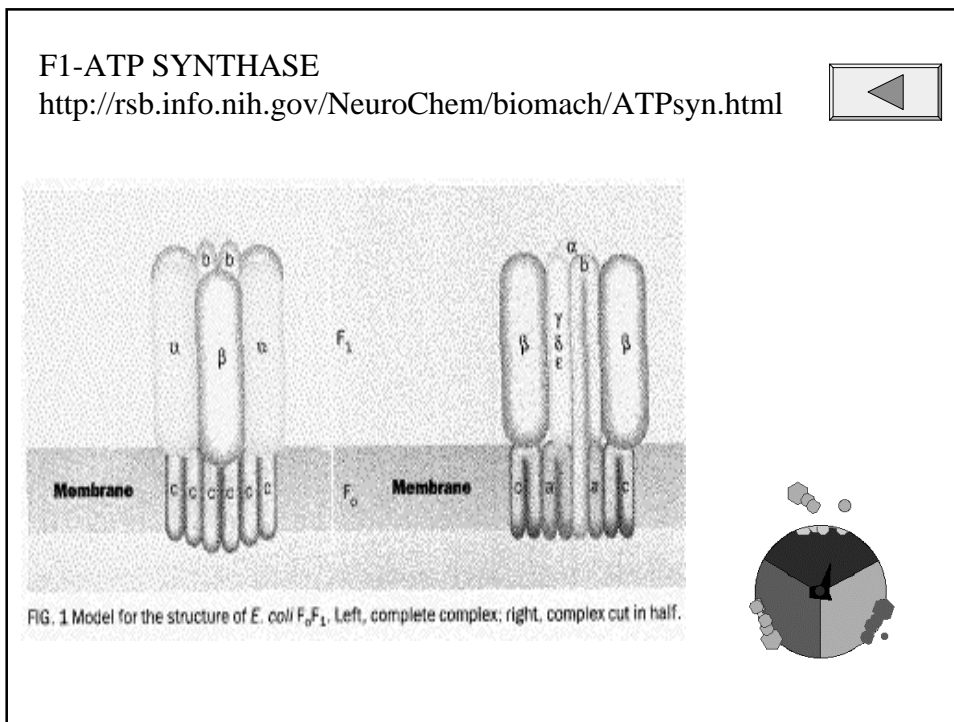


F1_ATPsynthase animation



F1_ATPsynthase Motor driving fluorescent rod







Consensus Phosphorylation site

.....DSGLDS.....

Recognized by GSK-3beta

Aberle et al EMBO J 16:,3797-3804

see also F-Box listing in PFAM

<http://pfam.wustl.edu/cgi-bin/getdesc?name=F-box>



MOTIFS AND DOMAIN SITES

Pfam: <http://pfam.wustl.edu/>

Prosite:

<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC50001>

SH2 ...SRC HOMOLOGY DOMAIN

PROSITE CROSS REFERENCES

[PS50001:SH2](#)

The Src homology 2 (SH2) domain is a protein domain of about 100 amino-acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps [1]. Similar sequences were later found in many other intracellular signal-transducing proteins [2]. SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and strictly phosphorylation-dependent manner [3,4,5,6].

Techniques for the Prediction of Tertiary Structure

- *ab initio* Modelling
- Comparative Modelling
- Fold Recognition (Threading)



See Sternberg et al

Curr Opin Struct Biol vol 9:368-373, 1999

(<http://www.bmm.icnet.uk>)

& CASP4 Home Page

Third Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction

<http://predictioncenter.llnl.gov/casp4/>

<http://predictioncenter.llnl.gov/casp5/Casp5.html>

C
A
S
P
4



ab initio Secondary Structure Prediction



Target (top row): helices - PURPLE , strands - GREEN

Model (second row): helices - PURPLE , strands - GREEN

Residues not predicted (third row): BLACK



MOTIF SEARCH ...PFAM

---><http://pfam.wustl.edu/cgi-bin/getdesc?name=EGF>

Pfam entry: EGF

Accession number: PF00008
Definition: EGF-like domain
Author: Bateman A, Sonnhammer ELL
Alignment method of seed: Clustalw_manual
Source of seed members: Swissprot_feature_table
Gathering cutoffs: 24.7 10.6
Trusted cutoffs: 24.70 10.60
Noise cutoffs: 24.60 39.80
HMM build command line: hmmbuild -F HMM SEED
HMM build command line: hmmscalibrate HMM
HMM build command line: hmmscalibrate --seed 0 HMM
Reference Number: [1]
Reference Medline: [96222301](#)
Reference Title: Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders.
Reference Title: syndrome and other genetic disorders.
Reference Author: Downing AK, Knott V, Werner JM, Cardy CM, Campbell ID,
Reference Author: Handford PA;
Reference Location: Cell 1996;85:597-605.



Sequence has 3D-coordinate entry in "PDB"



Research Collaboratory for

STRUCTURAL BIOINFORMATICS

Rutgers • SDSC • NIST

[Comments](#) [Home](#)

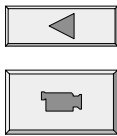
Protein Coordinate Databases

<http://www2.ebi.ac.uk/dali/>

<http://scop.berkeley.edu/>

<http://www.rcsb.org/pdb/>

SCOP ENTRY FOR EGF



IEGF Growth Factor **date** Oct 01, 1991

authors G.T.Montelione, K.Wuthrich, H.A.Scheraga

compound **source**

Mouse (Mus Musculus)

method NMR

note is a representative structure

Data retrieval

Asymmetric unit, PDB entry: [[header only](#)] [[complete with coordinates](#)] [[Save to disk](#)] [Likely Quarternary Molecular Structure](#) file(s) for IEGF [NMR restraints](#) (1egf.mr.Z), size 17 Kb Retrieve IEGF in [mmCIF](#) format


View IEGF in 3D

[MDL's Chemscape Chime](#) ([Install Chime freeware](#)) This is good for easiest viewing of basic structure.
[RasMol](#) ([Install RasMol freeware](#)) Here's help on [how to use RasMol](#).

Visual 3D analysis of IEGF

STING, shows the **sequence** and maps the locations of sequence residues or residue ranges onto the 3D structure.

Functional Searching



MVEMLEPTAIL LVLAVSVVAK DNATCDGPCG LRFRQNPQGG VRIVGGKAAQ HGAWPVMVSL

QIFRYNSHRY HTC GGSL LNS RWVLTAA **H**CF VGKNNVHDWR LVFGAKEITY GNNKPVKAPL

QERYVEKIII HEKYNSATEG N **D**IALVEITP PISCGRFIGP GCLPHFKAGL PRGSQSCWVA


GWGYIEEKAP RPSSILMEAR VDLIDLDCN STQWYNGRVQ PTNVCAGYPV GKIDTCQGD **S**

GGPLMCKDSK ESAYVVVGIT SWGVGCALAK RPIYATATWP YLNWIASKIG SNALRMIQSA

TPPPPTTRPP PIRPPFSHPI SAHLPWFQ PPRPLPPRP AAQPPPPSP PPPPPPPAS

LPPPPPPPP TPSSTTKLPQ GLSFAKRLQQ LIEVLKGY SDGKNHYDME TTELPILTSTS*

PFAM-->PROSITE <http://pfam.wustl.edu/>


Pfam 4.0 (St. Louis) : [Home](#) | [Analyze a sequence](#) | [Browse alignments](#) |
[Text search](#) | [Swisspfam](#) | [Help](#) | 

Starting search. Estimated time: 13 seconds. Please wait...

Pfam HMM search results

[\[Go here for an explanation of the format of the results\]](#)

Model	Seq-from	Seq-to	HMM-from	HMM-to	Score	E-value
!!	trypsin	43	285	1	258	323. 3.2e-93

trypsin: domain 1 of 1, from 43 to 285: score 323.1, E = 3.2e-93 

```

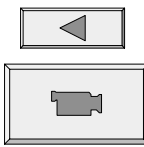
-> IvGGreaqpgsfGpwqvsIqirggsg.srkhfCGGsLisenwVLTAA
IvGG+ aq+g++ Pw+vslqi + +++ r+h+CGGsL+++wVLTAA
query 43 IVGKAAQHGAW-PWMVSLQIFR-YNshRYHTCGGSLlnSRWVLTAA 87

HCvsgssakassyrvsvrlGehnlstsegt....eqkfpveryiivHpn
HC+ g + +++r+ ++G+ +++ + ++ + + q ++ve+ ii+H++
query 88 HCFVG-KNNVHDWRL--VFgAKEITYGNNKpvkapLQERYVEK-IIIHEK 133

YnsdtldngaYdnDiAlLkLsspcvtlsdtrpicLPstvtasddlpvGt
Yns t + nDiAL++++p ++++ +++p cLP + + + +
query 134 YNSATEG-----NDIALVEITPP-ISCGRFIGPGCLPHF--KAGLPRGSQ 175

tctvsGWGrrtknngaplsdvLqevnvplvsneeCrsayeyggrdkkefv
+c+v+GWG+ +++ +++s++L+e+ v l++ + C+s ++y+gr v
query 176 SCWVAGWGY-IEEKAPRPSSILMEARVDLIDLdLcNSTQWYNGR-----V 219

```



[LIVM]-[ST]-A-[STAG]-H-C

acrosin


<http://www.expasy.ch/cgi-bin/niceprot.pl?P10323>

FT ACT_SITE 88 88 CHARGE RELAY SYSTEM

FT ACT_SITE 142 142 CHARGE RELAY SYSTEM

FT ACT_SITE 240 240 CHARGE RELAY SYSTEM

MOTIF AND DOMAIN IDENTIFICATION SITES




Subcellular localization and transmembrane signals

PSORT <http://psort.ims.u-tokyo.ac.jp/> signal seq

NNPSL <http://predict.sanger.ac.uk/nnpsl/> ...signal seq

SignalP <http://www.cbs.dtu.dk/services/SignalP> signal seq

TMHMM <http://www.cbs.dtu.dk/services/TMHMM> transmemb seq




PREDICTION OF SIGNAL SEQUENCE CLEAVAGE SITES

MAGPLRAPLLLLAILAVALAVSPAAGSSPGKPPRLVGGPMDASV
EEEGVRRALDFAVGEYNKASNDMYHSRALQVVRARKQIVAGV
NYFLDVELGRITCTKTQPNLDNCPFHDPHLKRKAFCFQIYAVP
WQGTMTLSKSTCQDA

↓

SignalP V1.1
World Wide Web Prediction Server
Center for Biological Sequence Analysis
Nielsen H ,Brunak S & von Heijne G
Protein Engineering vol12:3-9(1999)



SignalP V1.1 World Wide Web Server

•Explanation of the output


Human cystatin c seq

SignalP predictions *****

Using networks trained on euk data

>WEBSEQUENCE length = 146

#	pos	aa	C	S	Y
	1	M	0.012	0.935	0.000
	2	A	0.012	0.958	0.000
	3	G	0.015	0.959	0.000
	4	P	0.012	0.959	0.000
	5	L	0.012	0.961	0.000
	6	R	0.011	0.970	0.000
	7	A	0.014	0.972	0.000
	8	P	0.014	0.972	0.005
	9	L	0.012	0.977	0.017
	10	L	0.012	0.978	0.024
	11	L	0.013	0.975	0.034
	12	L	0.014	0.978	0.044
	13	A	0.015	0.978	0.053
	14	I	0.021	0.974	0.072
	15	L	0.015	0.963	0.067
	16	A	0.025	0.965	0.092
	17	V	0.072	0.946	0.171
	18	A	0.025	0.967	0.106
	19	L	0.189	0.952	0.312



SignalP V1.1 World Wide Web Server


Human A33 seq

- [Explanation of the output](#)

SignalP predictions *****
Using networks trained on euk data

>WEBSEQUENCE length = 120

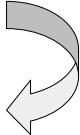
#	pos	aa	C	S	Y
	1	M	0.012	0.939	0.012
	2	V	0.012	0.954	0.015
	3	G	0.012	0.950	0.019
	4	K	0.015	0.947	0.029
	5	M	0.012	0.955	0.033
	6	W	0.012	0.959	0.041
	7	P	0.011	0.945	0.047
	8	V	0.012	0.944	0.054
	9	L	0.012	0.965	0.059
	10	W	0.012	0.985	0.064
	11	T	0.012	0.988	0.070
	12	L	0.017	0.960	0.088
	13	C	0.015	0.964	0.087
	14	A	0.033	0.903	0.136
	15	V	0.019	0.789	0.107
	16	R	0.020	0.794	0.112



Unkown N-terminal sequence

RPCQCNDNIDPNAVGNCRNLTGECLKCIYNTAGFYCDRC
KEGFFGNPLAPNPADKCKA

<http://pfam.wustl.edu/hmmsearch.shtml>



Domain structure of proteins in the laminin_Nterm Seed alignment
LMG1_MOUSE P02468 LAMININ GAMMA-1 CHAIN PRECURSOR
 (LAMININ B2 CHAIN).

<http://www.rcsb.org/>
 Search for laminin

Select 1TLE

Glycoprotein **date** Jan 26, 1996

title Le (Laminin-Type Egf-Like) Module Giii4 In Solution At Ph 3.5 And 290 K, Nmr, 14 Structures

authors R.Baumgartner, M.Czisch, U.Mayer, E.Poschl, R.Huber, R.Timpl, T.A.Holak

compound **source**

Laminin effects cell morphology, movement and survival

A

30 μm 20 μm 5 μm
 30 μm 10 μm 40 μm

Integrin signalling

B

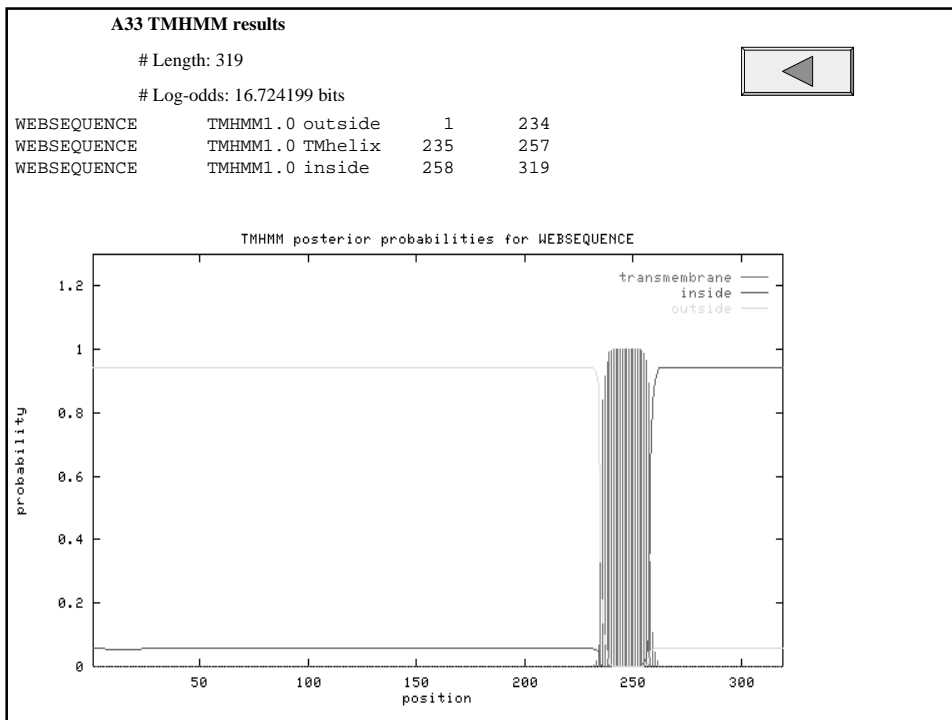
Adhesive island area (μm^2)	Apoptosis (% labeled nuclei)	DNA synthesis (% labeled nuclei)
0	~20	~0
~200	~5	~2
~500	~3	~5
~1000	~2	~10
~1500	~1	~25
~2000	~0.5	~45
~2500	~0.2	~55

Geometric Control of **Cell** Life and Death Christopher S. Chen, Milan Mrksich, Sui Huang, George M. Whitesides, and Donald E. Ingber *Science* 1997 May 30; 276: 1425-1428.

Search for Transmembrane Helix.....

<http://www.cbs.dtu.dk/services/TMHMM-2.0>

MVGKMWPVLW TLCAVRVTVD AISVETPQDV LRASQGKSVT
LPCTYHTSTS SREGLIQWDKLLLTHTERVV IWPFSNKNYI
HGELYKNRVS ISNNAEQSDA SITIDQLTMA DNGTYECSVS
LMSDLEGNTK SRVRLVLVP PSKPECGIEG ETIIGNNIQL
TCQSKEGSPT PQYSWKRYNILNQEQLAQP ASGQPVSLKN
ISTDTSGYYI CTSSNEEGTQ FCNITVAVRS PSMNVALYVG
IAVGVVAALI IIGIIIYCCC CRGKDDNTED KEDARPNREA
YEEPPEQLRE LSREREEEDDYRQEEQRSTG RESPDHL DQ



PSORT II.....ABELSON TYROSINE KINASE

ABLI_HUMAN

Input Sequence

QUERY (1130 aa)

MLEICLKLVG CKSKKGLSSS SSCYLEEALQ RPVASDFEPQ GLSEARWNS
KENLLAGPSE NDPNLFVALY DVFASGDNTL SITKGEKLRV LGYNHNGEWC
EAQTKNGQG WPSNYITPVN SLEKHSWYHG PVS RNAAEYL LSSGINGSFL
VRESESSPGQ RSISLRYEGR VYHYRINTAS DGKLYVSSSES RFNTLAEVLH
HHSTVADGLI TTLHPAPKR NKPTVYGVSP NYDKWEMERT DITMKHKLGG
GQYGEVYEGV WKKYSLTVAV KTLKEDTMEV EEFLKEAAM KEIKHPNLVQ
LLGVCTREPP FYIITEFMTY GNLLDYLREC NRQEVNAVVL LYMATQISSA
MEYLEKKNFI HRDLAARNCL VGENHLVKVA DFGLSRLMTG DTYTAHAGAK
FPIKWTAPES LAYNKFSIKS DVWAFGVLLW EIATYGMSPY PGIDLSQVYE
LLEKDYRMER PEGCPEKVYE LMRACWQNP SDRPSFAEIH QAFETMFQES
SISDEVEKEL GKQVVRGAVS TLLQAPELPT KTRTSRRAAE HRDTPVPEM
PHSGKQGESD PLDHEPAVSP LLPRKERGPP EGGLNEDERL LPKDKKTNLF
GEEEGGSSS KRFLRSCSAS CVPHGAKDTE WRSVTLPRDL QSTGRQFDSS
SALIKKKKKT APTPPKRSSS FREMDGQPER RGAGEEEGRD ISNGALAFTP
LDTADPAKSP KPSNGAGVFN GALRESGGSG FRSPHLWKKs STLTSSRLAT
TFGHHKSEKP ALPRKRAGEN RSDQVTRGTV TPPPRLVKKN EEAADVEFKD
IMESSPGSSP PNLTTPKPLRR QVTVPASGL PHKEEAEEKS ALGTPAAAE
VTPTSKAGSG APGGTSKGA EESRVRHKKH SSESFGRDKG KLSRLKPAPP
PPPAASAGKA GPKPSQSPSQ EAAGEAVLGA KTKATSLVDA VNSDAAKPSQ
PGEGLKPVLPATPKPQSAK PSQTPISPAP VPSTLPSASS ALAGDQPSST
AFIPLISTRV SLRKRQPPPE RIASGAIKGG VVLDSTEALC LAISRNSEQM
ASHSAVLEAG KNLYTFCVSY VDSIQQMRNK FAFREAINKL ENNLRELQIC
PATAGSGPAA TQDFSKLLSS VKEISDIVQR

Results of the k-NN Prediction

k = 9/23

87.0 % nuclear

4.3 %: plasma membrane

4.3 %: vesicles of secretory system

4.3 %: cytoskeletal

prediction for QUERY is

nucleus

PSORTII.....FUMARATE HYDRATASE

FUMH_HUMAN

Input Sequence

QUERY (510 aa)

MYRALRLLAR SRPLVRAPAA ALASAPLGG AAVPSFWPPN AARMASQNSF
RIEYDTFGEL KVPNDKYYGA QTVRSTMNFK IGGVTERMPT PVIKAFGILK
RAAAEVNQDY GLDPKIANAI MKAADVAEG KLNDFPLVV WQTGSGTQTN
MNVNEVISNR AIEMLGELG SKIPVHPNDH VNKSQSSNDT FPTAMHIAAA
IEVHEVLLPG LQKLHDALDA KSKEFAQIIK IGRTHQDAV PLTLGQEFSG
YVQVQKYAMT RIKAAMPRIY ELAAGGTAVG TGLNTRIGFA EKVAAKVAAL
TGLPPVTAPN KFEALAAHDA LVELSGAMNT TACSLMKIAN DIRFLGSGPR
SGLGELILPE NEPGSSIMPG KVNPTQCEAM TMVAAQVMGN HVAVTVGGSN
GHFELNVFKP MMIKNVLHSA RLLGDASVSF TENCVVGIQA NTERINKLMN
ESLMLVTALN PHIGYDAAK IAKTAHNGS TLKETAIELG YLTAEQFDEW
VKPKDMLGPK

Results of the k-NN Prediction

78.3% mitochondrial

8.7 %: cytoplasmic

8.7 %: nuclear

4.3 %: plasma membrane

prediction for QUERY is


mitochondrial

Results of Subprograms

PSG: a new signal peptide prediction method
N-region: length 10; pos.chg 3; neg.chg 0
H-region: length 1; peak value -9.10
PSG score: -13.50

GvH: von Heijne's method for signal seq. recognition
GvH score (threshold: -2.1): -3.16
possible cleavage site: between 31 and 32

>>> Seems to have no N-terminal signal peptide



PSORTII.....ALCOHOL DEHYDROGENASE GAMMA1
ADHG_HUMAN

Input Sequence
QUERY (374 aa)

```
STAGKVIKCK AAVLWELKKP FSIEEVEVAP PKAHEVRIKM VAAGICRSDE
HVVSGNLVTP LPVILGHEAA GIVESVGEGV TTVKPGDKVI PLFTPQCGKC
RICKNPESNY CLKNDLGNPR GTLQDGTRRF TCSGKPIHFF VGVSTFSQYT
VVDENAVAKI DAASPLEKVC LIGCGFSTGY GSAVKVAKVT PGSTCAVFGL
GGVGLSVVMG CKAAGAARII AVDINKDKFA KAKELGATEC INPQDYKKPI
QEVLKEMTDG GVDFSFVIG RLDTMASLL CCEACGTSV IVGVPPDSQN
LSINPMLLLT GRTWKAIFG GFKSKESVPK LVADFMAKKF SLDALITNIL
PFEKINEGFD LLRSGKSIRT VLTF
```

Results of Subprograms
*** Warning: 1st aa is not methyonine

PSG: a new signal peptide prediction method
N-region: length 10; pos.chg 3; neg.chg 0
H-region: length 5; peak value -1.94
PSG score: -6.34


GvH: von Heijne's method for signal seq. recognition
GvH score (threshold: -2.1): -10.00
possible cleavage site: between 48 and 49

Results of the k-NN Prediction

33.3 %: endoplasmic reticulum
22.2 %: mitochondrial
22.2 %: cytoplasmic
11.1 %: Golgi
11.1 %: nuclear

Prediction for QUERY is

endoplasmic reticulum



Metabolic Databases

<http://biocyc.org/>

See Trends in Biochemical Sciences 23:114-11(1998)

EcoCyc <http://biocyc.org/ecocyc/>

KEGG <http://www.genome.ad.jp/kegg/>

ENZYME <http://www.expasy.org/enzyme/>


METACYC <http://biocyc.org/metacyc/>

See also: Cell signaling alliance

<http://www.cellularsignaling.org/>

Signal Transduction Knowledge Environment

<http://stke.sciencemag.org/>

Molecular Pathology Resources 

<http://www.icr.ac.uk/molcarc/teams/molpath.html>

National Centre for Biotechnology Information


<http://www.ncbi.nlm.nih.gov/>

Genes and Diseases [OMIM]

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>


<http://www.ncbi.nlm.nih.gov/disease/>

Cancer

<http://www.ncbi.nlm.nih.gov/disease/Cancer.html> 

Colon Cancer <http://www.ncbi.nlm.nih.gov/disease/Colon.html>

MLH1...DNA Mismatch repair enzyme

OMIM [Online Mendelian Inheritance in Man] 

Review of Colon Cancer Biology and Medicine

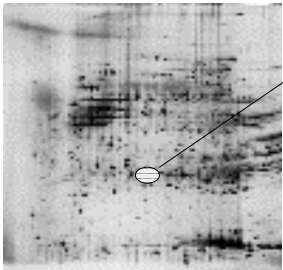
<http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmm?114500>

PROTEOME : 2D-GEL DATABASES

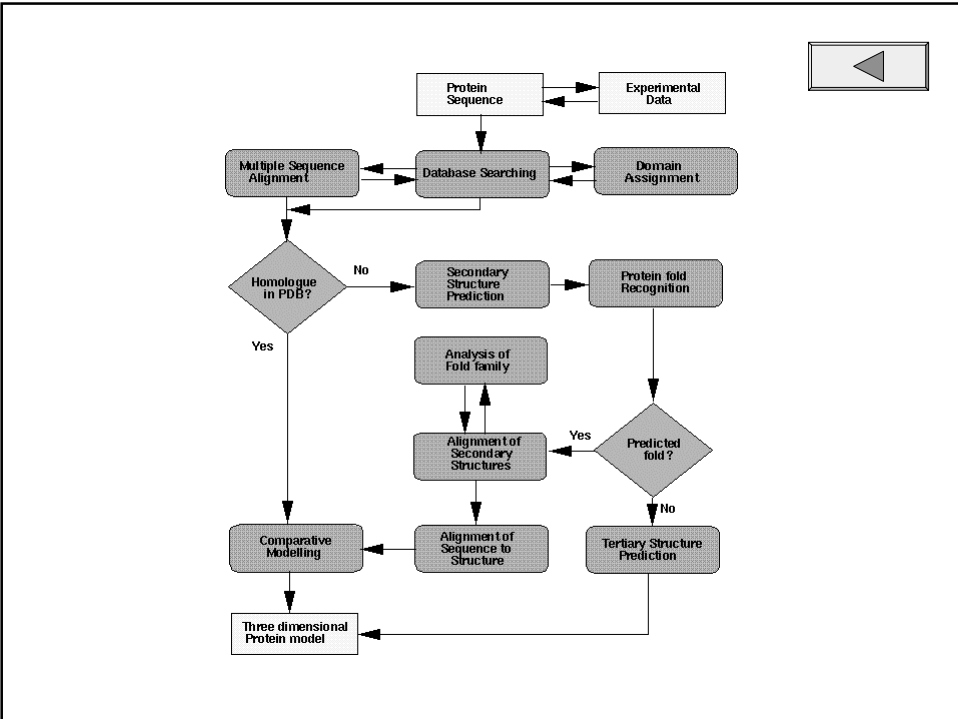
<http://www.expasy.ch/cgi-bin/map1>

SWISS-2DPAGE Map Selection

Colorectal_epithelial_cells

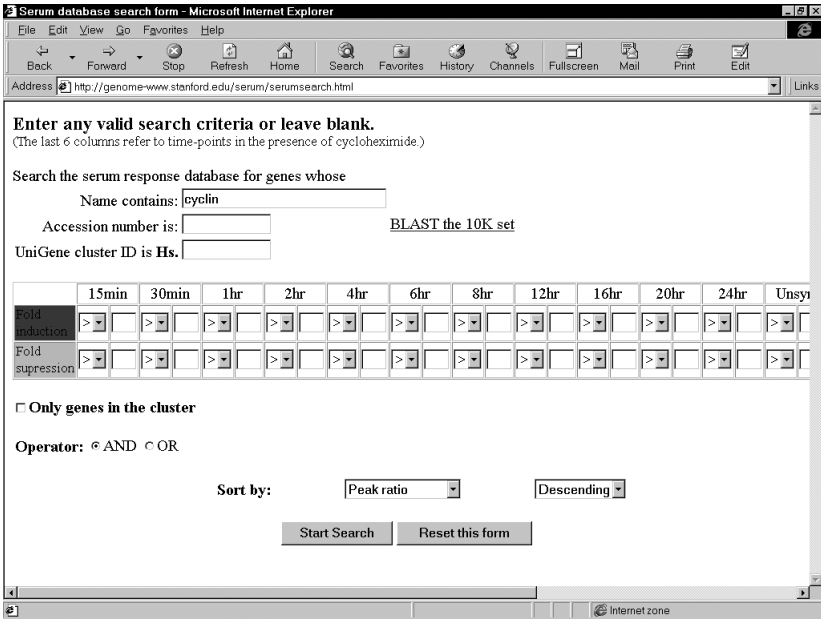



Protein Relational Database



Regulatory DNA Sequences
Given a single DNA sequence common applications are - to identify restriction enzyme cutting sites, to aid in various cloning experiments; to identify transcription factor sites to further the understanding of gene regulation; and to translate any open reading frames into their amino acid equivalents. Long genomic sequences derived during sequencing projects require analysis to separate the intron and exon regions and identify genes

- Restriction enzyme analysis
- Transcription initiation (factors and sites) analysis ←
- Miscellaneous sequence features and utilities
- Translation and complementation
- RNA secondary structure
- Finding genes, coding regions and ORFs
- Genetic codes and codon usage tables
- PCR methods
- Other lists for DNA/RNA sequence analysis
-
- **Restriction enzyme analysis**



Enter any valid search criteria or leave blank.
(The last 6 columns refer to time-points in the presence of cycloheximide.)

Search the serum response database for genes whose

Name contains:

Accession number is: [BLAST the 10K set](#)

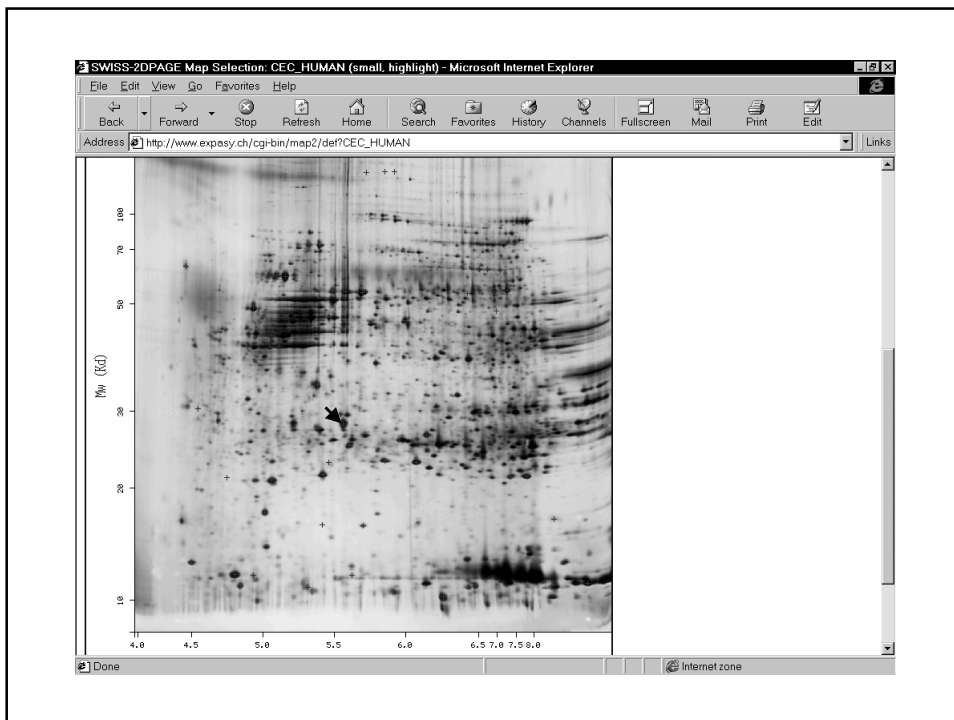
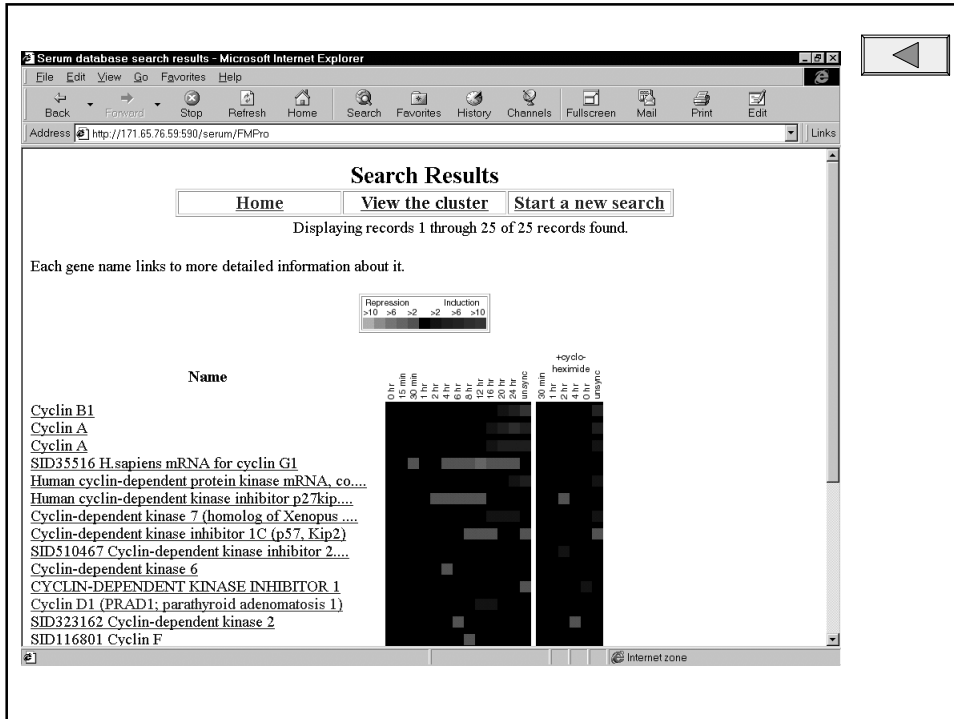
UniGene cluster ID is Hs.

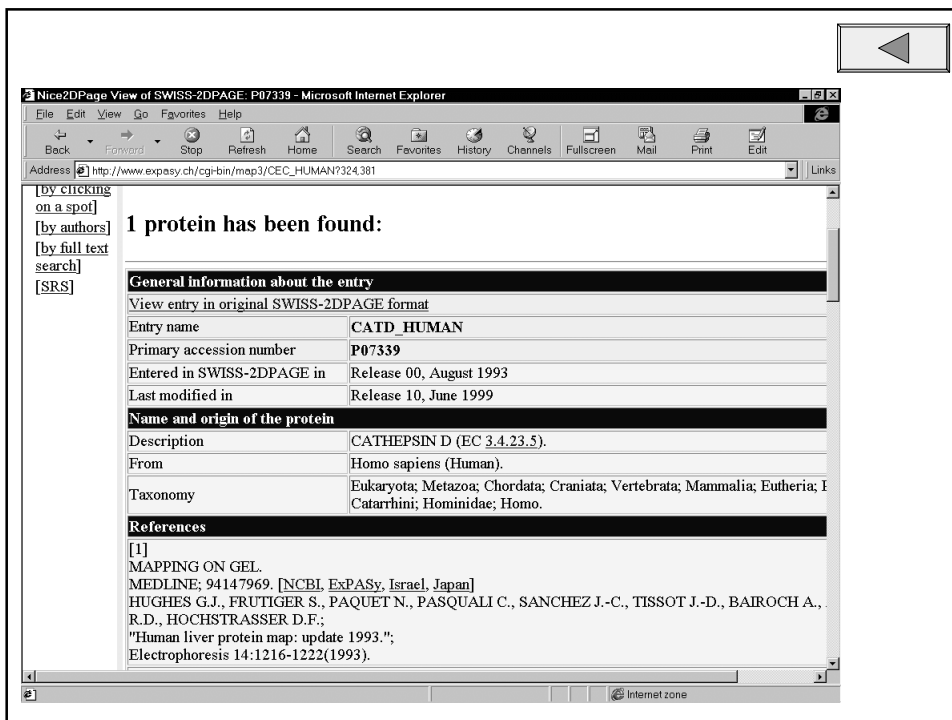
	15min	30min	1hr	2hr	4hr	6hr	8hr	12hr	16hr	20hr	24hr	Unsyn
Fold induction	>	>	>	>	>	>	>	>	>	>	>	>
Fold suppression	>	>	>	>	>	>	>	>	>	>	>	>

Only genes in the cluster

Operator: AND OR

Sort by:





Future Directions

- Completion of genetic maps
- Automated mutation detection
- Transcriptional Profiling
- Bioproteomics including
- Drug Design
- Prediction of shape/structure....**Homology**---->ab initio
- Interaction/ Pathway Databases ((<https://www.biomedcentral.com/news/20010315/01/>)
- Biological Response Algorithms....Molecular
 - Cellular
 - Physiology/Pathology/Medicine
- Simulation of Molecular and Cellular Differentiation and Proliferation