



251 300  
TGQYVGKVL L SQNGTTPRAV GVEFGTHKGN THNVYAKHEV LLAAGSAVSP  
TGQMVGKVL F KQTASGPQAV GVNFGTNKAV NFDVFAKHEV LLAAGSAISP  
TGQ.VGKVL. .Q....P.AV GV#FGT.K.. ..#V%AKHEV LLAAGSA!SP

•••• •••• ••••••

YYYYYY

301 350  
TILEYSGIGM KSILEPLGID TVVDLPVGLN LQDQTTATVR SRITSAGAGQ  
LILEYSGIGL KSVLDQANVT QLLDLPVGIN MQDQTTTTVS SRASSAGAGQ  
.ILEYSGIG\$ KS!L#...!. ...DLPVG.N \$QDQTT.TV. SR..SAGAGQ

•  
YYYYYYYYYY

351 400  
GQAAWFATFN ETFGDYSEKA HELLNKLEQ WAEAVARGG FHNTTALLIQ  
GQAVFFANFT ETFGDYAPQA RDLLNKLKLDQ WAEETVARGG FHNVTALKVQ  
GQA..FA.F. ETFGDY...A .#LLNKL#Q WAE.VARGG FHN.TAL.!Q

• ••• ••••

401 450  
YENYRDWIVN HNVAYSELF L DTAGVASFDV WDLLPFTRGY VHILDKDPYL  
YENYRNWLLD EDVAFAELFM DTEGKINF DL WDLIPFTRGS VHILSSDPYL  
YENYR#W..# .#VA%.ELF\$ DT.G...FD. WDL.PFTRG. VHIL..DPYL

• • •

451 500  
HHFAYDPQYF LNELDLLGQA AATQLARNIS NSGAMQTYFA GETIPGDNL A  
WQFANDPKFF LNEFDLLGQA AASKLARDLT SQGAMKEYFA GETLPGYNL V  
..FA.DP.%F LNE.DLLGQA AA..LAR#.. ..GAM..YFA GET.PG.NL.

• •••••

501 550  
YDADLSAWTE YIPYHFRPNY HGVGTCSMMP KEMGGVVDNA ARVYGVQGLR  
QNATLSQWSD YVLQNF RPNW HAVSSCSMMS RELGGVVDAT AKVYGTQGLR  
.#A.LS.W.# Y!...FRPN. H.V..CSMM. .E\$GGVVD.. A.VYG.QGLR

• •••

551 588  
VIDGSIPPTQ MSSHVMTVFY AMALKISDAI LEDYASMQ  
VIDGSIPPTQ VSSHVMTIFY GMALKVADAI LDDYAKSA  
VIDGSIPPTQ .SSHVMT!FY .MALK!.DAI L#DYA...

•• •••••

•

## *Identifizierung von Sequenzmustern mit Prosite*

Die Prosite Datenbank enthält eine Zusammenstellung von Sequenzmustern, die einer definierten Funktion zugeordnet werden können oder in Proteinfamilien konserviert sind.

### Literatur

1. M. A. S. Saqi, Identification of protein sequence motifs, in: Protein Structure Prediction (Ed. M. J. E. Sternberg), Oxford University Press, Oxford 1996.
2. K. Hofmann et al., The PROSITE database, ist status in 1999, Nucleic Acids res. 27, 215 (1999).

### Syntax

- Jede Sequenzposition im Motiv ist durch einen Bindestrich getrennt
- Ein Buchstabe steht für eine Aminosäure, die in dieser Position vorkommt.
- {...} steht für einen Satz von verbotenen Aminosäuren
- [...] steht für einen Satz von erlaubten Aminosäuren
- (n) steht für eine n-fache Wiederholung
- (n,m) steht für eine n bis m-fache Wiederholung
- x eine beliebige Aminosäure

### Beispiel

[AG]-x(4)-G-K-[ST]

### Signifikanz der Prosite Resultate

(YA)-x-D-T-D-S-(LIVM)

Relative Häufigkeit der Aminosäuren in Sequenzdatenbanken

Y	Tyr	0.032	A	Ala	0.077
D	Asp	0.052	T	Thr	0.058
S	Ser	0.071	L	Leu	0.091
I	Ile	0.054	V	Val	0.065
M	Met	0.023			

Die Wahrscheinlichkeit für das zufällige Auftreten dieses Motivs beträgt

$$(0.032+0.077) \cdot 0.052 \cdot 0.058 \cdot 0.052 \cdot 0.071 \cdot (0.091 + 0.054 + 0.065 + 0.023) = 2.83 \cdot 10^{-7}$$

Wenn man eine Datenbank mit  $10^9$  Sequenzen durchsucht, beträgt die erwartete Anzahl von Treffern 283.

### *Prosites Pattern Search für die Glucose Oxidase*

N-glycosylation site.

N-{P}-[ST]-{P}.

Randomized probability: 5.138e-03 .

Site : 43 to 46 NISV. Identity.  
Site : 89 to 92 NQTA. Identity.  
Site : 161 to 164 NASC. Identity.  
Site : 168 to 171 NGTV. Identity.  
Site : 258 to 261 NGTT. Identity.  
Site : 355 to 358 NETF. Identity.  
Site : 388 to 391 NTTA. Identity.  
Site : 473 to 476 NISN. Identity.

-----  
Protein kinase C phosphorylation site.

[ST]-x-[RK].

Randomized probability: 1.423e-02 .

-----  
Casein kinase II phosphorylation site.

[ST]-x(2)-[DE].

Randomized probability: 1.482e-02 .

-----  
N-myristoylation site.

G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.

Randomized probability: 1.397e-02 .

-----  
Regulator of chromosome condensation (RCC1) signature 2.

[LIVMFA]-[STAGC](2)-G-x(2)-H-[STAGLI]-[LIVMFA]-x-[LIVM].

Randomized probability: 1.710e-06 .

-----  
GMC oxidoreductases signature 1.

Prosite access number: PS00623

Prosite documentation access number: PDOC00543

[GA]-[RKN]-x-[LIV]-G(2)-[GST](2)-x-[LIVM]-N-x(3)-[FYWA]-x(2)-  
[PAG]-x(5)-[DNESH].

Randomized probability: 8.116e-11 .

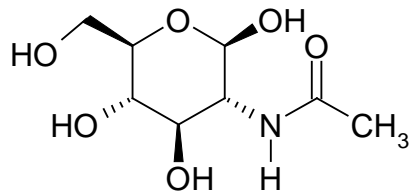
Site : 97 to 120 GNGLGGSTLVNGGTWTRPHKAQVD. Identity.  
-----

GMC oxidoreductases signature 2.  
 Prosite access number: PS00624  
 Prosite documentation access number: PDOC00543  
 [GS]-[PSTA]-x(2)-[ST]-P-x-[LIVM](2)-x(2)-S-G-[LIVM]-G.  
 Randomized probability: 1.087e-09 .

Site : 290 to 304 GSAVSPTILEYSGIG. Identity.

-----  
 7 different patterns found.

N-Acetylglucosamin ist an 7 von potentiellen 8 der Glycosilierungspositionen der Glucose Oxidase gebunden



#### GMC-Oxidoreduktasen

Glucose Oxidase :	Glucose + O <sub>2</sub>	->	Gluconolacton + H <sub>2</sub> O <sub>2</sub>
Methanol Oxidase:	Methanol + O <sub>2</sub> CH <sub>3</sub> OH	->	Acetaldehyd + H <sub>2</sub> O <sub>2</sub> CH <sub>3</sub> CHO
Cholin Dehydrogenase:	Cholin + ? <sup>+</sup> N(CH <sub>3</sub> ) <sub>3</sub> -(CH <sub>2</sub> ) <sub>2</sub> -OH	->	Betain Acetaldehyd + ? <sup>+</sup> N(CH <sub>3</sub> ) <sub>3</sub> -CH <sub>2</sub> -CHO
Glucose Dehydrogenase:	Glucose + ?	->	Gluconolacton + ?

## *Bisherige Ergebnisse der Sequenzanalyse der Glucose Oxidasen*

### Paarweises Sequenzalignment

- Die FAD-bindenden Aminosäuren sind weitgehend konserviert, ebenso die Aminosäuren, die an der Substraterkennung bzw. am Mechanismus beteiligt sind.
- Die Sequenzidentität liegt bei 65%.
- Wegen der hohen Sequenzidentität lassen sich die funktionell relevanten Aminosäuren nur schwer identifizieren.

### Prosite Pattern Search

- Die Proteine lassen sich den GMC-Oxidoreduktasen zuordnen.
- Die GMC-Signaturen überlappen teilweise in der Sequenz mit den den FAD-bindenden Aminosäuren.
- In der *Aspergillus niger* Sequenz finden sich 8 potentielle Glycolysierungsstellen, die zu 50% auch in der *Penicilium amagasakiense* Sequenz vorkommen.
- Aminosäuren, die am Mechanismus und an der Substratbindung beteiligt sind, werden nicht erkannt.
- Es werden weitere Sequenzmuster gefunden, deren Funktion im Zusammenhang mit der Glucose Oxidase bisher nicht fest steht.



201 250  
 ALOX\_PICAN QRPCN..... ..NRDLHGF DGPIKVS... FGNYTYPTCQ DFLRAAESQG  
 BETA\_ECOLI ..DMG..... ..ENDYHGG DGPVSVTT.. SKPGVNPLFE AMIEAGVQAG  
 DHGL\_DROPS ELDAV..... ..GTEYHAK GLLLPVGK.. F.PYNPPLSY ALLKAGEEMG  
 GOX\_ASPNG APNAKQIAAG HYFNASCHGV NGTVHAGPRD TGDDYSPIVK ALMSAVEDRG  
 Consensus ...a..... ..n...Hg. .G.v.vg... .g....P... al..A.e..G



251 300  
 ALOX\_PICAN IPVVDLEDF KTSHGAEHWL KWINRDLGRR SDSAHAYVHP TMRNKQSLFL  
 BETA\_ECOLI YPRTDDLNGY QQEGFGP..M DRTVTPQGRR ASTARGYLDQ A.KSRPNLTI  
 DHGL\_DROPS FS.VQDLNGQ NSTGFMI..A QMTAR.NGIR YSSARAFLRP A.RMRNNLHI  
 GOX\_ASPNG VPTKKDFGCG DPHGVSM..F PNTLHEDQVR SDAAREWLLP N.YQRPNLQV  
 Consensus .p...Dl... ..g..... ..t....g.R sd.Ar..l.p ....rpnL..

301 350  
 ALOX\_PICAN ITSTKCDKVI IEDGKAVAVR TVPMKPLNPK KPVSRTFRAR KQIVISCGTI  
 BETA\_ECOLI RTHA.MTDHI IFDG.K...R AVGVWLEGD STIPTRATAN KEVLLCAGAI  
 DHGL\_DROPS LLNTTVTKVL IHPGTK...N VGVVEVSDQF GSM.RKILVK KEVIVSGGAV  
 GOX\_ASPNG LTGQYVGKVL LSQNGTTP.R AVGVVEFGTHK GNT.HNVYAK HEVLLAAGSA  
 Consensus lt...v.kvl i..g.....r aVgve....k g.....ak k#!ll.aG..



YY

351 400  
 ALOX\_PICAN SSPLVLQRSG IGAHHLSV GVKPIVDLPG VGENFQDHYC FFTPYVVKPD  
 BETA\_ECOLI ASPQILQRSG VGNAELLAEF DIPLVHELPG VGENLQDHLE MYLQYECKEP  
 DHGL\_DROPS NSPQILLLSG VGPKEQLQKV NVRPVHLLPG VGKNLQNHVA YFTNFFIDDA  
 GOX\_ASPNG VSPTILEYSG IGMKSILEPL GIDTVVDLP. VGLNLQDQTT ATVRSRITSA  
 Consensus .SP.!L..SG !G.k..L... g!..!vdLPg VG.NlQ#h.. .....i..a

●  
 YYYYYYYYYYYYYY

401 450  
 ALOX\_PICAN VPTFDDFVRG DPVAQKAAFD QWYSNKDGPL TTNGIEAGVK IRPTEELAT  
 BETA\_ECOLI VSLYPALQWW N...QPKIGA EWLFGGTGVG ASNHFEAGGF IRSREEFAWP  
 DHGL\_DROPS DTAPLNWATA M...EYLLFR DGLMSGTGI. ..SDVTAKMA TRWADRPNLP  
 GOX\_ASPNG GAGQGQAAWF .....ATFN ETF..GDYSE KAHELLNTKL EQWAEAVAR  
 Consensus .....aw. ....a.f. #....gdg.. .....a... .rwa#e..a.



451 500  
 ALOX\_PICAN ADEDFRGYA EYFENKPKDP LMHYSVISGF FGDHTKIPNG KFMTMFHFLE  
 BETA\_ECOLI NIQYHFLPVA INYNG..... ..SNA VKE.....H GFQCHVGS MR  
 DHGL\_DROPS DLQLYFGGYL ASCAR..... ..TGQ VGELLSNNSR AIQIFPAVLN  
 GOX\_ASPNG GGFHNTTALL IQYENYRDWI VNHNVAISEL FLD.....TA GVASFDVWDL  
 Consensus .....l i.yen..d.. ..h....s.. f.#..... g...f.....



	501				550
ALOX_PICAN	YPFSRGEFVRI	TSANPY.DAP	DFDPGFLNDE	RDLWPMVWAY	KKSRETARRM
BETA_ECOLI	SP.SRGHVRI	KSRDPH.QHP	AILFNYMSHE	QDWQEFRDAI	R.....
DHGL_DROPS	PK.SRGYITL	RSADPL.DPP	RIFANYLTDE	RDVKTLEVEGI	K.....
GOX_ASPNG	LPFTRGYVHI	LDKDPYLHHF	AYDPQYFLNE	LDLLGQAAAT	Q.....
Consensus	.pfsRGy!.i	.s.#Py..hp	a.dp.%...E	.Dl.....a.	.....

●

	551				600
ALOX_PICAN	ESFAGEVTSH	HPLFKVDSPA	RARDLDLETC	SAYAGPKHLT	ANLYHGSWTV
BETA_ECOLI	..ITREIMHQ	PAL.....	.....	.....	.DQYRGR...
DHGL_DROPS	..FAIRLSQT	SPL.....	.....	.....	.KQYGMRLDK
GOX_ASPNG	..LARNISNS	GAM.....	.....	.....	.QTYFAGETI
Consensus	...ar.is..	.a\$......	.....	.....	...Y....t.

●●●●

	601				650
ALOX_PICAN	PIDKPTPKND	FHVTSNQVQL	HSDIEYTEED	DEAIVNYIKE	HTETTWHCLG
BETA_ECOLI	EISPGVECQ.	.....	.....T	DEQLDEFVRN	HAETAHFHPCG
DHGL_DROPS	TVVKGCESHA	F.....	.....AS	DAYWECAVRQ	NTGPENHQAG
GOX_ASPNG	PGDNLAYDAD	L.....	.....	.SAWTEYIPY	HFRPNYHGVG
Consensus	p.d.....d	.....	.....	d.aw.ey!..	h..p..H..G

● ●●●

	651				700
ALOX_PICAN	TCSMAPREGS	KIAPKGGVLD	ARLNyGVQn	LKVADLSVCP	DNVGCNTYST
BETA_ECOLI	TCKMG..YDE	M.....SVVD	GEGRVHGLEG	LRVVDASIMP	QIITGNLNAT
DHGL_DROPS	SCKMGPSHDP	M.....AVVN	HELRVHGVRG	LRVMDTSIMP	KVTAGNTHAP
GOX_ASPNG	TCSMMPKEM.	.....GGVVD	NAARVYGVQG	LRVIDGSIPP	TQMSSHVMTV
Consensus	tCsM.p.e..	.....ggVv#	...rVyGvqg	LrV.D.S!.P	.....n....

●●

●●●●

●

	701				743
ALOX_PICAN	ALTIGEKAAT	LVAEDLGYSG	SDLDMTIPNF	RLGTYEETGL	ARF
BETA_ECOLI	TIMIGEKIAD	MIRGQEALPR	STAGYFVANG	MPVRACK...	...
DHGL_DROPS	AVMIAEKGAY	LLKRAWGAKV	.....	.....	...
GOX_ASPNG	FYAMALKISD	AILEDYASMQ	.....	.....	...
Consensus	...iaeKiad	.i.ed.a...	.....	.....	...

●

## Ergebnisse des multiplen Alignments der GMC-Oxidasen

- 9% der Aminosäuren sind konserviert.
- weitere 24% der Aminosäuren sind in mindestens 2 Enzymen konserviert.
- Ein großer Anteil davon besteht aus Aminosäuren, die bei der Glucos Oxidase das FAD binden.
- Die Glykolisierungsmuster sind nicht konserviert und haben daher keine allgemeine Bedeutung für die GMC-Oxidoreduktasen.
- Substratbindende Aminosäuren sind meist nicht konserviert. Unterschiedliche Substrate bzw. Mechanismen erfordern unterschiedliche Funktionalitäten und Geometrien im aktiven Zentrum.