Protein Engineering vol.3 no.5 pp.433–442, 1990

Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering

Ylva Gavel and Gunnar von Heijne[1]

Research Group for Theoretical Biophysics, Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm and [1]Department of Molecular Biology, Karolinska Institute, Center for Biotechnology, NOVUM, S-141 52 Huddinge, Sweden

In *N*-glycosylated glycoproteins, carbohydrate is attached to Asn in the sequence Asn-X-Ser/Thr, where X denotes any amino acid. However, the presence of this consensus peptide does not always lead to glycosylation. We have compiled an extensive collection of glycosylated and non-glycosylated Asn-X-Thr/Ser sites and present a statistical study based on this data set. Our results indicate that non-glycosylated sites tend to be found more frequently towards the C termini of glycoproteins, and that proline residues in positions X and Y in the consensus Asn-X-Thr/Ser-Y strongly reduce the likelihood of *N*-linked glycosylation. Beyond this, there are no obvious local sequence features that seem to correlate with the absence or presence of *N*-linked glycosylation. These findings are discussed in terms of the prediction and engineering of glycosylation sites in secretory proteins.

Key words: glycoproteins/*N*-glycosylation/statistical study

Introduction

The attachment of *N*-linked carbohydrates to proteins is thought to occur during or shortly after translocation of the nascent chain into the lumen of the endoplasmic reticulum (Kaplan *et al.*, 1987; Lennarz, 1988; Hubbard and Ivatt, 1981). The oligosaccharide chain is transferred by the enzyme oligosaccharyl transferase to the asparagine in the consensus tripeptide Asn-X-Thr/Ser, where X is any amino acid (Marshall, 1972). Most putative acceptor sites that become exposed on the lumenal side of the endoplasmic reticulum (ER) membrane are efficiently glycosylated, but some are never used. It has long been known that glycosylation is blocked when X is a proline (Mononen and Karjalainen, 1984), but this rule accounts for only a minor portion of all known non-glycosylated consensus sites.

Here, we present a study based on a data set of carefully selected glycosylated (gs^+) and non-glycosylated (gs^-) Asn-X-Thr/Ser sites. All gs^- sites included in this set have been checked in the literature. Sites from homologous proteins have been removed, both from the gs^+ and gs^- sets. Statistical methods have been used in order to test the significance of the results.

Our analysis indicates that glycosylation is strongly inhibited by proline residues both in positions X and Y in the consensus Asn-X-Thr/Ser-Y. Also, non-glycosylated sites tend to be found more frequently towards the C termini of the proteins in our sample, whereas glycosylated sites are rare in this region. These observations allow prediction of gs^+ sites to be made with ~95% confidence, whereas only some 25% of all gs^- sites can be reliably predicted from the primary sequence.

Materials and methods

Sequences of glycoproteins were collected from the literature and from the NBRF-PIR database (George *et al.*, 1986).

Preparation of data sets

Earlier studies of the sequence patterns associated with *N*-linked glycosylation suffer from a number of methodological shortcomings, e.g. small sample sizes, inclusion of sites from homologous proteins, no statistical analysis. But the most obvious weakness is that too little attention has been paid to the collection of proper sets of both gs^+ and gs^- sites, thus precluding any useful comparisons between the two classes of sequences.

gs^- sites must be picked with some care. Many proteins are non-glycosylated merely because they are never exposed to the carbohydrate-transferring enzyme. Asn-X-Thr/Ser sequences from such proteins should not be included in the gs^- set. Since *N*-glycosylation is thought to occur in the lumen of the ER (Kaplan *et al.*, 1987; Lennarz, 1988), gs^- sites from cytoplasmic proteins must be excluded. The same is true of sites from intracellular and transmembrane parts of membrane proteins. Furthermore, sites from proteins produced by cells unable to carry out *N*-glycosylation must be avoided. For these reasons, we have restricted the data set to gs^- sites found in lumenal domains within the sequences of proteins that also contain gs^+ sites and thus are certain to have been exposed to the oligosaccharyl transferase.

In an extensive literature search, we found a total of 55 gs^- sites in proteins known to be *N*-glycosylated. A total of 48 gs^- sites were included in our final data set. The rest (Robinson and Appella, 1979; Takahashi *et al.*, 1984; Van Den Berg *et al.*, 1976, 1977; Beintema, 1985; Havinga and Beintema, 1980) were from highly homologous proteins, mainly in pancreatic ribonucleases.

We also collected ~600 gs^+ sites from the NBRF database and from the literature. Again, obviously homologous proteins were removed in order to avoid distortions of the statistics. The final version of our data set contained 417 gs^+ sites.

All the gs^+ sites were explicitly stated to be glycosylated in the literature or in the database. Some references reported non-glycosylated sites as well. For the rest of the gs^- sites, the absence of sugar could be inferred from experimental data presented in the literature. Some potential glycosylation sites were located in carbohydrate-free tryptic peptides and therefore were not glycosylated. In other cases, it was possible to make assignments based on the results from sequence determinations. With most sequencing methods, a glycosylated residue cannot be detected; instead, a blank appears in the sequence, and the amino acid in this position has to be identified by other means. Therefore, if some of the asparagines found in the Asn-X-Thr/Ser sequences of a protein show up as blanks whereas others do not, those which give an Asn signal can be assumed to be non-glycosylated.

The sites included in the data set are given in Table I. In some

Table I. Data set of *N*-glycosylation sites

Proteins containing both gs^+ and gs^- sites	gs^+	gs^-	Experimental evidence	Reference
Alkaline extracellular protease (*Y.lipolytica*)	123		1	Matoba *et al.* (1988)
		292	3, 5	Davidow *et al.* (1987)
		330	5	
Alkaline phosphatase (human placenta)		122	3	Millan (1986)
	249		2	
Alpha1-antitrypsin (human)	46		1	Carrell *et al.* (1982)
	83		1	Mega *et al.* (1980)
		247	1	Carrell *et al.* (1981)
		390	5	
Alpha-lactalbumin (rabbit)	45		1	Hopp and Woods (1979)
		84	5	
Alpha-lactalbumin (bovine)	42[a]		1	Hopp and Woods (1979)
		73	3	Struck *et al* (1978)
Aspartic proteinase (*Rhizomucor miehei*)	79		1	Boel *et al.* (1986)
	188		1	
		313	3, 5	
Beta-hexosaminidase beta-chain (human)	142		1	Mahuran *et al.* (1988)
		323	4	
	327		1	
Cathepsin D (porcine spleen)	70		1	Shewale and Tang (1984)
	192		1	Takahashi and Tang (1983)
		282	3	
Ceruloplasmin (human)	119		2	Takahashi *et al.* (1984)
		208	3	
	339		2	
	378		2	
		569	3	
	743		2	
		907[a]	3	
Cholinesterase (human serum)	17		1	Lockridge *et al.* (1987)
	57		1	
	106		1	
	241		1	
	256		1	
	341		1	
	455		1	
	481		1	
		485	3	
	486		1	
Deoxyribonuclease (bovine pancreas)	18		1	Liao *et al.* (1973)
		103	5	Catley *et al.* (1969)
				Salnikow *et al.* (1973)
Epidermal growth factor receptor (human)	104		2	Ullrich *et al.* (1984)
	151		2	
	172		2	
		247	4	
	328		1, 2	
	337		2	
	389		2	
	420		2	
	504		2	
	543		2	
	579		2	
	599		2	
Folate-binding protein (cow's mik)	49		1	Hansen *et al.* (1984)
	141		1	
		181	4	
		212	4, 5	
Follistatin (porcine)		95	3, 4	Shimasaki *et al.* (1988)
	259		1	Esch *et al.* (1987)
Glucoamylase G1 (Aspergillus niger)	169		1	Svensson *et al.* (1983)
		180	3, 4, 6	
	393		1	

Table I. Continued

Proteins containing both gs^+ and gs^- sites	gs^+	gs^-	Experimental evidence	Reference
Glycophorin (porcine erythrocyte)		16	4	Honma *et al.* (1980)
	19		1	
	39		1	
Immunoglobulin A, alpha (heavy) chain:		129	5	Taylor and Wall (1988)
(mouse myeloma J558, MOPC 511)	162		1	Tucker *et al.* (1988)
		223	3, 5	Robinson and Appella (1980)
		402	3,4	
	419		1	
Immunoglobulin G, gamma3 (heavy) chain	6		2	Frangione *et al.* (1980)
(human deletion mutant Wis)	140		2	
		235	5	
Immunoglobulin M, mu (heavy) chain	57		1	Kehry *et al.* (1979)
(murine myeloma MOPC 104E)	171		1	Kehry *et al.* (1982)
		263	3	
	332		1	
	364		1	
	402		1	
	563[a]		1	
Immunoglobulin M, mu (heavy) chain		74	3	Shimizu *et al.* (1971)
(human Ou, Waldenstroms)	170[a]		1	Putnam *et al.* (1973)
	332[a]		1	Aubert *et al.* (1976)
	395		1	
	402[a]		1	
	563[a]		1	
Inter-alpha-trypsin inhibitor, 2nd component		42	3	Gebhard *et al.* (1988)
(human)	64		2	
	617		2	
Lactotransferrin	81		1	Metz-Boutige *et al.* (1981)
		167	4, 5	
	268		1	
		334	4, 5	
Leucine-rich alpha2-glycoprotein (human)	44		2	Takahashi *et al.* (1985)
	151		2	
	234		2	
		271	3	
	290		2	
Lipase (porcine pancreas)	166		1, 2	Bianchetta *et al.* (1979)
		409	4	De Caro *et al.* (1981)
NCA-50 (human)	118		1	Grunert *et al.* (1988)
	139		1	
		254	3, 4	
	258		1	
Ovalbumin (hen)	292		1, 2	Nisbet *et al.* (1981)
		311	3	
Ovotransferrin (hen)	473[a]		1	Williams *et al.* (1982)
		618	5	Kingston and Williams (1975)
		672	5	
Pepsinogen (hen)	113		1	Baudys and Kostka (1983)
		218	3	
Peroxidase (horseradish)	13		1	Welinder (1976)
	57		1	
	158		1	
	186		1	
	198		1	
	214		1	
	255		1	
	268		1	
		286	3	
Ribonuclease (coypu pancreas)	34		1	Van Den Berg *et al.* (1976)
		94	3	
Ribonuclease B (guinea-pig pancreas)	21[a]		1	Van Den Berg *et al.* (1977)
	34		1	
		94[a]	3	

Table I. Continued

Proteins containing both gs^+ and gs^- sites	gs^+	gs^-	Experimental evidence	Reference
Ribonuclease (hippopotamus pancreas)		21	3	Havinga and Beintema (1980)
	34[a]		1	
		34[b]	3	
		62	3	
		76	3	
Ribonuclease (horse pancreas)		22[a]	3	Beintema (1985)
	22[a,b]			
	34[a]		1	
	62		1	
S8-glycoprotein (*Brassica campestris*)[c]	15		1	Takayama *et al.* (1987)
	33		1	
	83		1	
	90		1	
	119		1	
	214		1	
	230		1	
		284	3, 4	
		359	3, 4	
Sex steroid binding protein (human)		244	4, 5	Walsh *et al.* (1986)
	351		1	
	367		1	
Thy-1 glycoprotein (rat brain)	23		1	Campbell *et al.* (1981)
	74		1	
		93	3	
	98		1	
Tissue-type plasminogen activator (human)	117		2	Hansen *et al.* (1988)
	184		2	Pennica *et al.* (1983)
		218	3	
	448		2	
von Willebrand factor (human)	94		1	Titani *et al.* (1986)
	384		1	
		452	3, 4	
	468		1	
	752		1	
	811		1	
	1460		1	
	1527		1	
	1594		1	
	1637		1	
	1783		1	
	1822		1	
		1872	3, 4	
	2027		1	

Proteins containing gs^+ sites only	gs^+ sites excluded due to homology	Reference
120 kd lysosomal membrane glycoprotein (rat)		Howe *et al.* (1988)
		Lewis *et al.* (1985)
7s nerve growth factor gamma chain (mouse)		NBRF
Alpha1-acid glycoprotein (human)		NBRF
Alpha1-B glycoprotein (human)		NBRF
Alpha2-macroglobulin (human)		NBRF
Antithrombin-III (human)		NBRF
Apolipophorin-III (migratory locust)		Kanost *et al.* (1988)
Apolipoprotein D (human)		Drayna *et al.* (1986)
Avidin (chicken)		NBRF
Beta2-glycoprotein I (human)		NBRF
Biliary glycoprotein I (human)		Hinoda *et al.* (1988)
C1 inhibitor (human)		Bock *et al.* (1986)
Calsequestrin (rabbit)		Fliegel *et al.* (1987)
Carboxypeptidase Y (yeast)		Svendsen *et al.* (1982)
Cathepsin B (rat)		NBRF
Cathepsin H (rat)		NBRF

Table I. *Continued*

Proteins containing gs^+ sites only	gs^+ sites excluded due to homology	Reference
Cathepsin L (rat)		Ishidoh *et al.* (1987)
Chorionic gonadotrophin α-subunit (human)		Morgan *et al.* (1975)
Chorionic gonadotrophin β-subunit (human)		Morgan *et al.* (1975)
Coagulation factor VII (bovine)		Takeya *et al.* (1988)
Coagulation factor IX (bovine)	261[a]	NBRF
Coagulation factor X (bovine		NBRF
Coagulation factor XI (human)		NBRF
Coagulation factor XIIa (human)		McMullen and Fujikawa (1985)
		Fujikawa and McMullen (1983)
Collagen II alpha1 chain (chicken)		NBRF
Collagen IV alpha1-subunit (human)		Brazel *et al.* (1987)
Collagenase (human)		NBRF
Complement component C1q A-chain (human)		Reid *et al.* (1982)
Complement component C1r A-chain (human)		Tosi *et al.* (1987)
		Spycher *et al.* (1986)
Complement component C3 (human)		de Bruijn and Fey (1985)
Complement component C5		NBRF
Complement factor B (human)		NBRF
C-reactive protein 1.1 (*Limulus polyphemus*)		NBRF
Erythropoietin (human)		Jacobs *et al.* (1985)
Favin beta-subunit (*Vicia faba*)		Hemperley *et al.* (1979)
		Hopp *et al.* (1982)
Fibrinogen beta chain (human)		NBRF
Fibrinogen gamma-A chain (human)		NBRF
Fibrinogen gamma chain (lamprey)		Strong *et al.* (1985)
Fibrinopeptide B (sea lamprey)		NBRF
Glucosylceramide activator protein (human)		NBRF
Glycophorin A (human)		NBRF
Glycophorin C (human)		NBRF
gp71A protein (Friend leukaemia virus)		NBRF
H-2Db class I antigen alpha chain (mouse)		Reyes *et al.* (1982)
		Maloy and Coligan (1982)
Haemagglutinin HA1 chain (influenza)		Ward and Dopheide (1981)
		Ward *et al.* (1980)
Haemagglutinin HA2 chain (influenza)		Ward and Dopheide (1981)
		Ward *et al.* (1980)
Haptoglobin-1 (human)		NBRF
Hemocyanin a chain (lobster)		NBRF
Hemopexin (human)		NBRF
Hepatic lectin (chicken)		NBRF
HLA class I histocompatibility antigen alpha chain (human)		NBRF
HLA class II histocompatibility antigen DC β-chain (human)		NBRF
HLA class II histocompatibility antigen DR α-chain (human)		NBRF
Hydroxymethylglutaryl CoA reductase (hamster)		Liscum *et al.* (1985)
Immunoglobulin D, delta (heavy) chain (human)		Takahashi *et al.* (1982)
Immunoglobulin, heavy chain V-III region (human)		NBRF
Immunoglobulin, κ-light chain V-I region (human)		NBRF
Immunoglobulin G, lambda (light) chain (human Sm)		Garver *et al.* (1975)
Immunoglobulin G, lambda (light) chain (human NEI)		Garver and Hilschmann (1972)
Immunoglobulin G1, heavy chain V-II region (human)		NBRF
Inhibin beta A chain precursor (pig)		NBRF
Insulin receptor (human)		Ebina *et al.* (1985)
		Ullrich *et al.* (1985)
Interferon gamma (human)		NBRF
Invertase (yeast)		NBRF
J-chain (human)		NBRF
Lectin (*Dolihus biflorus* seed)		Schnell and Etzler (1987)
Leukocyte adhesion glycoprotein Mac-1 (human)		Corbi *et al.* (1988)
Link protein (rat)		Neame *et al.* (1986)
Lymphotoxin (human)		NBRF
Lysosomal membrane glycoprotein lamp-2 (human)		Fukuda *et al.* (1988)
Lysosomal membrane glycoprotein lamp-A (human)		Viitala *et al.* (1988)
M1-1 protoxin (yeast)		NBRF
Macrophage-lymphocyte Fc receptor (mouse)		Lewis *et al.* (1986)
MRC OX-2 antigen (rat)		Clark *et al.* (1986)
		NBRF

Table I. *Continued*

Proteins containing gs^+ sites only	gs^+ sites excluded due to homology	Reference
Mylein P0 protein (rat)		NBRF
Neutrophil elastase (human)		Takahashi *et al.* (1988)
Nidogen (mouse)		Mann *et al.* (1989)
Peroxidase (turnip)		NBRF
Phaseolin (*Phaseolus vulgaris*)		Slightom *et al.* (1985)
Pheromone prepro-alpha-factor (yeast)		Waters *et al.* (1988)
		Emter *et al.* (1983)
		Julius *et al* (1984)
		Kurjan and Herskowitz (1982)
Phospholipase A2 (honeybee)		NBRF
Phospholipase A2 (Australian taipan)		NBRF
Plasma kallikrein (human)		NBRF
Plasminogen (bovine)		NBRF
Platelet-derived growth factor A chain (human)		NBRF
Platelet glycoprotein Ib alpha chain (human)		Titani *et al.* (1987)
Poly-immunoglobulin receptor (human)		NBRF
Postheparin plasma hepatic triglyceride lipase (human)		Martin *et al.* (1988)
Prolactin receptor (rat)		Boutin *et al.* (1988)
Proopiomelanocortin (*Xenopus laevis*)		Martens (1986)
Prostatein (rat)		Peeters *et al.* (1981)
Protective protein (human)		Galjart *et al.* (1988)
Protein C light chain (bovine)		Fernlund and Stenflo (1982)
Protein C heavy chain (bovine)		Stenflo and Fernlund (1982)
Protein Z (bovine)		NBRF
Proteoglycan core protein (human)		Krusius and Ruoslahti (1986)
Prothrombin (bovine)		MacGillivray and Davie (1984)
Retinoblastoma-associated protein (human)		NBRF
Rhodopsin (bovine)		Ovchinnikov (1982)
		Hargrave (1977)
Riboflavin-binding protein (chicken)		Hamazume *et al.* (1984)
Ribonuclease, secretory (human urine)	34[a], 76[a]	Beintema *et al.* (1988)
Ribonuclease (porcine pancreas)	21[a], 3[a]	Jackson and Hirs (1970)
Ricin D, B chain (castor bean)		NBRF
Stellacyanin (Japanese lacquertree)		NBRF
Structural glycoprotein E1 (SFV)	141[a]	Garoff *et al.* (1980)
Structural glycoprotein E2 (SFV)	262[a]	Garoff *et al.* (1980)
Structural glycoprotein E1 (Sindbis virus)		Rice and Strauss (1981)
Structural glycoprotein E2 (Sindbis virus)		Rice and Strauss (1981)
Structural glycoprotein E3 (Sindbis virus)		Rice and Strauss (1981)
Stylar glycoprotein 2 (winged tobacco)		NBRF
T-cell receptor alpha chain C region (mouse)		NBRF
T-cell receptor alpha chain V region (human)		NBRF
T-cell receptor beta chain V region (mouse)		NBRF
T-cell surface glycoprotein CD4 (human)		NBRF
T-cell surface glycoprotein CD8 (mouse)		NBRF
Thyroxine-binding globulin (human)		Flink *et al.* (1986)
		Zinn *et al.* (1978)
Tissue factor (human)		Spicer *et al.* (1987)
Transferrin receptor (human)		Schneider *et al.* (1984)
Transforming growth factor-alpha (human)		Bringman *et al.* (1987)
Transmembrane protein E1 (coronavirus)		Laude *et al.* (1987)
Urokinase-type plasminogen activator (human)		NBRF
Urokinase-type plasminogen activator (pig)		NBRF
Variant surface glycoprotein (*Trypanosoma*)		Bangs *et al.* (1988)
		Rice-Ficht *et al.* (1981)
Vitellogenin (chicken)		van het Schip *et al.* (1987)
Vitronectin (human)		NBRF

(1) According to experimental evidence presented in the reference.
(2) According to experimental evidence cited in the reference.
(3) In the reference, the absence of carbohydrate at this site is explicitly mentioned.
(4) PTH-Asn was detected.
(5) The relevant portion of the protein did not contain carbohydrate.
(6) The Asn(180)-Gly(181) bond was susceptible to cleavage with hydroxylamine.
[a]Among the other sites listed, there is at least one located in a sequence highly homologous to this one. Therefore, this site has not been included in the *sequence statistics*.

[b]In some molecules, due to amino acid substitution.
[c]In those positions where the amino acid was not identified, the corresponding amino acid of S6-glycoprotein has been used instead.
(NBRF) The NBRF database. See George *et al.* (1986).
When the protein is known to contain a cleaved N-terminal signal sequence, the residues of the prepeptide are given negative numbers, i.e. amino acid number one corresponds to the N terminus of the mature protein.

sequences there are additional potential glycosylation sites that could not be assigned or were discarded because they were located in transmembrane or cytoplasmic domains of the integral membrane proteins. These sites are not mentioned. Known partially glycosylated sites have been counted in the gs^+ set. If a protein contains sites that had to be excluded from the statistics owing to homology, this is noted in the table.

Results

Non-standard sites

In a small number of cases, the sequence around the reported *N*-glycosylated asparagine did not agree with the Asn-X-Thr/Ser consensus. As was also noted by Nakai and Kanehisa (1988), three Asn-X-Cys patterns have been reported as gs^+ sites in the NBRF database. The possibility of carbohydrate attachment at such sites was predicted by Bause and Legler (1981). The Asn-X-Cys sites were found in bovine and human protein C and in human von Willebrand factor. However, we did not find any experimental evidence for glycosylation of the Asn-X-Cys site in the reference given for human protein C (Foster *et al.*, 1985).

We also found another unusual *N*-glycosylation site: in murine IgM heavy chain, carbohydrate is found bound to asparagine in the sequence Asn-Gly-Gly-Thr. A similar site has been reported for egg yolk phosvitin, a protein derived from vitellogenin. In this case, the sequence at the point of attachment was reported as Asn-Ser-Gly-Psr, where Psr is phosphoserine (Shainkin and Perlmann, 1971). However, the nucleotide sequence (van het Schip *et al.*, 1987; Byrne *et al.*, 1984) indicates that the site is of the normal Asn-Gly-Ser type.

Thus, although some of the putative non-standard sites may have been erroneously identified, at least a couple remain that seem to be authentic (Titani *et al.*, 1986; Kehry *et al.*, 1979; Stenflo and Fernlund, 1982). In exceptional cases, then, *N*-linked glycosylation does not seem to require the Asn-X-Ser/Thr consensus.

Statistical analysis

As can be seen in Table I, gs^+ sites are far more common in glycoproteins than are gs^- sites. Apparently, if the oligosaccharyl transferase is present, the Asn-X-Thr/Ser signal leads to glycosylation approximately nine times out of 10.

In order to compare gs^+ and gs^- sequences, we extracted 33-residue segments centred around the glycosylation signals listed in Table I. Amino acid distributions were calculated for gs^+ and gs^- sites separately. The results for the residues immediately surrounding the consensus tripeptide are shown in Table II.

According to previous statistical studies, Pro is very rare or even absent in position +1 of gs^+ sites (Mononen and Karjalainen, 1984). The statistical significance of this observation is confirmed by our data (Figure 1; $P < 2 \times 10^{-8}$ as estimated from a binomial distribution with $P = 0.0558$, i.e. the mean frequency of Pro outside positions 0 to +3). Actually, the frequency of glycosylated Asn-Pro-Thr/Ser sites may be even lower since the Pro-containing site in thyroxine-binding globulin may have been erroneously identified as a gs^+ site (an Asn-Cys-Thr acceptor site in position 233 seems to have been overlooked; cf. Flink *et al.*, 1986; Zinn *et al.*, 1978).

In addition to the under-representation of Pro in position +1 of the gs^+ sites, there is another significant ($P < 2 \times 10^{-4}$) frequency drop in position +3. In contrast, for the gs^- sites, Pro is significantly enriched in position +1 ($P < 6 \times 10^{-5}$). The second most proline-rich position in gs^- sequences is +3, although the over-representation in this position is not statistically significant. The +3 pattern has not been noted previously, but

Table II. (a) Amino acid distributions around gs^+ sites

Amino acid	Position										
	−5	−4	−3	−2	−1	0	+1	+2	+3	+4	+5
A	5.8	5.1	5.5	5.0	6.2	0.0	8.4	0.0	7.2	7.7	6.7
C	3.4	5.3	3.4	1.9	1.7	0.0	2.6	0.5	4.3	2.2	3.8
D	4.4	5.3	3.4	3.1	3.8	0.0	4.3	0.0	4.8	4.3	5.5
E	4.6	6.1	4.6	3.6	5.0	0.0	2.9	0.0	6.5	5.8	4.3
F	4.1	4.1	3.6	6.5	4.8	0.0	5.0	0.0	3.6	5.0	4.1
G	6.8	5.6	9.2	5.3	8.6	0.0	11.0	0.2	5.8	6.0	7.2
H	2.7	3.4	2.7	4.1	1.7	0.0	2.4	0.0	2.2	2.6	2.4
I	4.8	4.6	5.1	3.8	2.4	0.0	5.8	0.0	5.5	4.6	5.3
K	4.4	5.6	5.3	6.5	3.8	0.0	4.8	0.0	3.6	4.3	5.0
L	11.4	9.7	9.6	7.2	8.9	0.0	8.4	0.0	9.6	8.2	9.1
M	4.4	1.7	1.7	1.9	2.2	0.0	1.7	0.0	2.2	1.9	1.4
N	3.9	4.4	2.7	5.8	4.8	100.0	4.8	0.0	3.8	3.6	4.1
P	4.8	5.8	6.5	7.0	5.8	0.0	0.5	0.0	1.9	6.5	6.7
Q	3.6	4.4	5.5	4.6	4.3	0.0	2.2	0.0	4.3	2.6	3.4
R	4.6	3.4	5.1	6.0	4.3	0.0	4.8	0.0	4.3	3.6	4.8
S	9.0	6.5	7.2	7.9	9.8	0.0	9.8	33.6	8.9	8.2	6.2
T	6.5	5.8	5.8	5.8	6.7	0.0	6.2	65.7	5.3	9.8	6.2
V	5.6	8.5	5.8	7.9	7.7	0.0	9.1	0.0	10.3	7.9	7.9
W	1.5	1.0	2.4	1.4	1.9	0.0	1.0	0.0	1.4	1.7	1.7
Y	3.9	3.9	5.1	4.6	5 5	0.0	4.3	0.0	4.6	3.6	3.8

(b) Amino acid distributions around gs^- sites

A	6.2	4.2	0.0	10.4	4.2.	0.0	2.1	0.0	4.2	2.1	4.3
C	4.2	6.2	4.2	4.2	10.4	0.0	6.2	0.0	6.2	2.1	6.4
D	0.0	2.1	8.3	16.7	2.1	0.0	8.3	0.0	10.4	6.2	0.0
E	4.2	8.3	6.2	6.2	12.5	0.0	4.2	0.0	2.1	2.1	2.1
F	2.1	8.3	8.3	0.0	0.0	0.0	8.3	0.0	2.1	4.2	4.3
G	4.2	6.2	6.2	6.2	6.2	0.0	6.2	0.0	8.3	10.4	6.4
H	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	0.0
I	2.1	2.1	4.2	4.2	4.2	0.0	8.3	0.0	2.1	4.2	2.1
K	6.2	6.2	10.4	0.0	4.2	0.0	0.0	0.0	4.2	16.7	2.1
L	8.3	10.4	6.2	4.2	4.2	0.0	6.2	0.0	8.3	10.4	6.4
M	6.2	2.1	2.1	2.1	2.1	0.0	2.1	0.0	2.1	2.1	4.3
N	10.4	4.2	8.3	2.1	0.0	100.0	6.2	0.0	4.2	6.2	10.6
P	10.4	2.1	0.0	8.3	6.2	0.0	22.9	0.0	12.5	8.3	8.5
Q	4.2	0.0	0.0	8.3	8.3	0.0	0.0	0.0	4.2	4.2	4.3
R	6.2	4.2	2.1	2.1	6.2	0.0	0.0	0.0	4.2	2.1	6.4
S	16.7	12.5	16.7	6.2	10.4	0.0	2.1	52.1	6.2	6.2	10.6
T	2.1	0.0	6.2	6.2	2.1	0.0	6.2	47.9	6.2	0.0	12.8
V	4.2	10.4	6.2	10.4	12.5	0.0	4.2	0.0	4.2	2.1	4.3
W	2.1	2.1	0.0	0.0	0.0	0.0	0.0	0.0	6.2	0.0	2.1
Y	0.0	8.3	4.2	2.1	4.2	0.0	6.2	0.0	2.1	8.3	2.1

Frequencies are given as percentages. Position 0 is the carbohydrate attachment point.

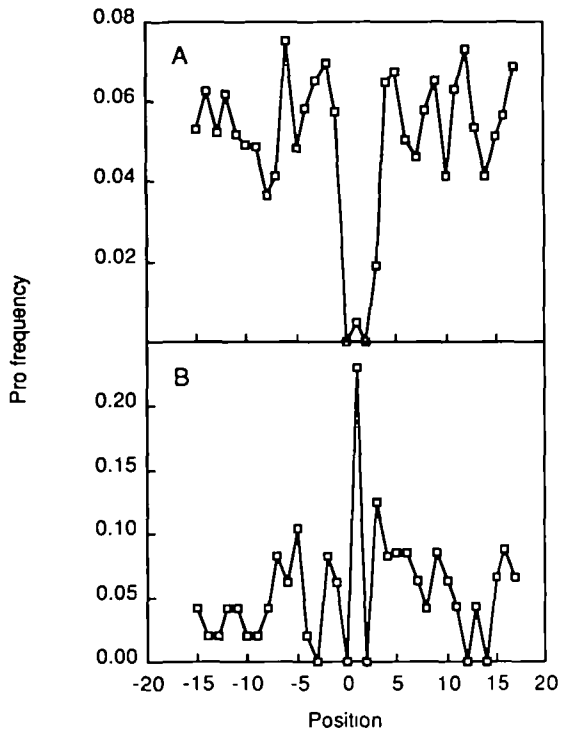

Fig. 1. Frequency of proline residues as a function of position relative to glycosylated (A) and non-glycosylated (B) Asn-X-Thr/Ser sites. The Asn residue is in position zero.

is consistent with the findings of Bause (1983) and Roitsch and Lehle (1989) that peptides containing a potential glycosylation site cannot be glycosylated if the site has Pro in position +1 or +3.

Low counts in position +1 of gs^+ sites have been reported for other amino acids besides Pro. For example, Cys and Trp (Kaplan *et al.*, 1987; Lennarz, 1988) as well as the acidic residues, Glu and Asp (Mononen and Karjalainen, 1984) have been claimed to be rare in glycosylated tripeptides. Our data do not substantiate this, and there are no residues besides Pro for which a high frequency in some position close to gs^- sites is matched by a low frequency close to gs^+ sites, or vice versa.

For proteins in general, i.e. including cytoplasmic ones, the frequencies of Asn-X-Thr and Asn-X-Ser tripeptides are equal. However, the sequence Asn-X-Thr has been reported to be about three times as frequent as Asn-X-Ser for gs^+ sites (Kaplan *et al.*, 1987; Lennarz, 1988). In agreement with this, experiments by Bause (1984) indicate that the tripeptide Asn-X-Thr is glycosylated more rapidly than Asn-X-Ser. Our data essentially confirm these results. As expected, the frequencies of Ser- and Thr-containing tripeptides are approximately equal for the gs^- sites, and the difference of occurrence in gs^+ sites is ~2-fold.

In order to look for correlations between glycosylation tendency and position in the protein, we have tabulated the relative positions of gs^+ and gs^- sites in the sequence and the absolute distances to nearest gs^+ sites as well as to the N and C termini of the protein. From these data, we find that the frequency of gs^- sites is higher towards the C terminus, whereas that of gs^+ sites is lower. The distributions (Figure 2) differ significantly ($P < 10^{-4}$ by χ^2 test).

In terms of absolute distances, we have found one gs^+ site only four residues away from the C terminus of a mature glycoprotein and one only one residue away from the N terminus.

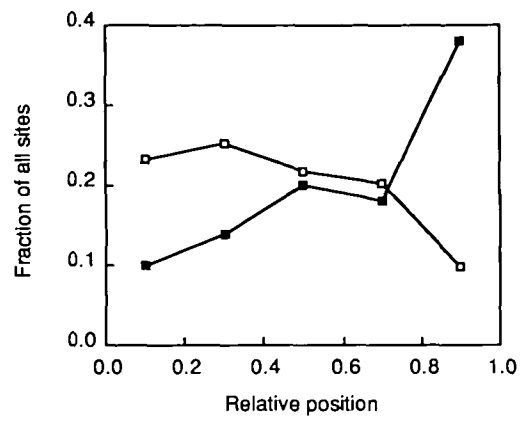

Fig. 2. Incidence of glycosylated (open squares) and non-glycosylated (solid squares) Asn-X-Thr/Ser sites as a function of relative position in the protein chain (N- to C-terminal direction).

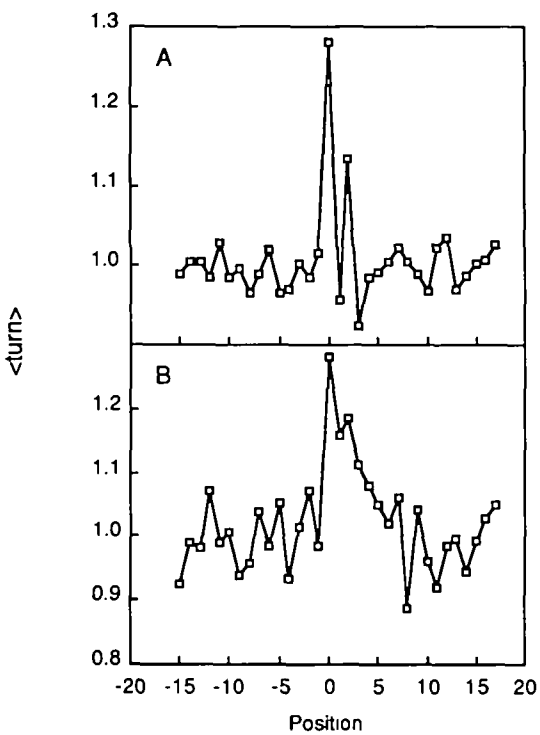

Fig. 3. Mean turn propensities as a function of position relative to glycosylated (A) and non-glycosylated (B) Asn-X-Thr/Ser sites.

We have also calculated distributions for the distances between neighbouring gs^+ sites. The smallest separation in our database is four residues (in haptoglobin-1 with the sequence Asn-His-Ser-Glu-Asn-Ala-Thr). It may be that steric hindrance prevents more closely spaced sites from being glycosylated at the same time; thus, a site with the sequence . . . Asn-Asn-Ser-Thr . . . with the second but not the first Asn glycosylated is found in human cholinesterase (see Table I).

Beeley (1977) as well as Aubert *et al.* (1976) have noted that gs^+ sites are often situated in beta turns or other loop structures. We have thus calculated the mean turn, β-structure, and α-helix potentials as a function of position using the scale of Levitt (1978). We find no indications that gs^+ sites are more turn-prone than gs^- sites (Figure 3), confirming earlier results (Mononen and

Karjalainen, 1984). Since both asparagine and the hydroxy amino acids have fairly high turn propensities, turn probability should not be expected to differ much between gs^+ and gs^- sites. Actually, the mean turn potential of gs^- sites is a little higher than that of gs^+ sites because of the high frequency of proline residues. Thus, the importance of turn conformations for glycosylation remains conjectural.

Conclusions

Our data demonstrate that potential glycosylation sites that are located near the C terminus or contain Pro in position +1 or +3 often are non-glycosylated. These observations allow a clear prediction of the likelihood for glycosylation to be made in some cases, but not in others. A site with Pro_{+1} is non-glycosylated in at least 90% of all cases. For Pro_{+3}, this value is ~50%. Thus, one can safely assume a site with Pro_{+1} to be gs^- and a site lacking Pro_{+1} and Pro_{+3} to be gs^+ (correct 93% of the time). Sites with Pro_{+3} are best left undecided. Thus, out of a total of 465 sites, 10/11 sites (90%) with Pro_{+1} are correctly predicted as gs^-, 369/400 sites (92%) not having Pro_{+1} or Pro_{+3} are correctly predicted as gs^+ and 6/465 sites with Pro_{+3} (1%) cannot be predicted. Beyond this, no simple local patterns have been found to correlate with glycosylation; in particular, secondary structure predictions are of no use in differentiating between gs^+ and gs^- sites.

Experiments with ovalbumin have shown that co-translational glycosylation cannot take place until ~45 amino acids have been added beyond the carbohydrate attachment site (Glabe *et al.*, 1980). Therefore, the bias of gs^+ sites away from the C terminus may indicate that glycosylation occurs more easily when the nascent polypeptide chain is still attached to polyribosomes and spanning the ER membrane. N-Terminally located sites spend more time in the neighbourhood of the lumenal face of the ER membrane, where the sugar-adding enzyme is situated. Also, C-terminal sites may be more quickly pulled into an already almost fully folded structure, thus becoming inaccessible for glycosylation.

In conclusion, we find that gs^- sites tend to be found more frequently towards the C termini of glycoproteins, and that proline residues in positions X and Y in the consensus Asn-X-Thr/Ser-Y strongly reduce the likelihood of *N*-linked glycosylation. For protein engineering purposes, these results suggest that prolines should be avoided near newly introduced acceptor sites, and that they can be used to block an otherwise modifiable site.

Acknowledgements

This work was supported by a grant from the Swedish Natural Sciences Research Council to G.vH.

References

Aubert,J.-P., Biserte,G. and Loucheux-Lefebvre,M.-H. (1976) *Arch. Biochem. Biophys.*, **175**, 410–418.

Bangs,J.D., Doering,T.L., Englund,P.T. and Hart,G.W. (1988) *J. Biol. Chem.*, **263**, 17697–17705.,

Baudys,M. and Kostka,V. (1983) *Eur. J. Biochem.*, **136**, 89–99.

Bause,E. (1983) *Biochem. J.*, **209**, 331–336.

Bause,E. (1984) *Biochem. Soc. Trans.*, **12**, 514–517.

Bause,E. and Legler,G. (1981) *Biochem. J.*, **195**, 639–644.

Beeley,J.G. (1977) *Biochem. Biophys. Res. Commun.*, **76**, 1051–1055.

Beintema,J.J. (1985) *FEBS Lett.*, **185**, 115–120.

Beintema,J.J., Blank,A., Schieven,G.L., Decker,C.A., Sorrentino,S. and Libonato,M. (1988) *Biochem. J.*, **255**, 501–505.

Bianchetta,J.D., Bidaud,J., Guidoni,A.A., Bonicel,J.J. and Rovery,M. (1979) *Eur. J. Biochem.*, **97**, 395–405.

Bock,S.C., Skriver,K., Nielsen,E., Thogersen,H.-C., Wiman,B., Donaldson, V.H., Eddy,R.L., Marrinan,J., Radziejewska,E., Huber,R., Shows,T.B. and Magnusson,S. (1986) *Biochemistry*, **25**, 4292–4301.

Boel,E., Bech,A.-M., Randrup,K., Draeger,B., Fiil,N.P. and Foltmann,B. (1986) *Proteins*, **1**, 363–369.

Boutin,J.-M., Jolicoeur,C., Okamura,H., Gagnon,J., Edery,M., Shirota,M., Banville,D., Dusanter-Fourt,I., Djiane,J. and Kelly,P.A. (1988) *Cell*, **53**, 69–77.

Brazel,D., Oberbäumer,I., Dieringer,H., Babel,W., Glanville,R.W., Deutzmann,R. and Kühn,K. (1987) *Eur. J. Biochem.*, **168**, 529–536.

Bringman,T.S., Lindqvist,P.B. and Derynck,R. (1987) *Cell*, **48**, 429–440.

Byrne,B.M., van het Schip,A.D., van de Klundert,J.A.M., Arnberg,A.C., Gruber,M. and Ab,G. (1984) *Biochemistry*, **23**, 4275–4279.

Campbell,D.G., Gagnon,J., Reid,K.B.M. and Williams,A.F. (1981) *Biochem. J.*, **195**, 15–30.

Carrell,R.W., Jeppson,J.-O., Vaughan,L., Brennan,S.O., Owen,M.C. and Boswell,D.R. (1981) *FEBS Lett.*, **135**, 301–303

Carrell,R.W., Jeppson,J.-O., Laurell,C.-B., Brennan,S.O., Owen,M.C., Vaughan,L. and Boswell,D.R. (1982) *Nature*, **298**, 329–334.

Catley,B.J., Moore,S. and Stein,W.H. (1969) *J. Biol. Chem.*, **244**, 933–936.

Clark,M.J., Gagnon,J., Williams,A.F. and Barclay,A.N. (1986) *EMBO J.*, **4**, 113–118.

Corbi,A.L., Kishimoto,T.K., Miller,L.J. and Springer,T.A. (1988) *J. Biol. Chem.*, **263**, 12403–12411.

Davidow,L.S., O'Donnell,M.M., Kaczmarek,F.S., Pereira,D.A., DeZeeuw,J.R. and Franke,E. (1987) *J. Bacteriol.*, **169**, 4621–4629.

De Caro,L., Boudouard,M., Bonicel,J., Guidoni,A., Desneulle,P. and Rovery,M. (1981) *Biochim. Biophys. Acta*, **671**, 129–138.

de Bruijn,M.H.L. and Fey,G.H (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 708–712.

Drayna,D., Fielding,C., McLean,J., Baer,B., Castro,G., Chen,E., Comstock,L., Henzel,W., Kohr,W., Rhee,L., Wion,K. and Lawn,R. (1986) *J. Biol. Chem.*, **261**, 16535–16539.

Ebina,Y., Ellis,L , Jarnagin,K., Edery,M., Graf,L., Clauser,E., Ou,J., Masiarz,F., Kan,Y.W. and Rutter,W.J. (1985) *Cell*, **40**, 747–758.

Emter,O., Mechler,B., Achstetter,T., Müller,H. and Wolf,D.H. (1983) *Biochem. Biophys. Res. Commun.*, **116**, 822–829.

Esch,F.S., Shimasaki,S., Mercado,M., Cooksey,K., Ling,N., Ying,S., Ueno,N. and Guillemin,R. (1987) *Mol. Endocrinol.*, **1**, 849–855.

Fernlund,P. and Stenflo,J. (1982) *J. Biol. Chem.*, **257**, 12170–12179.

Fliegel,L., Ohnishi,M., Carpenter,M.R., Khanna,V.K., Rethmeier,R.A.F. and MacLennan,D.H. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1167–1171.

Flink,I.L., Bailey,T.J , Gustafson,T.A., Markham,B.E. and Morkin,E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7708–7712.

Foster,D.C., Yoshitake,S. and Davie,E.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4673–4677.

Frangione,B., Rosenwasser,E., Prelli,F. and Franklin,E.C. (1980) *Biochemistry*, **19**, 4304–4308.

Fujikawa,K. and McMullen,A. (1983) *J. Biol. Chem.*, **258**, 10924–10933.

Fukuda,M., Viitala,J., Matteson,J. and Carlsson,S.R. (1988) *J. Biol. Chem.*, **263**, 18920–18928.

Galjart,N.J., Gillemans,N., Harris,A., van der Horst,G.T.J., Verheijen,F.W., Galjaard,H. and d'Azzo,A. (1988) *Cell*, **54**, 755–764.

Garoff,H., Frischauf,A.-M., Simons,K., Lehrach,H. and Delius,H. (1980) *Nature*, **288**, 236–241.

Garver,F.A. and Hilschmann,N. (1972) *Eur. J. Biochem.*, **26**, 10–32.

Garver,F.A., Chang,L., Mendicino,J., Isobe,T. and Osserman,E. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 4559–4563.

Gebhard,W., Schreitmüller,T., Hochstrasser,K. and Wachter,E. (1988) *FEBS Lett.*, **229**, 63–67.

George,D.G., Barker,W.C. and Hunt,L.T. (1986) *Nucleic Acids Res.*, **14**, 11–15.

Glabe,C.B , Hanover,J.A. and Lennarz,W.J. (1980) *J. Biol. Chem.*, **255**, 9236–9242.

Grunert,F., Kolbinger,F., Schwartz,K., Schwaibold,H. and von Kleist,S. (1988) *Biochem. Biophys. Res. Commun.*, **153**, 1105–1115.

Hamazume,Y., Mega,T. and Ikenaka,T. (1984) *J. Biochem.*, **95**, 1633–1644.

Hansen,L., Blue,Y., Barone,K., Collen,D. and Larsen,G.R. (1988) *J. Biol. Chem.*, **263**, 15713–15719.

Hansen,S.I., Holm,J. and Lyngbye,J. (1984) *Carlsberg Res. Commun.*, **49**, 123–131.

Hargrave,P.A. (1977) *Biochim. Biophys. Acta*, **492**, 83–94.

Havinga,J. and Beintema,J.J. (1980) *Eur. J. Biochem.*, **110**, 131–142.

Hemperley,J.J., Hopp,T.P., Becker,J.W. and Cunningham,B.A. (1979) *J. Biol. Chem.*, **254**, 6803–6810.

Hinoda,Y., Neumaier,M., Hefta,S.A., Drzeniek,Z., Wagener,C., Shively,L., Hefta,L.J.F., Shively,J.E. and Paxton,R.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6959–6963.

Honma,K., Tomita,M. and Hamada,A. (1980) *J. Biochem.*, **88**, 1679–1691.

Hopp,T.P. and Woods,K.R. (1979) *Biochemistry*, **18**, 5182–5191.

Hopp,T.P., Hemperley,J.J. and Cunningham,B.A. (1982) *J. Biol. Chem.*, **257**, 4473–4483.
Howe,C.L., Granger,B.L., Hull,M., Green,S.A., Gabel,C.A., Helenius,A. and Mellman,I. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7577–7581.
Hubbard,S.C. and Ivatt,R.J. (1981) *Annu. Rev. Biochem.*, **50**, 555–583.
Ishidoh,K., Towatari,T., Imajoh,S., Kawasaki,H., Kominami,E., Katunuma,N. and Suzuki,K. (1987) *FEBS Lett.*, **223**, 69–73.
Jackson,R.L. and Hirs,C.H.W. (1970) *J. Biol. Chem.*, **245**, 637–653.
Jacobs,K., Shoemaker,C., Rudersdorf,R., Neill,S.D., Kaufman,R.J., Mufson,A., Seehra,J., Jones,S.S., Hewick,R., Fritsch,E.F., Kawakita,M., Shimizu,T. and Miyake,T. (1985) *Nature*, **313**, 806–810.
Julius,D., Shekman,R. and Thorner,J. (1984) *Cell*, **36**, 309–318.
Kanost,M.R., Boguski,M.S., Freeman,M., Gordon,J.I., Wyatt,G.R. and Wells,M.A. (1988) *J. Biol. Chem.*, **263**, 10568–10573.
Kaplan,H.A., Welply,J.K. and Lennarz,W.J. (1987) *Biochim. Biophys. Acta*, **906**, 161–173.
Kehry,M., Sibley,C., Fuhrman,J., Schilling,J. and Hood,L.E. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2932–2936.
Kehry,M.R., Fuhrman,J.S., Schilling,J.W., Rogers,J., Sibley,C.H. and Hood,L.E. (1982) *Biochemistry*, **21**, 5415–5424.
Kingston,I.B. and Williams,J. (1975) *Biochem. J.*, **147**, 463–472.
Krusius,T. and Ruoslahti,E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7683–7687.
Kurjan,J. and Herskowtiz,I. (1982) *Cell*, **30**, 933–943.
Laude,H., Rasschaert,D. and Huet,J.-C. (1987) *J. Gen. Virol.*, **68**, 1687–1693.
Lennarz,W.J. (1988) In Op den Kamp,J.A.F. (ed.), *Membrane Biogenesis*, NATO ASI Series H:16. Springer Verlag, Berlin, pp. 287–306.
Levitt,M. (1978) *Biochemistry*, **17**, 4277–4285.
Lewis,V., Green,S.A., Marsh,M., Vihko,P., Helenius,A. and Mellman,I. (1985) *J. Cell. Biol.*, **100**, 1839–1847.
Lewis,V.A., Koch,T., Plutner,H. and Mellman,I. (1986) *Nature*, **324**, 372–375.
Liao,T.-H., Salnikow,J., Moore,S. and Stein,W.H. (1973) *J. Biol. Chem.*, **248**, 1489–1495.
Liscum,L., Finer-Moore,J., Stroud,R.M., Luskey,K.L., Brown,M.S. and Goldstein,J.L. (1985) *J. Biol. Chem.*, **260**, 522–530.
Lockridge,O., Bartels,C.F., Vaughan,T.A., Wong,C.K., Norton,S E. and Johnson,L.L. (1987) *J. Biol. Chem.*, **262**, 549–557.
MacGillivray,R.T.A. and Davie,E.W. (1984) *Biochemistry*, **23**, 1626–1634.
Mahuran,D.J., Neote,K., Klavins,M.H., Leung,A. and Gravel,R.A. (1988) *J. Biol. Chem.*, **263**, 4612–4618.
Maloy,L. and Coligan,J.E. (1982) *Immunogenetics*, **16**, 11–22.
Mann,K., Deutzmann,R., Aumailley,M., Timpl,R., Raimondi,L., Yamada,Y., Pan,T., Conway,D. and Chu,M. (1989) *EMBO J.*, **8**, 65–72.
Marshall,R.D. (1972) *Annu. Rev. Biochem.*, **41**, 673–702.
Martens,G.J.M. (1986) *Nucleic Acids Res.*, **14**, 3791–3798.
Martin,G.A., Busch,S.J., Meredith,G.D., Cardin,A.D., Blankenship,D.T., Mao,S.J.T., Rechtin,A.E., Woods,C.W., Racke,M.M., Schafer,M.P., Fitzgerald,M.C., Burke,D.M., Flanagan,M.A. and Jackson,R.L. (1988) *J. Biol. Chem.*, **263**, 10907–10914.
Matoba,S., Fukayama,J., Wing,R.A. and Ogrydziak,D.M. (1988) *Mol. Cell. Biol.*, **8**, 4904–4916.
McMullen,B.A. and Fujikawa,K. (1985) *J. Biol. Chem.*, **260**, 5328–5341.
Mega,T., Lujan,E. and Yoshida,A. (1980) *J. Biol. Chem.*, **255**, 4057–4061.
Metz-Boutige,M.-H., Mazurier,J., Jollès,J., Spik,G., Montreuil,J. and Jollès,P. (1981) *Biochim. Biophys. Acta*, **670**, 243–254.
Millan,J.L. (1986) *J. Biol. Chem.*, **261**, 3112–3115.
Mononen,I. and Karjalainen,E. (1984) *Biochim. Biophys. Acta*, **788**, 364–367.
Morgan,F.J., Birken,S. and Canfield,R.E. (1975) *J. Biol. Chem.*, **250**, 5247–5258.
Nakai,K. and Kanehisa,M. (1988) *J. Biochem.*, **104**, 693–699.
Neame,P.J., Christner,J.E. and Baker,J.R. (1986) *J. Biol. Chem.*, **261**, 3519–3535.
Nisbet,A.D., Saundry,R.H., Moir,A.J.G., Fothergill,L.A. and Fothergill,J.E. (1981) *Eur. J. Biochem.*, **115**, 335–345.
Ovchinnikov,Y.A. (1982) *FEBS Lett.*, **148**, 179–191.
Peeters,B., Rombauts,W., Mous,J. and Heyns,W. (1981) *Eur. J. Biochem.*, **115**, 115–121.
Pennica,D., Holmes,W.E., Kohr,W.J., Harkins,R.N., Vehar,G.A., Ward,C.A., Bennett,W.F., Yelverton,E., Seeburg,P.H., Heyneker,H.L. and Goeddel,D.V. (1983) *Nature*, **301**, 214–221.
Putnam,F.W., Florent,G., Paul,C., Shinoda,T. and Shimizu,A. (1973) *Science*, **182**, 287–291.
Reid,K.M.B., Gagnon,J. and Frampton,J. (1982) *Biochem. J.*, **203**, 559–569.
Reyes,A., Schöld,M. and Wallace,R.B. (1982) *Immunogenetics*, **16**, 1–9.
Rice,C.M. and Strauss,J.H. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2062–2066.
Rice-Ficht,A.C., Chen,K.K. and Donelson,J.E. (1981) *Nature*, **294**, 53–57.
Robinson,E.A. and Appella,E. (1979) *J. Biol. Chem.*, **254**, 11418–11430.
Robinson,E.A. and Appella,E. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4909–4913.
Roitsch,T. and Lehle,L. (1989) *Eur. J. Biochem.*, **181**, 525–529.
Salnikow,J., Liao,T.-H., Moore,S. and Stein,W.H. (1973) *J. Biol. Chem.*, **248**, 1480–1488.
Schneider,C., Owen,M.J., Banville,D. and Williams,J.G. (1984) *Nature*, **311**, 675–678.
Schnell,D.J. and Etzler,M.E. (1987) *J. Biol. Chem.*, **262**, 7220–7225.
Shainkin,R. and Perlmann,G.E. (1971) *J. Biol. Chem.*, **246**, 2278–2284.
Shewale,J.G. and Tang,J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3703–3707
Shimasaki,S., Koga,M., Esch,F., Cooksey,K., Mercado,M., Koba,A., Ueno,N., Ying,S., Ling,N. and Guillemin,R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4218–4222.
Shimizu,A., Paul,C., Köhler,H., Shinoda,T. and Putnam,F.W. (1971) *Science*, **173**, 629–633.
Slightom,J.L., Drong,R.F., Klassy,R.C. and Hoffman,L.M. (1985) *Nucleic Acids Res.*, **13**, 6483–6497.
Spicer,E.K., Horton,R., Bloem,L., Bach,R., Williams,K.R., Guha,A., Kraus,J., Lin,T.-C., Nemerson,Y. and Konigsberg,W.H. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5148–5152.
Spycher,S.E., Nick,H. and Rickli,E.E. (1986) *Eur. J. Biochem.*, **156**, 49–57.
Stenflo,J. and Fernlund,P. (1982) *J. Biol. Chem.*, **257**, 12180–12190.
Strong,D D., Moore,M., Cottrell,B.A., Bohonus,V.L., Pontes,M., Evans,B., Riley,M. and Doolittle,R.F. (1985) *Biochemistry*, **24**, 92–101.
Struck,D.K., Lennarz,W.J. and Brew,K. (1978) *J. Biol. Chem.*, **253**, 5786–5794.
Svendsen,I., Martin,B.M., Viswanatha,T. and Johansen,J.T. (1982) *Carlsberg Res. Commun.*, **47**, 15–27.
Svensson,B., Larsen,K., Svendsen,I. and Boel,E. (1983) *Carlsberg Res. Commun.*, **48**, 529–544.
Takahashi,T. and Tang,J. (1983) *J. Biol. Chem.*, **258**, 2819–2830.
Takahashi,N , Tetaert,D., Debuire,B., Lin,L.-C. and Putnam,F.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2850–2854.
Takahashi,N., Ortel,T.L. and Putnam,F.W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 390–394.
Takahashi,N., Takahashi,Y. and Putnam,F.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1906–1910.
Takahashi,H., Nukiwa,T., Yoshimura,K., Quick,C.D., States,D.J., Holme,M.D., Whang-Peng,J., Knutsen,T. and Crystal,R.G. (1988) *J. Biol. Chem.*, **263**, 14739–14747.
Tayakama,S., Isogai,A., Tsukamoto,C., Ueda,Y., Hinata,K., Okazaki,K. and Suzuki,A. (1987) *Nature*, **326**, 102–105.
Takeya,H., Kawabata,S., Nakagawa,K , Yamamichu,Y., Miyata,T., Iwanaga,S., Takao,T. and Shimonishi,Y. (1988) *J. Biol. Chem.*, **263**, 14868–14877
Taylor,A.K. and Wall,R. (1988) *Mol. Cell. Biol.*, **8**, 4197–4203.
Titani,K., Kumar,S., Takio,K., Ericsson,L.H., Wade,R.D., Ashida,K., Walsh,K.A., Chopek,M.W., Sadler,J.E. and Fujikawa,K. (1986) *Biochemistry*, **25**, 3171–3184.
Titani,K., Takio,K., Handa,M. and Ruggeri,Z.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5610–5614.
Tosi,M., Duponchel,C., Meo,T. and Julier,C. (1987) *Biochemistry*, **26**, 8516–8524.
Tucker,P.W., Slightom,J.L. and Blattner,F.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7684–7688.
Ullrich,A., Coussens,L., Hayflick,J.S., Dull,T.J., Gray,A., Tam,A.W., Lee,J., Yarden,Y., Libermann,T.A., Schlessinger,J., Downward,J., Mayes,E.L.V., Whittle,N., Waterfield,M.D. and Seeburg,P.H. (1984) *Nature*, **309**, 418–425.
Ullrich,A., Bell,J.R., Chen,E.Y., Herrera,R., Petruzzelli,L.M., Dull,T.J., Gray,A., Coussens,L., Liao,Y.-C., Tsubokawa,M., Mason,A., Seeburg,P.H., Grunfeld,C., Rosen,O.M. and Ramachandran,J. (1985) *Nature*, **313**, 756–761.
Van Den Berg,A., Van Den Hende-Timmer,L. and Beintema,J.J. (1976) *Biochim. Biophys. Acta*, **453**, 400–409.
Van Den Berg,A., Van Den Hende-Timmer,L., Hofsteenge,J., Gaastra,W. and Beintema,J.J. (1977) *Eur. J. Biochem.*, **75**, 91–100.
Van het Schip,F.D., Samallo,J., Broos,J., Ophuis,J., Mojet,M., Gruber,M. and Ab,G. (1987) *J. Mol. Biol.*, **196**, 245–260.
Viitala,J., Carlsson,S.R., Siebert,P.D. and Fukuda,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3743–3747.
Walsh,K.A., Titani,K., Takio,K., Kumar,S., Hayes,R. and Petra,P.H. (1986) *Biochemistry*, **25**, 7584–7590.
Ward,C.W. and Dopheide,T.A. (1981) *Biochem. J.*, **193**, 953–962.
Ward,C.W., Gleeson,P.A. and Dopheide,T.A. (1980) *Biochem. J.*, **189**, 649–652.
Waters,M.G., Evans,E.A. and Blobel,G. (1988) *J. Biol. Chem.*, **263**, 6209–6214.
Welinder,K.G. (1976) *FEBS Lett.*, **72**, 19–23.
Williams,J., Elleman,T.C., Kingston,I.B., Wilkins,A.G. and Kuhn,K.A. (1982) *Eur. J. Biochem.*, **122**, 297–303.
Zinn,A.B., Marshall,J.S. and Carlson,D.M. (1978) *J. Biol. Chem.*, **253**, 6761–6767.

Received on October 27, 1989; accepted on January 15, 1990