ALMA MATER STUDIORUM
UNIVERSITA' DEGLI STUDI DI BOLOGNA

Dottorato di Ricerca in Biotecnologie Cellulari e Molecolari

XXIV CICLO

Settore scientifico disciplinare BIO14

IDENTIFICATION OF THE N-LINKED GLYCOSYLATION SITES OF THE TRANSCRIPTION FACTOR REST AND EFFECT OF GLYCOSYLATION ON DNA BINDING AND TRANSCRIPTIONAL ACTIVITY

Presentata da:

Dott.ssa Gioia Carbonari

Coordinatore Chiar.mo Prof.
Santi Spampinato

Relatore Chiar.mo Prof.
Santi Spampinato

ANNO ACCADEMICO 2010-2011

TABLE OF CONTENTS

ABSTRACT

REST is a zinc-finger transcription factor implicated in several processes such as maintenance of embryonic stem cell pluripotency and regulation of mitotic fidelity in non-neuronal cells [Chong et al., 1995]. The gene encodes for a 116-kDa protein that acts as a molecular platform for co-repressors recruitment and promotes modifications of DNA and histones [Ballas, 2005]. REST showed different apparent molecular weights, consistent with the possible presence of post-translational modifications [Lee et al., 2000]. Among these the most common is glycosylation, the covalent attachment of carbohydrates during or after protein synthesis [Apweiler et al., 1999]

My thesis has ascertained, for the first time, the presence of glycan chians in the transcription factor REST. Through enzymatic deglycosylation and MS, oligosaccharide composition of glycan chains was evaluated: a complex mixture of glycans, composed of N-acetylgalactosamine, galactose and mannose, was observed thus confirming the presence of O- and N-linked glycan chains. Glycosylation site mapping was done using a ^{18}O-labeling method and MS/MS and twelve potential N-glycosylation sites were identified. The most probable glycosylation target residues were mutated through site-directed mutagenesis and REST mutants were expressed in different cell lines. Variations in the protein molecular weight and mutant REST ability to bind the *RE-1* sequence were analyzed. Gene reporter assays showed that, altogether, removal of N-linked glycan chains causes loss of transcriptional repressor function, except for mutant N59 which showed a slight residual repressor activity in presence of IGF-I.

Taken togheter these results demonstrate the presence of complex glycan chians in the transcription factor REST: I have depicted their composition, started defining their position on the protein backbone and identified their possible role in the transcription factor functioning. Considering the crucial role of glycosylation and transcription factors activity in the aetiology of many diseases, any further knowledge could find important and interesting pharmacological application.

INTRODUCTION

1. THE TRANSCRIPTION FACTOR REST

Despite a genetic homogeneity, cells in multicellular organisms are structurally and functionally heterogeneous. The diversity of cell phenotype exists due to differential transcription programs precisely regulated by specific nuclear factors and induced upon differentiation: in fact the fate of a cell is ultimately determined by the way its genetic material and its protein scaffold are modified.

The differences in gene expression programs arise during development and become heritable during cell proliferation: it becomes of crucial interest understanding the molecular mechanisms by which central nervous system-specific genes are expressed in the nervous system and repressed in other tissues.

During neurogenesis, progenitor cells must select between neuronal versus non-neuronal fates; in the former case they must also select among a large repertoire of neuronal subtypes fates: these developmental decisions require the action of transcriptional regulatory proteins. Gene expression can be controlled by positively acting cell-type-specific or tissue-specific transcription factors; however also negative-acting DNA regulatory elements can participate in preventing the expression of neuronal genes in non-neuronal cell types or in inappropriate neuronal subtypes: these elements can be recognized by silencers, transcription factors that behave similar as enhancers but which have an opposite effect on transcription [Schoenherr and Anderson, 1995].

Cells in the nervous system acquire distinct fates in response to extrinsic signals that activate repertoires of transcription factors in a region- and cell-type-specific manner, endowing each type of neuron with unique characteristics and determining the generation. Transcription factors play pivotal roles in regulating the generation of neural precursors, the regionalization of the nervous system and subsequent differentiation of specific cell types within this region.

The loss of function or misexpression of a single transcription factor can be sufficient to alter neuron morphology, synaptic input, projection pattern or neurotransmitter phenotype [Bang and Goulding, 1996].

In 1995 two groups independently identified a gene encoding a zinc-finger protein that was suggested to function as a master regulator of the neuronal phenotype. The transcription factor REST, an *RE-1*-silencing transcription factor [Chong et al., 1995] also known as neuron-restrictive silencing factor NRSF [Schoenherr and Andreson 1995], blocks transcription of its target genes by binding to a specific consensus sequence that is present in the target genes' regulatory region.

The original proposed role for REST was that of a factor responsible for restricting neuronal gene expression to the nervous system, by silencing expression of these genes in non-neuronal cells: since its discovery as a negative regulator of neuronal differentiation, REST has been implicated in several processes such as maintenance of embryonic stem cell pluripotency, self-renewal and regulation of mitotic fidelity in non-neuronal cells.

By fluorescence in situ hybridization *REST* gene (3365 base pairs) was assigned to the chromosomic region 4q12 [GeneBank NM_005612] [Cowan et al., 1996]. The gene encodes for a protein constituted by 1097 aminoacids: it is a 116-kDa Krüppel family zinc finger protein that contains a DNA-binding domain, composed of eight zinc-fingers, and one zinc finger at the carboxy terminal. The zinc finger domain, which includes 2 conserved cysteine and 2 conserved histidine residues, have the ability to bind to both RNA and DNA. The DNA-binding domain of REST binds to a 21-bp consensus DNA sequence, the *RE-1* binding site/neuron-restrictive silencer element (*RE-1*/NRSE), present in the target gene's regulatory regions. In addition to the DNA binding domain (DBD), REST is composed of an N-terminal repressor domain and a C-terminal repressor domain, which contains a single zinc finger motif [Ballas and Mandel, 2005]: deletion to aminoacidic residues 76 or 44 totally abolished repression, suggesting that a repressor domain is located between residues 44 and 83. Aminoacid residues 1061-1097 containing the C-terminally located zinc finger are also essential for the repression activity of REST, defining aminoacids 989-1097 as a second repression domain (Figure 1.1).

REST protein is present at its highest level in the nuclei of pluripotent embryonic stem (ES) cells; as ES cells differentiate to neural progenitors REST mRNA levels stay relatively constant but REST protein is downregulated post-translationally to minimal levels. The release of REST from neuronal gene chromatin as progenitor differentiate into neurons is concomitant with its own transcriptional repression.

By acting as an evolving molecular platform to which diverse factors may be recruited, REST promotes dynamic modifications of DNA, histones, nucleosomes and higher-order chromatin codes and helps maintaining genomic stability.

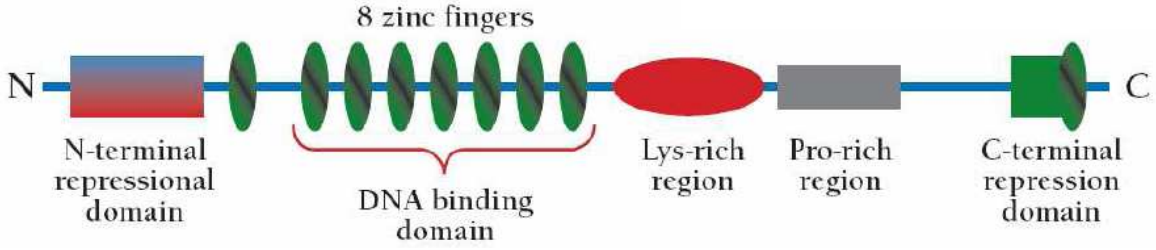

Figure 1.1: Schematic representation of REST protein domains.

REST is obligatory for the correct development of vertebrates, since perturbations of its expression or function in the developing embryo results in the ectopic expression of neuronal genes in non-neuronal tissues and early embryonic lethality: mice homozygous null for REST showed brain malformation at embryonic day E9.5, followed by extensive apoptosis and lethality by E11.5. These results suggested an important role for REST in brain development where it acts as a molecular restraint for precocious neuronal differentiation [Chen et al., 1998].

Although the molecular weight calculated on REST sequence is about 116 kDa, using western blot this transcription factor appears to have distinct apparent molecular weights (Table 1.1): this difference could be explained by post-translational modifications of the proteins.

In addition, several splice variants of REST were discovered encoding proteins with five or four zinc finger motifs [Palm et al., 1998]: two variants, termed REST4 and REST5, were only detected in neuronal tissues and are conserved in human, mouse and rat [Palm et al., 1999]. These transcripts are generated by alternative splicing of a neuron-specific exon (exon N) located between exons V and VI. REST4 in particular have an insertion of 16 nucleotides followed by an inframe stop codon, so it retains the N-terminal repression domain and five of the eight zinc fingers. The biological function of REST4 is controversial: although it does not bind to the *RE-1* itself, it acts as a dominant negative by inhibiting the binding of REST to the *RE-1* sequence and regulating the repressor activity of REST [Shimojo et al., 1999].

Both REST and REST4 interact with RILP, a LIM domain protein, necessary for nuclear traslocation, suggesting that REST action could also depend on its cellular environment [Shimojo and Hersh 2003]. REST protein was found to migrate on SDS-PAGE as a protein of ~200 kDa, while the predicted molecular weight, based on aminoacid sequence, is 116 kDa; the isoforms REST4 and REST5 migrated on SDS-PAGE as 53 kDa and 50 kDa proteins, respectively, while the predicted molecular weights are 37 kDa and 33 kDa, consistent with glycosylation of the protein backbone. REST4, and presumably REST,

was found to be O-glycosylated between residues 87 and 152: however full length REST could contain additional glycosylation sites [Lee et al., 2000].

REFERENCE	CELL LINE	MOLECULAR WEIGHT (KDA)
Chong et al. 1995	PC12	116
Wood et al. 2003	PC12 and JTC-19	120
Cheong et al. 2005	Murine endothelial cells	
Qiang et al. 2005	Primary cortical neurons (mouse)	150
Thiel et al.1998	NS20Y, NG108-15, CHO-K1	190-210
Zuccato et al.2003	Neuronal murine cells	200 and 160-180
Pance et al. 2006	PC12(A35C), K562	
Grimes et al. 2000	Transfected HEK293	200
Jia et al. 2006	Rat cortical neurons	
Westbrook et al. 2005	Colorectal cancer cell lines	
Di Toro et al. 2005	SH-SY5Y	
Kim et al.2006	HEK293, SH-SY5Y, HB1, F3, PC12, C6, C17.2	
Kuratomi et al. 2007	Rat cardiomyocytes	

Table 1.1: REST molecular weight in different cell lines.

Additional layers of flexibility and specificity are conferred on the activity of the REST complex through several parameters that include its affinity for binding to different *RE-1* sites, the modulation via non-coding RNAs, and emerging post-transcriptional events.

1.1 REST BINDING SITE

REST binds to a 21 base-pair sequence termed *RE-1* (Repressor Element-1)/*NRSE* (Neuron Restrictive Silencer Element) and found in the regulatory regions of its target genes. Several approaches, from bioinformatics to genome-wide binding experiments and study of diseases, have been used to define the *RE-1* element and to identify REST target

genes: these strategies allowed the identification of 1892 putative *RE-1*s in the human genome.

Starting from the Repressor Element-1 originally found in the 5' flanking region of the *voltage-gated sodium type II channel* and *superior cervical ganglion 10* genes, a 21-bp element has been identified in the regulatory regions of many neuron-specific genes. These genes include those encoding ion channels (e.g., NaV1.3, Kv3.4, and Cavl.3), neurotransmitter-synthesizing enzymes and receptors (e.g., M4 muscarinic, D3 dopamine, and γ-aminobutyric acid type β3 receptors), transporters (e.g., γ-aminobutyric acid transporter 4), neurotrophic receptors and proteins involved in vesicular trafficking (e.g., synaptosomal-associated protein SNAP25; synaptotagmins IV, V and VII; syntaxin 8; Rab3), neuronal cytoskeleton (tubulin β3) and axonal guidance (e.g., SCG10, stathmin 3, netrin-2, roundabout, semaphorin 5A, L1 cell-adhesion molecule L1CAM). However among REST identified targets there are also many genes not related to obvious neuron-specific functions, such as those involved in cellular metabolic processes (peroxisome and proteosome components) or in immune and inflammatory responses; additionally, there are genes that specify proteins that perform neuronal functions but are also required in non-neuronal tissues, including those involved in the regulation of cardiovascular tone (endothelial nitric oxide synthase, vasoactive intestinal peptide, atrial natriuretic peptide, brain natriuretic peptide, and KCNH2). [Bruce et al., 2004].

In addition there is an in vivo hierarchy in which distinct subsets of *RE-1* interact with endogenous levels of REST, whereas others function as *bona fide* transcriptional control elements only in the presence of elevated levels of REST.

To identify the genes effectively regulated by REST *in vivo* efforts have turned to chromatin occupancy studies, by using new methods such as Serial Analysis of Chromatin Occupancy (SACO) and large-scale chromatin immunoprecipitation (ChIP Seq). Interestingly these studies revealed that the canonical *RE-1* sequence results from a more complex motif, that can be both compressed or expanded, and that nucleotides 10 and 11 of *RE-1* were two of the least-highly conserved nucleotides across mammalian species: the new expanded and compressed motifs increased the number of potential REST-binding sites. [Otto et al., 2007].

The canonical REST-binding motif of 21 bp can be naturally subdivided into two nonidentical, nonpalindromic half sites; left and right half-site motifs are separated by a "spacer" sequence that increases the center-to-center distance from the canonical 11 bp to 16 to 19 bp, or decreases it by 1 bp to 10 bp (Figure 1.2) [Johnson et al., 2007]. Although REST could bind to half-sites of the *RE-1* element, neither the left nor the right

half-site by itself is an effective repressor, indicating that half-site binding is not as strong or functional as full-site binding [Jothi et al., 2008].

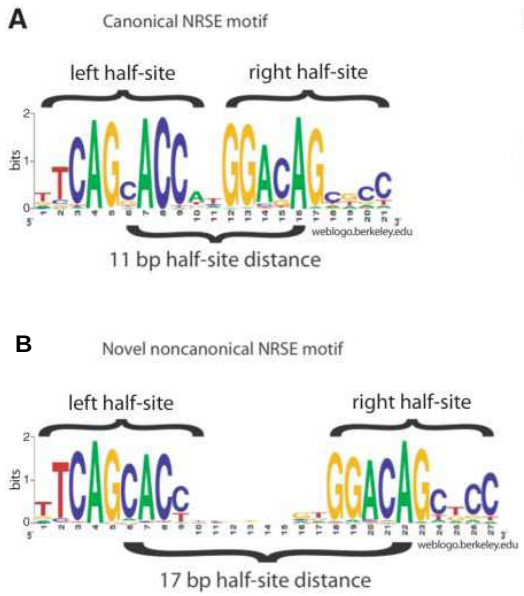

Figure 1.2: A) Canonical REST-binding motif: its left and right half sites have a center-to-center distance of 11 bp; B) WebLogo of noncanonical *RE-1* with half-site distance of 17, showing the lack of conservation in the spacer nucleotides.

While canonical sequence motifs facilitate strong REST binding and control functional classes of target genes that are common to all cell types, non-canonical motifs participate in weak interactions and control those targets which are cell- or tissue-specific [Bruce et al, Genome Research 2009].
Genome-wide analysis also revealed that variability in the occupancy of the *RE-1* elements could be a reflection of actual variation in REST binding to its target genes based on its affinity for the DNA element or based on the presence or absence of coactivators/corepressors of the process, or could result from nucleosome structure, or occupancy of adjacent sequences by other DNA-binding proteins.

1.2 REST AND COREPRESSORS

In multipotent neural stem cells (NSCs) neuronal genes are repressed, but these cells have the capacity for subsequent expression in response to a developmental signal: the recruitment of chromatin modifiers to genomic loci by DNA binding protein is crucial for the establishment of epigenetic modifications, representing distinct stages of

differentiation. REST provides a paradigm for elucidating the link between epigenetic mechanisms and neurogenesis: it orchestrates a set of epigenetic modifications that are distinct between ES cells, non-neuronal cells that give rise to neurons and those that are destined to remain as nervous system outsiders.

REST-mediated gene repression is achieved by the recruitment of two separate corepressor complexes, mSin3 and CoREST. The amino terminal repressor domain of REST interacts with mSin3, a corepressor found in all eukaryotes that recruits two class I histone deacetylases (HDAC1 and HDAC2) and the retinoblastoma-associated proteins RbAp48 and RbAp46, thought to interact with histones [Grimes et al., 2000; Roopra et al., 2000; Naruse et al., 1999]. The mSin3–HDAC complex, however, is associated primarily with a dynamic mode of repression that can alternate between repression and activation and therefore by itself would probably be inadequate for long-term silencing of neuronal genes. This conundrum was solved by the discovery of the corepressor CoREST, which interacts directly with the carboxy terminal repressor domain of REST [Ballas et al., 2001] and forms a complex containing HDAC1 and HDAC2, the histone H3 lysine 4 (H3-K4) demethylase LSD1, the chromatin-remodeling enzyme BRG1, the histone H3–K9 methyltransferase G9a and the methyl-CpG-binding protein MeCP2.

In neuronal genes outside the nervous system REST recruits chromatin modifiers necessary for a long-term silencing: their activities are interdependent and occur in an ordered stepwise fashion to mediate efficient gene repression. First BRG1 is recruited by the C-terminal repression domain of REST, as part of the CoREST complex: it stabilizes REST binding on the *RE-1* and allows REST to gain better access to its chromatin targets [Ooi et al., 2006]. Once associated with the chromatin, REST mediates repression through histone deacetylase, histone demethylase and histone methylase activities. HDACs deacetylate H3-K9 and H3-K14, a process that is required for subsequent demethylation of H3-K4 by LSD1 and dimethylation of H3-K9 by G9a: the addition of methyl groups to H3-K9 prevents the re-acetylation and is necessary for the subsequent recruitment of the Heterochromatin protein 1 (HP1), which mediates chromatin condensation.

In long-term silencing this continued repression seems to be facilitated by the action of MeCP2 and of the histone H3–K9 methyltransferase SUV39H1: recent studies revealed that CpG dinucleotides spanning the REST binding site and surrounding regions of neuronal genes are methylated in differentiated non-neuronal cells [Ballas et al., 2005]. The overall effect of these modifications is manifested in the removal of several marks that are associated with transcriptional activation followed by their replacement with marks that are associated with transcriptional repression; moreover the recruitment

of silencing machinery by the REST–CoREST complex might result in the propagation of silencing across a large chromosomal interval, containing also neuronal genes that do not have their own *RE-1*, suggesting a relationship between higher order chromatin structure and patterns of gene expression (Figure 1.3).

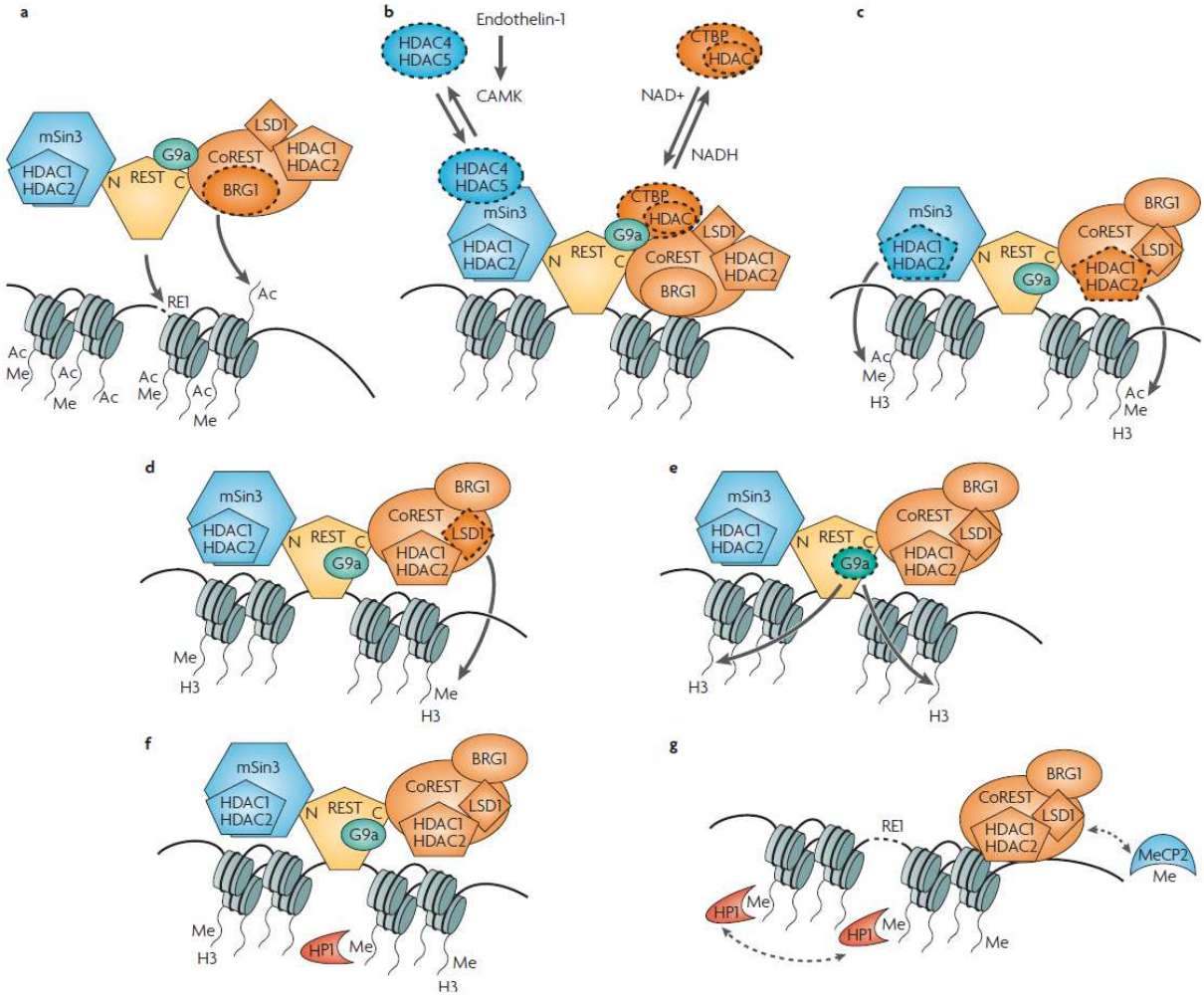

Figura 1.3: Schematic representation of long-term silencing mediated by REST.

In pluripotent ES cells, instead, neuronal gene chromatin is programmed by the REST machinery to stay in an inactive but poised status, ready for subsequent activation: in this state REST is bound to the *RE-1* motif and its corepressors, CoREST, mSin3, HDAC and MeCP2, are also present, but the *RE-1* motif and the surrounding sequences in neuronal genes are not methylated. The absence of histone H3-K9 methylation and the enrichment of di- and tri-methylated H3-K4 coincide with actively transcribed genes and, in addition, RNA polymerase II (Pol II) is present on *RE-1* sites in the 5'-untranslated regions of several neuronal genes, accompanied by very low levels of transcription. As ES cells proceed in the differentiation program towards neural progenitors, REST protein is

downregulated post-translationally; however terminal differentiation of progenitors to mature neurons occurs only with the transcriptional repression of the *REST* gene itself, causing the loss of the REST repressor complex from the *RE-1* site of neuronal genes (Figure 1.4).

Even if the transition from ES to neural stem/progenitor cell involves post-translational degradation of REST, its co-repressors, mSin3 and CoREST, persisted in the mature neurons: CoREST remains complexed with HDACs providing a platform for the dynamic assembly of repressor complexes required for plasticity even in absence of REST [Ballas et al., 2005]. However, since REST target genes exhibit different affinities, decreasing its concentration could result in differential gene expression: based on this observation REST-regulated genes can be divided in two classes. Class I genes are expressed by default, relying solely on the dissociation of the repressor complex from the *RE-1* site for maximal expression; in contrast class II genes, typified by *BDNF* and *Calbindin*, show, in addition to the REST complex at the *RE-1* site, the persistence of CoREST and MeCP2 (methyl-CpG-binding-protein-2) complexes on a distinct site of mCpGs in their promoters throughout differentiation and relatively low levels of transcription even during differentiation into cortical neurons.

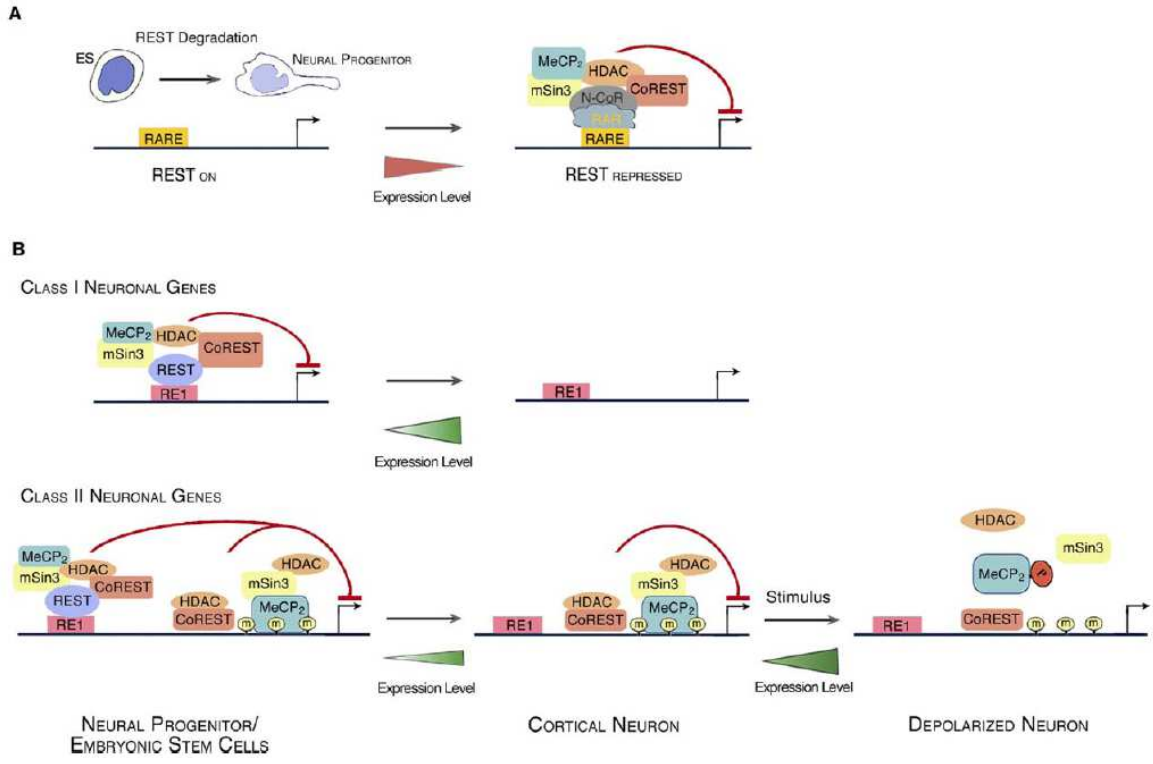

Figure 1.4: Schematic models for the differential regulation of REST and its target genes during development and in mature cortical neurons. (A) REST is downregulated by two different mechanisms during neural development. (B) Two classes of *RE-1*-containing genes are regulated differentially in postmitotic cortical neurons.

The molecular basis for differential recruitment of REST cofactors to *RE-1* is the strength of the REST association: the "complex" sites, where all the corepressors are recruited, are usually more likely to conform to the strong *RE-1* consensus motif, whereas the REST "alone" sites, where only REST is present, match less well with the consensus binding motif and form less stable complexes [Yu et al., 2011].
Epigenetic regulation is a compelling mechanism for controlling development events, which allows inheritance of distinct patterns of gene expression by chromatin modifications that do not involve changes in DNA sequence.

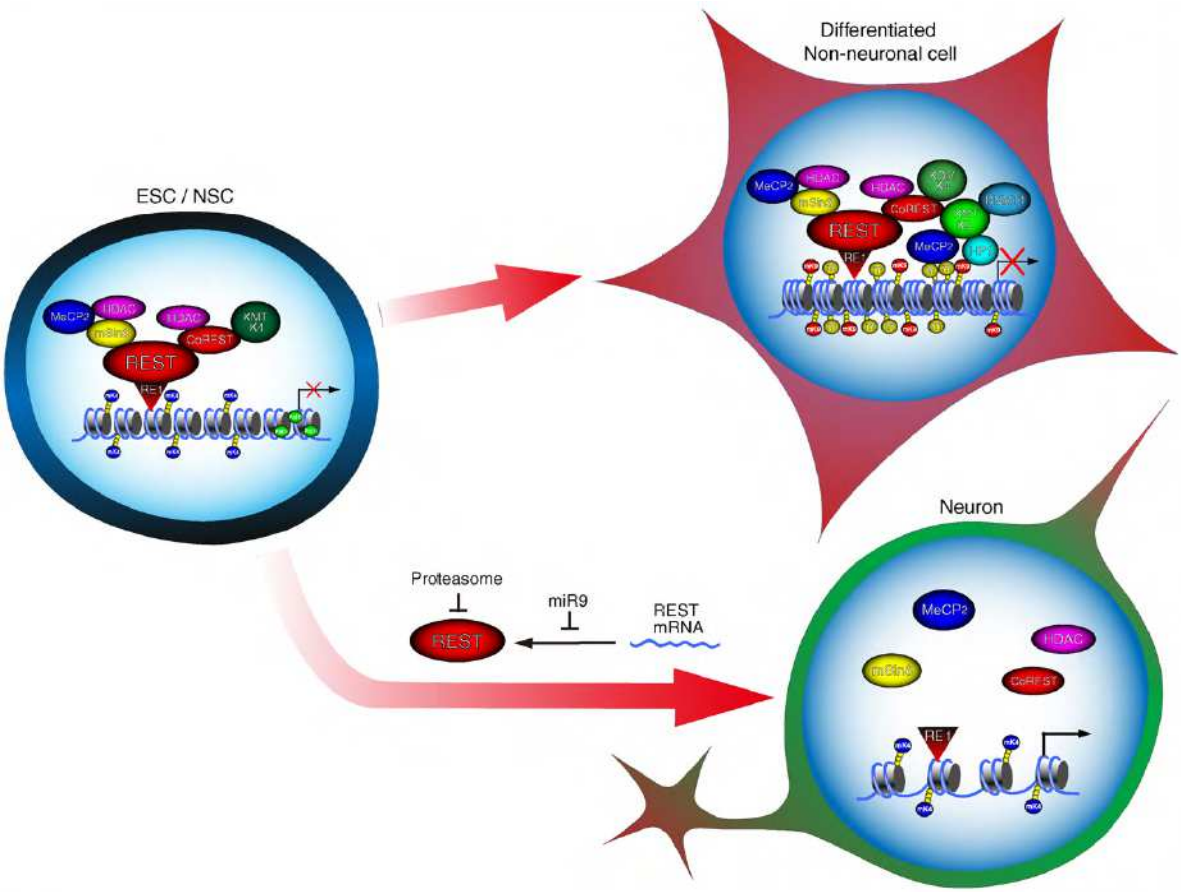

Figure 1.5: Chromatin remodeling of neuronal genes by the REST complex.

In concert with DNA methylation and non-coding RNA-mediated processes, histone modifications such as acetylation and methylation regulates gene expression: in general an increase of histone acetylation by HATs (Histone Acetyl Transferases) induces the remodelling of chromatin from a tightly to a loosely packed configuration, which facilitates transcriptional activation; conversely a decrease of histone acetylation by HDACs (Histone DeACetylases) yields a condensed chromatin structure and thus transcriptional silencing. Patterns of histone methylation are less strainghtforward: for

example histone H3 lysine 4 (H3-K4) methylation is a well-known marker of transcriptionally active chromatin, while methylated H3-K9 and H3-K27 mark transcriptionally inactive chromatin [Juliandi et al., 2010].
Histone acetylation is at least partially involved in the progression of ES cells to NSCs and neurons: ES cells appear to have a higher global level of histone acetylation than lineage-restricted stem cells and differentiated cells, which is consistent with the higher levels of transcription and more open chromatin configuration. Neuronal gene chromatin in ES cells is not completely silenced but is kept in a poised state: during neuronal differentiation the key switch for releasing neuronal genes from their poised state into an active one is the disappearance of REST by proteasomal degradation.
Perturbation of HDAC activity may also contribute to the alleviation of the poised state, as HDAC inhibition results in an enhancement of neuronal differentiation of NSCs by upregulation of the proneuronal NeuroD gene (Figure 1.5).

1.3 STEMNESS AND PLURIPOTENCY

In multicellular organisms the molecular basis for functional diversity of the various cell types is specific gene expression.
Derived from the inner cell masses blastocysts, embryonic stem (ES) cells retain two peculiar characteristics: self renewal, that is the ability to generate exact copies of itself, and pluripotency, namely the capacity to differentiate into virtually all cell types of the body. The precise mechanisms that regulate ES self-renewal and pluripotency are largely unknown: current evidences suggest that ES cells maintain their pluripotent state by expressing a battery of transcription factors including Oct4 and Nanog. Oct4 regulates cell fate in a quantitative fashion and must be maintained at a critical concentration to sustain ES cells self-renewal, through interactions with other factors, such as Sox2 and FoxD3; in contrast Nanog, originally proposed as a transcription repressor of genes important for cell differentiation, primarily acts to activate those genes involved in maintaining stem cell pluripotency. Despite their role in maintaining stem cell pluripotency, little is known about how these transcription factors can regulate the activity of each other: it has been recently hypothesized that a negative feedback loop anchors Oct4, Nanog and FoxD3. In fact, Oct4 regulates Nanog activity in a dose-dependent manner, since Oct4 activates Nanog promoter when expressed below a steady-state and represses it at or above the steady-state concentration; FoxD3 is an activator of Nanog as well, apparently to counteract the repressive activity of Oct4 at the steady-state. On the

other hand Nanog and FoxD3 were found to activate Oct4 promoter, and Oct4 appears to serve as its own repressor to limit its own quantity, thus exerting a negative feedback pressure on the Oct4-Nanog-FoxD3 regulatory loop [Pan et al., 2006].
REST is expressed at high levels in embryonic stem cells, coinciding with the presence of critical regulators such as Oct4, Nanog and Sox2 and confirming that it has a unique role as a protector of self-renewal and pluripotency. It has been shown recently that in mouse ES cells REST is bound to the gene chromatin of a set of miRNAs that potentially interfere with the expression of critical self-renewal regulators such as Oct4, Nanog and Sox2 [Singh et al., 2008]. At least one of these miRNAs, miR-21, markedly decreased the self-renewing capacity of ES cells: interestingly earlier studies indicated that miR-21 expression was higher in differentiated compared to undifferentiated ES cells. Take together these data indicate that REST regulates the self-renewal of ES cells via repressing the transcription of a specific miRNA, miR-21, which suppresses self-renewal through the loss of expression of the critical regulators Oct4, Nanog and Sox2. It is noteworthy that Oct4, Nanog and Sox2 were found to co-occupy REST promoter/enhancer sequences, apparently as an activator complex, suggesting that REST can be a major component of the Oct4-Nanog-Sox2 network in ES cells [Zhou et al., 2007].

1.4 REST AND NEUROGENESIS

Neural stem cells (NSCs) are a subset of undifferentiated precursors that retain the ability to proliferate and self-renew, and have the capacity to give rise to neuronal and glial lineages in the nervous system: neurogenesis, initiated from NSCs and resulting in functional new neurons, is a fundamental process for both embryonic development and adult brain plasticity. Neuron differs from any other cells by containing a specific set of proteins that are critical for execution of specialized functions and are encoded by genes that must be expressed in a neuron-specific manner: the process of creating a properly functional neuron, including NSCs self-renewal and fate specification, neuronal migration, maturation and integration, is regulated by the dynamic interplay between transcription factors, epigenetic control, miRNA regulators and cell-extrinsic signals from the niche where stem cells reside.
Generally thought of as positive regulators of transcription, one can imagine that by recognizing binding sites in the promoter of various markers on neuronal activity and inducing their expression, transcription factors encourage immature cells along the path of neuronal differentiation; a deeper consideration of neurogenesis however, reveals that

differentiation is as much dependent on upregulation of proneural genes as on inhibition of repressors. A master switch for differentiation is NeuroD2, which inhibits the expression of REST, thus linking two important factors within the transcriptional landscape of a maturing neuron. In particular NeuroD2 downregulates REST transcription indirectly, through the activity of the repressor Zfhx1a [Ravanpay et al., 2010].
Regulation of REST plays a fundamental role in the progression of pluripotent cells to lineage-restricted neural progenitors: it functions very effectively as a transcriptional repressor at a distance and is able to repress transcription despite location or orientation of the binding site within a gene.
REST is believed to be a major transcriptional repressor of neurogenesis and activation of its target genes was found to be sufficient to convert NSCs cells to neuronal phenotypes [Su et al., 2004]: Sox4 and Sox11 are transcriptional activators that can induce the expression of neuronal markers in self-renewing precursors and are vital for the establishment of neuronal protein expression. The expression of Sox4/11 is directed to post-mitotic neural cells by a combinatorial action of pro-neural proteins and REST: in fact in addition to its ability to directly repress neuronal gene enhancers, REST also appears to prevent precocious expression of neuronal proteins in undifferentiated neural cells, through its capacity to suppress the expression of Sox4/11 [Bergsland et al., 2006].
While it is important to understand how positive markers of a particular cells fate are induced during differentiation, it is also important to learn how other cellular fates are actively repressed such that a cell does not get lost along the differentiation axis.

1.5 ADULT NEUROGENESIS

When REST binding to various targets in ES, neural stem and hippocampal cells was evaluated it was observed that distinct REST complexes occupied *RE-1* elements at various target genes, supporting the hypothesis that the interaction of REST with its target genes was distinct in these cell types. Adult neurogenesis, a multistep process supporting neuronal turnover throughout life, represents a natural paradigm for maintaining tissue homeostasis, from a resident population of quiescent NSCs. In the mammalian hippocampal subgranular zone and subventricular zone that lines the lateral ventricles, both early and late events of adult neurognesis are governed by multiple signalling pathways and neurotransmitter signalling systems.

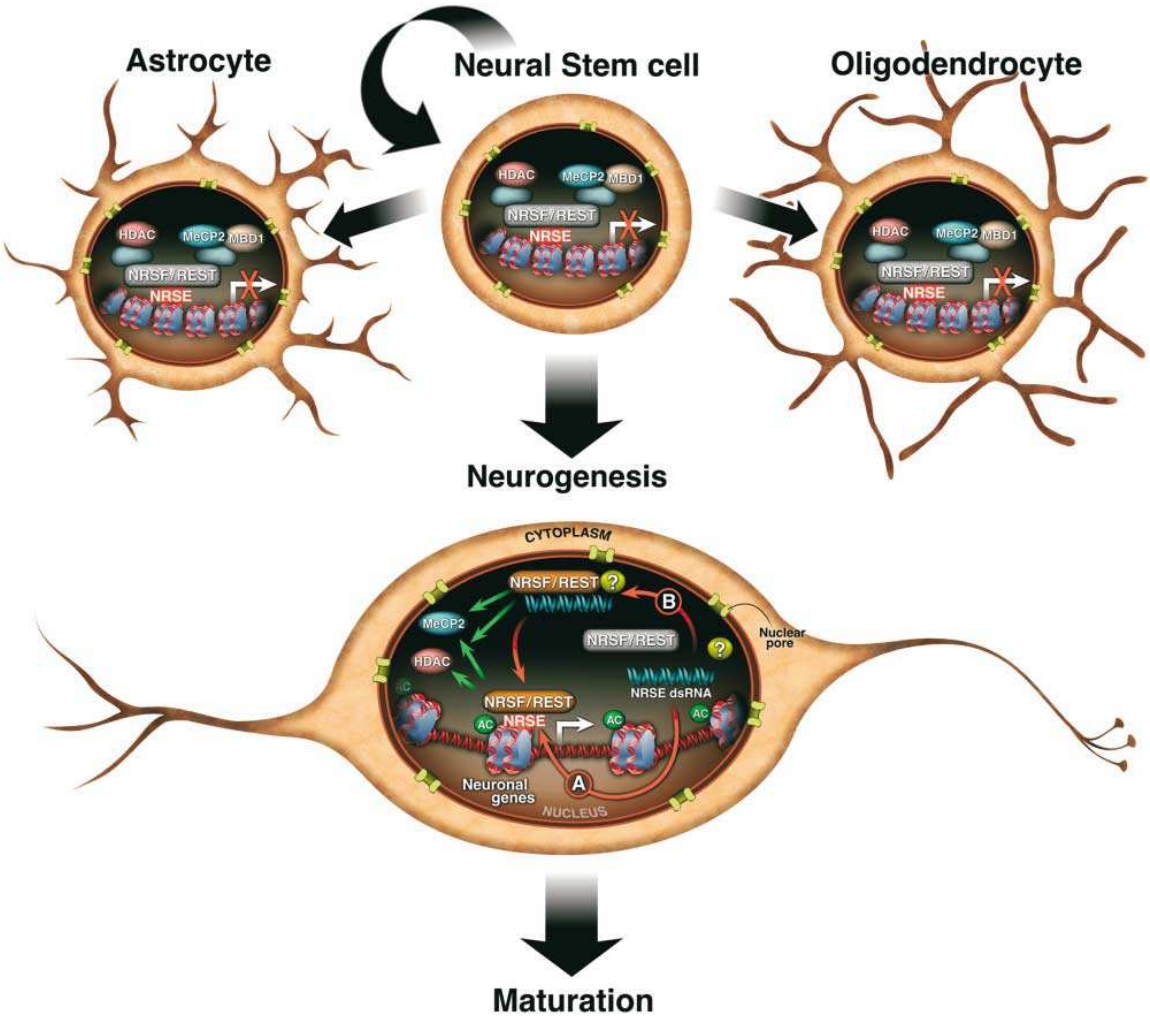

Figure 1.6: Schematic representation of REST expression and action during neurogenesis.

During the course of adult neurogenesis, NSCs get through a number of stages, starting from cells that retain the ability to proliferate, going to slowly dividing "quiescent" NSCs, transit-amplifying progenitors and neuroblasts. It has been difficult to determine whether REST plays distinct roles during embryonic and adult stages of neurogenesis: it has been demonstrated that REST is required to maintain the adult neural stem cell pool and orchestrates stage-specific differentiation [Gao et al., 2011]. Consistent with REST being a transcriptional repressor in non-neuronal cells, its expression has been verified in astorcytes and oligodentrocytes; on the contrary, loss of REST in NSCs leads to the exit of quiescence and loss of neurogenic capacity, since it resulted in de-repression of a cohort of REST target genes and precocious neuronal differentiation (Figure 1.6).

1.6 REST AND ncRNAs

To understand the CNS, scientists have investigated a variety of molecules, including proteins, lipids, and various small molecules: however one large class of molecules, noncoding RNAs (ncRNAs), has been relatively unexplored. ncRNAs function directly as structural, catalytic, or regulatory molecules, rather than serving as templates for protein synthesis. Examples include the small nucleolar RNAs (snoRNAs), which guide site-specific RNA modifications, the miRNAs, which post-transcriptionally repress gene expression, and the rasiRNAs (repeat associated small interfering RNAs), which direct heterochromatin formation at centromeres.

MicroRNAs (miRNAs) are endogenous ~21-nt single-stranded RNAs processed by Dicer, a double-stranded RNA(dsRNA)-specific endonuclease, from one arm of ~ 70-nt stemloop precursors (pre-miRNAs); pre-miRNAs, in turn, are excised from longer primary transcripts (pri-miRNAs) by another dsRNA-specific endonuclease named Drosha. miRNAs function as negative regulators of gene expression by base-pairing with, in many cases, the 3'-untranslated region (3'-UTR) of target mRNAs in the form of ribonucleoprotein complexes, known as RNA-induced silencing complexes (RISCs). Because the interaction between a miRNA and its target requires only partial complementarity, one miRNA could regulate hundreds of mRNAs [Cao et al., 2006].

Many mammalian miRNAs are expressed in a temporal and/or tissue-specific manner. The brain, in particular, hosts a diverse collection of miRNAs; approximately 70% of experimentally detectable miRNAs are expressed in the brain, and during brain development, the expression levels of some miRNAs change dynamically.

High-throughput studies revealed that REST might have links to some ncRNAs; although they were previously known to be involved in the regulation of neurogenesis, the underlying mechanism remained unclear, until recent researches added some of these molecules to the portfolio of REST targets.

miR-124 is the most abundant miRNA in the adult mammalian brain, accounting for 25%–48% of all brain miRNAs: conserved sequences complementary to the seed region of miR-124 are prevalent in the 3'-UTRs of mammalian transcripts, with more than 1100 genes being predicted to be putative targets of miR-124. During neurogenesis miR-124 expression is undetectable or expressed at low levels in progenitor cells and is upregulated in differentiating and mature neurons: previous studies demonstrated that overexpression of miR-124 in HeLa cells led to decreased expression of many non-neuronal genes, to reflect a gene expression program more similar to that of neuronal cells [Lim et al., 2005]. Blocking miR-124 activity in mature neurons leads selectively to

increased levels on non-neuronal transcripts. In addition one of the earliest miRNAs identified as a direct REST target based on binding studies and expression analysis was miR-124, revealing that the establishment and maintenance of neuronal identity requires both derepression of REST-regulated genes as well as post-transcriptional downregulation of non-neuronal transcripts by a miRNA, that is also under REST control: in NSCs and non-neuronal cells REST represses the expression of the proneural miR-124, preventing neurogenesis and facilitating the continued expression of non-neural genes [Conaco et al, 2006]. Upon onset of neuronal differentiation declining REST levels associate with global de-repression of neuronal genes and miR-124, which leads to *en-masse* downregulation of non-neuronal transcripts: together these events alter the balance in favour of the neuronal phenotype. Moreover miR-124 may act in turn by silencing REST and switching on neuron-specific splicing by targeting PTBP1, a repressor of pre-mRNA splicing in non-neuronal cells, allowing the transition form a progenitor state to a differentiated neuronal phenotype (Figure 1.7).

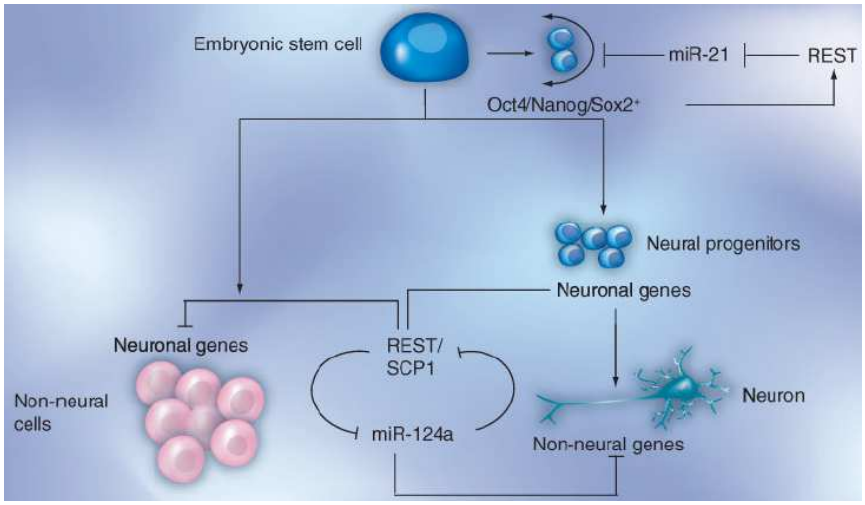

Figure 1.7: Overview of REST function in ES cells self-renewal and neurogenesis in neural and non-neural cells. In ES cells REST represses the expression of miR-21, a negative regulator of proteins involved in self-renewal including Sox2, Nanog and Oct4. In the absence of miR-21, the expression of Sox2, Nanog and Oct4 in ES cells is maintained, which in turn feeds back and activates REST expression and contributes to self-renewal. In non-neural cells, the REST/SCP1 complex represses the expression of neuronal genes and the pro-neuronal miR-124. In neurons, REST is absent and, consequently, miR-124 is expressed: this not only promotes neuronal gene expression, but also represses non-neuronal gene expression.

Thus the intricate transcriptional and post-transcriptional regulatory controls for neuronal gene expression are highly integrated and mediated, in part, by double-negative feedback loops between REST and the miRNAs.

Since these initial reports, the list of putative REST-target miRNAs has grown substantially: computational analyses, genome-wide binding studies and ChIP analyses have collectively identified several neuron-specific miRNAs, including miR9-1, miR9-3, miR-29a/miR-29b, miR-95, miR-132, miR-133, miR-135b, miR-139, miR-153, miR-218-212, miR-346, miR-375 and miR-455, to have the potential for regulation by REST.

miR-9 has been proposed as one of the crucial regulators of neuronal development, physiology and pathology in several organisms: in mammals it is induced upon embryonic stem cell neuronal differentiation and blocking its function during this process decreases neuronal differentiation at the expense of astrocytes. In addition it has onco-suppressor properties, with a crucial role in repressing cell proliferation of human neuroblastoma and medulloblastoma. In proliferating cells, where it is highly expresses, REST is bound to the miR9 promoter preventing transcriptional activity; upon neuronal differentiation REST dismissal and the concomitant phosphorylation of the already bound neuronal activator CREB activate transcription. Noteworthy REST mRNA is also targeted by miR-9 during neuronal differentiation [Laneve et al., 2010].

Although the disappearance of REST repressor complex in differentiating neurons leads to the expression of *RE-1*-containing genes during development, a reduced presence of REST has been observed even in adult neurons: one mechanism to cancel its repressor effect has been demonstrated by the discovery of a nuclear localized small modulatory RNA (smRNA), which appears transiently at the early period of neuronal differentiation in cultured adult neural stem cells. This double-stranded smRNA is about 20 bp in length, carrying the same sequence as *RE-1/NRSE* and hence designated as *RE-1/NRSE*-smRNA [Kuwabara et al., 2004]. *RE-1/NRSE*-smRNA binds REST as well as the *RE-1* dsDNA: the dsRNA captures REST as a decoy and release the genome from repression. Binding the *RE-1/NRSE*-smRNA to REST prevents the association of corepressor proteins, such as HDACs and methyl-DNA binding proteins, and, possibly through additional mechanisms, converts REST to a transcriptional activator. After the *RE-1/NRSE*-smRNA appears, transcriptional activation of *RE-1*-containing genes is triggered.

Not surprisingly, perturbations of REST activity and related deregulation of ncRNAs have also been implicated in the molecular pathophysiology of diverse disorders, that range from cancer to neurodegenerative and neurodevelopmental diseases.

1.7 REGULATION OF REST TRANSCRIPTION

Despite the extensive number of studies to elucidate the function of REST, relatively little has been done to determine how expression of *REST* gene itself is regulated. Analysis of the human, mouse and rat gene structures has shown that the exon and intron structure is conserved and that *REST* transcripts begin from one of three alternatively used 5' exons (indicated by black boxes in Figure 1.8). The presence of three alternatively used 5' exons, called exon A, B and C, suggests that *REST* contains at least three promoters: promoter A has been shown to be the most active in non-neuronal cells, while promoter B in neuronal cells. Promoter C, on the contrary, showed no activity in neuronal cells and only weak activity in non-neuronal and neural progenitors, suggesting that it acts as a potent silencer of the other two *REST* promoters: these findings suggest that the cell-specific regulation of any of the three *REST* promoters by adjacent or distal enhancers or repressors is likely to depend on the action of trans-acting factors that bind to these regions in particular cell types.

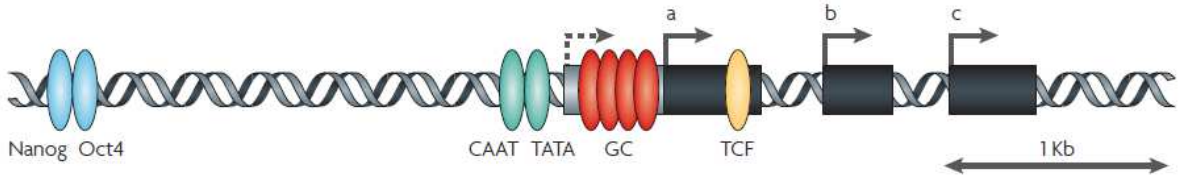

Figure 1.8: Schematic representation of *REST* promoter region.

Human and mouse *REST* promoters have multiple transcription start sites and contain GC boxes that recruit the Sp1 family of transcription factors; other transcription factors involved in the regulation of *REST* expression are PBX, AP-1, CREB and Sox [Koenigsberger et al., 2000]. Moreover *REST* expression is regulated by the Wnt signalling pathway which is important in many tissues, particularly during development: ectopic expression of Wnt1 or β-catenin induces *REST* in chicken embryonic spinal cord through a T-cell factor (TCF)-binding site, present in exon A and conserved between humans and mice. *REST* was also identified as a target for the transcription factors Nanog and Oct4 in mouse and human embryonic stem cells and in fact both Nanog and Oct4 have been shown to be simultaneously bound to a region about 2 kb from exon A.
Chromatin immunoprecipitation analysis of the *REST* promoter revealed also the presence of an HDAC complex that contained mSin3, CoREST and MeCP2. Intriguingly the Retinoic Acid Receptor (RAR) and its corepressor, N-CoR, were part of this repressor complex, bound to a region containing a Retinoic Acid Response Element (RARE), located 400 bp upstream of the transcriptional start site [Ballas et al., 2005]. During

terminal differentiation an unliganded RAR repressor complex is recruited to a RARE site located upstream from the transcriptional start site, causing the repression of REST transcription and exit from the cell cycle and the onset of terminal differentiation.
Although REST mRNA levels are known to change in response to several stimuli, it is not known whether this is mediated by changes in transcription or by changes in mRNA turnover, possibly mediated by miRNAs.

1.8 REST DEGRADATION

The transition from stem or progenitor cell to a post-mitotic neuron requires disarming REST: several groups have demonstrated that REST levels are significantly different in ES cells, NSCs and non-neural cells. In particular REST levels appear to be regulated by transcriptional and post-transcriptional mechanisms depending upon the cell type (Figure 1.9): during cortical differentiation, post-translational degradation of the REST protein precedes both its dismissal from *RE-1* sites and transcriptional inactivation of the *REST* gene itself.

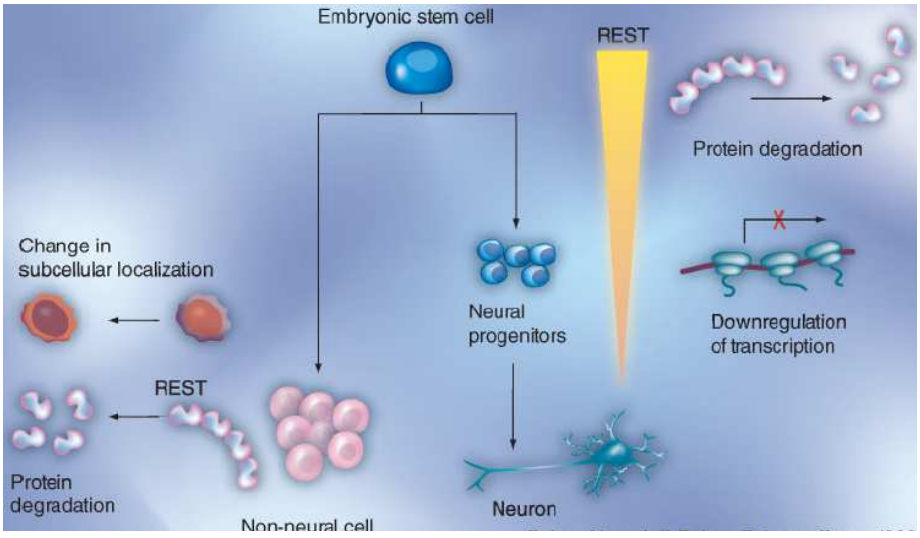

Figure 1.9: Regulation of REST by transcriptional and post-transcriptional mechanisms. REST is expressed at high levels in ES cells; a decline is observed during their specification into NSCs. In neural progenitors REST is transcriptionally downregulated by the RAR; in non-neural cells REST function is also regulated by changes in its subcellular localization mediated by the Huntingtin/RILP complex.

REST was identified as a substrate for a member of the Skp1-Cullin1-F-box (SCF) β-TrCP family ubiquitin ligases E3s and in fact REST levels are regulated through its

proteasomal degradation, involving β–TrCP E3 ubiquitin ligase. $SCF^{\beta-TrCP}$ binds and ubiquitinates its substrates through a conserved phospho-degron (DSGXXS), in which the serine residues are phosphorylated: REST has two adjacent yet distinct similar motifs at its C-terminus and degron-mutant REST was shown to be substantially more stable that wild-type REST, thus attenuating differentiation [Westbrook et al., 2008].

Any disruption of the cell-cycle-, β-TrCP-dependent regulation of REST in non-neuronal cells contributes to genetic instability; since REST is a transcriptional repressor, its degradation in the G2 phase of the cell cycle might be necessary to de-repress genes involved in mitosis. In particular $SCF^{\beta-TrCP}$-dependent degradation of REST during G2 is required for the fidelity of mitosis, since it allows transcriptional de-repression of Mad2, an essential component of the spindle assembly checkpoint [Guardavaccaro et al., 2008]. Mutant REST, unable to bind β-TrCP, inhibiting Mad2 expression causes faulty activation of the spindle checkpoint, with specific defects such as shortened mitosis, premature sister-chromatid separation, chromosome bridges and mis-segregation in anaphase.

REST proteolysis must be accurately controlled to avoid subjecting neuronal tissues to cancer risk: in fact increased REST levels resulting from its over-expression, as observed in human neuronal tumours, would both inhibit differentiation and generate chromosomal instability, two mechanisms that contribute to tumour development. On the other hand hyper-activation of the pathways priming REST degradation may be oncogenic in epithelial tissues, thus serving as novel therapeutic targets in cancers with compromised REST function.

However reciprocal mechanisms that stabilize REST are needed. The deubiquitylase HAUSP (Herpesvirus-Associated-Ubiquitin-Specific Protease 7) antagonizes β-TrCP-mediated ubiquitylation of REST and prevents NSCs differentiation: the balance between ubiquitylation and deubiquitylation may determine the net REST protein levels, defining the maintenance of stemness or initiation of neuronal differentiation. When the HAUSP-mediated deubiquitylation overrides β-TrCP-mediated ubiquitylation REST is stabilized to suppress differentiation and promote NSCs maintenance; in contrast, when β-TrCP-mediated ubiquitylation exceeds HAUSP-mediated deubiquitylation, REST is targeted for degradation, which promotes cell differentiation by releasing repression of differentiation-associated genes [Huang et al., 2011].

1.9 REST INVOLVEMENT IN SEVERAL PATHOLOGIES

REST was found to regulate several genes that may impact important biological processes: most of the understanding about the role of REST in normal adult tissue comes from studies of disease states, in which REST function is altered.

Increased levels of REST were found in rat hippocampal and cortical neurons in response to epileptic seizures and ischaemia: these insults de-repress REST mRNA and protein in dying neurons, and this was suggested to be a critical mechanism of insult-induced neuronal-death [Calderone et al, 2003]. REST was also found to repress the μ-opioid receptor in neuronal cells, and thus may have a neuroprotective role in opium addiction. Similarly, REST was found to repress the serotonin 1A receptor, which is implicated in depression and anxiety [Lemonde et al., 2004].

Potential roles for REST and its target genes have also been implicated in the pathogenesis of Huntington's disease. REST and huntingtin protein were found to form a complex that forced REST to localize in the cytoplasm in normal neuronal cells, inhibiting its function; in patients with Huntington's disease this interaction was ablated by the huntingtin mutation, resulting in the translocation of REST in the nucleus and causing blockade of neuronal gene expression. One of REST targets is the neuronal survival factor BDNF (Bran Derived Neurotrophic Factor), low levels of which are thought to contribute to neuronal degeneration [Zuccato et al., 2003].

Among non-neuronal tissues REST was found to be present in normal ventricular myocites, having a role in maintaining normal cardiac structure and function: in cardiac dysfunctions and arrhythmogenesis *REST* expression is inhibited, resulting in the expression of fetal cardiac genes.

Alterations of REST are evident even in several tumors, although it appears to be achieved through diverse mechanisms: several genomic alterations have been identified in *REST* coding region, all located in exon 3, which may cause changes in the biochemical properties of the REST protein, and affect its binding to target DNA.

In some neural cells REST can behave as an oncogene: increased REST expression has been implicated in neuroblastomas and medulloblastomas. Medulloblastoma is one of the most malignant pediatric brain tumors and is believed to arise from undifferentiated neuroectodermal stem cells in the cerebellum: these cells have a primitive embryonal phenotype and show the capacity for divergent differentiation. To date the mechanism of medulloblastoma tumorigenesis is still unknown. Medulloblastoma cells express REST at high levels, compared with either neuronal progenitor cells or fully differentiated neurons, accompanying a failure to differentiate: in contrast to the reduction of REST during

normal neuronal differentiation, this generates proliferating cells where many target genes characteristic of post-mitotic neurons are silenced. A recombinant transcription factor, REST-VP16, constructed by replacing the repressor domains of REST with the activation domain of the herpes simplex virus protein VP16, can compete with endogenous REST for DNA binding, activate REST target genes and trigger apoptosis through the activation of caspase 3 cascade [Fuller et al., 2005]. However medulloblastoma tumorigenesis starts when two conditions occur in NSCs: lack of differentiation, as a result of REST expression, and overall increased proliferation, as a result of the Myc oncogene expression [Su et al., 2006].

On the other hand a variety of tumors, including those arising in breast, ovary, lung and colon, activates expression of neuron-specific genes: the expression of neuronal genes out of their normal context causes the permanence of malignant cells in the cell cycle. Abnormal expression of neuronal genes results from the loss of REST activity, which normally represses their expression and leads to oncogenic conversion of these cells, perhaps through the aberrant expression of REST target genes, resulting in an abnormal neuroendocrine phenotype. Moreover inactivating mutations of the *REST* gene were identified in colorectal cancer: deletions of human chromosome 4q12, on which REST is located, are frequent events in this type of cancer. A frameshift mutation in the *REST* coding region, resulting in a premature termination codon in the center of the coding sequence and yielding a truncated protein that lacks the C-terminal repression domain, was also identified in primary colon adenocarcinoma: this dominantly acting truncation is capable of transforming epithelial cells through an increased of PI(3)K-dependent signalling, in both intensity and duration [Westbrook et al., 2005]. Finally, reduced REST function caused by increased expression of a splice variant that lacks some of the DNA-binding domain and the C-terminal repression domain has been associated with small-cell lung carcinomas (SCLCs): this event leads to the expression of neuronal markers, like L1-cell adhesion molecule (L1-CAM) and neural cell adhesion molecule (NCAM), enhanced cell proliferation and anchorage independent cell growth [Kreisler et al., 2010]. In some instances the inappropriate expression of neural genes elicits an autoimmune response that culminates in neurological disorders, collectively known as Paraneoplastic Neurologic Degenerations (PNDs).

Thus, depending on the cell type REST may both have tumorigenic and tumor-suppressor effects, which would indicate that REST is a major biological regulator of normal and abnormal development: in epithelial cells REST functions as a tumor-suppressor, while in differentiating or differentiated neuronal cells REST repressor activity has an oncogenic function.

2. PROTEIN GLYCOSYLATION

Fundamental to the normal functioning of a single cell or multicellular organism is control over their entire metabolism. This control is important for the central dogma of molecular biology that DNA leads to RNA, which in turn leads to protein. In this context, control is related to decision-making about gene expression, transcription and translation.
In eukaryotes, the process of going from DNA to functional protein usually consists, at the most basic level, of at least six discreet steps: transcription, splicing, polyadenylation, transport, translation, and post-translational modification. At the level of RNA, alternative transcriptional start sites and alternative mRNA splicing may increase complexity; regulation of mRNA levels and differential protein stability can modulate protein levels. External stimuli also contribute to the uniqueness of a cell's proteome through specific expression patterns. Although the most common way that diversity is generated in proteins is, of course, alteration of the order and composition of their constituent amino acids, additional diversity is generated by covalent modification of amino acids by the addition of post-translational modifications (PTMs), such as phosphates or glycans, or by disulfide bond formation. While only 20 amino acids are used in protein synthesis, as many as 140 different amino acids are found in proteins due to covalent post-translational modification [Vosseller et al., 2001].
PTMs are ubiquitous in the cell: they regulate the function of proteins and provide a mechanism for increasing the diversity of protein structures. In addition, such modifications may result in the localization of proteins to specific cellular organelles, with incorrect targeting being associated with a number of diseases. It is often believed that PTMs may regulate protein activity by changing the protein surface and by introducing new functional groups that act as a signaling tag for binding to other molecules. These modifications may also modulate the thermodynamic and kinetic features of proteins (e.g. stability, structural flexibility, and folding rate): thus, they efficiently and economically increase the complexity and dimensions of the primary gene products. Various PTMs have been shown to affect protein biophysics [Shental-Bechor and Levy, 2009]. For example, phosphorylation contributes with negative charges to the protein surface and can induce conformational changes, while methylation and acetylation mask charges and play an important role in protein–DNA interactions.

2.1 GLYCOSYLATION

Among the various PTMs the most common is the complex process of glycosylation, that is the covalent attachment of one or more carbohydrates, also termed "glycans", that occurs during or after the process of protein synthesis. It is a complex process that commonly involves different monosaccharides, attached to any of eight types of aminoacid residues and is assisted by many enzymes: although the number of different types of glycosidic bond seems to be increasing at a frenzied rate, N- and O-linked glycans, where the amino acid residues involved are asparagine and serine or threonine, still remain the most abundant. Glycosylation is possible involving about 1% of human genes [Lowe and Marth, 2003]; furthermore, according to estimates based on the SwissProt database, more than one half of all proteins are glycosylated [Apweiler et al., 1999].
Glycosylation imparts many properties to proteins and might be important for their function. One of the initial functions of glycosylation in a given protein is to direct the protein to the appropriate cellular location: for example, many lysosomal proteins harbor a mannose-6-phosphate moiety that behaves as a signaling molecule. In addition, the calnexin–calreticulin cycle in the endoplasmic reticulum (ER), a quality-control mechanism that prevents misfolded proteins from being transported further, is dependent upon the N-glycosylation of the proteins in question (see paragraph 2.3.1). Among the roles ascribed to glycosylation, the most important is the recognition of glycosylated proteins by carbohydrate-binding proteins called lectins: the lectin–sugar interaction mediates several important processes, including host-cell recognition by pathogens. A benchmark example is the docking of the influenza virus on host cells by binding to the terminal sialic acids of the host-cell surface glycoproteins. Glycosylation can modulate physico-chemical properties of proteins, such as thermal stability and hydrophobicity. In most instances, glycosylation of a protein begins even as the protein is being synthesized, suggesting that glycosylation might have an important role in protein folding and oligomerization: inhibition and/or suppression of glycosylation often results in aggregation or misfolding of proteins. Also, certain glycosylation sites in a given protein are more important compared with other sites, suggesting that glycosylation exerts local effects on protein conformation in addition to more global effects.
Glycoproteins have been found in essentially all living organisms, ranging from eubacteria to eukaryotes, from unicellular organisms, such as yeast and trypanosomes, to the highly differentiated tissues of the animal and plant kingdoms. Thirteen different monosaccharides and eight amino acid types participate in these bonds so that at least 31 sugar–amino acid combinations exist: the aminoacids involved in glycosylation process

known so fare are asparagine, arginine, serine, threonine, proline, hydroxyproline, tryptophan and tyrosine. Formation of the sugar–aminoacid linkage is a crucial event in the biosynthesis of the carbohydrate units of glycoproteins. The modification of proteins through enzymatic glycosylation is an event that reaches beyond the genome and is controlled by factors that differ greatly among cell types and species. Many elaborate glycosylation routes have been identified in various organisms that lead to the mature carbohydrate units on glycoproteins that are secreted by cells or become components of its membranes, cytoplasm, or nucleus [Spiro, 2002].

2.2 FOLDING OF GLYCOPROTEINS

Glycosylation may affect many protein activities: it has been shown to facilitate folding kinetics and to affect the biophysical properties of proteins, which are a consequence of the underlying energy landscapes that govern their folding. It may also modulate a protein's characteristics by affecting differently the energetics (enthalpy or entropy) of the unfolded and/or folded state. However, the role of glycans in attaining the correct fold is ambiguous. There is evidence for protein misfolding and aggregation in the absence of glycans; in other cases, however, elimination of some or all glycans has no effect on folding or protein function, implying that some glycosylation sites are more crucial to protein folding or function than others and that the effect of glycans on folding is likely to be local.

A protein folds by navigating through an energy landscape that is globally funneled toward a structurally defined native state: the funneled nature of energy landscapes is a product of the evolutionary selection of sequences that are minimally frustrated, in that they have fewer interactions that are in conflict with the native state and thus exhibit relatively high conformational preferences. The defining event in the biogenesis of peptide-linked oligosaccharides is clearly the formation of the sugar–aminoacid bond: this, in most instances, determines the nature of the carbohydrate units that will subsequently be formed by the cellular enzymatic machinery, which in turn influences the protein's biological activity. A multitude of diverse sugar–aminoacid combinations have been described.

Several factors may govern the effect of glycosylation on the protein energy landscape and, consequently, on the biophysical characteristics of the protein. From the oligosaccharide perspective, the size of the glycans, their flexibility and structure, the number of branches, may affect the properties of the protein. From the protein

perspective, the location and chemical surroundings of the glycosylation sites, as well as the number of occupied glycosylation sites, may influence the effect of glycosylation on the protein. Quantifying how glycosylation can affect protein biophysics by modulating the protein's energy landscape may formulate a 'glycosylation code': deciphering such a molecular code, however, is a difficult task for two main reasons. First, oligosaccharides are complex and heterogeneous systems, and structurally more complex than proteins (although they lack folding capability), since they can form various topological trees via various linkage positions, while the proteins are linear polymers. Second, the number of glycoproteins having a resolved 3D structure is very limited (while ~50% of all proteins are glycosylated only about 3,5% of the proteins in the Protein Database (PDB) carry covalently bound glycan chains and even fewer have full structure of the glycans), mainly because the conformational heterogeneity of the oligosaccharides often inhibits their crystallization or makes them barely visible in X-ray crystallography or nuclear magnetic resonance spctroscopy [Apweiler et al., 1999].

Glycans, which are bulky hydrophilic polymers, often increase solubility of the protein to which they are attached as well as its resistance to proteolysis; moreover, the covalent binding of glycans to protein surfaces may enhance their thermodynamic and kinetic stability. Greater thermodynamic stability is expressed by an increase in the melting temperature of the glycosylated protein or in the free energy difference between the unfolded and folded state ensembles, being greater for the glycosylated protein than for its nonglycosylated wild-type counterpart.

Many glycoproteins contain several glycosylation sites, which may be occupied alternatively or concomitantly: it is therefore of great interest to understand how the biophysical characteristics of glycoproteins are affected by the number of oligosaccharides covalently attached to the proteins and whether there is a cooperative effect between the various glycans: the degree of folding increases significantly with the degree of glycosylation, by about 4% per added glycan. The attachment of a sugar can modulate protein stability depending on the position of the glycosylation site: glycosylation at a more-structured region results in destabilization, while glycosylation at a more-disordered region increases protein stability. The degree of glycosylation also affects protein folding kinetics: the unfolding rate decreases as the degree of glycosylation increases, indicating kinetic stabilization of the protein [Shental-Bechor and Levy, 2009]. In some structural studies of glycoproteins, specific interactions between the sugar and aminoacids of the folded protein were observed: the stabilization effect in these cases is enthalpic in origin [Hanson et al., 2009]. Earlier studies suggested that the origin of glycoprotein's thermal stabilization is largely entropic, rather than enthalpic, proposing a 'chaperone-like' activity

of glycans [Wang et al., 1996]. Both the enthalpy and the entropy of the folded state decrease as the degree of glycosylation increases: the reduced entropy induced by glycosylation is caused by the suppression of protein fluctuations in the folded state while its conformation remains unchanged. The glycans can therefore be viewed as imposing pressure on the protein that reduces the flexibility of the folded state: the rigidified structure induced by glycosylation is characterized by tertiary interactions that are more localized in their optimal distances, thus reducing enthalpy.
These evidences demonstrate that there is a fine cross-talk between the carbohydrate moiety and the peptide backbone of a given protein: mutating the glycosylation sites of a protein *in vivo* reduces the efficiency of folding and increases degradation [De Virgilio et al., 1999].

2.3 N-GLYCOSYLATION

N-linked glycosylation results from the attachment of the glycan to an asparagine (Asn) side chain of the protein: five different N-glycan linkages have been reported, of which N-acetylglucosamine to asparagine (GlcNAc-β-Asn) is the most common. The process occurs on the lumenal side of the endoplasmic reticulum (ER) membrane while the protein moiety is being synthesized on ER-bound ribosomes and is translocating through the translocon in the ER membrane: the fact that N-glycosylation starts during protein synthesis, when the glycan is added to the still unfolded protein, indicates that it may assist in obtaining the correct fold, following the recruitment of the lectins calnexin and calreticulin. Processing of the N-glycans is coupled with the fates of glycoproteins in cells: they play a crucial role in keeping order in the 'protein society' by controlling folding, transport, and degradation of their carrier polypeptide chains [Roth et al., 2010].
The biosynthesis of all eukaryotic N-glycans begins on the cytoplasmic face of the ER membrane with the transfer of N-acetylglucosamine-phosphate (GlcNAc-P) from UDP-GlcNAc to the lipid-like precursor dolichol-phosphate (Dol-P) to generate dolichol-pyrophosphate N-acetylglucosamine (Dol-P-P-GlcNAc). Fourteen sugars are sequentially added to Dol-P before *en bloc* transfer of the entire glycan to the protein that is being synthesized and translocated through the ER membrane by the Sec61 translocon complex: subsequently an oligosaccharyltransferase (OST), which is a hetero-oligomeric membrane protein, in association with a translocon in the lumen of the ER, catalyzes the transfer of the $Glc_3Man_9GlcNAc_2$ procursor from dolichol pyrophosphate to the Asn residue in the consensus sequence. The protein-bound N-glycan is subsequently remodeled in the ER

and Golgi by a complex series of reactions catalyzed by membrane-bound glycosidases and glycosyltransferases: these reactions encompass the sequential removal of all three glucose residues of the branch A and some of the mannose residues of the branch B and C, as well as branch A (Figure 2.1). The glycan processing leads to the formation of high-mannose, hybrid or complex type oligosaccharides. Oligomannose glycans are composed of solely mannose residues attached to the core $Man_3GlcNAc_2$; in complex glycans, instead, "antennae" initiated by N-acetylglucosaminyltransferases (GlcNAcTs) are attached to the core; hybrid glycans contain only mannose residues attached to the Manα1–6 arm of the core and one or two antennae on the Manα1–3 arm. [Cummings, 1992]. The glycans and their final processing vary with the protein and the species in question. For example, although mammalian and yeast systems start with the same kind of N-glycans that undergo glucose and mannose removal by glucosidase I or II and α-mannosidase I, yeast systems go on to form high-mannose structures, whereas mammalian N-glycans lose mannose residues to form $Man_5GlcNAc_2$-Asn by the action of further mannosidases [Mitra et al., 2006].

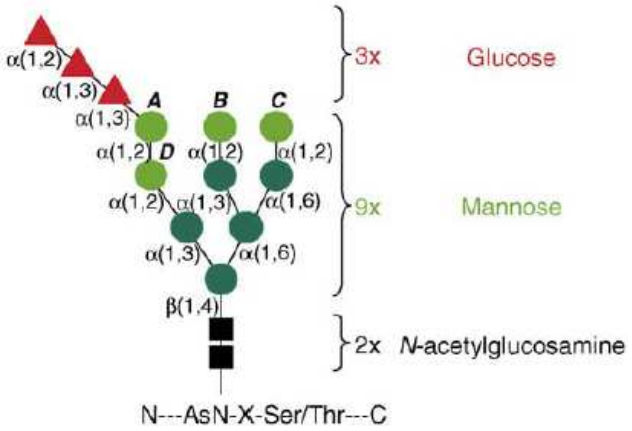

Figure 2.1: Structure of the $Glc_3Man_9GlcNAc_2$ precursor linked to an Asn residue.

In addition, a strict consensus sequence, Asn-Xxx-Ser/Thr, was postulated, where Xxx can be any aminoacid except proline (Pro): it has been proposed that the Ser or Thr is required for a hydrogen-bond donor function in enzyme binding and in oligosaccharide transfer; the negative effect of Pro as the Xxx aminoacid has been attributed to its interference with the ability of the peptide chain to adopt and stabilize a loop conformation [Bause E., 1983].

Variation in the degrees of saturation at available glycosylation sites results in heterogeneity in the mass and charge of glycoproteins; each glycosylation site may further exibit variable occupancy and each occupied site, in turn, exhibits heterogeneity in the attached glycan structures. The composition and structure of the carbohydrate attached to

a glycoprotein are determined by the glycosylation machinery available in a specific cell type; there changes can significantly alter the structure of the protein. Variable sequon occupancy can give rise to glycoforms: this term essentially describes this structural diversity, which is a result of cell specific biosynthesis, and leads to functional diversity [Devasahayam, 2007].

Potential N-glycosylation sites can be identified by the presence of the Asn-Xxx-Ser/Thr sequon in peptide sequence databases; however it is very difficult to determine which of these sites are occupied. Structural analysis by X-ray crystallography can provide direct and unambiguous evidence for the occupancy of a glycosylation site; on the contrary, evidence from X-ray crystallography for the unoccupancy of a site is more ambiguous because the absence of the glycan is only one reason for the absence of resolved electron density [Petrescu et al., 2004].

Comparative analysis of the properties of the protein around glycosylation sites can be helpful in identifying important glycosylation features. It has been shown that N-glycosylation sequon occupancy is modulated by local aminoacid sequence: there is a marked preference for hydrophobic aminoacids at either side of the glycosylation site, with an increased probability of finding small hydrophobic amino acids at position +1 and larger hydrophobic amino acids at position +3, whereas there is an increase in the probability of finding aromatic residues immediately before a glycosylation site. There is a significant above average occurrence of aromatic and Ser/Thr residues and a significantly below average occurrence of basic residues within 5 Å of a glycosylation site; there is also a large preference for Thr, as opposed to Ser, in position +2, due to a different affinity for the recruitment of the OST. In addition, the Asn-GlcNAc linkage only adopts one significant populated conformation: glycosylation, in fact, reduces the overall conformational freedom of the side chain and of the protein backbone (see paragraph 2.2). Analysis of the secondary structures found around glycosylated Asn residues revealed that occupied glycosylation sites can occur on all forms of secondary structure, with a bias in favor of turns and bends (~27% and 17,5% respectively): the lowest incidence of glycosylation sites is on helices, with 10,5% incidence. The majority of sites are found in convex or flat regions of the surface; there is also a large number of glycosylation sites located on the edge of a groove, with the glycan filling partially or totally the depression. Finally about 10% of the glycans are situated in deep, narrow recesses in the protein surface: the positions of glycosylation sites may have evolved to act as landmarks for ending or starting regions of regular secondary structure.

As well as stabilizing a particular local fold, glycans may play a role in organizing the folding process by promoting changes in backbone conformation of the folding

intermediates. Given the hydrophilic nature of monosaccharides, one might have expected to be there fewer occupied glycosylation sites with lower accessibilities and more with higher: the small number of potential sites with a very high accessibility is consistent with the idea that protein-glycan contacts are functionally significant and have been selected for during evolution [Petrescu et al., 2004].

2.3.1 N-GLYCOSYLATION AND PROTEIN QUALITY CONTROL

N-glycosylation plays a more general role early in the life of proteins, which is in the quality control of protein folding: the rough ER (rER) is not only the site of synthesis of secretory and membrane proteins but also of initial steps of protein N-glycosylation, and provides an environment promoting protein folding. Thus, the rER is also the main organelle in which the quality control of protein folding takes place. Both, the carbohydrate biosynthetic process and the resulting oligosaccharide are evolutionary conserved from yeast to mammals.

The α1,2-linked outermost glucose residue is removed from the oligosaccharide, probably immediately after its transfer to the polypeptide, by oligosaccharyltransferase (Figure 2.2 A). The reaction is catalyzed by glucosidase I, which is a neutral processing α1,2-exoglucosidase of the glycosyl hydrolase family 63 with a type II membrane protein topology: removal of the outermost glucose residue by glucosidase I prevents further interaction of the oligosaccharide with the OGT. The two inner α1,3-linked glucose residues are subsequently removed by glucosidase II: this neutral processing α1,3-exoglucosidase belongs to the glycosyl hydrolase family 31 and is a luminal glycoprotein composed of two subunits, a catalytic α subunit and a regulatory β subunit.

In general terms, di-glucosylated oligosaccharides generated by glucosidase I appear to represent a glyco-code promoting protein N-glycosylation, by influencing the oligosaccharyltransferase-mediated reaction, whereas mono-glucosylated oligosaccharides represent a glyco-code promoting protein folding through the calnexin/calreticulin cycle (Figure 2.2 B).

Like many chaperones present in the rER that shelter glycoproteins to prevent their aggregation and assist them in folding, the calnexin/calreticulin (CNX/CRT) cycle provides an environment for productive protein folding [Caramelo and Parodi, 2008]. Calnexin and calreticulin are highly related legume lectins with calnexin being a type I membrane protein and calreticulin being a soluble luminal ER protein: both are retained

in the ER through specific localization signals at their C-termini and have been found to form complexes with ERp57, a thiol-disulfide oxidoreductase.

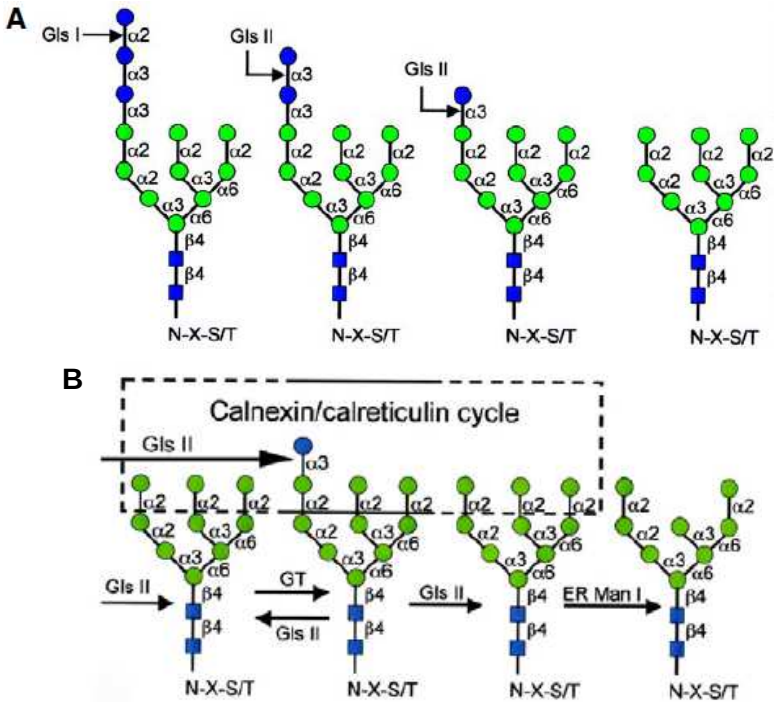

Figure 2.2: A) Schematic representation of the sequential glucose trimming by glucosidases I and II; B) The calnexin/calreticulin cycle.

The carbohydrate-binding specificity of calnexin and calreticulin appears to be identical: they both require the innermost α1,3-linked glucose residue present on the branch A of the oligosaccaride. Regardless, the first trimming reaction by glucosidase II resulting in the $Glc_1Man_9GlcNAc_2$ oligosaccharide permits entry of the glycoprotein in the CNX/CRT cycle: the ensuing lectin-carbohydrate interaction acts in a chaperone-like fashion by sequestering the mono-glucosylated glycoproteins. Exit from the CNX/CRT cycle requires glucosidase II as well, which removes the innermost glucose residue: the fate of de-glycosylated glycoproteins after their exit from the CNX/CRT cycle strongly depends on their conformation. The folding sensor UDP-glucose:glycoprotein glycosyltransferase (UGGT) scans the structure of the polypeptide for exposed non-native determinants: proteins with a native conformation can exit the ER, while if misfolded domains are detected then UGGT adds a glucose residue to the terminal mannose of branch A, thus returning the immature polypeptide to the CNX/CRT chaperone system, for another cycle [Herbert et al., 2005]. UGGT is a soluble luminal glycoprotein, which is present throughout the smooth and rER and concentrated in the pre-Golgi intermediates, and acts in a two-fold manner: first as a folding sensor and

second as a glycosyltransferase [Trombetta and Parodi, 2003]. It can bind to non-native conformers most probably through interaction with exposed hydrophobic amino acid patches; this is followed by the re-glucosylation reaction and re-entry in the CNX/CRT cycle. However glycosylation by UGGT is transient, since the glucose residue will be removed by glucosidase II for exit from the CNX/CRT. For the majority of glycoproteins expressed in mammalian cells deletion of UGGT does not affect maturation, since most of them acquire native structure in a single round of association with calnexin and calreticulin.

Re-glycosylation by UGGT occurs most efficiently when all nine mannoses are present and is less efficient for glycoproteins with $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$ oligosaccharides: some mannose trimming will occur between repeated CNX/CRT cycles, which may affect subsequent processing by UGGT and glucosidase II. The importance of mannosidase activities in the ER and their role in delaying and eventually preventing re-entry of not correctly folded glycoproteins in the CNX/CRT cycle, has only recently been elucidated: both the folding state of a glycoprotein and the extent to which mannose residues have been removed seem to be important factors for the ER-Associated Degradation (ERAD) of not correctly folded glycoproteins. The substrate for ER mannosidase I is the $Man_9GlcNAc_2$ oligosaccharide, generated by the combined action of glucosidase I and II: mannosidase I cleaves a single α1,2 mannose residue to yield the $Man_8GlcNAc_2$. Additional mannose trimming yields $Man_7GlcNAc_2$ oligosaccharides, due to the removal of a mannose residue of the C branch: the irreversible cleavages of B and C branches serve as "Man timer", while trimming of an additional α1,2-linked mannose residue to yield $Man_6GlcNAc_2$ sends the glycoprotein to ERAD. To further complicate the situation also an ER mannosidase II exists in mammalian cells: it also trims the $Man_9GlcNAc_2$ oligosaccharide, but more excessively to a $Man_5GlcNAc_2$ oligosaccharide, negatively affecting the rate of re-glucosylation by UGGT required for re-entry in the CNX/CRT cycle [Avezov et al., 2008]. Proteins that are fully defective for folding are ratined in the ER and targeteted to the ERAD pathway.

An additional α-mannosidase I–like protein that lacks enzyme activity is also a resident of the ER: it is called EDEM (ER Degradation-Enhancing α-Mannosidase I–like protein) and it has an important role in the recognition of misfolded glycoproteins, thereby targeting them for ER degradation. Overexpression of EDEM resulted in accelerated release of misfolded glycoproteins from the CNX/CRT cycle and their subsequent degradation. EDEM binds selectively to not correctly folded glycoproteins: after exit from the CNX/CRT cycle, ER mannosidase I processes the oligosaccharides on not correctly folded glycoproteins to $Man_8GlcNAc_2$, which in turn is converted by EDEM to

$Man_7GlcNAc_2$ to provide the oligosaccharide for interaction with OS-9. The latter, a glycoprotein that contains a conserved mannose 6-phosphate receptor homology domain, selectively binds to non-native conformers of glycoproteins and routes them in the ERAD pathway through athe Hrd1/Hrd3 ubiquitin ligase complex. Various mechanisms, which involve distinct ubiquitin-ligase complexes, have been proposed to function in the management of ERAD substrates depending whether the misfolded protein is luminal (ERAD-L), cytosolic (ERAD-C) or intramembrane (ERAD-M) [Denic et al., 2006]. A glycoprotein that fails to be correctly folded during the Man trimming in the ER is retrogradely translocated in the cytosol through the protein-conducting channel, ubiquitinated by ubiquitin ligases, deglycosykated by peptide:N-glycanase (PNGase) and finally subjected to proteasomal degradation [Suzuki and Lennarz, 2003]. Among ERAD-linked ubiquitin ligases it has been identified the sugar-recognizing Skp1-Cullin1-F-box (SCF) protein complex, in which the lectin activity is carried by the F-box protein (Figure 2.3).

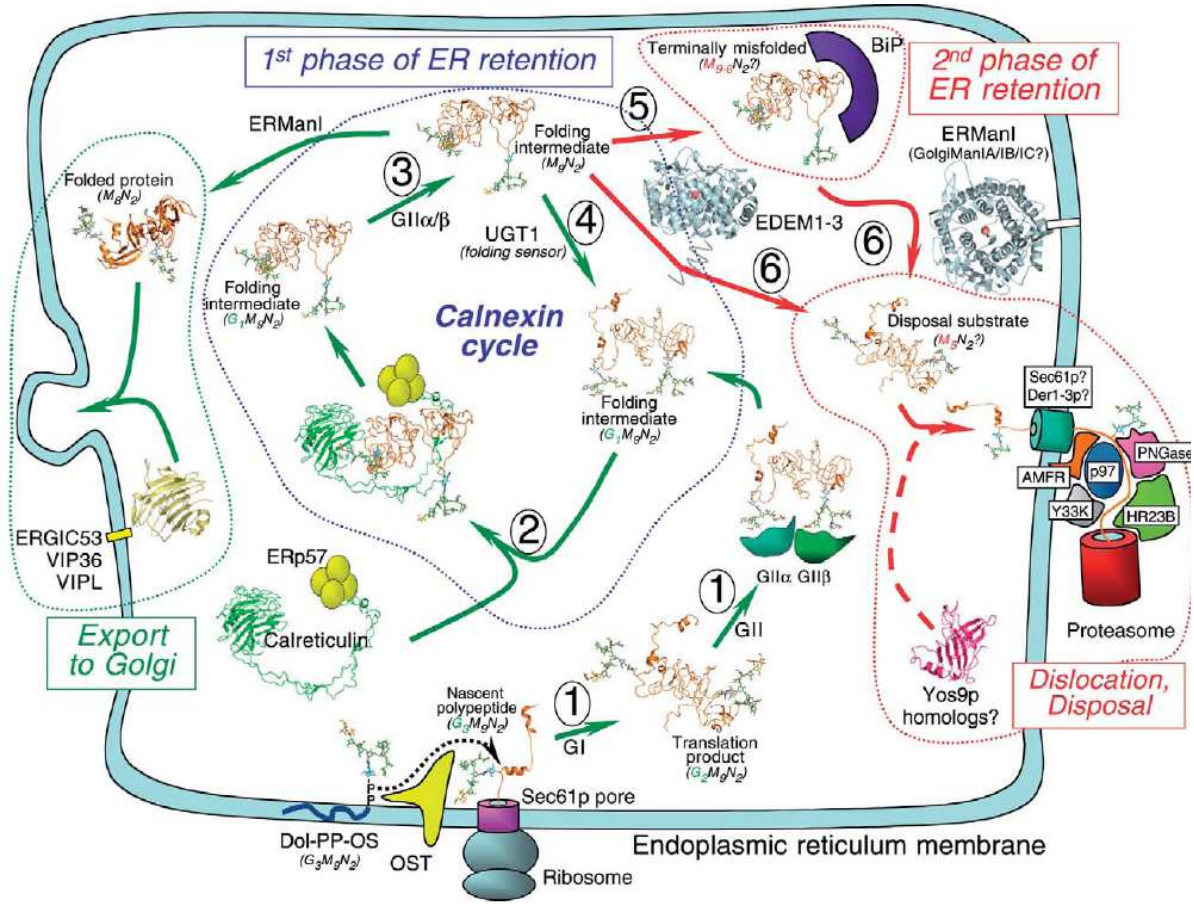

Figure 2.3: The role of glycan processing in regulating access to the lectin chaperone system or targeting for ERAD.

PNGase is a cytosolic enzyme that cleaves the β-aspartyl-glucosamine bond and removes the carbohydrate moieties from the misfolded glycoproteins, thereby facilitating proteasomal degradation: the carbohydrate moieties are bulky and therefore postulated to preclude proteasomal degradation of ubiquitinated substrates, since the active sites for proteolysis are within a narrow channel of the proteasome. In conclusion it is the trimming of carbohydrates and the elongation of polyubiquitin chains that target the protein to degradation [Kato and Kamiya, 2007].
The targeting of folding-defective polypeptides into the cytosol and proteasome-mediated destruction is crucial for mantenance of cell homeostasis: diverse proteins may not fold correctly because of disease-causing mutations and will be retained in the ER because of the quality control. Misfolded glycoproteins in addition can form self-aggregates or interact with proteins in addition to chaperones: since most disorders caused by protein misfolding are chronic diseases, ultrastructural changes may be observed only late [Moremen and Molinari, 2006].

2.4 O-GLYCOSYLATION

O-linked glycans often have lower mass thn N-linked structures, but can be more abundant and heterogeneous. In O-linked glycosylation the glycan is attached to the serine/threonine side chain, but unlike N-glycosylation, no consensus sequence exists. Linkages in which the sugar is attached to an aminoacid containing a hydroxyl group occur in great variety of proteins: every aminoacid with a hydroxyl functional group (Ser, Thr, Tyr, Hyp [hydroxyproline], Hyl [Hydroxylysine]) has been implicated. Through databases of O-glycosylated protein and statistical analysis on the sequences around target sites, preferential motifs for O-glycosylation have been identified leading to the development of algorithms for O-glycosylation prediction [Christlet and Veluraja, 2001]. Several O-glycosylation prediction softwares are available on the Internet, such as NetOGlyc 3.1 prediction server, which produces neural network predictions of 'mucin-type' O-glycosylation sites in mammalian proteins [Julienus et al., 2005].
The O-glycosylation process produces an immense multiplicity of chemical structure: each monosaccharide has three or four attachment sites for linkeage of other sugar residues and can form glycosydic linkage in an α or β configuration, allowing glycan structures to form branches.

2.4.1 'MUCIN-TYPE' O-GLYCOSYLATION

O-glycosylation usually starts with an N-acetylgalactosamine (GalNAc) linked to Ser/Thr: this linkage has been considered a hallmark of mucin glycoproteins, which form part of the epithelial secretions in, for example, the gut, cervix and lungs. However a wide variety of glycoproteins contain this linkage [Hanisch et al., 2001]. At least nine GalNAc-transferases exist: these enzymes work in concert in a hierarchical manner to form the clustered Ser/Thr-linked oligosaccharides that frequently occur in the 'mucin-type' glycoproteins. They are generally homologous to each other and although they act on characteristic peptide regions, no specific consensus sequence has been identified, mainly due to the multiplicity of the GalNAc-transferases [Clausen and Bennett 1996].

The O-GalNAc-α-Ser/Thr linkage is found in clusters of residues with a β-turn near proline and at a distance from charged aminoacids: in addition *in vitro* studies suggest that Thr is favoured over Ser for O-glycosylation [Elhammer et al., 1993]. This modification generally occurs in the cis-Golgi: however the multiplicity of enzymes composing this family makes it possible that some act also in a pre-Golgi or ER compartment [Hirschberg et al., 1998].

The biosynthesis of 'mucin-type' O-glycans can be divided into 3 stages. In the first stage nucleotide sugars are synthesized in the cytoplasm; in the second stage they are transported into the ER or the Golgi. In the third stage specific glycosyltransferases attach the sugars to a protein or to a glycan in the ER or Golgi. Monosaccharides used for the biosynthesis of nucleotide sugars derive from dietary sources and salvage pathways. Glucose and fructose are the major carbon sources in humans from which all other monosaccharides can be synthesized: series of phosphorylation, epimerization and acetylation reactions convert them into various high-energy nucleotide sugar donors, whose biosynthesis takes place in the cytosol. As the nucleotide sugars are synthesized in the cytosol they must be translocated into the lumen of the ER and/or Golgi: since they cannot cross the membrane lipid bilayer, specific transport mechanisms are responsible for their translocation. Two transport mechanisms can be distinguished: the first mechanism is the entrance of mannose and glucose through binding to the lipid carrier dolichol phosphate (Dol-P). The second transport mechanism exploits specific nucleotide sugar transporters (NSTs): these NSTs are antiporters in which nucleotide sugar entry into the ER and Golgi is coupled to equimolar exit of the corresponding nucleoside monophosphate from the lumen [Hirschberg et al., 1998]. After that a glycosyltransferase will transfer the monosaccharide to a glycan by cleaving off the nucleotide part: the nucleoside diphosphates are converted to dianionic nucleoside monophosphates, used for

the antiporter, and inorganic phosphate by a nucleoside diphosphatase. Nucleoside di- and monophosphates can inhibit the nucleotide sugar transport process and the activity of glycosyltransferases.

The activity of glycosyltransferases can be influenced by different factors. It is known, for example, that for optimal action some of the glycosyltransferases require divalent cations, such as Mn^{2+} or Mg^{2+}. Furthermore, it was recently discovered that human core 1 β3-galactosyltransferase (core 1 β3-Gal-T), which is involved in the formation of 'mucin-type' O-glycans, requires a molecular chaperone for its functioning: this molecular chaperone is called Cosmc (core 1 β3-Gal-T-specific molecular chaperone) and is an ER-localized type II transmembrane protein. A third factor that might influence glycosyltransferase activity is the structure of the protein substrate. However, how proteins are recognized by glycosyltransferases remains largely unknown.

Finally, glycosyltransferase activity can be dependent on heterocomplex formation: glycosyltransferases involved in the linkage of monosaccharides to the protein backbone and those involved in the core processing of 'mucin-type' O-glycans are specific and not involved in other classes of glycoconjugates, whereas most glycosyltransferases involved in the elongation, branching, and termination of glycans are not specific for one glycoconjugate class. Given the sequential and competing nature of glycosyltransferases, the precise localization of these enzymes within the Golgi is of great importance: glycosyltransferases are arranged in an assembly line in the Golgi, whereas early-acting transferases are localized in the cis-Golgi, intermediate-acting transferases in the medial-Golgi, and terminating transferases in the trans-Golgi [Opat et al., 2001].

In general, O-linked glycans have been found to function in protein structure and stability, immunity, receptor-mediated signaling, modulation of enzyme activity and signaling molecules, protein interactions, expression and processing.

Since O-glycosylation is a post-folding event, only Ser and Thr residues that are exposed on the protein surface will be glycosylated. Just like N-glycans, O-glycans can influence protein structure: the glycan can break the α-helicity of peptides; can have a role in the tertiary protein structure and in aggregation. Subsequently, O-linked glycans maintain protein stability, heat resistance, hydrophilicity, and protease resistance by steric hindrance [Wopereis et al., 2006]; another important function of O-linked sugars is to mediate immunologic recognition, such as recognition of glycopeptides by the MHC complex or by antibodies. Also the effects of O-linked glycosylation on the bioactivity of many signaling molecules, particularly hormones and cytokines, and a relatively small number of enzymes, have been described.

Many other O-linked glycosylation modification can occur: among them, GlcNAc-β-Ser/Thr represents an increasingly important linkage, that is widely dispersed among eukaryote, from protozoa to higher mammals. It is distinctive since it can be found in nuclear and cytoskeletal proteins and indeed represents the first reported example of glycosylation found outside of the secretory channels [Hart, 1997].

2.4.2 O-β-GlcNAcYLATION

The discovery of O-linked-β-N-acetylglucosamine (O-GlcNAc) more than 20 years ago disproved the long-held dogma that protein glycosylation is restricted to the luminal compartments of the secretory machinery and to the cell surface [Torres and Hart, 1984]. Thus far, O-GlcNAcylation has been documented in all metazoans, including *Caenorhabditis elegans*, insects, and plants. Even though O-GlcNAcylation is very abundant and widespread on most of the cell's regulatory proteins, progress in understanding its roles in the cell has been slow, since it is generally undetected by commonly used analytical protein methods. In fact addition of the sugar does not generally affect migration of the glycopeptide in gel electrophoresis or in isoelectric focusing; second, the sugar modification is rapidly hydrolyzed by cellular hexosaminidases upon cellular damage or during protein isolation. Third, O-GlcNAc is labile by conventional mass spectrometric methods and it is lost at the source under conditions typically used in electrospray mass spectrometry.
Subcellular fractionation of rat liver indicated that O-GlcNAcylated proteins are present in the cytoplasm and virtually in all subcellular organelles: many cytosolic enzymes, including kinases, glycolytic enzymes, cytoskeleton regulatory proteins, and cytoskeleton proteins themselves are modified. These same studies showed that the nucleoplasm, nuclear envelope and chromatin fractions are particularly enriched in O-GlcNAcylated proteins [Holt and Hart, 1986]; however the highest concentration of O-GlcNAcylated proteins occurs in the nuclear pore complex. Here glycoportiens are not required for the assembly of the nuclear envelope, for assembly of the major structures af the pore, or for formation of the transport channel, but rather are essential for recognition of the nuclear localization signals and transport [Holt et al., 1987]. Moreover, several transcription factors are O-GlcNAcylated, with the modification being involved in nuclear transport, DNA binding, assembly into multimeric complexes or regulation of phosphorilation [Jackson and Tjian, 1988]; in addition to transcription factors, the largest subunit of RNA polymerase II is also multiply O-GlcNAcylated on its C-terminal domain (CTD). It has

been shown that during transcription initiation the CTD is O-GlcNAcylated and dephosphorylated, whereas upon transcriptional elongation it is deglycosylated and phopshotylated: this suggests a role for glycosylation in the assembly of the pre-initiation complex; subsequent to initiation, the O-GlcNAc moieties are removed and multiple phosphorylation allows elongation to occur. Histones, as well, are O-GlcNAcylated *in vivo* and O-GlcNAcylation changes during mitosis [Sakabe et al., 2010].
O-GlcNAc could act as a protective signal against proteasomal degradation, both by modifying target substrates in PEST sequences (Pro-Glu-Ser-Thr sequences, known to mediate protein degradation), and by reversibly inhibiting the proteasome itself, through the O-GlcNAcylation of the 19S subunit [Guinez et al, 2008]: this mechanism couples proteasomes to the general metabolic state of the cell, since it allows the organism to respond to its metabolic needs by controlling the availability of aminoacids and regulatory proteins. In fact O-GlcNAcylation is an end point of the hexosamine biosynthetic pathway, which culminates in the production of UDP-N-acetylglucosamine (UDP-GlcNAc), the high-energy donor substrate for the O-GlcNAc transferase (OGT): virtually all metabolic pathways influence the cellular concentrations of UDP-GlcNAc: depending upon the cell type, as much as 2% to 5% of glucose is metabolized via the hexosamine biosynthetic pathway. Because the biosynthesis of UDP-GlcNAc is affected and regulated by nearly every metabolic pathway in the cell and because OGT-catalyzed O-GlcNAcylation is sensitive to insulin, to nutrients, and to cellular stress, it has been proposed that O-GlcNAcylation serves primarily to modulate cellular signalling and transcription regulatory pathways in response to nutrients and stress (Figure 2.4). O-GlcNAcylation of a large number of proteins in fact rapidly increases when cells are exposed to stress, including heat, high salt, heavy metals, UV light, hypoxia, and others: increased O-GlcNAcylation mediates stress tolerance through the upregulation of chaperones and heat shock proteins, both by increasing their expression and decreasing their turnover.

2.4.3 ENZYMES INVOLVED IN O-GlcNAc CYCLING

While Ser and Thr phosphorylation is determined by the action of over 300 different genetically encoded kinases, each with its own peptide selectivity [Manning et al., 2002], O-β-GlcNAc is regulated only by two enzymes: a glycosylatransferase and a glucosaminidase.

Uridine diphospho-N-acetylglucosamine:peptide β-*N*-acetylglucosaminyltransferase (OGT) is comprised of two subunits with molecular weights of 110 kDa (α-subunit) and one of 78 kDa (β-subunit). OGT is present in all cells, but it is most aboundant in the glucose-sensing cells of the pancreas and in the brain. Although there is no absolute consensus sequence, approximately one-half of the known O-GlcNAc sites contain a PVS (Pro-Val-Ser) motif, similar to that recognized by proline-directed kinases; in addition, many of the known O-GlcNAc sites have high PEST scores, PEST being a sequence that is associated with rapid degradation when phosphorylated. OGT is a bifunctional protein with a catalytic C-terminal domain, which apparently evolved from the glycogen phosphorylase superfamily, and an N-terminal protein-protein interaction domain constituted by several tetratricopeptide repeats (TPRs) [Iyer and Hart, 2003]: its activity is sensitive to the concentration of the donor substrate, UDP-GlcNAc. OGT interacts with histone deacetylase complexes by binding to the corepressor mSin3A: the latter acts as a scaffold for the HDAC subunits and directly binds transcription factors, suggesting a direct between histone deacetylation and chromatin remodeling in the repression of transcription. OGT activity, recruited on transcriptionally inactive regions via mSin3A, might trigger the disassembly of the activation complex, shifting the equilibrium of the gene expression from an active to a repressive state [Yang et al., 2002].

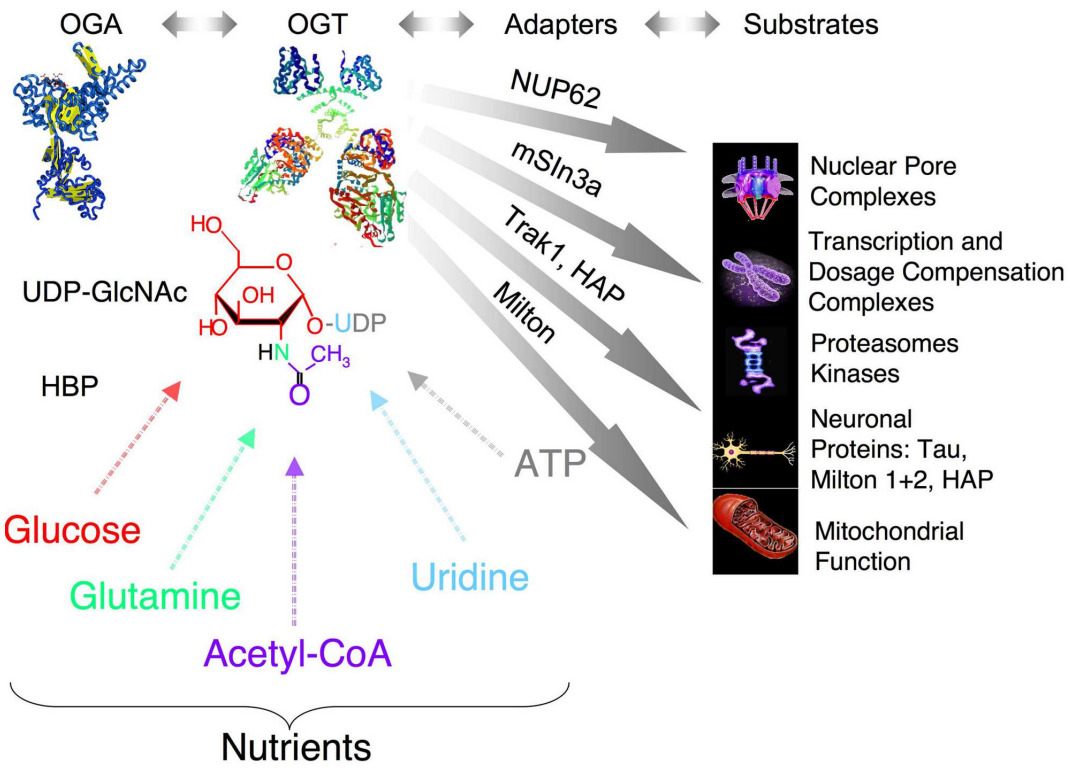

Figure 2.4: The O-GlcNAc code.

O-GlcNAc-specific β-N-acetylglucosaminidase (O-GlcNAcase) is a nucleocytoplasmic, neutral enzyme, composed of one 54 kDa-subunit (α subunit) and one 51 kDa-subunit (β subunit), which removes O-GlcNAc from peptide substrates glycosylated by OGT: like OGT, O-GlcNAcase is highly conserved and is expressed at the highest levels in pancreas, brain, and thymus. O-GlcNAcase is also a bifunctional protein with both a catalytic glycosydase domain at the N-terminus and an histone acetyltransferase (HAT) domain at the C-terminus: HAT domain allows the interaction of O-GlcNAcase with the transcription machinery, linking low-density O-GlcNAcylated proteins to transcription active regions. However within the nucleus and particularly at sites of transcription, the two O-GlcNAc cycling enzymes are often found within the same complex.

2.4.4 'YIN-YANG' MODEL

Recent studies have shown that the modulation of cellular pathways and functions by O-GlcNAcylation involves a very extensive cross talk with the pathways and mechanisms that are also regulated by protein phosphorylation signaling cascades: the system is not binary, with an 'on' and 'off' state for each signalling molecule; rather, the combination of modifications, differentially regulated, creates enormous molecular diversity. This cross-talk has been defined as the 'YinYang' model [Hart et al., 1995].
Many proteins displayed reciprocal occupancy at the same Ser or Thr residue under different conditions, such as at sites on the c-Myc oncogene protein, estrogen receptor β, some sites on RNA polymerase II. However, a majority of transcription factors, such as p53 and FOXO1, displayed reciprocal occupancy of the two modifications at proximal sites on the polypeptide, but not at the same residue. Competition between O-GlcNAcylation and phosphoryation for occupancy of Ser and Thr sites occurs by several distinct mechanisms. First, studies with synthetic O-GlcNAc peptides have suggested that addition of an O-GlcNAc moiety induces a β-turn conformation, while addition of a phosphate tends to open up the peptide conformation. In addition O-GlcNAcylation and phosphorylation dynamically modify the enzymes controlling each other's cycling on polypeptides [Hart, 2011].

2.5 GLYCOSYLATION AND DISEASES

Alterations in any step of the N- and O-glycosylation process can be involved in the aetiology of many human congenital diseases, collectively referred to as congenital disorders of glycosylation (CDG). Moreover, O-GlcNAcylation dysregulation contributes to the aetiology of important human diseases, particularly diabetes and neurological disorders (Table 1).

Table 1 | Pathologies associated with aberrant *O*-GlcNAcylation

Tissue	Proposed mechanism	Pathology
Muscle, fat	Decreased GLUT4 trafficking Decreased insulin signalling Decreased glycogen synthesis	Insulin resistance
Vasculature	Decreased eNOS activity Increased SP1 activation	Vascular disease Erectile dysfunction Retinopathy
Kidney	Increased TGF-β expression	Glomerular sclerosis
Cardiac muscle	Decreased SERCA2a expression	Cardiomyopathy
Brain	Increased tau phosporylation Decreased tau O-GlcNAcylation	Alzheimer's disease

Table 1: Pathologies associated with aberrant O-GlcNAcylation.

Given O-GlcNAc's role as a nutrient sensor and its extensive cross talk with phosphorylation, it is not surprising that the protein modification plays fundamental roles in chronic diseases, particularly diabetes, neurodegeneration, cardiovascular disease, and cancer. In addition, many viruses that infect eukaryotic cells have key regulatory proteins that are also modified, presumably by the host O-GlcNAc cycling enzymes. It has long been postulated that the hexosamine biosynthetic pathway played a role in the etiology of diabetes and glucose toxicity. There are numerous genetic and other studies in insulin-sensitive tissues, such as those of muscle, liver, and fat, that strongly support a direct involvement of O-GlcNAcylation and the hexosamine biosynthetic pathway in insulin resistance. Hyper-O-GlcNAcylation not only alters phosphorylation-mediated signalling cascades, but also alters the regulation of transcription [Slawson et al., 2010].

2.5.1 CONGENITAL DISORDERS OF GLYCOSYATION

Any step of the N-glycosylation process can be involved in the have crucial roles in the onset of congenital disorders of glycosylation (CDG). CDG form a group of autosomal recessive metabolic disorders caused by defects in the biosynthesis of protein-linked glycans. The primary defect of these disorders is not necessarily localized in one of the glycan-specific transferases, but can likewise be found in the biosynthesis of nucleotide sugars, their transport to the ER and Golgi, and in Golgi trafficking: defects can lead to a severe autosomal recessive multisystem syndrome with neurologic involvement, whereas some defects do not produce a clinical phenotype.
It is becoming increasingly evident that also defects in the O-glycosylation process can be ascribed to CDG. The clinical variations among the different errors of O-glycan metabolism can be enormous: as O-glycans are involved in numerous processes, it is inevitable that defects in O-glycan biosynthesis might lead to severe abnormalities for cellular functioning.
The majority of patients with 'classical CDG' have a defect in N-glycan biosynthesis. They have common symptoms such as muscle hypotonia, central nervous system abnormalities, growth delay, feeding problems, coagulation defects and liver disease, and frequently show specific signs such as abnormal fat distribution and inverted nipples, which help with the early clinical diagnosis. In contrast, patients with O-glycosylation disorders commonly have involvement of only one organ or one organ system and do not have the general symptoms that are suggestive for an inborn error of metabolism. Most of the disorders of O-glycan biosynthesis seem to have very specific tissue expression, whereas N-glycans are expressed ubiquitously. Another remarkable difference between N- and O-glycan deficiencies is that N-glycan deficiencies generally have recessive inheritance, whereas in some of the O-glycan biosynthesis diseases inheritance is autosomal dominant. Patients with combined defects in protein N- and O-glycosylation often have a phenotype that is a mixture of the features of inborn errors in combination with congenital malformations [Wopereis et al., 2006].

2.5.2 GLYCOSYATION AND DIABETES

O-GlcNAc transferase (OGT) uses the donor sugar UDP-GlcNAc: concentrations of UDP-GlcNAc, which is also the end product of the hexosamine biosinthetic pathway, are highly sensitive to ambient glucose levels (see paragraph 2.4.2). Increased flux through the hexosamine pathway under hyperglycemic conditions leads to elevated levels of O-

GlcNAc modified proteins in skeletal muscle and in pancreatic beta cells. In the case of muscle cells, reduced insulin receptor substrate (IRS) –1 and –2 signaling are associated with their increased O-GlcNAc modification and decreased phosphorylation, causing attenuated insulin stimulated signaling. For example, hyperglycemic conditions increase O-GlcNAcylation of the transcription factor Sp1, which correlates with increased Sp1-dependent TGFβ1 expression. Thus, altered O-GlcNAc levels under hyperglycemic conditions appear to perturb normal signaling events required for insulin-mediated homeostasis [Vosseller et al., 2001].

Hyperglycemia is associated with insulin resistance, although the underlying molecular mechanisms for this effect are still largely unknown. O-GlcNAc has a key role in the regulation of insulin signalling and as a mediator of glucose toxicity [Buse, 2006]: increasing O-GlcNAcylation in adipocytes or muscle blocks insulin signalling at several points. Elevated O-GlcNAcylation on insulin receptor substrate seems to reduce its interaction with phosphatidylinositol-3-OH kinase, thus blocking insulin signalling at an early stage. Moreover many aspects of glucose toxicity require its metabolism into glucosamine, which, in turn, elevates UDP-GlcNAc and increases O-GlcNAcylation. Hyperglycemia, hyperlipidamia and hyperinsulinemia all result in abnormal increases in O-GlcNAcylation, which disturb the normal dynamic balance between O-GlcNAcylation and O-phosphorylation that controls signalling, transcription and other cellular functions.

2.5.3 NEURODEGENERATIVE DISEASES

O-GlcNAcylation plays an important role in both normal brain functions and the etiology of neurodegeneration, contributing to transcriptional regulation, neuronal communication and neurodegenerative disease. O-GlcNAcylation was recently shown to play a role in neuronal plasticity, regulating the number of axonal filopodia and the ability of axons to elaborate branches: over-expression of O-GlcNAcase promoted the formation of axon branches, even In the context of injury and regeneration [Francisco et al., 2009]. However unlike diabetes, in these cases higher O-GlcNAcylation appears to prevent disease, and impaired glucose metabolism, with the resulting lower O-GlcNAcylation in brain, leads to hyperphosphorylation which contributes to neuronal cell death.

Approximately, one quarter of the neuronal proteins are O-GlcNAcylated, involving important processes such as transcription, neuronal signaling and synaptic plasticity. These proteins can be transcription factors, such as cAMP-responsive element binding protein (CREB), coactivators and corepressors, as the steroid receptor coactivator-1

(SRC1), and cytoskeletal proteins such as neurofilaments, microtubule-associated proteins and the β-amyloid precursor protein (APP). An intriguing interplay between O-GlcNAc glycosylation and phosphorylation has been observed in cerebellar neurons, wherein activation of certain kinase pathways reduces O-GlcNAc levels on cytoskeleton-associated proteins. Finally, recent studies suggest that O-GlcNAc is dynamically regulated by excitatory stimulation of the brain *in vivo* [Rexach et al., 2008].

There is considerable indirect evidence that O-GlcNAc may play a role in neurodegenerative disorders. Interestingly, OGT maps to the X-linked Parkinson dystonia locus and O-GlcNAcase maps to the Alzheimer's disease locus on chromosome 10 [Wells et al., 2003].

Evidence has been mounting connecting defects in glucose metabolism in the brain with neurodegenerative diseases, such as Alzheimer's disease (AD) and Parkinson's Disease [Dias and Hart, 2007]. Virtually all proteins involved in AD can be both O-GlcNAcylated and phosphorylated. Neuron glucose metabolism declines with age and this decline is more marked in AD neurons, concomitant with reduced O-GlcNAc [Liu et al., 2009]. One normal function of O-GlcNAc seems to be the prevention of excessive phosphorylation at sites at which this would be deleterious.

Tau protein is required for microtubule polymerization and stability in neurons, and is the major component of neurofibrillary tangles associated with the disease: in normal brain it is extensively O-GlcNAcylated at more than 12 sites and is hyperphosphorylated in AD brain, which in turn causes it to aggregate into the paired helical filaments that constitute the visible neurofibrillary tangles characteristic of the disease [Liu et al., 2004].

The amyloid-β precursor protein (APP) is O-GlcNAcylated and its cytoplasmic tail is phosphorylated, which is known to affect its proteolytic processing. Abnormal proteolysis of APP gives rise to the toxic β1–42 peptide fragment, which forms amyloid plaques that are found in neurons in Alzheimer's disease. Neurofilaments are hypo-O-Glc-NAcylated in neurons in a rat model of amyotrophic lateral sclerosis. O-GlcNAcylation of the clathrin assembly proteins AP-3 and AP-180 declines in Alzheimer's disease, suggesting that reduced O-GlcNAcylation contributes to the loss of synaptic vesicle recycling [Hart et al., 2007].

Finally, O-GlcNAc glycosylation has been demonstrated to inhibit the proteasome, thus providing a mechanism to couple ubiquitin-mediated protein degradation to the general metabolic state of the cell: blocking the removal of O-GlcNAc from the proteasome leads to increased protein ubiquitination and possibly neuronal apoptosis. Proteasomal dysfunction and ubiquitinated inclusion bodies are found in the diseased tissue of

individuals with Amyotrophic lateral sclerosis (ALS), Parkinson's disease, Huntington's disease and Alzheimer's disease [Ciechanocer and Brundin, 2003].

2.5.4 GLYCOSYATION AND CANCER

In every organism all cellular functions can be distinguished in two general typologies: in fact there are cells endowed with 'social' functions, such as cell-cell interactions and cell-matrix interactions, in contrast to cells with 'housekeeping' functions, such as gene expression and cell replication. These different functions are regulated by glycosylation and phosphorylarion, respectively: tumor cell motility and metastasis are complex processes which result mainly by the alteration of 'social' functions. Glycans esposed on tumour cells can behave both as adhesion molecules, that can alter cell motility, and as transmembrane signaling molecules, able to carry signals across the plasma membrane and regulate tumour growth. Glycosylation also affects the trafficking of cell adhesion molecules that are important for metastasis: for example, O-GlcNAcylation of β-catenin not only regulates its nuclear localization and transcriptional activity, but also blocks its association with E-cadherin, a cell adhesion molecule critical to epithelial cell adhesion.
Aberrant glycosylation observed in tumour cells can become a good method for diagnosis and prognosis of tumour progression: in fact there is a clear correlation between aberrant glycosylation in primary tumours and their invasiveness [Hakomori, 1996].
As noted above many oncogenes or tumour suppressors play a direct role in regulating transcription, and nearly all such proteins are modified by glycosylation such as c-Myc, retinoblastoma (Rb), β-catenin, estrogen receptors, and others. Given the extensive cross-talk between O-GlcNAcylation and phosphorylation, and the known roles of phosphorylation in mechanisms underlying cancer, it is not surprising that O-GlcNAcylation is also involved in cancer aetiology [Kamemura and Hart, 2003]: O-GlcNAc modification adds another level of regulation, which could allow for exquisite control of cell regulatory mechanisms. Disruptions of either of these post-translational modifications may interfere with critical control mechanisms, leading to the transformed phenotype [Comer and Hart, 2000]. For example, the c-myc oncogene is glycosylated in its transactivation domain at Thr-58, which is also the mutational hotspot found in a large percentage of Burkitt's lymphomas in patients; interestingly, this glycosylation site is also an important phosphorylation site that regulates c-myc transcriptional activity. The tumour suppressor, p53, which is the most commonly mutated gene in a wide range of

human cancers, is also O-GlcNAc-modified: there is preliminary evidence that the O-GlcNAc on p53 regulates its binding to DNA.

2.6 THERAPEUTIC GLYCOPROTEINS

Glycosylation patterns of recombinant proteins are relevant for the immunogenicity, the pharmacological activity, pharmacokinetic profile, solubility and stability of the protein. Processing of oligosaccharides is specified by the processing enzymes present in the cell. Hence the glycoform populations for a glycoprotein are both species and cell specific. The importance of therapeutic proteins has grown rapidly since the emergence of the biotechnology industry: currently 64 products have been approved by regulatory agencies, with some 500 product candidates in clinical and preclinical development. Of these, approximately 70% are glycoproteins [Walsh, 2005]. Therapeutic proteins were initially derived from human sources; however, concerns over product purity and consistency, the potential for viral contamination and the emergence of genetic engineering tools has shifted their production into recombinant expression systems, such as mammalian cell lines, bacteria, yeast and insect cells. However these expression systems vary widely, with regard to their ability to incorporate the post-translational modifications found on native human proteins: because many proteins of therapeutic importance require N-glycosylation for biological activity, expression systems with N-glycosylation capability have become essential for therapeutic glycoprotein production. In fact not-correct glycoforms can adversely affect pharmacokinetic properties of the recombinant glycoprotein and raise immunogenicity and safety concerns: thus, the therapeutic use of glycoproteins derived from expression systems that do not perform human glycosylation could lead to rapid clearance, complement activation, and enhanced immunogenicity by targeting to antigen-presenting cells [Sethuraman and Stadheim, 2006].
Glycans contribute significantly to the hydrodynamic volume and charge of glycoproteins. In particular, the content of sialic acid contributes to the net negative charge, and improves the pharmacokinetics of glycoproteins such as erythropoietin (EPO) [Macdougall, 2002]. Although glycosylation can improve plasma half-life, the presence of certain carbohydrate moieties can trigger lectin-mediated clearance, thereby reducing plasma half-life: in fact lectins are differentially expressed in tissues so that therapeutic proteins with distinct glycans can be developed for cell- and tissue-specific targeting. An expression system that allows for control over the glycosylation of therapeutic proteins has, until recently, been unavailable. Through the customization of glycan profiles, such

as that recently demonstrated in glycoengineered yeast, it is now possible to systematically probe for glycosylation-dependent therapeutic effects: increased availability of glycoengineered expression systems will enable a greater understanding of structure-function relationships among glycoforms, which should lead to the development of novel therapeutics [Sethuraman and Stadheim, 2006].

3. AIM OF THE STUDY

Cells of a multicelluar organism are structurally and functionally heterogeneous: the diversity of cell phenotype is due to differential transcription programs precisely regulated by specific nuclear factors and induced upon differentiation. As regards neuronal cells, it thus becomes of crucial interest to understand the molecular mechanisms regulating central nervous system-specific genes expression in the nervous system and their repression in other tissues. Gene expression can be controlled by positively acting cell-type-specific or tissue-specific transcription factors; however also negative-acting silencers participate in preventing the expression of neuronal genes in non-neuronal cell types or in inappropriate neuronal subtypes.

As previously described, the original role proposed for REST was of transcription factor responsible for restricting or limiting neuronal gene expression to the nervous system, by silencing the expression of the regulated genes in non-neuronal cells. However, since its discovery as a negative regulator of neuronal differentiation, REST has been implicated in several processes such as maintenance of embryonic stem cell (ESC) pluripotency, self-renewal and regulation of mitotic fidelity in non-neuronal cells.

REST gene is expressed in both embryonic stem cells (ESCs) and neural stem cells (NSCs). In ES cells, consistent levels of nuclear REST occur and the REST-corepressor complex strongly represses target genes. As the ES cells start to differentiate into NSCs, most of the REST is degraded through the proteasomal pathway whereas lower amounts of REST mRNA are transcribed by the gene transcriptional machinery. It has been hypothesized that in NSCs, as the levels of REST decrease, the REST-corepressor complex may still occupy the RE-1 sequence in target genes but represses them less substantially than in ES cells, causing them to be poised for transcription. However, the expression of multiple neuronal differentiation genes requires both the absence of REST function and the presence of other promoter/enhancer-specific positive activators. Therefore, as the NSCs exit the cell cycle and differentiate into mature neurons, REST gene transcription is decreased. In the absence of active REST transcription, REST protein levels diminish, resulting in the expression of numerous genes. Taken together, these findings suggest that at ESC level, are already present and finely regulated cellular processes that modulate post-traslational mechanisms that may contribute to modify its stability as well as its interaction with other transcription factors or with regulatory sequences of the target genes.

The aim of this study was to investigate if one the most relevant processes involved in the post-traslational regulation of proteins: N-linked glycosylation, may contribute to REST folding, stability and interaction with other proteins and DNA.

REST gene encodes a 116-kDa protein that contains a DNA-binding domain, composed of eight zinc-fingers, an N-terminal repressor domain and one zinc finger at the C-terminal repressor domain [Ballas and Mandel, 2005]. By acting as an evolving molecular platform to which diverse factors may be recruited, REST promotes dynamic modifications of DNA, histones, nucleosomes and higher-order chromatin codes and helps maintaining genomic stability. Although REST molecular weight, calculated on aminoacid sequence, is about 116 kDa it was found to migrate on SDS-PAGE as different molecular weight protein (see Table 1.1); thus, suggesting the possible presence of post-translational modifications of the protein backbone. However, up to now, only the presence of O-glycan chains has been evidenced; thus indicating the contribution of O-lined glycosylation to REST functions [Lee et al., 2000] whereas is unknown ay contribution of N-linked glycosylation.

Although the most common way to generate protein diversity is, of course, alteration of the order and composition of their constituent amino acids, additional diversity is generated by covalent modification of amino acids by post-translational modifications: while only 20 amino acids are used in protein synthesis, as many as 140 different amino acids are found in proteins due to covalent post-translational modification [Vosseller et al., 2001]. Glycosylation, that is the covalent attachment of one or more carbohydrates that occurs during or after the process of protein synthesis, causing stable and long-term changes of the native protein, is another relevant and only partially known event. Although the number of different types of glycosidic bond is continuously increasing, N- and O-linked glycans still remain the most abundant. The process of glycosylation involves about 1% of human genes [Lowe and Marth, 2003]; furthermore, according to estimates based on the SwissProt database, more than one half of all proteins are glycosylated [Apweiler et al., 1999]. Glycosylation may affect many protein activities: it has been shown to facilitate folding kinetics and to affect the biophysical properties of proteins. The observation that N-glycosylation starts during protein synthesis, when the glycan is added to the still unfolded protein, indicates that it may assist in obtaining the correct fold [Roth et al., 2010]. On the contrary O-linked glycans often have lower mass than N-linked structures but they can be more abundant and heterogeneous. O-linked glycans have been found to function in protein structure and stability, immunity, receptor-mediated signaling, modulation of enzyme activity and signaling molecules, protein interactions, expression and processing. Moreover the discovery of O-linked-β-N-

acetylglucosamine (O-GlcNAc) more than 20 years ago disproved the long-held dogma that protein glycosylation is restricted to the luminal compartments of the secretory machinery and to the cell surface: several transcription factors are O-GlcNAcylated, with the modification being involved in nuclear transport, DNA binding, assembly into multimeric complexes or regulation of phosphorylation [Torres and Hart, 1984]. Alterations in any step of the N- and O-glycosylation process can be involved in the aetiology of many human disorders, such as congenital diseases, diabetes, degeneration and cancer [Wopereis et al., 2006].

REST, as well, has been found to be involved in the establishment of the same diseases. Particularly, it has been proposed that it may have a double role in the pathogenesis of cancer, since it can behave both as an oncogene and an oncosuppressor.

Moving from these observations, we hypothesized the presence of glycan chains in the aminoacid backbone of REST and investigated the presence of glycosylation chains in the aminoacidic backbone of the REST protein with the aim to relate this post-translational modifications to its repressor transcriptional activity. Particularly, our goal is to identify and characterize the N-lined glycosylation pattern of REST to understand any correlation between alteration of glycosylation and protein misfunction and to translate this correlation in any pathological conditions where such post-traslational events could be modified.

First of all, using different deglycosylating enzymes we investigated the composition of the glycan chains present in REST. Thereafter, we ascertained their composition through mass spectrometry analysis and confirmed the presence of complex glycan chains in REST by expressing the protein *in vitro* in suitable prokaryotic and eukaryotic systems.

The second step was to identify REST aminoacid residues bearing N-linked glycan chains. Using suitable post-translational prediction software, we identified aminoacids that could be target of N-linked glycosylation events; afterwards, these aminoacids were mutated by site-directed mutagenesis. We generated nine plasmids bearing REST mutant forms where a sequential removal of glycan chains was obtained. The plasmids were expressed in cell lines of embryonic or neuronal origin and we analyzed molecular weight and ability of REST mutated forms to bind to the *RE-1* DNA consensus sequence by gel-shift analysis. Finally, we investigated the contribution of glycan chains in the repressor transcriptional activity by employing gene reporter assays.

In conclusion, the present study aims to contribute to a more accurate comprehension of the post-traslational processes modulating REST activity and useful for a more detailed knowledge of crucial biological events like neurogenesis and oncogenesis. Moreover, considering the important role of protein glycosylation and of transcription factors in the

aetiology of many disease states, any further knowledge in this field may become relevant for future pharmacological therapies.

4. MATERIALS AND METHODS

4.1 MATERIALS

Cell culture media, fetal bovine serum (FBS) and horse serum (HS) were from Lonza (Basel, Switzerland). EXGEN 500 Transfection Reagent was from Fermentas (Hanover, MD, USA). The rabbit polyclonal antibody anti-REST purchased from Upstate-Millipore (Darmstadt, Germany) is raised against residues 801-1097 of the protein; the polyclonal rabbit anti-REST antibody from Bethyl Laboratories (Montgomery, USA) is raised against residues 1050-1097. Monoclonal anti-FLAG M5 antibody was from Sigma-Aldrich (St. Louis, MO, USA). The peroxidase-conjugated secondary antibodies goat anti-mouse IgG and goat anti-rabbit IgG were from Santa Cruz Biotechnology Inc. (Santa Cruz, CA, USA).

Glycoprotein Deglycosylation kit was purchased from Calbiochem-Merck (Darmstadt, Germany), while Native Protein Degycosylation kit from Sigma-Aldrich (St. Louis, MO, USA). Rabbit Reticulocyte Lysate (TnT T7 Coupled Reticulocyte Lysate System) and *Escherichia coli* lysates (S30 T7 High-Yield Protein Expression System) were purchased from Promega (Madison, WI, USA).

Restriction enzymes were obtained from Roche (Mannheim, Germany) or Fermentas (Hanover, MD, USA). NE-PER Extraction Reagent kit and LightShift Chemiluminescent Electrophoretic Mobility Shift Assay kit were from Pierce (Rockford - IL, USA). Biotin-labeled oligonucleotides were from Sigma-Aldrich (St. Louis, MO, USA).

Recombinant human IGF-I, Reporter Lysis Buffer, Bright-Glo Luciferase Assay System and Beta-Glo Assay System were all purchased from Promega.

All other reagents were of analytical grade or of the highest purity available, purchased from Sigma-Aldrich (St. Louis, MO, USA).

4.2 CELL COLTURE

SH-SY5Y cells were obtained from the European Collection of Cell Cultures (Salisbury, UK) and routinely grown as monolayers in MEM and Ham's F12 (1:1) medium, supplemented with 15% (v/v) fetal bovine serum (FBS), L-glutamine (2mM), 1% non-essential amino acids and 1× antibiotic-antimycotic solution in a humidified environment

containing 5% CO_2 and 95% air. Cells were grown in 75-cm^2 flasks, detached with a cell scraper and viability was determined using trypan blue staining. Cells were analysed in an inverted phase contrast microscope (Eclipse TS100; Nikon).
HEK293 cells were obtained from European Collection of Cell Cultures (Salisbury, UK) and routinely grown as monolayers in EMEM (EBSS) supplemented of 2mM glutamine, 1% non essential amino acids, 10% FBS (v/v) and 1× antibiotic-antimycotic solution in a humidified environment containing 5% CO_2 and 95% air. Cells were grown in 75-cm^2 flasks, or alternatively were plated in 100mm dishes, and detached with a 10% Trypsin/EDTA solution. Cells, at 50–60% confluence, were transiently transfected with pCMV-Tag2B+REST plasmid, which contains REST coding sequence fused to FLAG epitope, or with plasmids encoding for different REST mutants (10μg/dish) using the EXGEN 500 Transfection Reagent (Fermentas, Hanover, MD, USA). 72h post-transfection, cells were collected and nuclear and cytoplasmic proteins were extracted.
Rat pheochromocytoma PC-12 cells (European Collection of Cell Culture) were grown in RPMI 1640 medium supplemented with 10% heat-inactivated (55°C, 30 min) horse serum, 5% FBS and 1x antibiotic-antimycotic solution. Cells were cultured at 37°C in a humidified atmosphere of 5% CO_2.
For REST over-expression and gene reporter experiments, PC-12 cells were seeded in 100mm dishes and in 24-well plates, respectively, coated with Collagen type IV from human placenta 0,1 mg/mL (Sigma-Aldrich), transiently transfected through EXGEN 500 Transfection Reagent and harvested 72 or 48h later, respectively, with a 10% Trypsin/EDTA solution.

4.3 WESTERN BLOTTING

Nuclear and cytoplasmic protein fractions were obtained using the commercial NE-PER Extraction Reagent kit (Pierce): lysis and resuspension buffers contained a mixture of protease inhibitors composed of 0,5 mg/mL Benzamidine, 2 μg/mL Aprotinin, 2 μg/mL Leupeptin and 0,75 mM phenylmethlysulfonyl fluoride.
Cells were scraped off in 10mL cold PBS, pelleted and resuspended in 200mL of CER I buffer: after 10min incubation on ice, 11mL of CER II buffer were added and the suspension was vortexed, incubated on ice for 1min and then separated by centrifugation at 12000g for 5min. To obtain the nuclear extract, the cell pellet was resuspended in 100mL of NER buffer and incubated on ice for 40min: the soluble proteins in the lysate

were separated by centrifugation at 12000g for 10min at 4 C. Protein samples were then kept at – 80°C.
Protein concentration was determined using the Lowry method [Lowry et al., 1951].
Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting were performed according to standard procedures: proteins from the nuclear extract (70 μg for SH-SY5Y and HEK-293 cells and 100 μg for PC-12 cells) were separated by 7% SDS-PAGE and transferred onto a Hubond-ECL nitrocellulose membrane (GE Healthcare). As a molecular weight standard we used the Magic Marker XP (Invitrogen).
For detection of REST protein, the blots were blocked with 3% non-fat milk in TBS (10mM Tris-HCl, pH8, containing 150mM NaCl) for 1,5h at room temperature and then incubated overnight at 4°C with the rabbit polyclonal IgG directed against REST, used at a dilution of 1:2000 (Upstate) or 1:2500 (Bethyl Lab). Alternatively, the recombinant FLAG-REST protein was detected by blocking the membrane with 5% non-fat milk in TBS (10mM Tris-HCl, pH8, containing 150mM NaCl) plus 0,1% Tween 20 for 1,5h at room temperature, and then incubating the membrane 5% milk in TBS-T 0,1% with the anti-FLAG M5 (Sigma) mouse monoclonal IgG directed against FLAG, a 1:2000 dilution was used, for 1,5h at room temperature. The membranes were incubated with peroxidase-conjugated secondary antibodies at a dilution of 1:8000, for 1,5h at room temperature. Detection of immunoreactive bands was done by the enhanced chemiluminescent method (Immobilon Western Chemiluminescent HRP Substrate – Millipore) and a FUJILAS3000 acquisition camera.

4.4 ENZYMATIC DEGLYCOSYLATION

In order to characterize glycan chains present in REST we performed an enzymatic removal of O-linked oligosaccarides using Glycoprotein Deglycosylation kit, following manufacturer instruction for the non-denaturing protocol (Calbiochem). Briefly, reaction mixtures contained 0,5 μL each of several enzymes, alone or in combination (*PNGase F*, *α2-3,6,8,9-neuraminidase*, *Endo-α-N-acetylgalactosaminidase*, *β1,4-galactosidase* and *β-N-acetylglucosaminidase*), 100 μg of proteins, and 5X reaction buffer (250mM sodium phosphate buffer, pH 7.0): the mixtures were incubated at 37°C for 3 days and then analyzed using western blotting.
N-linked glycan chains removal was performed using the Native Protein Degycosylation kit from Sigma-Aldrich (St. Louis, MO, USA): 100 μg of proteins were incubated with 1

μL each of different enzymes (*Endoglycosidase F1*, *Endoglycosidase F2* and *Endoglycosidase F3*) and 10 μL of 5X reaction buffer (250mM sodium phosphate buffer, pH 5.5)and then incubated at 37°C for 1 hour. Subsequently they have been analyzed by western blotting.

4.5 CELL-FREE PROTEIN SYNTHESIS

The cell-free translation system employed were an extract from rabbit reticulocytes or *Escherichia coli*: they use DNA as a template, from which RNA is transcribed and subsequently translated without any purification.

4.5.1 ESCHERICHIA COLI PROTEIN EXPRESSION

The S30 T7 High-Yield Protein Expression System (Promega) is an *E. coli* extract-based cell-free protein synthesis system: it contains all the macromolecular components (70S ribosomes, tRNAs, aminoacyl-tRNA synthetases, initiation, elongation and termination factors) required for transcription and translation of exogenous DNA. To ensure efficient translation, each extract must be supplemented with amino acids, energy sources (ATP, GTP), energy regenerating systems (phosphoenol pyruvate and pyruvate kinase), and co-factors like $Mg^{2+}$and K^{+}.

This *E. coli*-based protein expression system uses the bacteriophage T7 RNA polymerase and a T7 promoter, and permits high-level expression of REST recombinant protein: the T7 RNA polymerase is typically supplied by inducing a recombinant T7 polymerase gene contained on a lambda lysogen in the host strain. During transcription the 5' end of the RNA becomes available for ribosomal binding and undergoes translation, while its 3' end is still being transcribed: this early binding of ribosomes to the RNA maintains transcript stability and promotes efficient translation.

2 μg of template plasmid DNA (pSP72+REST) was incubated with 20 mL S30 premix plus and 18 mL T7 S30 extract: the tube was then vortexed and centrifuged for 5 seconds, to force the reaction mixture to the bottom of the tube. In vitro transcription and translation occurred for 60 min at 37°C with vigorous shaking: the reaction was stopped by incubating the tube for 5 min on ice and samples have been denatured for 2 min at 95°C, just before loading the sample for SDS-PAGE.

4.5.2 RETICULOCYTE LYSATE PROTEIN EXPRESSION

Rabbit Reticulocyte Lysates (Promega) are a highly efficient *in vitro* eukaryotic protein synthesis system which combines the T7 prokaryotic phage RNA polymerase and promoter with eukaryotic extracts, in presence of energy regenerating systems such as creatine phosphate and creatine phosphokinase, to synthesize proteins from exogenous DNA templates. *In vivo* reticulocytes are highly specialized cells primarily responsible for the synthesis of hemoglobin, which represents more than 90% of the proteins that they produce. Reticulocytes are immature red cells which have already lost their nuclei: they have to be purified to remove contaminating cells, which could otherwise alter the translational properties of the extract, and then treated with micrococcal nuclease to destroy endogenous globin mRNA, thus reducing background translation to a minimum.
Many factors affect translation efficiency of a specific mRNA in Rabbit Reticulocyte Lysates like, for example, the presence of certain nucleic acid sequence elements such as poly(A) tails, 5´-caps, 5´-untranslated regions and the sequence context, which can have profound effects on initiation fidelity and translation efficiency. In addition, Rabbit Reticulocyte Lysates may contain a variety of post-translational processing activities, including acetylation, isoprenylation, phosphorylation and glycosylation.
The reaction was assembled in a 1,5 mL tube, by adding 25 μL of TnT Rabbit Reticulocyte Lysate, 2 μL of Reaction Buffer, 0,5 μL of each T7 RNA Polymerase Aminoacid mixture (-Met and -Leu), 1 μg of template DNA (pSP72+REST), 1 μL of T7 RNA Polymerase and 1 μL of Ribonuclease Inhibitor: the mixture was resuspended and centrifuged briefly, then incubated at 30°C for 90 min.
As a positive control of *in vitro* protein synthesis, a transcription/translation reaction of the firefly luciferase coupled to a luciferase activity assay was set: this is an excellent control for *in vitro* translation because only full-length luciferase is active. The luciferase synthesis was evaluated by measuring light intensity, which depends on the rate of catalysis of luciferin by synthetic luciferase, through a luminometer. Assembly of the control reaction mixture is similar to the one described above, except that the DNA template is a plasmid encoding firefly luciferase.

4.6 IMMUNOPRECIPITATION

FLAG peptide (N-DYKDDDDK-C) is a synthetic polypeptide which is absent from any natural peptide sequence: REST coding sequence has been previously cloned downstream

of the FLAG peptide coding sequence, so that FLAG could be used for purification and identification assays. With the aim of affinity-purifying REST from nuclear protein extracts we optimized an immunoprecipitation protocol, which exploits the binding affinity of the monoclonal anti-FLAG M2 IgG covalently attached to agarose beads.
2 mg of nuclear protein extract, denatured at 95°C for 5 min, were incubated with 40 µL of FLAG M2 agarose resin, previously washed twice with 0,5 mL TBS (50mM TrisHCl, 150mM NaCl, pH 7.4) and centrifuged at 12000g for 1 min at 4°C: the protein and resin were incubated over-night at 4°C on a roller shaker. The next day the resin was centrifuged and the supernatant kept as the *wash* fraction, then washed three more times with TBS: FLAG-REST was eluted in 50 µL of a 200 µg/mL FLAG peptide solution, incubated with vigorous shaking at 4°C for 30 min. After centrifugation, the supernatant containing purified FLAG-REST was immediately loaded for SDS-PAGE or stored at -80°C.

4.7 MASS SPECTROMETRY ANALYSIS

Mass spectrometry (MS) is an ideal method for protein mass determination: the two primary methods for ionization of proteins are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). In a MALDI source, proteins typically carry a charge state of +1 and since mass spectrometers measure the m/z ratio (mass/charge state), proteins typically appear with their mass plus the mass of a single hydrogen ion in a MALDI TOF spectrum.

4.7.1 GLYCAN COMPOSITION ANALYSIS

MS, with its low sample requirement and high sensitivity, has been the predominantly technique used in glycobiology for characterization and elucidation of glycan structures, since provides a complementary method to HPLC for the analysis of glycans. However, the composition of native oligosaccharides is variable due to the presence of N-acetylated and acidic residues, and this variation affects ionization capabilities of the oligosaccharides, sometimes preventing them from being detected by MS. For this reason, derivatization of oligosaccharides by chemical means such as permethylation is often performed before MS analysis, because this chemical modification stabilizes the sialic acid residues in acidic oligosaccharides. Moreover, the addition of methyl groups to

oligosaccharides also enables them to become more uniformly ionized by reducing the variation of the chemical properties among the glycans, caused by the presence of highly polar –OH and $–NH_2$ groups in the molecule; methylated glycans ionize more efficiently than their native counterparts, and due to their hydrophobic nature, are easily separated from salts and other impurities that may affect the MS analysis. Additionally, the fragmentation of methylated glycans is more predictable than that of their native counterparts, leading to accurate structural assignments when MS/MS analysis is performed [Dell et al. 1994; Kang et al. 2005].

Purified nuclear protein extracts obtained from HEK-293 cells stably expressing recombinant FLAG-REST have been analyzed through MALDI/TOF mass spectrometry by the Glycotechnology Core Resource (University of California, UCSD): they were first denatured, to make linkages susceptible to enzymatic digestion, and then digested with *pronase*. The sample was treated with *PNGase F* to release the N-linked glycans, which were fractionated through a reversed-phase C18 column; subsequently oligosaccharides were permethylated and the modified glycan fractions have been characterized by MALDI/TOF mass spectrometry. The instrument used is the QSTAR XL qTOF Maldi-Mass Spectrometer, a hybrid quadrapole time-of-flight mass spectrometer ideally suited for the analysis of all classes of glycoconjugates, including free glycans, glycopeptides and for protein identification and peptide sequencing.

4.7.2 GLYCOSYLATION SITE MAPPING

The incontrovertible identification of REST aminoacidic residues target of glycosylation can be obtained only through mass spectrometry: purified protein extracts have been analyzed through MS/MS by the Complex Carbohydrate Research Center (University of Georgia). To elucidate N-glycosylation target sites ^{18}O-labeling method and LC-MS/MS were used.

Purified REST extracts were reduced with 25mM DTT for 1h at 55°C and carboxyamidomethylated with 90mM iodoacetamide in the dark for 45min; the dried dialyzed sample was resuspended in 50mM ammonium bicarbonate and digested with 2,5 μg of trypsin at 25°C for 20h. Following deactivation of *trypsin* at 100°C for 5min, the sample was then deglycosylated with 2 μg of *PNGase F* in 36 μL of ^{18}O water and 2 μL of 1M ammonium bicarbonate; labeled peptides were then analyzed by C18 capillary chromatography LC/MS. The mixture of peptides was resuspended with 200 μL of mobile phase A (0.1% formic acid in water) and the sample was loaded onto a nanospray

tapered capillary column/emitter (360x75x15μm, PicoFrit, New Objective, Woburn, MA) self-packed with C18 reverse-phase resin (10.5cm, Waters, Milford, MA) in a Nitrogen pressure bomb for 10min at 1,000psi (~5μL load), and then separated via a 160min linear gradient of increasing mobile phase B at a flow rate of~500nL/min directly into the mass spectrometer. LC-MS/MS analysis was performed on a linear quadrupole ion trap LTQ Orbitrap Discoverer mass spectrometer (Thermo Scientific) equipped with a nanospray ion source: the resulting data were searched against the REST sequence using the TurboSequest algorithm (Proteome Discoverer 1.1, Thermo Scientific): the SEQUEST parameters were set to allow 30.0 ppm of precursor ion mass tolerance and 0.8 Da of fragment ion tolerance with monoisotopic mass. Tryptic peptides were allowed with up to two missed internal cleavage sites, and the differential modifications of 57.02146 Da, 15.9949 Da and 2.98826 Da were allowed for alkylated cysteine, oxidation of methionines and ^{18}O-labeled aspartic acid, respectively.

4.8 PREDICTION OF GLYCOSYLATION TARGET SITES

NetOGlyc 3.1 server, available via internet http://www.cbs.dtu.dk/services/NetOGlyc/, is a revised non-redundant database of glycoproteins with O-linked glycosylation sites, which produces neural network predictions of mucin type N-acetylagalactosamine (GalNAc) O-glycosylation sites in mammalian proteins: the criteria for inclusion in that database are at least on experimentally verified O-glycosylation site. A method for predicting the location for mucin-type glycosylation sites was trained using a neural network approach. The best overall network used as input aminoacid composition, averaged surface accessibility predictions, together with substitution matrix profile encoding of the sequence; to improve prediction on isolated sites, networks were trained on isolated sites only. The final method combines predictions from the best overall network and the best isolated site network and this prediction method correctly predicted 76% of the glycosylated residues and 93% of the non-glycosylated residues.

O-linked-β-N-acetylglucosamine (O-GlcNAc) does not display a definite consensus sequence: however the glycosylation motif is marked by the close vicinity of proline residues (positions -4,-3,-2), valines (-1,+2,+4,+5) and a downstream tract of serines (+1,+4,+7), though leucines and glutamines are disfavoured. Moving from these considerations the YinOYang server was generated (http://www.cbs.dtu.dk/services/YinOYang/): it is a good predictor for O-GlcNAc attachment sites in eukaryotic protein sequences, based on neural networks on 40 experimentally

determined O-GlcNAc acceptor sites. In a cross-validation 72,5% of the glycosylated sites and the 79,5% of the non-glycosylated sites were identified by the server.
An N-glycosylation site predictor for human proteins, called NetNGlyc, is available at http://www.cbs.dtu.dk/services/NetNGlyc/: it examines the sequence context of Asn-Xaa-Ser/Thr sequons using artificial neural networks. Artificial neural networks are trained on the surrounding sequence context, in an attempt to discriminate between acceptor and non-acceptor sequons: in a cross-validated performance, the networks could identify 86% of the glycosylated and 61% of the non-glycosylated sequons, with an overall accuracy of 76%.

4.9 SITE-DIRECTED MUTAGENESIS

In vitro site-directed mutagenesis is an invaluable technique for characterizing the dynamic, complex relationships between protein structure and function. The QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies) was used since it is specifically optimized for large constructs, allowing site-specific mutation of the double-stranded plasmid DNA in a three-step procedure, with high mutation efficiency and decreased potential for random mutations. It is performed using *PfuUltra* high-fidelity DNA polymerase for mutagenic primer-directed replication of both plasmid strands: the procedure utilizes a supercoiled double-stranded DNA vector, with an insert of interest, and two synthetic oligonucleotide primers, both containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by *PfuUltra* DNA polymerase, without primer displacement; extension of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling the product is treated with *DpnI*, an endonuclease specific for methylated and hemimethylated DNA used to digest the parental DNA template and to select for mutation-containing synthesized DNA: finally the nicked vector DNA incorporating the desired mutations is transformed into XL-10 Gold ultracompetent cells.
REST mutant forms have been obtained starting from the wild-type form of REST, encoded by the plasmid vector pCMV-Tag2B+REST; concerning N-linked glycosylation, starting from the single mutant N59Q multiple mutants have been obtained, by mutating one predicted N-glycosylation each time and using the previous mutant plasmid as template.

4.9.1 MUTAGENIC PRIMER DESIGN

The mutagenic oligonucleotide primers used for site-directed mutagenesis must be designed individually according to the desired mutation and to the following rules [Zheng et al., 2004]:

- Both mutagenic primers must contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid;
- Primers should be between 25 and 45 bases in length, with a melting temperature ≥ 78°C;
- The desired mutation should be in the middle of the primer, with ~10-15 bases of correct sequence on both sides;
- The primers should have a minimum GC content of 40% and should terminate in one or more C or G bases;
- At the 3' ending of each primer there should be at least 8 non-overlapping bases, while the primers must perfectly overlap at their 5' end.

To estimate the T_m of primers the following formula must be used:

$$T_m = 81{,}5+0{,}4(\%GC)-675/N-\%mismatch,$$

where N is the primer length in bases.

Codons corresponding to serine (TCA) and threonine (ACA) have been converted to alanine (GCA) with a single nucleotide substitution, while codons corresponding to asparagine (AAT) have been converted to glutamine (CAA), with a double substitution. Table 4.1 summarizes the primers we designed and used for site-directed mutagenesis.

MUTANT NAME	TARGET NUCLEOTIDE	PRIMER SEQUENCE
T618A	a1852g	5'-GCCTGCTCCCGCAGAGGCGGTTCAGAAGGGGCCCG-3' 3'-GTCTACCCCGGACGAGGGCGTCTCCGCCAAGTCTT -5'
S971A	t2911g	5'-GAAGAACCAGTTGCACCAATGCTTCCCCCTTCAGCAGTAG-3' 3'-GTTGTCAACTTCTTGGTCAACGTGGTTACGAAGGGGGAAG-5'
N59Q	A175c t177a	5'-GGGGAAGTACAAGGCAGCTGCTGTGATTACCTGGTCGGTG-3' 3'-CCGTTTACACCGGAATTGACCCCTTCATGTTCCGTCGACG-5'
N83Q	a247c c249a	5'-GGGGATAACCAATTTTCAGATAGTGAAGAAGGAGAAGG-3' 3'CCGTCTTGACTACGGCCAACCCCTATTGGTTAAAAGTC-5'
N417Q	a1249c t1251a	5'-TCCTACTTGTCCTCAAAAAACAATGGATGTCTCAAAAGTG-3' 3'-GTTTAGATTCGTAGGATGAACAGGAGTTTTTTGTTACCTA-5'
N440Q	a1318c t1320a	5'-GACTTGCCTGATCAAATTACCAATGAAAAAACAGAAATAGAAC-3' 3'-GGTTTTTTGCTCTCCGACTGAACGGACTAGTTTAATGGTTACT-5'
N482Q	a1444c t1446a	5'-AGCCTTCTAATCAAGTGTCAGTGATCCAGGTGACTACCAG-3' 3'-CAGAGTTTTCTCTTTTTCGGAAGATTAGTTCACGTCACT-5'
N958Q	a2872c t2874a	5'-CACCCGAGAGCAACTCACTGGTATAAATTCAACAGTTGAAG-3' 3'-CCTGTGACTAGTCTTGTGTTCTCTCGTTGAGTGACCATATT-5'
N1000Q	a2998c t3000a	5'-ATGGCAGCACAAGAGTCTCAGGAAATTGATGAAGATGAAG-3' 3'-CCGTAGTGGAGGACGATGTTACCGTCGTGTTCTCAGAGTC-5'

Table 4.1: List of primers used for site-directed mutagenesis.

4.9.2 MUTAGENESIS REACTION

To synthesize the mutant plasmid the reaction was assembled as follows:

	VOLUME (µL)
10x REACTION BUFFER	5
template plasmid DNA	50 ng
Primer FORWARD (125ng)	2
Primer REVERSE (125ng)	2
dNTP MIX	1
Quik Solution	3
PfuUltra DNA polymerase	1
Nuclease-free water	To a final volume of 50 µL

Table 4.2: List of reaction components mixed for the mutant synthesis reaction.

Each site-directed mutagenesis reaction was run according to the protocol reported in Table 4.3

Segment	Cycles	Temperature	Time
1	1	95°C	1 min
2	18	95°C	50 sec
		62°C	50 sec
		68°C	8.5 min
3	1	68°C	7 min

Table 4.3: Cycling parameters for the QuikChange II XL method.

The PCR product obtained was digested with 1 μL of the *DpnI* restriction enzyme (10U/ μL), for 1h at 37°C; subsequently 2 μL of the *DpnI*-treated DNA were added to 45 μL of ultracompetent cells, previously thawed on ice and incubated with 2 μL of β-mercaptoethanol on ice for 10 minutes. The transformation reaction was incubated on ice for 30 min and heat-pulsed in 42°C water bath for 30 seconds: after that the reaction was incubated on ice for 2 minutes and then o,5 mL of pre-heated NZY+ broth (10 g/L casein hydrolysate, 5 g/L yeast extract, 5 g/L NaCl, pH 7.5, supplemented with 1M $MgCl_2$, 1M $MgSO_4$, 2M glucose) were added to the tube, leaving the transformation reaction at 37°C for 1 hour with gentle shaking (225-250 rpm). Finally, the transformation reaction was plated on agar plates containing Kanamicin for plasmid vector selection.

4.10 ELECTROPHORETIC MOBILITY SHIFT ASSAY

Electrophoretic mobility shift assays were performed using 15μg of nuclear protein extract obtained from HEK-293 cells or PC-12 cells transiently transfected with wild-type or different mutant forms of REST. The probe was a 60-bp biotin end-labeled, duplex oligonucleotide containing the 21-bp RE-1 sequence upstream of the human μ-opioid receptor gene (hMOR):
(sense,
5'CTTGGTTGCGGTGGGGCCGGGGTCAGTACCATGGACAGCAGCCGGCGG
CTGGCGGGGAGCA-3';
complementary sequence,
5'TCTCCCCGCCAGCCGCCGGCTGCTGTCCATGGTACTGACCCCGGCCCCACC
GCAACCAAG-3') [Bedini et al., 2010].

DNA binding reactions were set up employing the LightShift Chemiluminescent Electrophoretic Mobility Shift Assay kit (Pierce), the nuclear protein extract and 62.5fmol of biotin-labeled oligonucleotide duplex. For competition assays, a 50 or 100-fold molar excess of unlabeled oligonucleotide duplex were used as competitors (the sequence is the same reported above) and added just before the biotin end-labeled probe. The reactions were performed in the standard binding buffer supplied (10mM Tris-HCl, 50mM KCl, 1mM DTT, pH7.5) plus 5mM $MgCl_2$ and 50ng/μL poly(dI-dC) non-specific competitor DNA and were incubated for 20min at room temperature. In *supershift* samples, 15μg of nuclear protein extract were pre-incubated at room temperature with 5 μg of anti-REST polyclonal antibody (Upstate) or anti-FLAG M5 monoclonal antibody (Sigma) for 30min at room temperature before adding the probe. The reaction mixture was loaded onto a 5% non-denaturating polyacrylamide gel with 0.5× TBE buffer and electrophoresed for 60min at 100V. The gel was transferred to a nylon membrane (Nytran; Schleicher & Schuell Italia) in 0.5× TBE at ~350mA for 60min. The transferred DNA was cross-linked (120mJ/cm2 for 1min at 254nm) using a UV cross-linker and the detection was performed by a chemiluminescent procedure as indicated by the manufacturer.

4.11 REPORTER GENE ASSAY

PC-12 cells were plated in in collagen IV (Sigma)-coated 24-well dishes and at 50-60% confluence were transiently transfected with 1.4 μg/ well of hMOR promoter/luciferase reporter plasmid (pGL3+LNpMOR), 0.7 μg/well of the plasmid that encodes for wild-type or mutant FLAG-REST (pCMV-Tag2B+REST) and 0.4 μg/well of the normalization plasmid pSV-• Gal (Promega), using the EXGEN 500 Transfection Reagent (Fermentas, Hanover, MD, USA). Cells were cultured in serum-free medium for 12 h, then exposed to 10nM IGF-I or left untreated for 24 h. Finally, cells were lysed in Reporter Lysis Buffer (Promega) and samples were collected.

Luciferase and beta-galactosidase activities of cell lysates were measured with the Bright-Glo Luciferase Assay System or Beta-Glo Assay System (Promega), according to the manufacturer's instructions.

The reporter plasmid is based on the pGL3 Basic vector system (Promega) containing the firefly luciferase reporter gene and was obtained as reported elsewhere [Bedini et al., 2008].

5. RESULTS

5.1 ANALYSIS OF NATIVE REST PROTEIN

Although the molecular weight calculated on the basis of REST aminoacidic sequence is about 116 kDa, western blotting analysis shows that this transcription factor has higher apparent molecular weights (see Table 1.1). The difference could be explained by post-translational modifications of the protein. Although REST4 was found to be O-glycosylated between residues 87 and 152 [Lee et al., 2000], full length REST could contain additional glycosylation sites. We observed that REST protein migrates on SDS-PAGE with a specific pattern of bands, differing among various cell lines and between the nuclear and the cytosolic form of the protein.
To study the endogenous form of REST we evaluated the expression of this transcription factor in the neuroblastoma cell line SH-SY5Y and in the embryonic cell line HEK-293: to recogne REST protein, two antibodies have been used, both raised against the C-terminal region of the protein (see paragraph 4.3).

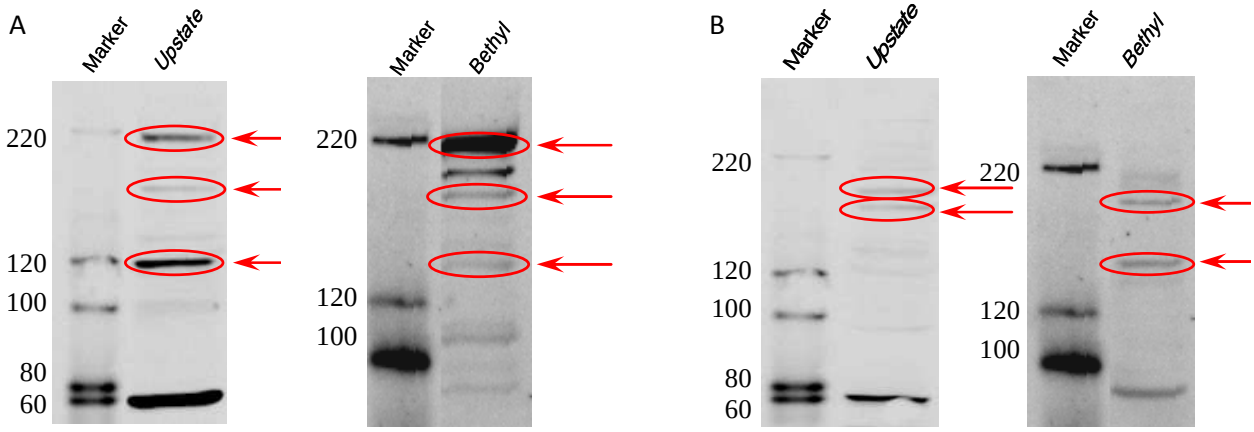

Figure 5.1.1: Western blotting analysis of endogenous REST expressed in neuroblasoma SH-SY5Y cells, from A) nuclear and B) cytoplasmic extracts. Detection was carried out using two different commercial antibodies, raised against the C-terminal region of REST (*Upstate* and *Bethyl*).

Nuclear protein extracts of neuroblastoma cells present at least three specific bands corresponding to REST, of apparent molecular weights of ~200, ~150 and ~120 kDa, consistent with different steps of post-translational modification of the protein backbone (Figure 5.1.1 A). On the contrary, cytoplasmic protein extracts migrate as two bands of

apparent molecular weight around 130-180 kDa (Figure 5.1.1 B). Native REST in embryonic HEK-293 cell nuclei migrates in SDS-PAGE with a specific pattern of apparent molecular weight of ~200, ~150 and ~120 kDa, as seen in neuroblastoma cells; cytoplasmic extracts, however, showed a double band of apparent molecular weight of ~130-150 kDa (Figure 5.1.2).

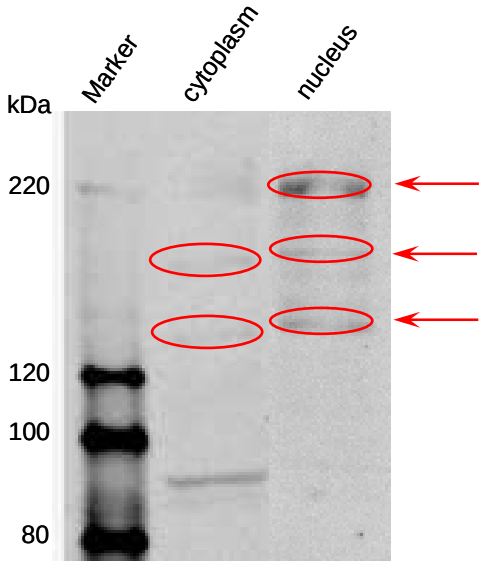

Figure 5.1.2: Western blotting analysis of endogenous REST protein from nuclear and cytoplasmic extracts, expressed in embryonic HEK-293 cells. Detection was carried out using *Upstate* anti-REST antibody.

5.2 ENZYMATIC DEGLYCOSYLATION OF REST

Moving from the observation that REST shows distinct patterns of bands in SDS-PAGE analysis, we started studying the composition of the glycan chains present in the protein: for this purpose we applied a strategy of enzymatic deglycosylation, using different combinations of enzymes capable of specifically remove the glucidic residues.

We used *N-glycosidase F* (*PNGase F*), one of the most effective method of removing all N-linked oligosaccharides from glycoproteins. Once released from the protein backbone, the oligosaccharide is left intact and therefore suitable for further analysis, while the asparagine residue from which the sugar was removed is deaminated to aspartic acid. The substrate of PNGase F is the glycosylated tripeptide Asn-X-Ser/Thr, but the presence of steric hindrance could inhibit enzyme activity (Figure 5.2.1) [Maley et al., 1989].

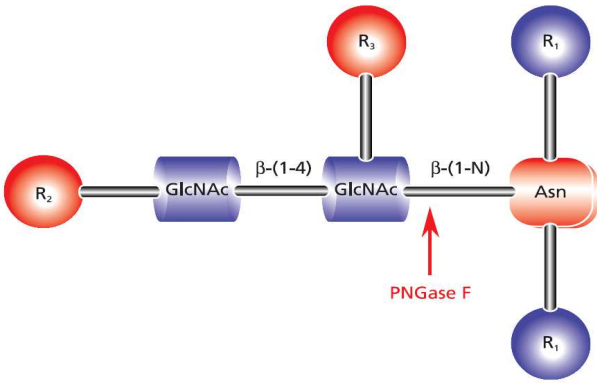

Figure 5.2.1: *PNGase F* enzymatic cleavage site. R1 = polipeptide backbone, R2 = carbhydrate or hydrogen atom, R3 = hydrogen atom.

Therefore, particular residues, due to their location in the native protein structure, are often resistant to the traditional deglycosylation with *PNGase F*: additional specific enzymes can be used for the removal of N-linked oligosaccharides from glycoproteins under native conditions.

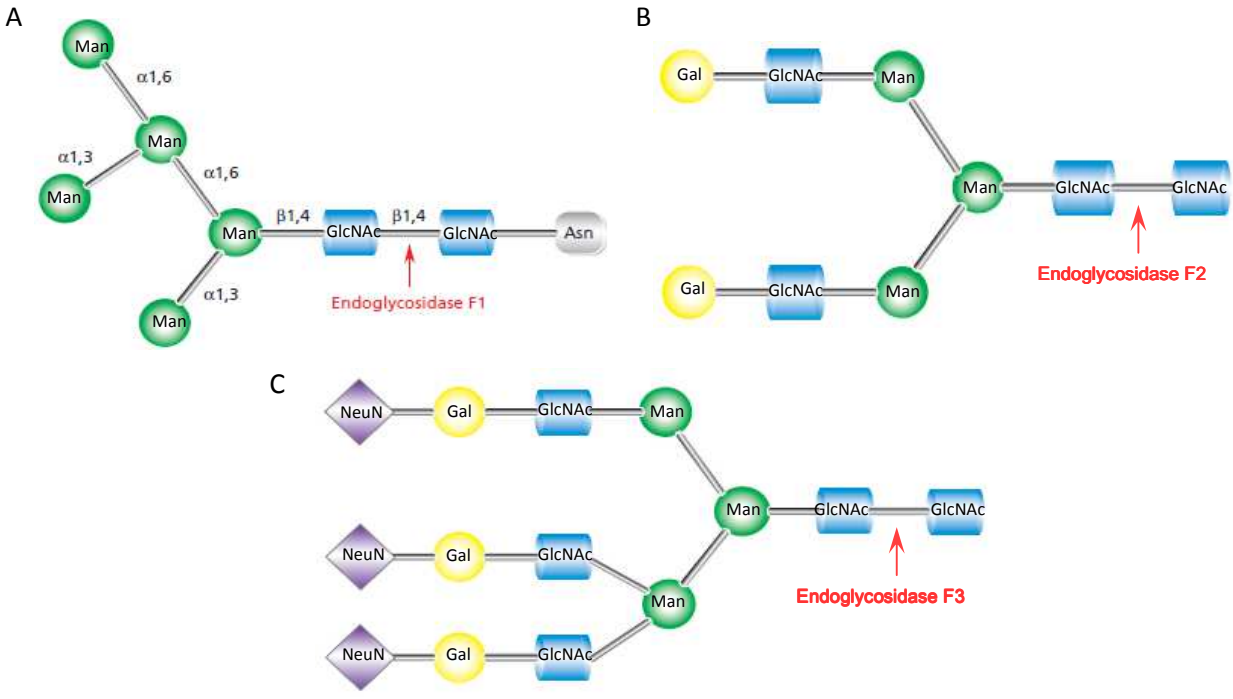

Figure 5.2.2: Schematic representation of the *Endoglycosidases* action; *Endo F1*, *Endo F2* and *Endo F3* specificity is indicated.

Endoglycosidase F1, *F2* and *F3* are less sensitive to protein conformation than *PNGase F* and more suitable for deglycosylation of native proteins. In particular oligomannose and hybrid structures can be removed by *Endoglycosidase F1*, or *Endo F1* (Figure 5.2.2 A), while biantennary and triantennary structures can be removed by *Endoglycosidase F2* and *F3*

respectively, also known as *Endo F2* and *Endo F3* (Figure 5.2.2 B and C). These enzymes cleave the oligosaccharide between two N-acetylglucosamine residues in the diacetylchitobiose core thus generating a truncated sugar molecule with one N-acetylglucosamine residue remaining on the asparagine. In addition, all complex oligosaccharides can be reduced to the trimannosylchitobiose core by treatment of the glycoproteins with *neuraminidase*, *β-galactosidase* and *N-acetylglucosaminidase*.
There is no enzyme comparable to *N-glycosidase F* for removing intact O-linked sugars. The enzyme *Endo-α-N-acetylgalactosaminidase* (*Endo-α*) can remove the intact core structure intact without any modification of the serine or threonine residues, but any modification of the core structure will block the action of the enzyme (Figure 5.2.3): however external monosaccharides must be removed by a series of exoglycosidases until only the Galβ1,3GalNAc core remains attached to serine or threonine.

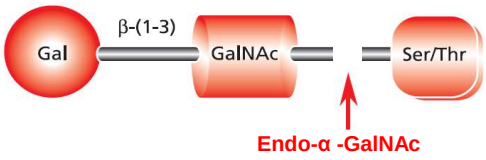

Figure 5.2.3: Schematic representation of the *Endo-α-N-acetylgalactosaminidase* cleavage site.

The most common modification of the core structure that can block the action of *Endo-α* is mono-, di-, or trisialylation: these residue can be easily removed by the action of *α2-3,6,8,9-neuraminidase*, since this enzyme is efficient at cleaving the NeuNAcα2,8-NeuNAc bond. A less common, but widely distributed O-linked hexasaccharide structure, contains β1,4-linked galactose and β1,6-linked N-acetylglucosamine: complete removal of this O-linked structures or its derivatives would require a *β1,4-galactosidase* and a *β-N-acetylglucosaminidase* (Figure 5.2.4).

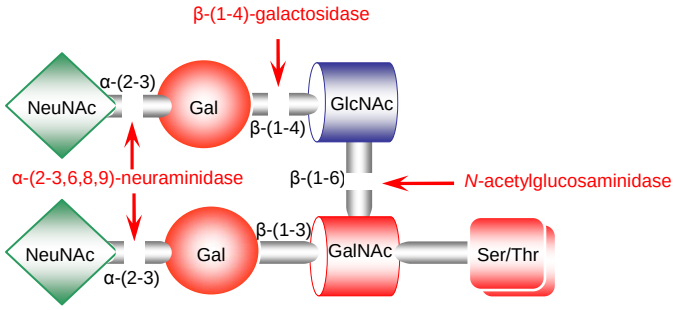

Figure 5.2.4: Schematic representation of the enzymatic removal of O-linked glycan chains from the Gal-β-(1-3)-GalNAc core of glycosylation.

Western blotting analysis of nuclear protein extracts from SH-SY5Y cells showed that the treatment with *PNGase F* did not cause any alteration in the migration of the bands identified as REST, even in denaturing conditions (Figure 5.2.5 A), suggesting that if N-glycan chains are present in the protein they can be buried in the protein core and unaccessible to the enzyme. The combination of all the enzymes necessary for O-glycosylation removal (*Endo-α* + *Galactosidase* + *Neuraminidase* + *β-N-acetylglucosaminidase*), caused the appearance of one additional band in the sample called *O-degly*, whose apparent molecular weight is about 180 kDa (Figure 5.2.5 A) and which is much more evident in the non-denaturing protocol (see paragraph 4.4). Also treatment with *Endo F1* caused the appearance of an additional band of an apparent molecular weight of about 180 kDa (Figure 5.2.5 B).

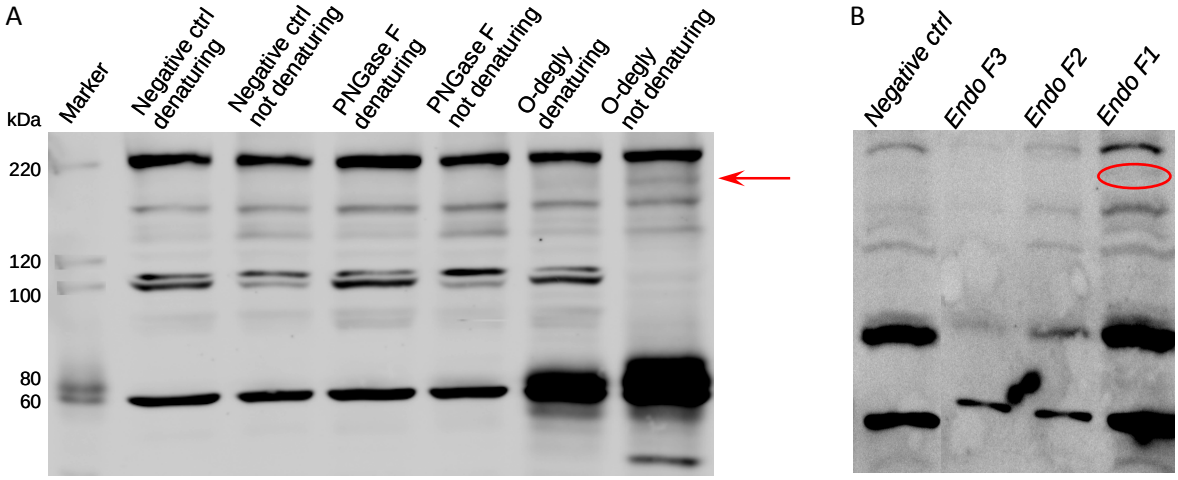

Figure 5.2.5: Western blotting analysis of endogenous REST nuclear protein expressed in neuroblasoma SH-SY5Y cells, deglycosylated by: A) *PNGase F* and the combination of *Endo-α*, *neuraminidase*, *galactosidase* and *N-acetylglucosaminidase* (O-degly); B) *Endo F1*, *F2* or *F3*.

To further understand the composition of the oligosaccharides we tested enzymatic deglycosylation abilities of the same set of enzymes both on endogenous REST expressed in HEK-293 cells and on the recombinant FLAG-REST protein, stably expressed in HEK-293 cells. To this aim recombinant REST has been cloned in a plasmid vector downstream of a sequence encoding for the FLAG peptide (N-DYKDDDDK-C), a synthetic polypeptide which is absent from any natural peptide sequence. First the over-expression of FLAG-REST has been verified through western blotting analysis: in Figure 5.2.6 is shown that transfection of the recombinant protein causes an increment in the intensity of the band corresponding to the fully glycosylated form of REST. Subsequently, nuclear protein extracts have been subjected to enzymatic deglycosylation: as can be seen in Figure 5.2.7, *PNGase F*, *Endo F2* and *Endo-α* alone cause no variations in band

migration, since the three bands corresponding to REST apparent molecular weights of approximately 200, 150 and 120 kDa did notchange with respect to the *Negative control* sample. On the contrary, treatment with the combination of *Endo*-α + *Galactosidase* or with all the enzymes necessary for O-glycosylation removal, also called *O-degly*, caused the appearance of one additional band, whose apparent molecular weight is about 180 kDa, and the increment in intensity of the band of 120 kDa, as highlighted by the red circles.

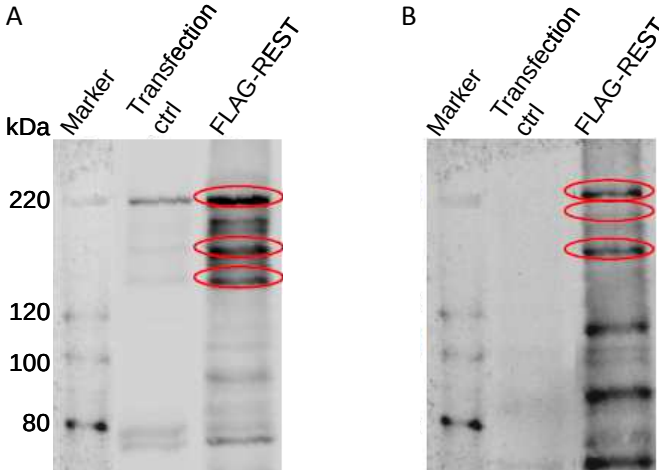

Figure 5.2.6: Western blotting analysis of nuclear protein extracts of the recombinant FLAG-REST protein expressed in HEK-293 cells. Detection was carried out using: A) *Upstate* anti-REST antibody; B)anti-FLAG M5 antibody.

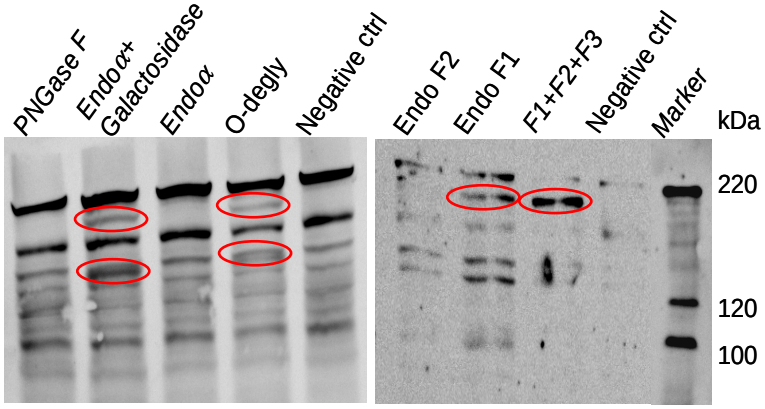

Figure 5.2.7: Western blotting analysis of the recombinant FLAG-REST after enzymatic deglycosylation with *PNGase F*, *Endo-α*, the combination of *Endo- α* and *β–galactosidase*, the combination of *Endo-α, neuraminidase*, *β -galactosidase* and *N-acetylglucosaminidase* (O-degly), *Endo F1*, *Endo F1* and the combination of *Endo F1*, *Endo F2* and *Endo F3*.

Similarly, treatment with *Endoglycosidase F1* caused the reduction of apparent molecular weight of the fully glycosylated form of REST and the appearance of an additional of

about 180 kDa. The same results were obtained also with the endogenous REST expressed in HEK-293 cells (data not shown).
Starting from these observations we hypothesized the presence of both N- and O-linked oligosaccharides in REST protein, but N-linked glycan chains seem to be buried in the protein structure, since *PNGase F* could not remove them.

5.3 CELL-FREE IN VITRO REST SYNTHESIS

In vitro synthesis of proteins in cell-free extracts is a useful tool in molecular biology and has a variety of applications, including the rapid identification of gene products, protein folding studies, protein interactions, protein structure and post-translational modifications analysis. The most frequently used cell-free translation systems consist of extracts from rabbit reticulocytes or *Escherichia coli*: both of them use DNA as a template, from which RNA is transcribed and subsequently translated without any purification.
The S30 T7 High-Yield Protein Expression System (Promega) is an *E. coli* extract-based cell-free protein synthesis system: it permits transcription and translation of DNA sequences cloned in plasmids downstream a T7 promoter, by providing an extract that contains a bacteriophage T7 RNA polymerase and all necessary components for translation, thus allowing the production of high levels of recombinant proteins. Large differences in the level of gene expression could be obtained due to the size of the protein, the gene of interest and the context in which the gene resides, especially relative to the ribosome binding site (RBS). RBS is generally located approximately seven bases upstream of the AUG start codon; in addition, many eukaryotic genes contain sequences within the protein-coding region that can function as ribosomal-binding sites when they precede a methionine codon.
Rabbit reticulocyte lysates (Promega) is a highly efficient *in vitro* eukaryotic protein synthesis system: it typically combines a prokaryotic phage RNA polymerase and promoter (T7 in our case) with eukaryotic extracts to synthesize proteins from exogenous DNA templates. For the highest protein yield and the best initiation fidelity, the DNA template must have a Shine-Dalgarno RBS upstream of the initiation codon. Rabbit Reticulocyte Lysate may contain a variety of post-translational processing activities, including acetylation, isoprenylation, phosphorylation and glycosylation.
As shown in Figure 5.3.1 synthetic REST protein produced in the *E.coli* cell-free system migrates in SDS-PAGE with an apparent molecular weight of about 120 kDa, comparable to the expected non-glycosylated form of endogenous REST. On the

contrary, the synthetic protein produced by the Rabbit Reticulocyte Lysate system shows an apparent molecular weight of about 220 kDa, similar to the fully glycosylated form of endogenous REST, expressed in SH-SY5Y cells and loaded in the SDS-PAGE as a positive control.

These results confirm our first hypothesis that REST is modified by the addition of complex glycan chains in the protein backbone, since only eukaryotic extracts were able to synthesize a protein with the exact folding and post-translational modifications, showing the proper apparent molecular weight on western blotting analysis.

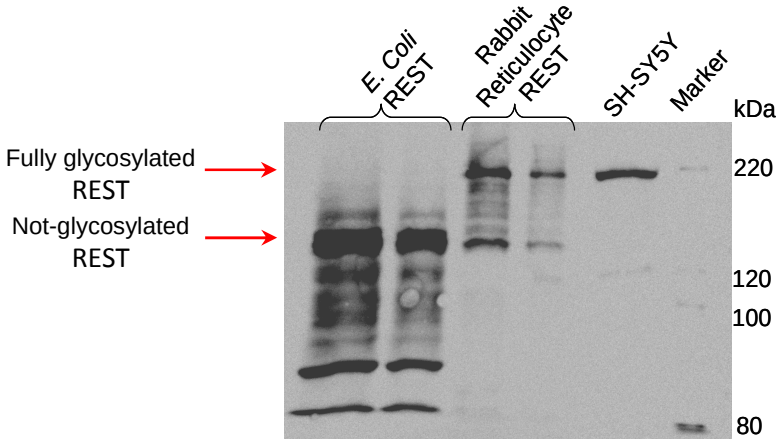

Figure 5.3.1 Western blotting analysis of the cell-free *in vitro* synthesized REST protein, produced in *E. coli* or in Rabbit Reticulocyte Lysate systems and compared to the endoguenous REST extracted from SH-SY5Y cells. *E.coli* REST migrates with an apparent molecular weight of ~120 kDa, comparable to the expected non-glycosylated form of endogenous REST; Rabbit Reticulocyte Lysate REST shows an apparent molecular weight of ~220 kDa, similar to the fully glycosylated endogenous REST.

5.4 MASS SPECTROMETRY ANALYSIS

To further confirm REST glycosylation we subjected recombinant REST protein extracts, obtained from HEK-293 cells stably transfected with the FLAG-REST plasmid construct, to mass spectrometry analysis, in order to depict the glycan chains in terms of oligosaccharide composition and linkage analysis. For this purpose, the protein extracts had to be purified by immuno-affinity, using an agarose resin conjugated with anti-FLAG antibody capable to bind FLAG-REST and purify it from any other protein present in the nuclear extract. Figure 5.4.1 shows that immunoprecipitated FLAG-REST (*IP REST*) is enriched in the fully glycosylated form of the protein, if compared with the non-

immunoprecipitated extract (*input*), while the non-glycosylated protein is removed by washing steps (see *wash* fraction).

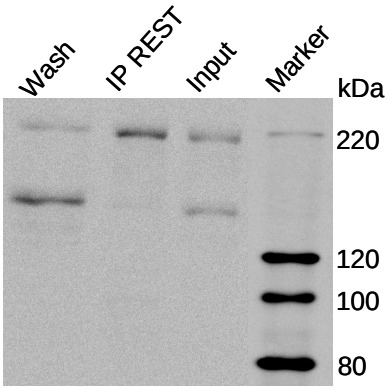

Figure 5.4.1: Western blotting analysis of the immunoprecipitated recombinant FLAG-REST protein stably expressed in HEK-293 cells, using anti-REST antibody.

5.4.1 GLYCAN COMPOSITION ANALYSIS

Protein extracts so far purified were analyzed through MALDI/TOF mass spectrometry by the Glycotechnology Core Resource (University of California, UCSD). Purified protein extracts were first denatured to make linkages susceptible to enzymatic digestion and the glycoprotein sample was then digested with *pronase* into peptides. The sample was treated with *PNGase F* to release the N-linked glycans, which were fractionated by passing the sample through a reversed-phase C18 column, which enables a significant reduction in sample complexity prior to mass spectrometry analysis without loss of important components. Subsequently permethylation of oligosaccharides was used to improve their signals by mass spectrometric analysis (see paragraph 4.7): released N-glycan fractions were characterized by MALDI/TOF mass spectrometry. REST has been shown to have a very complex mixture of glycans but the ion intensities of some species are at noise level; however, there are some dominant glycans, as showed in Figure 5.4.2, having higher peak intensities, which confirm the presence of N-linked glycan chains composed of N-acetylgalactosamine, galactose and mannose, as suggested by previous enzymatic deglycosylation analyses.

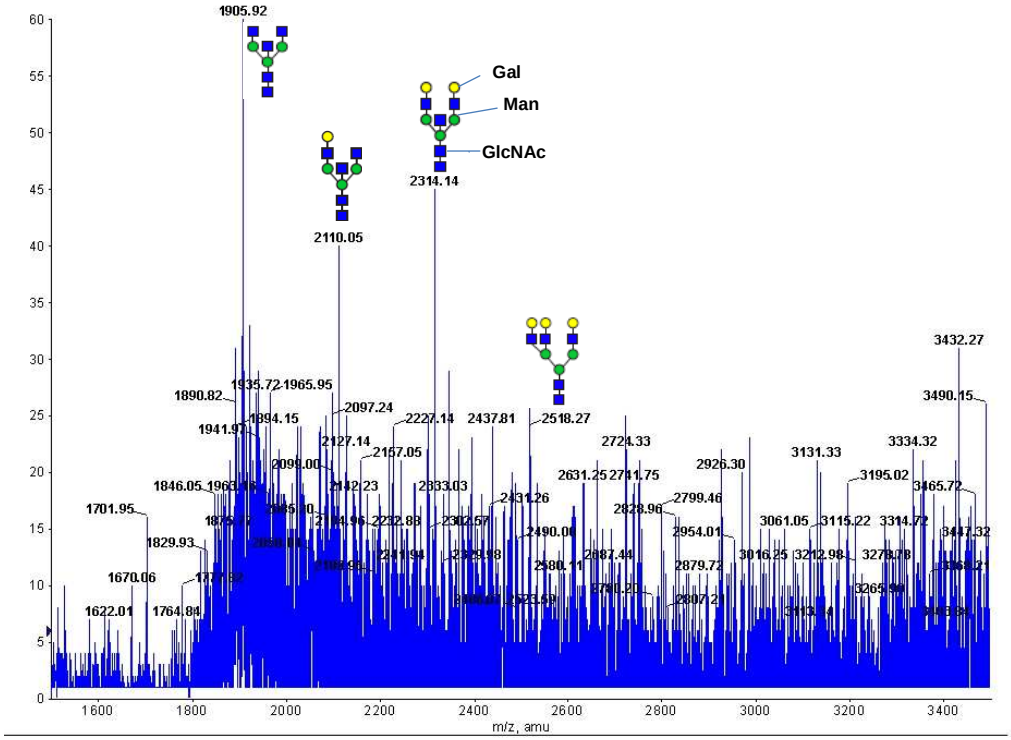

Figure 5.4.2: MALDI/TOF-MS spectrum of permethylated N-glycans released from immunoprecipitated FLAG-REST protein extracts.

5.4.2 GLYCOSYLATION SITE MAPPING

Additional information on REST glycosylation can be obtained by site mapping, since the incontrovertible identification of REST aminoacidic residues target of glycosylation can be obtained only through mass spectrometry. For this purpose purified protein extracts have been analyzed through MS/MS by the Complex Carbohydrate Research Center (University of Georgia). In particular, to elucidate N-linked sites a ^{18}O-labeling method was used. Samples are digested with *protease* and the glycopeptides deglycosylated with *PNGase F* in $H_2{}^{18}O$, converting the asparagine into an aspartic acid, are labeled with ^{18}O: the glycosylated peptide will show an increase of 3 Da mass compared to the non-glycosylated peptide. A mixture of peptides is then analyzed by C18 capillary liquid chromatography LC-MS: the data acquisition process included a full mass spectrometry scan followed by MS/MS fragmentation of the most intense ion selected from the full MS spectrum, by an LTQ (Linear Trap Quadrupole) instrument. Data obtained were searched against the REST sequence and general protein database (see paragraph 4.7). N-glycosylation site mapping of REST indicated twelve potential N-glycosylation sites: Asn59, Asn83, Asn417, Asn440, Asn482, Asn553, Asn586, Asn914, Asn958, Asn963,

Asn1000 and Asn1022: the potential N-glycosylation sites of REST are shown in Table 5.4.1.

Sequon	Digestion	Peptide position	Peptide sequence[a]	Theoretical (m/z)	Observed (m/z)	z
Asn^{59}	-	-	-	-	-	
Asn^{83}	Trypsin/Glu-C	76-88	(E)LMPVGDNN@FSDSE(E)	1427.5881	-	1
Asn^{417}	-	-	-	-	-	
Asn^{440}	Trypsin	434-445	(R)EADLPDN@ITNEK(T)	1361.6317	-	1
				681.3195	-	2
Asn^{482}	Trypsin	478-493	(K)KPSNN@VSVIQVTTRTR(K)	1802.9969	1802.9471	1
				902.0021	901.9772	2
Asn^{553}	Trypsin	553-564	(K)N@NSQEVPKGDSK(V)	1305.6167	1305.6347	1
				653.3120	653.3201	2
Asn^{586}	-	-	-	-	-	
Asn^{914}	Trypsin/Glu-C	905-915	(E)SLPGLAANIN@E(S)	1101.5672	-	1
Asn^{958}	Trypsin/Glu-C	958-967	(E)N@LTGINSTVE(E)	1050.5199	-	1
$Asn^{958, 963}$	Trypsin/Glu-C	958-967	(E)N@LTGIN@STVE(E)	1053.5082	-	1
Asn^{1000}	Trypsin/Glu-C	988-1001	(K)TALASPPATMAAN@E(S)	1347.6347	-	1
Asn^{1022}	Trypsin/Glu-C	1016-1025	(E)GSDLSDN@MSE(G)	1057.3876	-	1

Table 5.4.1: Potential N-linked glycosylation site mapping of *Trypsin* digests from REST sample. (@ indicates the site of N-glycosylation).

Asn482 and Asn553 were shown to be glycosylated by LC-MS/MS (Figure 5.4.3) and database searching using the TurboSequest algorithm (Figure 5.4.4): the sample showed doubly charged ions at m/z 901.98 (Figure 5.4.3 A) and at m/z 653.32 (Figure 5.4.3 B), which correspond to the ^{18}O labeled deglycosylated peptide containing Asn482 and Asn553. MS/MS spectra showing the localization of N-linked glycosylation sites for Asn482 with cross-correlation score (X corr = 0.70) and Asn553 (X corr = 0.45) on REST are presented in Figure 5.4.4; however, whether site Asn482 and Asn553 are truly glycosylated remains questionable, due to the fact that the X corr values of the ^{18}O-site mapping are under confident values (for a double charges ion peptide, its Sequest cross-correlation score must be >2 to be confident [Link et al., 1999]).

The other potential N-linked glycosylation sites of REST cannot be detected by ^{18}O-site mapping, suggesting that the protein could be highly glycosylated, which affected C18 binding ability and the elution time.

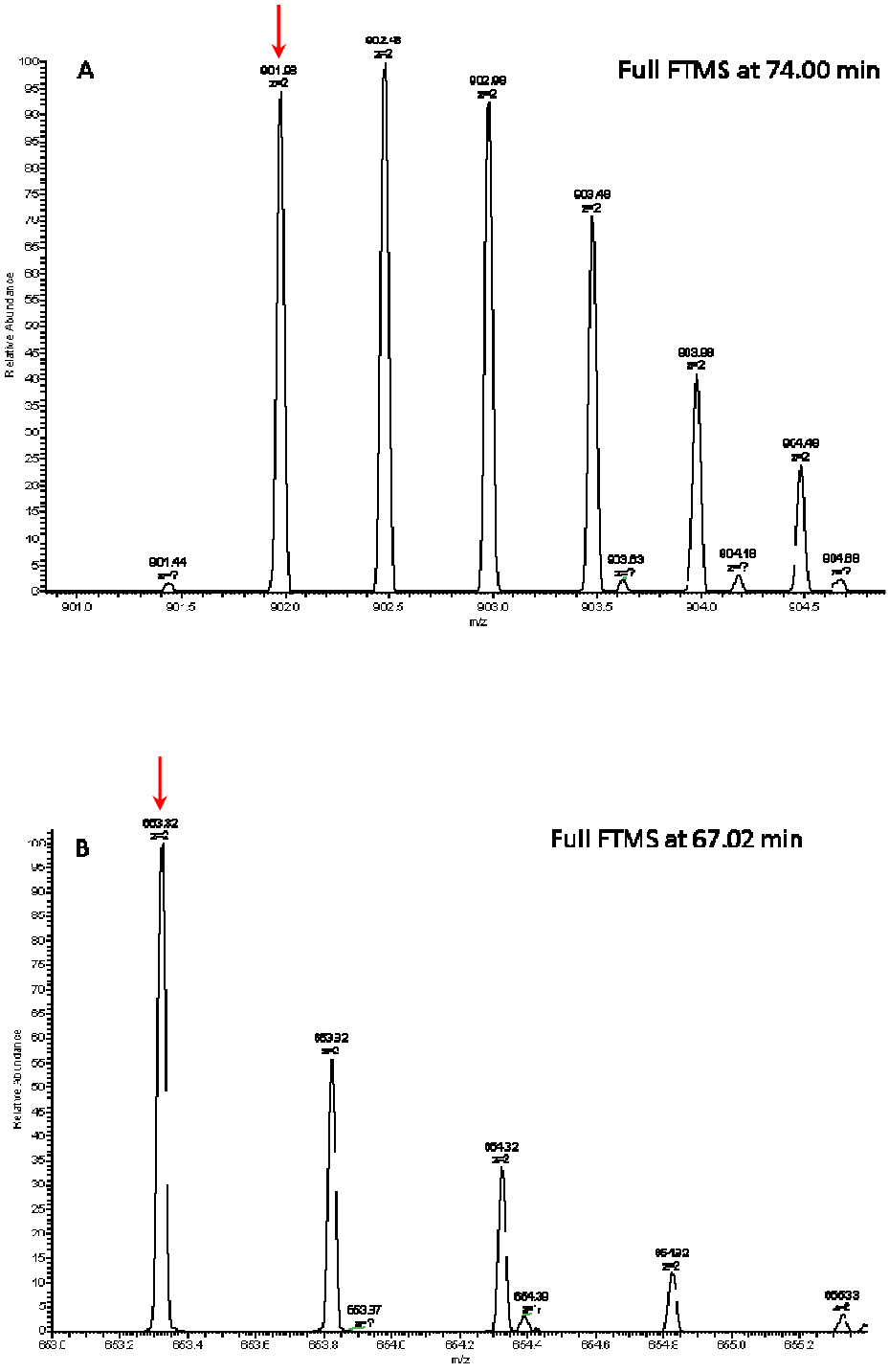

Figure 5.4.3: Full Fourier transform mass spectrometry (FTMS) spectra of ^{18}O labeled deglycosylated peptides of REST. A) doubly charged deglycosylated peptide ion (K)KPSNN@VSVIQVTTRTR(K), containing Asn482 , m/z 901.98 ; B) doubly charged deglycosylated peptide ion (K)N@NSQEVPKGDSK(V), containing Asn553 , m/z 653.32.

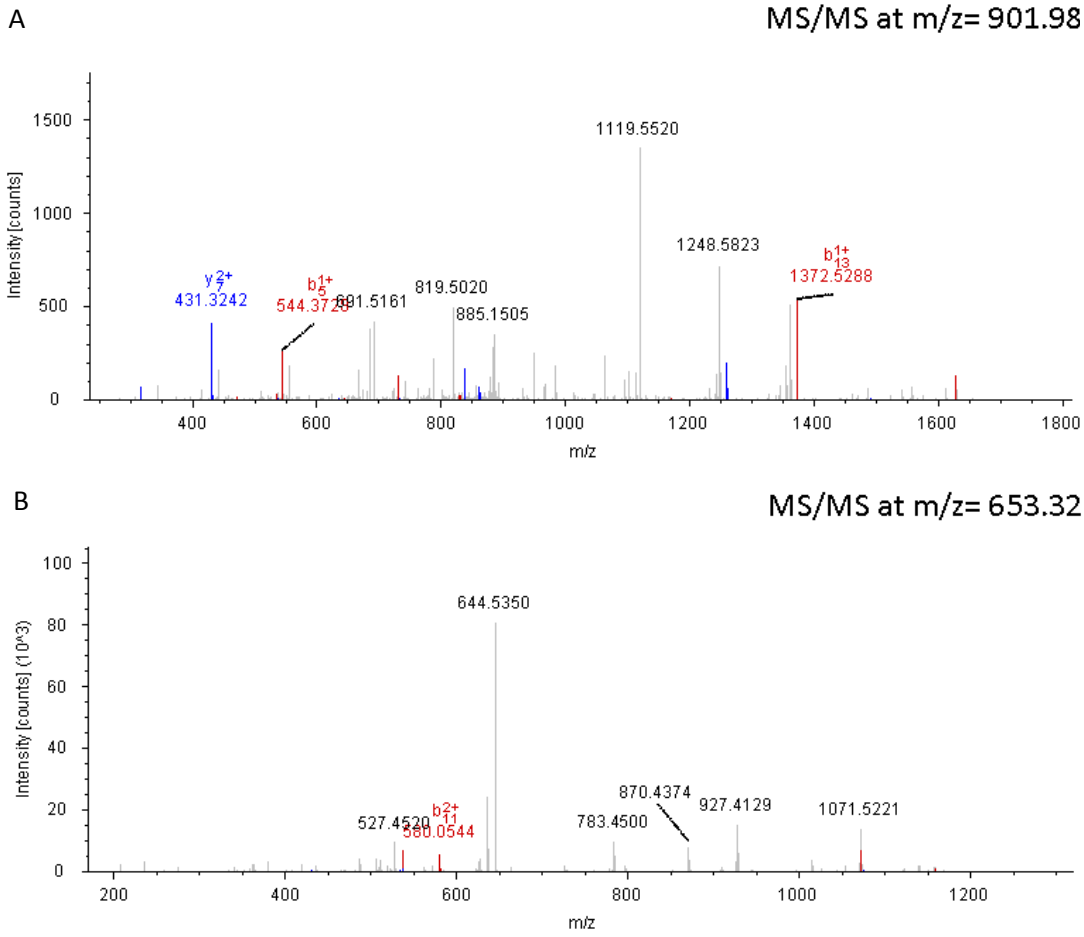

Figure 5.4.4: MS/MS spectra showing localization of glycosylation site on REST by *Trypsin* and *PNGase F* digestion and ^{18}O labeling of Asn residues carrying glycans. A) peptide containing Asn482 at m/z 901.98 [(K)KPSNN@VSVIQVTTRTR(K)]; B) peptide containing containing Asn553 at m/z 653.32 [(K)N@NSQEVPKGDSK(V)].

Thus, glycopeptides obtained after tryptic digestion have been O-deglycosyalted through β-elimination in mild conditions followed by Michael addition (BEMAD), in order to remove O-linked glycans, and then analyzed by mass spectrometry using a direct infusion technique: even after O-linked deglycosylation, REST glycopeptides were not confidently detected by LC-MS/MS, most likely because of un-appropriated sizes and solubility of their sequences.

5.5 SITE-DIRECTED MUTAGENESIS

To continue the characterization of the glucidic component present in the REST protein and its impact in the transcription factor's function, I applied a site-directed mutagenesis

strategy mutating several aminoacids predicted to be target of glycosylation. The aim was to verify variations in REST molecular weight, DNA binding activity or repressor function consequently to the removal of a single glycan chain or of all of them.

5.5.1 PREDICTION OF GLYCOSYLATION SITES IN REST

In order to investigate possible glycosylation sites on REST protein, different software have been used to submit REST sequence to the prediction aminoacidic residues target of both N- and O-linked glycosylation.

Although O-linked N-acetylgalactosamine is one of the main types of glycosylation in mammalian cells, no consensus sequence is known, making prediction methods necessary to bridge the gap between the large number of known protein sequences and the small number of proteins experimentally investigated with regard to glycosylation status. NetOGlyc is a revised non-redundant database of glycoproteins with O-linked glycosylation sites, rating 242 glycoprotein entries: the server produces neural network predictions of mucin type N-acetylagalactosamine (GalNAc) O-glycosylation sites in mammalian proteins. Examination of REST sequence with NetOGlyc found out several putative O-glycosylation sites: sequence analysis generated a graph in which the probability of each residue to be glycosylated was represented and compared to a set threshold (Figure 5.5.1 A). NetOGlyc analysis on REST protein sequence predicted one putative site modified with O-GalNAc: threonine 618.

Like 'mucin type' O-glycosylation, also O-linked-β-N-acetylglucosamine (O-GlcNAc) does not display a definite consensus sequence: however the motif is marked by specific characteristics that help defining region surrounding the glycosylation site. Moving from these considerations neural networks on 40 experimentally determined O-GlcNAc acceptor sites have been created, generating the YinOYang server, a good predictor for O-GlcNAc attachment sites in eukaryotic protein sequences: using this server eleven O-GlcNAc sites have been predicted in REST protein (Figure 5.5.1 B).

Furthermore prediction of N-glycosylation target sites has been performed. The consensus sequence, Asn-Xaa-Ser/Thr, is known to be a prerequisite for the modification; however not all of these sequons are modified and it is thus not discriminatory between glycosylated and non-glycosylated asparagines. The NetNGlyc prediction server predicts N-glycosylation sites in human proteins, using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequons, and allows the discrimination between acceptor and non-acceptor sequons: submitting REST protein

sequence to NetNGlyc, twelve Asn have been obtained as output, with seven of them being highly probable N-glycosylated (Figure 5.5.1 C).

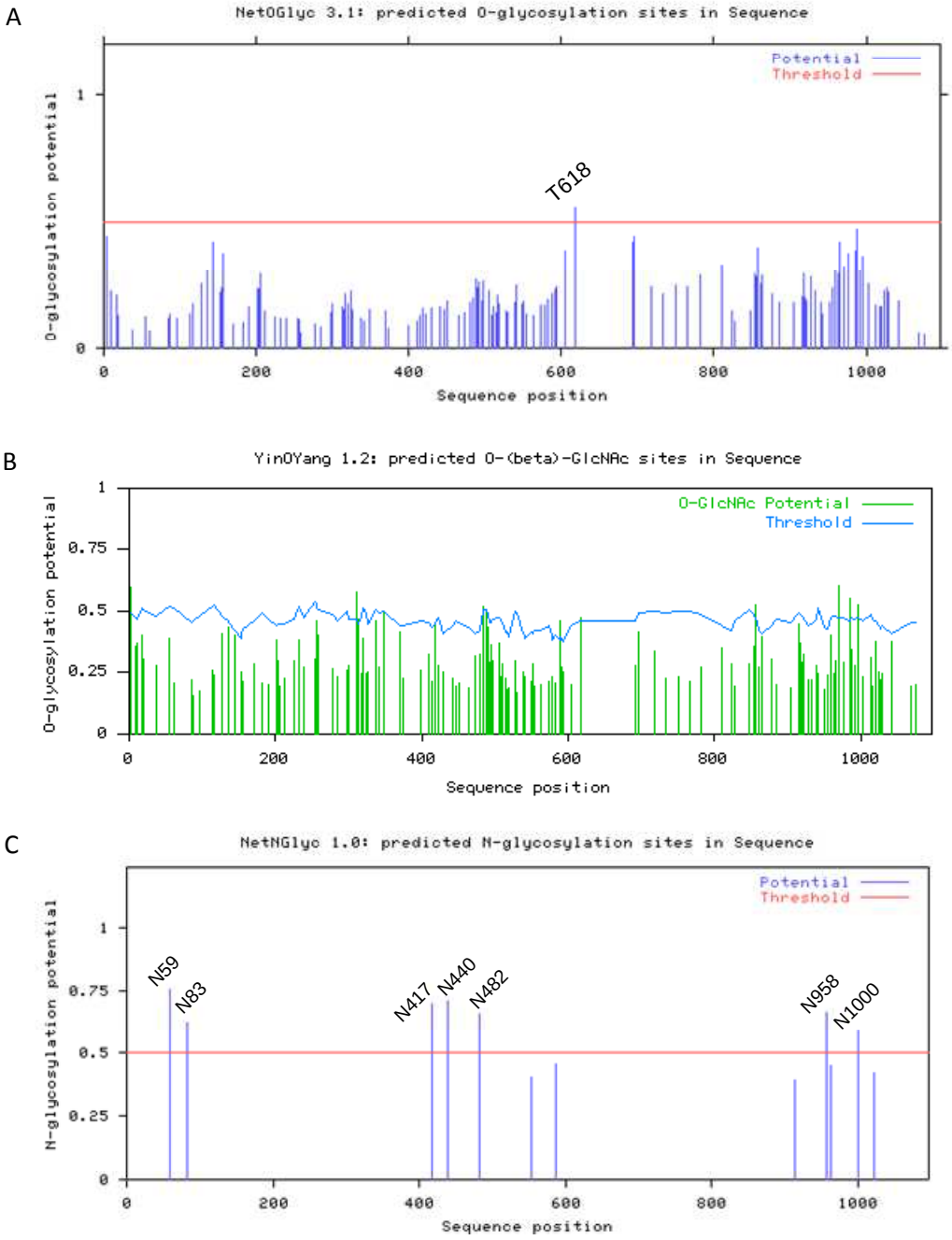

Figure 5.5.1: Prediction of putative glycosylation sites on REST sequence. A) 'mucin-type' O-glycosylation; B) O-linked-β-N-acetylglucosamine (O-GlcNAc); C) N-linked glycosylation.

Table 5.5.1 summarizes the prediction results obtained so far, regarding both O- and N-linked glycosylation.

GLYCOSYLATION	POSITION	RESIDUE	PROBABILITY	MUTATION
O-GalNAc	618	Thr		T618A
O-β-GlcNAc	3	Thr	+	
	312	Ser	+	
	313	Ser	+	
	484	Ser	+	
	588	Ser	+	
	589	Ser	+	
	618	Thr	+	
	857	Thr	+	
	971	Ser	++	S971A
	986	Ser	+	
	996	Thr	+	
N-glycosyaltion	59	Asn	+++	N59Q
	83	Asn	++	N83Q
	417	Asn	++	N417Q
	440	Asn	++	N440Q
	482	Asn	+	N482Q
	553	Asn	-	
	586	Asn	-	
	914	Asn	-	
	958	Asn	++	N958Q
	963	Asn	-	
	1000	Asn	+	N1000Q
	1022	Asn	-	

Table 5.5.1: Prediction of aminoacids target of glycosylation. The residues highlighted by coloured rectangles have been subjected to site-directed mutagenesis for subsequent analysis.

5.5.2 MUTAGENESIS OF PREDICTED GLYCOSYLATED AMINOACIDS

Aminoacids predicted to be glycosylated with the highest probabilities were subjected to site-directed mutagenesis, in order to convert them to targets that are not recognized by *glycosyltransferases* anymore. Site-directed mutagenesis allowed the characterization of the complex structure-activity relationship. In fact, inserting a point mutation in REST coding sequence it is possible to block the addition of oligosaccharides to the previously predicted aminoacids. The QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies) has been used, coupled with specific primers which contain the desired mutation themselves (see paragraph 4.9 and table 4.1 for primers).

In particular codons corresponding to serine (TCA) and threonine (ACA) residues, target of O-linked glycosylation, have been converted to alanine (GCA), while codons corresponding to asparagine (AAT), predicted to be target of N-glycosylation, have been converted to glutamine (CAA): we made these substitutions because although they prevent the addition of glycan chains in correspondence to the modified aminoacid, they allow the maintenance of physico-chemical properties of the protein. Regarding O-glycosylation, we generated a REST mutant form for 'mucin-type' glycosylation, where threonine 618 was converted to alanine (T618A REST). In addition, also some of the O-β-GlcNAc target residues have been subjected to site-directed mutagenesis, generating S128A, T144A and S971A mutants.
Moreover, all of the asparagines predicted to be target of N-glycosylation (Table 5.5.1) have been converted to glutamine, singularly or sequentially, thus obtaining mutant forms of REST on multiple residues.

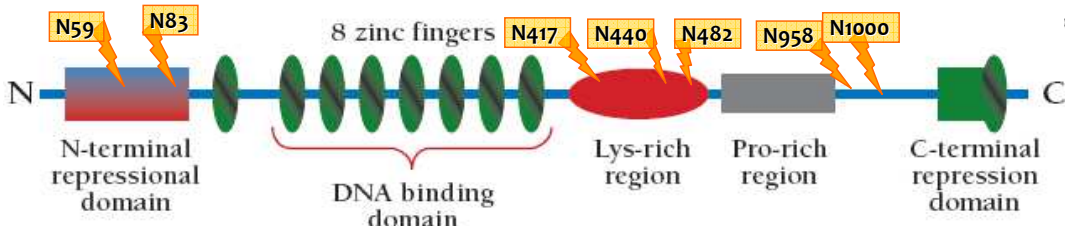

Figure 5.5.2: Schematic representation of potential N-glycosylation sites predicted in REST sequence.

REST single mutants have been obtained using as template REST wild-type, encoded by the plasmid pCMV-Tag2B+REST: the resulting mutant forms were S128A, T144A, T618A, S971A, N59Q, N83Q, N417Q, N440Q, N482, N958Q, N1000Q.
Concerning N-linked glycosylation, starting from the single mutant N59Q multiple mutants have been obtained, by adding first the N83Q mutation and, thus, obtaining the double mutant (2N); then the double mutant was used as template for the mutagenesis of the asparagine 417, giving the 3N mutant, and so on until all the seven asparagines highlighted in Figure 5.5.2 were mutated to glutamine, thus obtaining the 7N mutant. In addition we generated a mutant called N+C, where asparagines 59, 83 and 1000 were mutated to glutamine, in order to prevent glycosylation in amino- and carboxy-terminal domains of REST, known to be involved in the interaction with the co-represors (see paragraph 1.2 and Figure 1.3).
Figure 5.5.3 shows the schematic representation of the genetic map of the plasmid bearing REST coding sequence, where the mutated N-glycosylation sites have been

reported in purple and the restriction enzymes used for subsequent analysis of the mutants in black.

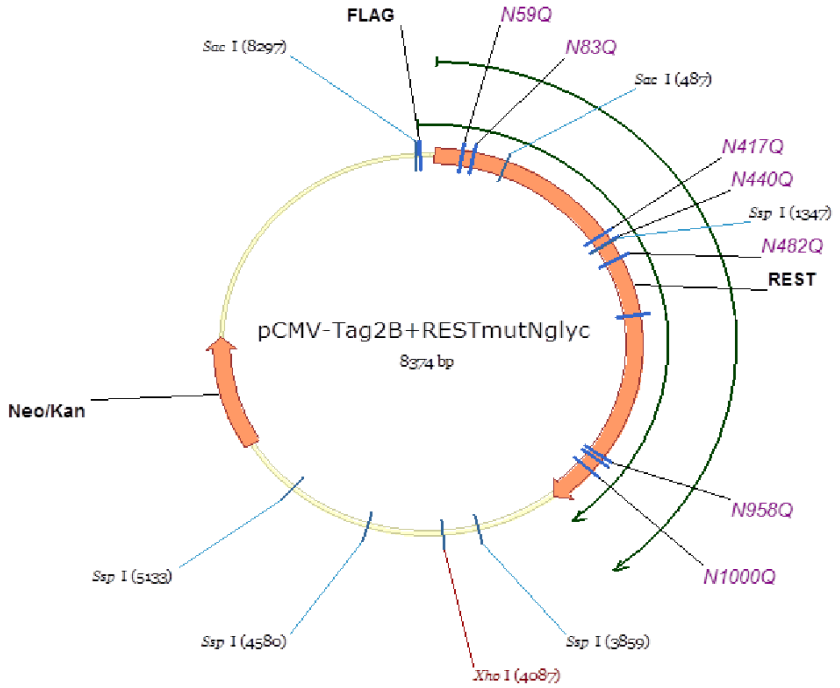

Figure 5.5.3: Genetic map of the pCMV-Tag2B+REST plasmid, showing the mutated sparagines target of N-glycosylation.

Once completed the mutagenesis procedure, the novel plasmids obtained, each encoding for a different mutant form of REST, were analyzed to confirm the presence of the desired mutation and the absence of any unwanted modification in the coding sequence. This was ascertained through restriction analysis and electrophoresis separation of the DNA fragments and, then, by sequencing the mutated plasmids to finally verify that the right mutation was obtained.

An example of restriction analysis and agarose gel electrophoresis of the mutant N1000Q is reported in Figure 5.5.4 A. I employed the enzyme *XhoI* to linearize the plasmid and obtain a linear fragment of 8374 base pairs (bp), the enzyme *SacI* to get two fragments of 7810 and 564 bp, and the enzyme *SspI* which cuts the plasmid into four fragments of 4588, 2512, 721 and 553 bp. Mutants containing the N440Q convertion, both single and multiple, showed a different pattern of restriction fragments due to the fact that one of the four restriction sites recognized by the enzyme *SspI* (5' -AATATT- 3') is lost during mutagenesis as it corresponds to the codon encoding asparagine 440. Therefore, as consequence of this mutation, only three fragments were obtained following restriction analysis, of 7100, 721 and 553 bp, the heavier of which results from the addition of the two fragments of 4588 and 2512 bp (Figure 5.5.4 B).

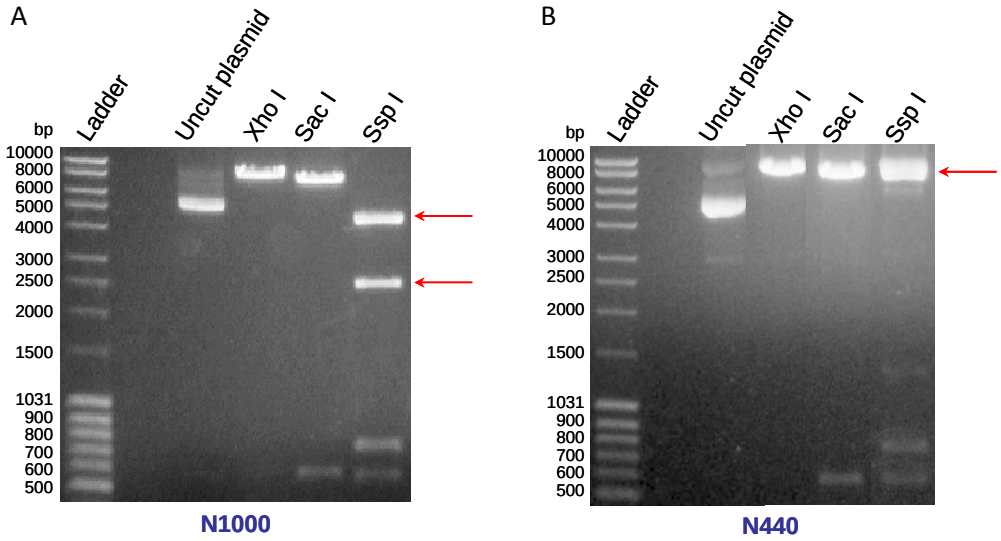

Figure 5.5.4: agarose gel electrophoresis of two representative mutant-REST plasmids, after restriction analysis with the enzymes *XhoI*, *SacI* and *SspI*. Red arrows highlight the difference of restriction fragments obtained after *SacI* treatment, in presence (B) or absence (A) of the N440Q mutation.

Finally, all the mutant plasmids so far obtained were subjected to sequencing analysis (BMR Genomics, Padua, Italy). In Figure 5.5.5 is reported the sequence analysis of N1000Q mutant as an example of the strategy we used. First we analyzed the chromatograms, to check peak size and the absence of any contaminating sequence; then we aligned the mutant sequence thus obtained to REST wild-type sequence, to confirm the presence of the desired mutation and the absence of any other unwanted modification of the plasmid DNA.

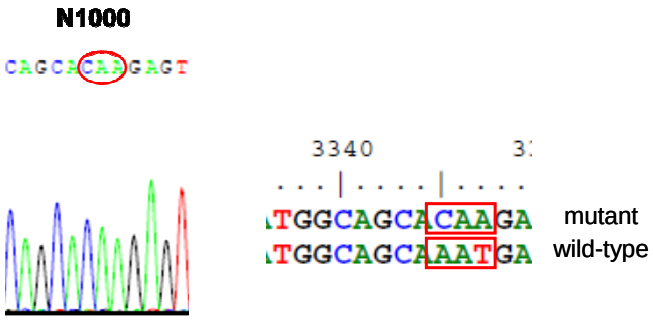

Figure 5.5.5: Sequence analysis of the mutated nucleotidic portion of REST after site-directed mutagenesis of N1000Q residue. A) Chromatogram obtained as a result of the sequencing analysis; B) Alignment of mutant and wild-type sequences. The mutant codon is highlighted in red.

5.6 FUNCTIONAL ANALYSIS OF THE MUTANT FORMS OF REST

Once the mutant REST coding sequences have been obtained and validated, the plasmids were transfected to transiently express mutant REST proteins in two different cell lines: PC-12 and HEK-293 cells, the first because among the neuronal cells is almost absent the expression of native REST. This property is particularly useful for the evaluation of mutant protein behaviour and activity. The second cell line it is a human embryonic model of epithelial origin that differs from the rat neuron-like model of PC-12 cells.

5.6.1 WESTERN BLOTTING ANALYSIS OF MUTANT REST

To verify the correct expression of recombinant mutant REST forms, nuclear protein extracts of transiently transfected cells were analyzed by western blotting, using both the anti-REST (Figure 5.6.1) and anti-FLAG (Figure 5.6.2) antibodies. By comparing band migration of wild-type and mutant REST samples in SDS-PAGE I have been able to observe that the removal of glycan chains, both N- and O-linked, did not alter REST apparent molecular weight.
Looking at the single mutants we first hypothesized that a small variation of apparent molecular weight, such the one consequent to the removal of a single N-linked oligosaccharide, could be hardly appreciated by western blotting; we expected multiple mutants, such as 7N REST, to show a significant reduction of apparent molecular weight of about 15 kDa, as suggested by mass spectrometry data (see paragraph 5.4.1 and Figure 5.4.2). None of the REST mutant forms that we generated and analyzed showed alterations of the apparent molecular weight after site-directed mutagenesis and glycan chains removal.

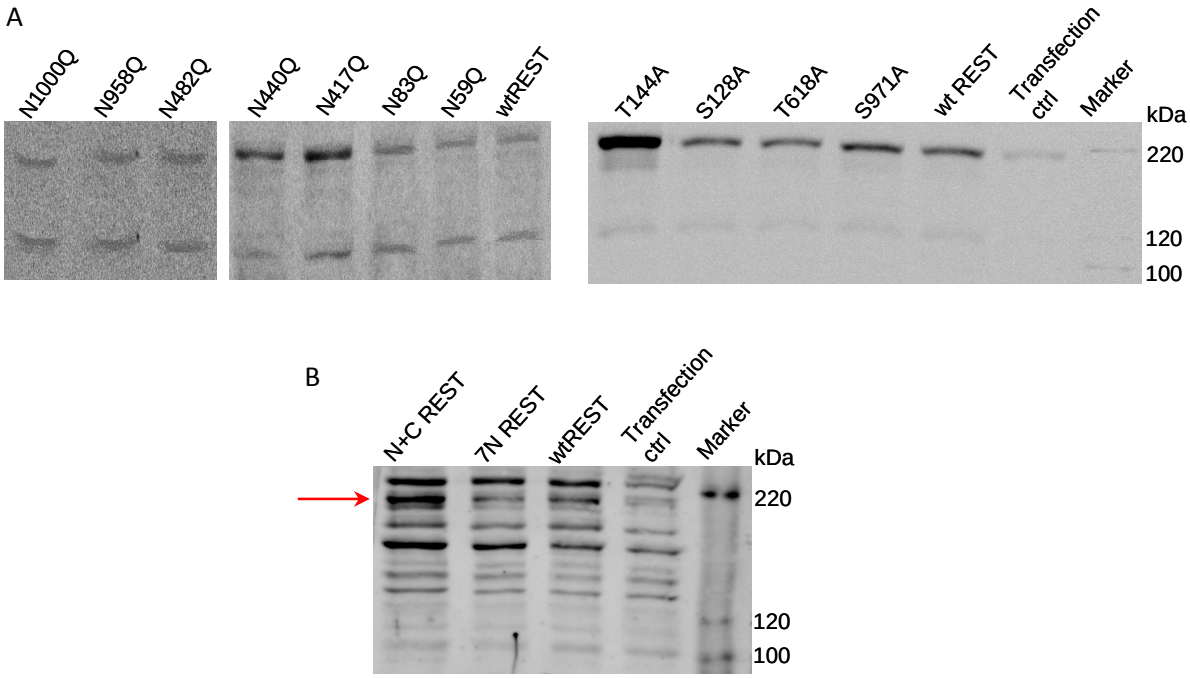

Figure 5.6.1: Western blotting analysis of nuclear protein extracts of the wild-type and mutant FLAG-REST protein expressed in PC-12 cells, using the anti-REST antibody. A) single mutant forms of REST; B) 7N and N+C mutant REST.

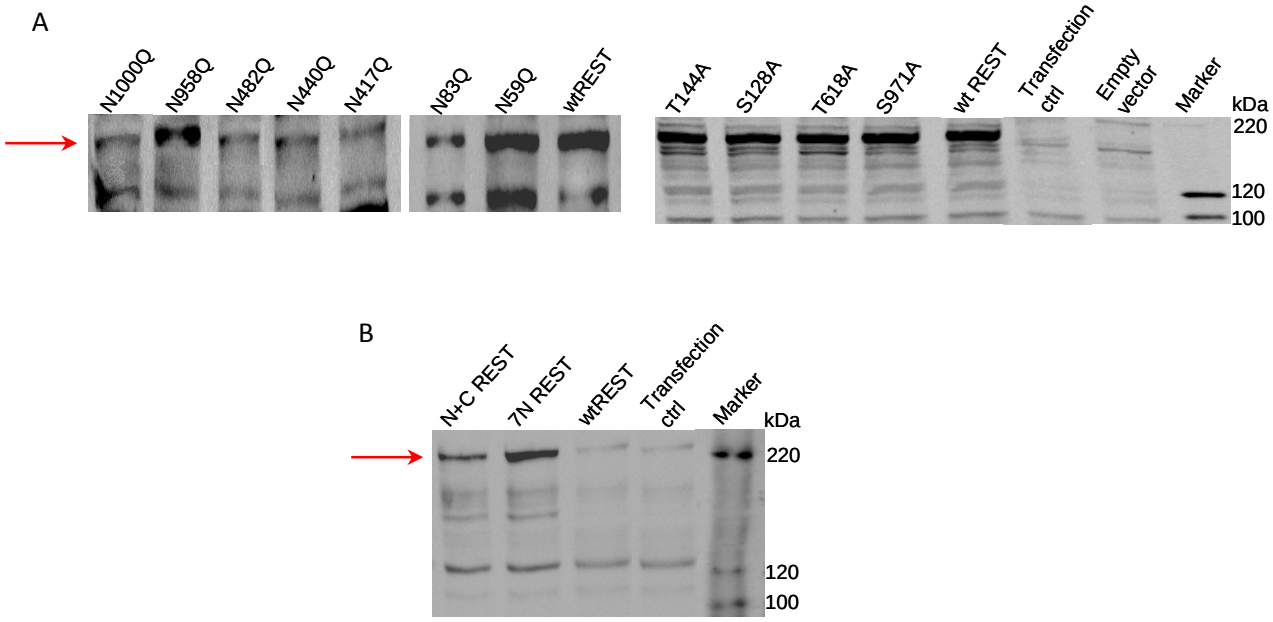

Figure 5.6.2: Western blotting analysis of nuclear protein extracts of the wild-type and mutant FLAG-REST protein expressed in PC-12 cells, using the anti-FLAG antibody. A) single mutant forms of REST; B) 7N and N+C mutant REST.

5.6.2 EVALUATION OF MUTANT REST DNA BINDING

REST mutant forms have been evaluated for their ability to recognize and bind the *RE-1* consensus sequence in comparison to the wild-type protein: the aim was that of testing the possible role of glycan chains in modulating binding affinity of REST to DNA, mimicking this interaction *in vitro* through Electrophoretic Mobility Shifting Assay (EMSA), by using a specific fragment of the human μ-opioid receptor gene promoter, already known to be regulated by REST activity [Bedini et al., Neurochemistry International 2009].

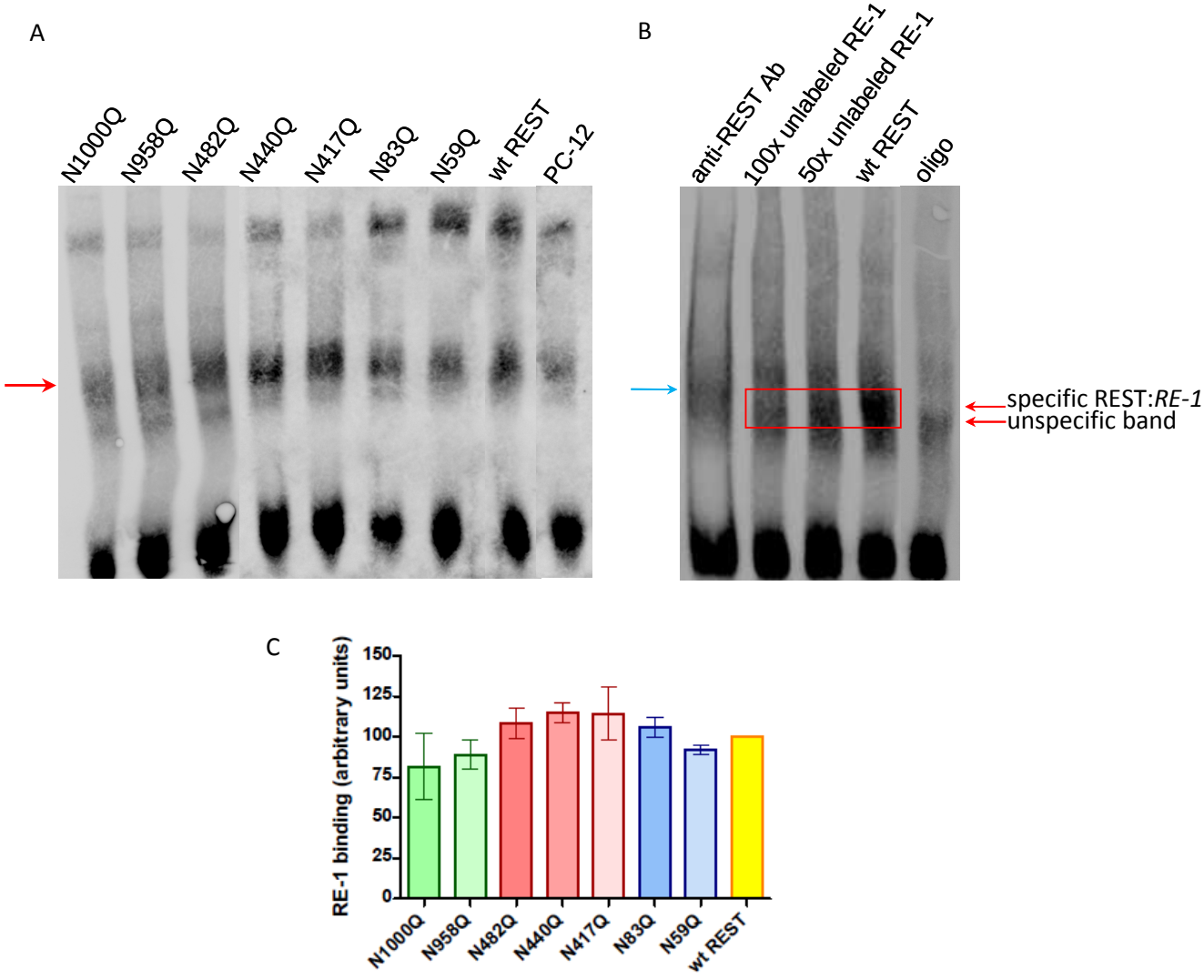

Figure 5.6.3: EMSA analysis of nuclear protein extracts of the wild-type and mutant FLAG-REST protein expressed in PC-12 cells. A) A representative experiment comparingf binding capability of single mutant REST forms and wt REST; B) A representative experiment showing the specificity of the complex, confirmed by the *unlabeled* and *supershift* samples; C) Densitometric analysis of three different experiments.

Nuclear protein extracts of N-linked single mutants, double mutant N59+N83 and 7N mutant transiently transfected in PC-12 cells have been evaluated, in parallel with nuclear protein extracts of native PC-12 cells (*PC-12* sample), showing low levels of endogenous REST, and of cells transfected with the wild-type protein (*wt REST*) (Figure 5.6.3 A).
In addition, to verify that REST binding to the oligonucleotide is selective and to discriminate the specific band corresponding to this interaction, EMSA was run in absence of any protein extract (*oligo*), in presence of an unlabeled oligonucleotide, functioning as a competitor for protein binding to biotinilated oligonucleotide (*unlabeled*), and in presence of a specific anti-REST antibody (Figure 5.6.3 B): in this latter case (*supershift*), the antibody causes a visible reduction in band intensity of about 30%, thus confirming the specificity of the interaction (see the blue arrow). Moreover, as highlighted in the red rectangle, the incubation of the complex with an excess of unlabeled competitor reduces REST band intensity in a way that is proportional to competitor concentration, of about 25% and 37% respectively.
In Figure 5.6.3 A it is reported that all mutant REST bands migrate at the same height as *wt* REST, indicating that no variations in REST:*RE-1* complex occurred; in addition, densitometric analysis of each band intensity revealed that removal of single glycan chains from the transcription factor did not alter its affinity to DNA (Figure 5.6.3 C).

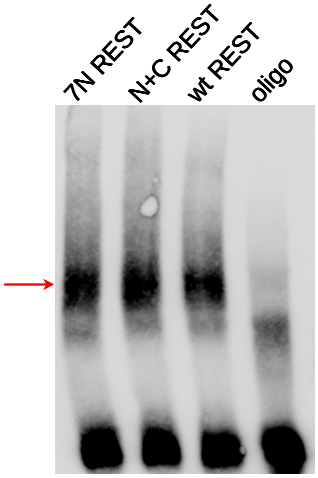

Figure 5.6.4: EMSA analysis of nuclear protein extracts of the wild-type, N+C and 7N mutant FLAG-REST protein expressed in PC-12 cells. The red arrow indicates the band corresponding to REST:*RE-1* complex.

Subsequently the ability of multiple mutant REST to bind DNA consensus sequence was assayed: again, 7N and N+C REST mutant forms did not show any variation in migration nor in band intensity, if compared to *wt* REST (Figure 5.6.4). I then decided to compare the interaction between DNA and mutant or wild-type REST forms expressed in HEK-

293 cells, to evaluate expression and binding affinity of the protein in a different cellular context (Figure 5.6.5 A).

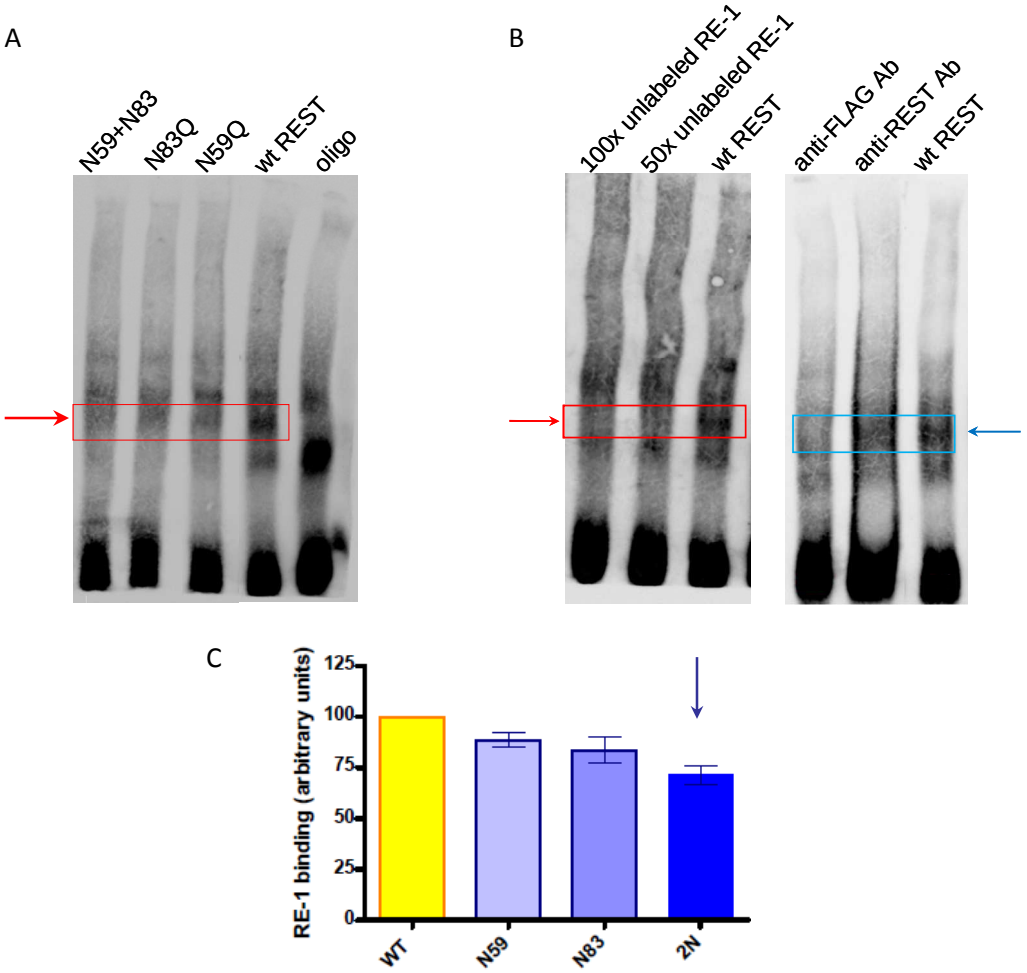

Figure 5.6.5: EMSA analysis of nuclear protein extracts of the wild-type and mutant FLAG-REST protein expressed in HEK-293 cells. A) Representative experiment comparing binding capability between mutant REST forms and wt REST; B) Representative experiment reporting the specificity of the complex has been confirmed by the *unlabeled* and *supershift* samples; C) Densitometric analysis of three different experiments.

To verify the specificity of REST binding to the oligonucleotide EMSA experiments were done in absence of any protein extract (*oligo*), in presence of an unlabeled oligonucleotide (*unlabeled*), and in presence of a specific anti-REST and anti-FLAG antibody (Figure 5.6.5 B): the incubation of the complex with an excess of unlabeled competitor reduces REST band intensity of about 20% and 30%, as the concentration of the competitor increases. The presence of the antibody causes a visible reduction in band intensity, of about 20% with the anti-REST antibody and 30% with the anti-FLAG antibody, although any *supershift* was observed.

The removal of single N-linked glycan chains from asparagine 59 and 83 of REST expressed in HEK-293 cells caused a slight but not statistically significant reduction in the affinity of protein:DNA complex formation, as shown by the densitometric analysis of band intensities (Figure 5.6.5 C); in addition all mutant REST bands migrate at the same height as *wt* REST, indicating that any variation in REST:*RE-1* complex occurred. However the double mutant N59+N83 (*2N* sample) showed a reduction in DNA binding of about 25%, as shown by the blue arrow in Figure 5.6.5 C.

5.6.3 TRANSCRIPTIONAL ACTIVITY OF MUTANT REST

Previous data indicate that neither apparent molecular weight nor DNA recognition and binding are affected by N-glycan chains removal after site-directed mutagenesis. However, I decided to test repressor activity of several mutant REST forms in comparison to the wild-type protein: this could be possible through the use of gene reporter assays, where the expression of a reporter gene, the firefly luciferase gene in this case, is driven by the activity of a fragment of the μ-opioid receptor gene promoter (hMOR), already known to be negatively regulated by REST [Bedini et al., Neurochemistry International 2009].
The assay was set in PC-12 cells, a cellular model chosen because of the minimal expression of endogenous REST protein: in such conditions luciferase transcription is regulated only by the activity of over-expressed recombinant REST, wild-type or mutant. PC-12 cells have been transiently co-transfected with three different expression plasmids: the reporter plasmid encodes for firefly luciferase cloned downstream of the human hMOR promoter (pGL3+LNpMOR); another plasmid encodes for the wild-type or mutant form of recombinant FLAG-REST, whose expression is driven by a strong viral promoter (pCMV-Tag2B+REST); the third plasmid has been chosen as a normalization control, since it encodes for the beta-galactosidase enzyme put under the control of the SV40 constitutive promoter (pSV-βGal).
The transcriptional activity of each construct is expressed as luciferase activity relative to the activity of non-transfected cells (*pGL3 empty* sample). Transfection efficiencies have been normalized to an internal control plasmid pSV-βGal, measuring β-galactosidase activity.

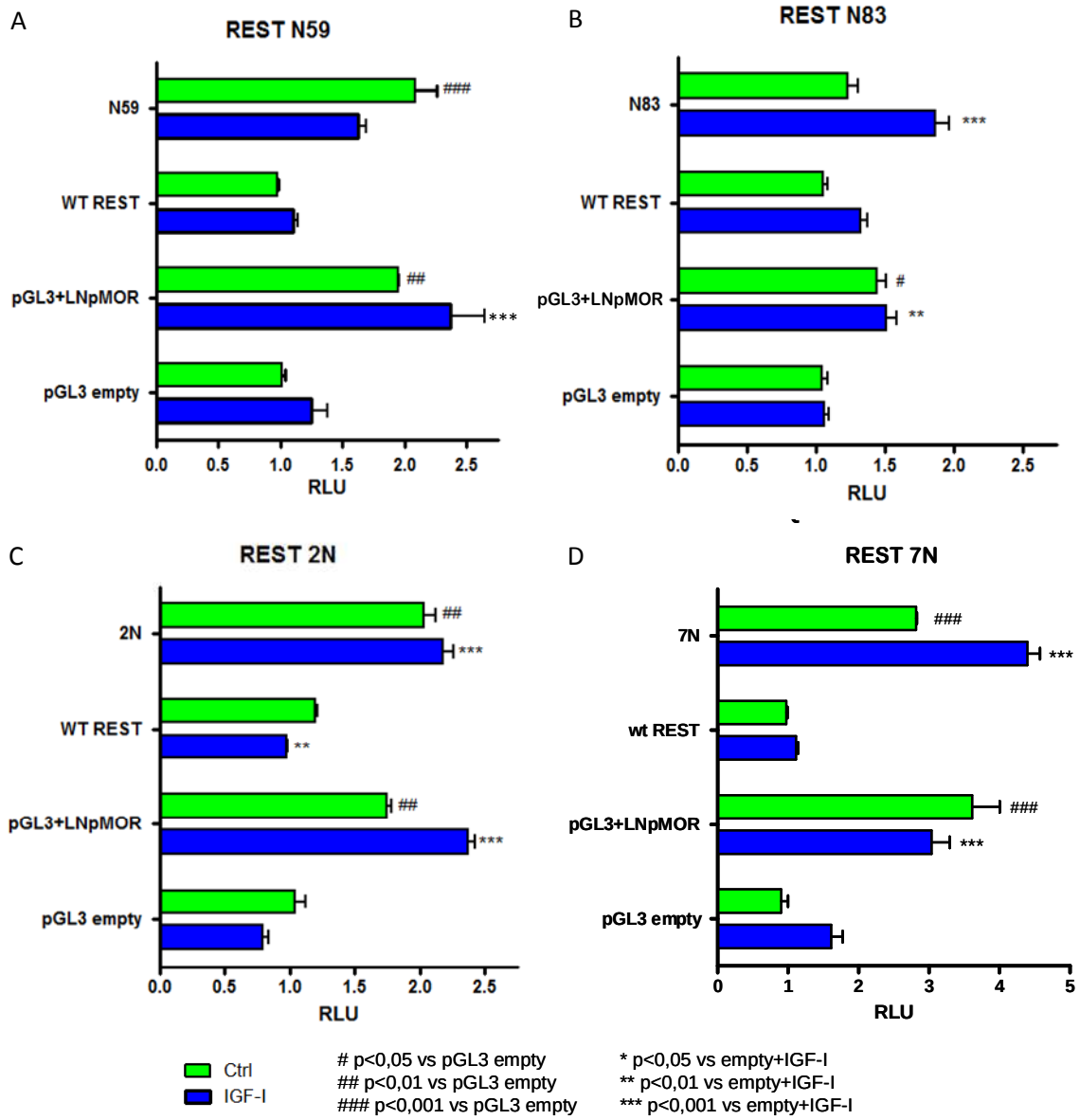

Figure 5.6.6: REST-dependent effects on the transcriptional activity of different hMOR promoter fragments in neuron-like PC-12 cells. Green bars represent control samples; blue bars represent IGF-I treated (10 nM, 24h) samples. Data are presented as mean ± SEM of four independent experiments run in triplicate.

As shown in Figure 5.6.6, in cells transfected with the reporter plasmid alone (*pGL3+LNpMOR* sample) the over-expression of luciferase is translated to an increment of relative luminescence units (RLU), as indicated by the statistically significant increase of the *pGL3+LNpMOR* green bars over the *pGL3 empty* sample.

The addition of wild-type REST-encoding plasmid to the reporter plasmid causes the concurrent over-expression of the transcriptional repressor, which can bind *RE-1* consensus sequence on hMOR promoter and thus it decreases luciferase activity. In fact,

as can be observed in the figure, luminescence of *wt REST* samples is reduced if compared to *pGL3+LNpMOR*, returning to RLU levels of the *pGL3 empty* samples (compare *wt REST* green bars to *pGL3+LNpMOR* and *pGL3 empty* ones). These data confirm the efficacy of the system and allowed to proceed with the analysis using mutant REST forms.

The co-transfection of luciferase reporter plasmid and mutant REST, instead of wild-type REST, showed RLU levels similar to cells transfected with *pGL3+LNpMOR* alone (see *N59, N83, 2N* or *7N* green bars): these results indicate that luminescence activity in presence of mutant REST is reverted to the one observed in samples where REST is absent, thus suggesting that removal of N-linked glycan chains causes loss of the transcriptional repressor function.

Moreover, I checked the effect of Insulin-like Growth Factor I (IGF-I) treatment on co-transfected PC-12 cells. This is a trophic factor that activates STAT3 transcription factor and thus stimulates hMOR transcriptional activity. IGF-I treatment (10 nM, 24 hours) increases luciferase activity and allows an easy detection of any slight variation in luciferase transcription and activity, thus permitting data analysis even in presence of a transcriptional repressor as REST. In cells co-transfected with the reporter plasmid in conjunction with different mutant REST forms and treated with IGF-I, RLU levels vary according to the mutation analyzed (see *N59, N83, 2N* or *7N* blue bars). Removal of N-linked glycan chain from asparagine 59 (Figure 5.6.6.A), for instance, causes a complete loss of repressor activity in control samples, but IGF-I treated samples maintain a slight residual repressor activity: this could be due to the fact that asparagine N59 is not crucial for REST-mediated transcriptional repression and any residual activity of the protein could be enforced by the activation of other transcription factors, mediated by IGF treatment, that could balance REST disfunction.

On the contrary other mutants analyzed, N83, 2N and 7N REST, show a complete loss of function even after IGF-I treatment and beave exactly as the control samples (Figure 5.6.6 B, C and D), suggesting that asparagine 83, which is mutated in each of the forms, may play a more critical role in determining REST transcriptional repressor function.

REST mutant forms analyzed so far have been chosen because of their position in the protein sequence, since asparagines 59 and 83 are known to be involved in the interaction with the co-repressor mSin3A, whose activity is crucial for the repressor activity (see paragrapch 1.2). These experiments confirm that N-glycosylation of the transcription factor REST plays a fundamental role in determining co-repressors recruitment.

6. DISCUSSION

Gene expression is regulated both by positively acting cell-type-specific transcription factors and negative-acting silencers, that can participate in preventing the expression of neuronal genes in non-neuronal cell types or in inappropriate neuronal subtypes.
REST [Chong et al., 1995], also known as NRSF [Schoenherr and Andreson 1995], is a zinc-finger transcription factor that blocks transcription of its target genes by binding to a 21-bp DNA consensus sequence (the repressor element 1, *RE-1*, or neuron-restrictive silencer element, *NRSE*) present in the target genes' regulatory region. REST has been implicated in several processes such as maintenance of embryonic stem cell pluripotency, self-renewal and regulation of mitotic fidelity in non-neuronal cells.
REST gene encodes for a 116-kDa protein that contains a DNA-binding domain composed of eight zinc-fingers, an N-terminal repressor domain and one zinc finger at the C-terminal repressor domain [Ballas and Mandel, 2005]: by acting as an evolving molecular platform to which diverse factors may be recruited, REST promotes dynamic modifications of DNA, histones, nucleosomes and higher-order chromatin codes and helps maintaining genomic stability.
Although REST molecular weight, calculated on aminoacid sequence, is about 116 kDa it was found to migrate on SDS-PAGE as a protein of different molecular weights (see Table 1.1), consistent with the possible presence of post-translational modifications in the protein backbone: it has been recently observed the presence of O-glycan chains in REST, suggesting the importance of glycosylation in modulating transcriptional activity [Lee et al., 2000]. The most common way that diversity is generated in proteins is alteration of the composition of the constituent amino acids; however additional diversity is generated by covalent modification of amino acids by post-translational modifications (PTMs) [Vosseller et al., 2001]. Among post-translational modifications the most common is the complex process of glycosylation, that is the covalent attachment of one or more carbohydrates that occurs during or after the process of protein synthesis: it has been estimated that more than one half of all proteins are glycosylated [Apweiler et al., 1999]. Glycosylation may affect many protein activities, can facilitate folding kinetics and affect the biophysical properties of proteins. The fact that N-glycosylation starts during protein synthesis indicates that it may assist in obtaining the correct fold [Roth et al., 2010]. On the contrary O-linked glycans, which often have lower mass than N-linked structures and higher heterogeneity, have been found to function in protein structure and

stability, immunity, receptor-mediated signaling, protein interactions, expression and processing. Moreover the discovery of O-linked-β-N-acetylglucosamine (O-GlcNAc) disproved the long-held dogma that protein glycosylation is restricted to the luminal compartments of the secretory machinery and to the cell surface: several transcription factors are O-GlcNAcylated, with the modification being involved in nuclear transport, DNA binding and assembly into multimeric complexes [Torres and Hart, 1984]. Alterations in any step of the N- and O-glycosylation process can be involved in the aetiology of many human disorders, such as congenital diseases, diabetes, neuronal degeneration and cancer [Wopereis et al., 2006]. REST, as well, has been found to be involved in the establishment of the same diseases: particularly it showed a double role in the pathogenesis of cancer, since it can behave both as an oncogene and an oncosuppressor [Coulson JM, 2005].

Moving from these observations, in this thesis I have demonstrated the presence of N-linked glycan chains in the aminoacidic backbone of REST and related this post-translational modification to the repressor transcriptional activity.

First of all, I have observed that REST protein migrates on SDS-PAGE with a specific pattern of bands, differing among various cell lines and between the nuclear and the cytosolic form of the protein. Then, I moved my attention to oligosaccharide composition of the glycan chains present in REST: for this purpose I applied a strategy of enzymatic deglycosylation, using different combinations of enzymes capable of specifically remove the glucidic residues from the endogenous protein. I evaluated the presence of N-linked oligosaccharides by the action of *PNGase F,* which removes intact N-linked oligosaccharides from glycoprotein, and *Endoglycosidases F1*, *F2* and *F3*, that cleave the native protein between the two N-acetylglucosamines. Furhtermore, I tested the activity of enzymes involved in the removal of O-linked glycans, such as *Endo-α-N-acetylgalactosaminidase*, *α2-3,6,8,9-neuraminidase, β1,4-galactosidase* and *β-N-acetylglucosaminidase*. Treatment of endogenous REST with *PNGase F* did not cause any alteration in the migration of the protein in the SDS-PAGE, while *Endoglycosidase F1* caused the appearance of an additional band of apparent molecular weight of about 180 kDa; similarly the combination of all the enzymes necessary for O-glycosylation removal (*Endo-α– N-acetylgalactosaminidase* + *Galactosidase* + *Neuraminidase* + *β-N-acetylglucosaminidase*), caused the appearance of one additional band, whose apparent molecular weight is about 180 kDa. Similar results have been observed for the recombinant FLAG-REST protein expressed in HEK-293 cells, with the appearance of the 180 kDa band subsequently to the treatment with the *Endo-α* + *Galactosidase* combination or with *Endoglycosidase F1*, and the increment of intensity of the band of 120 kDa, corresponding to the non-glycosylated

form of REST. Starting from these observations, I hypothesized the presence of both O-linked and N-linked glycan chains: the latter could be buried in the protein core structure and thus unaccessible for the enzyme.

To further confirm the hypothesis that REST is modified by the addition of complex glycan chains in the protein backbone, I evaluated apparent molecular weight of synthetic REST protein produced in a prokaryotic and an eukaryotic cell-free system: in fact only eukaryotic extracts are able to synthesize REST protein with the proper folding and post-translational modifications, while *E.coli* produced-REST migrated in SDS-PAGE with an apparent molecular weight of about 120 kDa, comparable to the non-glycosylated form of endogenous REST.

To finally depict the glycan chains in terms of oligosaccharide composition and linkage analysis, I have subjected recombinant REST protein extracts, obtained from HEK-293 cells stably expressed with the FLAG-REST plasmid construct, to mass spectrometry analysis: for this purpose the protein had to by affinity-purified using an agarose resin conjugated with anti-FLAG antibody, in order to enrich the fully glycosylated form. Purified protein extracts have been analyzed through MALDI/TOF mass spectrometry. REST has been shown to have a very complex mixture of glycans; however, there are some dominant glycans having higher peak intensities, which confirmed the presence of N-linked glycan chains composed of N-acetylgalactosamine, galactose and mannose, as suggested by previous enzymatic deglycosylation analyses.

Subsequently in order to fulfil glycosylation site mapping, purified protein extracts have been analyzed through MS/MS. In particular, to elucidate N-linked sites ^{18}O-labeling method has been used. Twelve potential N-glycosylation sites have been identified: Asn59, Asn83, Asn417, Asn440, Asn482, Asn553, Asn586, Asn914, Asn958, Asn963, Asn1000 and Asn1022, and Asn482 and Asn553 of REST were shown to be glycosylated by LC-MS/MS. However, whether sites Asn482 and Asn553 are truly glycosylated remains questionable, due to the fact that the X corr values of the ^{18}O-site mapping are under confident values; the other potential N-linked glycosylation sites of REST cannot be detected by ^{18}O-site mapping, suggesting that the protein might be highly glycosylated. Thus glycopeptides obtained after tryptic digestion have been O-deglycosyalted through β-elimination in mild conditions followed by Michael addition, in order to remove O-linked glycans, and then analyzed by mass spectrometry: even after O-linked deglycosylation, REST glycopeptides were not confidently detected by LC-MS/MS, most likely because of unappropriated sizes and solubility of their sequences.

Therefore, using post-translational prediction software I have identified the most probable REST residues target of both N- and O-linked glycosylation and mutated them

through site-directed mutagenesis. Only one residue was predicted to be modified with O-GalNAc in REST protein sequence, which is Thr618; on the contrary, eleven O-GlcNAc sites have been predicted in REST. Finally submitting REST protein sequence to the N-glycosylation prediction server, twelve Asn have been obtained as output, with seven of them being highly probable N-glycosylated (Asn59, Asn83, Asn417, Asn440, Asn482, Asn958 and Asn1000). All of these residues, in addition to Thr618 and Ser971, have been subjected to site-directed mutagenesis, singularly or sequentially, in order to convert these residues to aminoacids that are not recognized by *glycosyltransferases*. Site-directed mutagenesis allows the characterization of the complex structure-activity relationship: inserting a point mutation in REST coding sequence it is possible to block the addition of oligosaccharides to target aminoacids, maintaining the protein physico-chemical properties. The mutant plasmids obtained, each encoding for a different mutant form of REST, have been analyzed to confirm the presence of the desired mutation and the absence of any unwanted modification of the coding sequence, through restriction analysis and sequencing the mutated plasmids to finally verify that the right mutation was obtained.

To verify any variation of REST activity associated to glycan chains removal, REST mutant forms have been expressed in different cell lines, of embryonic and neuronal origin: first we analyzed variations in the protein molecular weight and observed that any of the REST mutant forms that were generated, showed alterations of the apparent molecular weight, probably because of the presence of additional glycosylation residues not predicted and thus not considered in the mutagenesis procedure.

Furthermore, to identify the role of glycan chains in modulating binding affinity of REST to DNA, I evaluated the ability of different REST forms to bind the *RE-1* consensus sequence, in presence or absence of glycosylation: any variation in REST:*RE-1* complex nor any statistically significant variation in band intensity occurred, except that the double mutant N59+N83 expressed in HEK-293 cells showed a reduction in DNA binding of about 25%.

Finally, I investigated the role of glycan chains in the modulation of the repressor transcriptional activity through gene reporter assays, testing the expression of the firefly luciferase driven by the activity of a fragment of the μ-opioid receptor gene promoter (hMOR), already known to be regulated by REST [Bedini et al., Neurochemistry International 2010]. The assay was set in PC-12 cells, where the expression of endogenous REST is barely detectable: in such conditions luciferase transcription is regulated only by the activity of over-expressed recombinant REST, wild-type or mutant.

The removal of N-linked glycan chains causes loss of transcriptional repressor function; in cells co-transfected with the reporter plasmid in conjunction with different mutant forms and treated with IGF-I, luciferase levels vary according to the mutation analyzed. Removal of N-linked glycan chain from asparagine 59 showed a slight residual repressor activity in presence of IGF-I: this could be possibly due to the fact that asparagine N59 is not crucial for REST-mediated transcriptional repression and any residual activity of the protein could be enforced by the activation of other transcription factors, mediated by IGF treatment, that can compensate for REST disfunction. On the contrary the other mutants analyzed, each containing N83 mutation, showed the complete loss of function even after IGF-I treatment, suggesting that asparagine 83 may play a more critical role in determining REST transcriptional repressor function.

My thesis has ascertained, for the first time, the presence of complex glycan chians in the transcription factor REST. I have depicted their composition, started defining their position on the protein backbone and identified their possible role in the transcription factor functioning.

REST mutant forms analyzed so far have been chosen because of their position in the protein sequence, since asparagines 59 and 83 are known to be involved in the interaction with the co-repressor mSin3A, whose activity is crucial for the repressor activity. My experiments confirmed that N-glycosylation of the transcription factor REST plays a fundamental role in determining co-repressors recruitment (Figure 6.1).

This study demonstrated the presence of a complex pattern of glycosylation on an nuclear protein, such as a transcription factor, event that seemed to be very unlikely: complex glycosylation is a well known and very common process for proteins that reside in the luminal compartments of the secretory machinery and to the cell surface; however the possible role for the presence of heavy and complex oligosaccharides in a nuclear protein remains to be elucidated.

Several transcription factors have been shown to be modified by the addition of O-linked-β-N-acetylglucosamine (O-GlcNAc), with the modification being involved in nuclear transport, DNA binding and co-repressors recruitment [Torres and Hart, 1984]. One example is Sp1, a ubiquitously expressed transcription factor particularly important for the regulation of TATA-less genes that encode housekeeping proteins, such as growth factors and receptors: it has been shown that Sp1 is multiply O-GlcNAcylated and its activity is regulated by glucose and glucosamine availability, with Sp1 being rapid proteolytically degraded when hypoglycosylated. Thus, O-GlcNAc modification of Sp1 may play a role as a nutritional checkpoint related to growth factor gene transcription: in

the absence of adequate nutrition, Sp1 becomes hypoglycosylated and is rapidly degraded, resulting in reduced general transcription [Han and Kudlow, 1997].

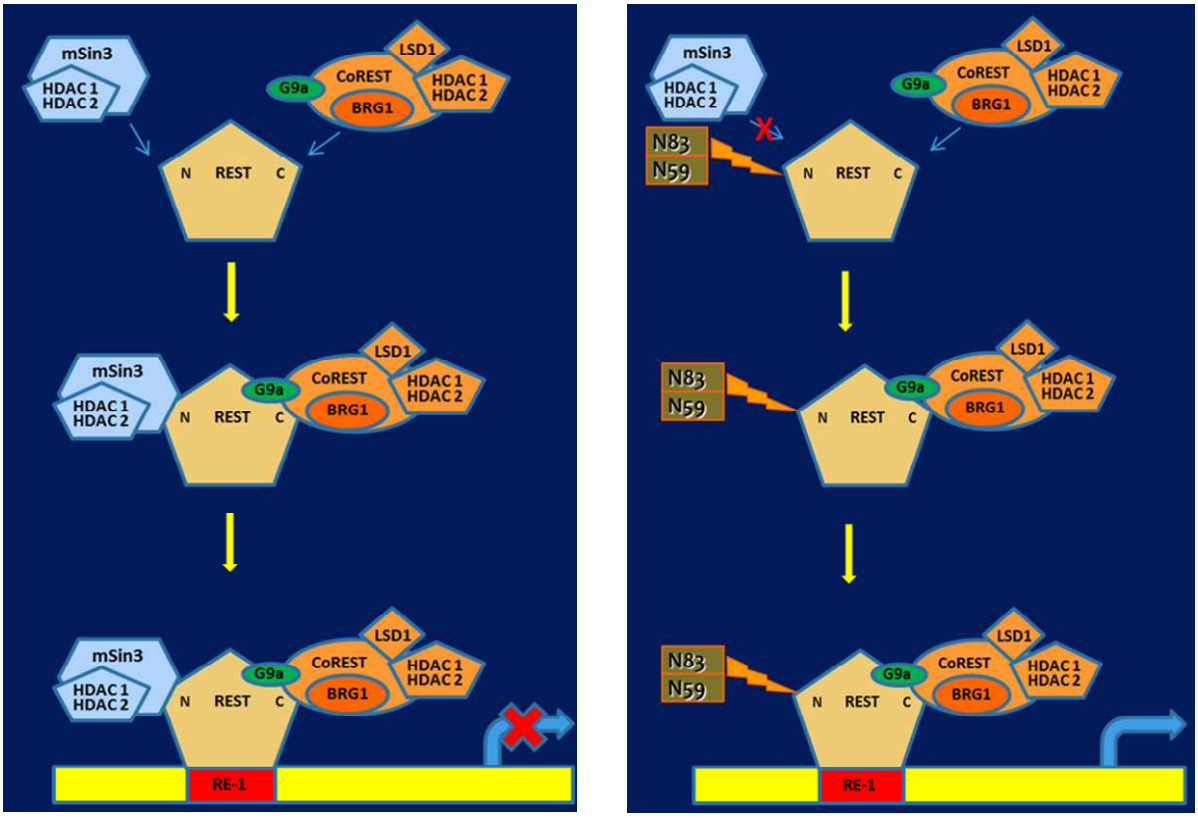

Figure 6.1: Schematic representation of potential role of N-glycan chains present in asparagines 59 and 83 of REST sequence. A) transcriptional repressive state maintained by REST binding on *RE-1* consensus sequence and by recruitment of the mSin3 and CoREST co-repressor complexes; B) loss of REST-mediated transcriptional repression after N59 and N83 N-glycan chains removal, due to the failed recruitment of mSin3 co-repressor complex.

CREB (cyclic AMP-responsive element binding protein), a transcription factor essential for long-term memory, is covalently modified by O-GlcNAc: it ha been demonstrated that the presence of O-GlcNAc impairs the ability of CREB to associate with TAFII130 and to activate transcription. These findings have important implications for the role of O-GlcNAc glycosylation in gene regulation, and provide the first link between O-GlcNAc and information storage processes in the brain [Lamarre-Vincent and Hsieh-Wilson, 2002].

On the contrary, the involvement of N-linked glycosylation in eukaryotic transcription factors' activity is scarcely known and only few transcription factors have been identified to be N-glycosylated so far. The Activating Transcription Factor 6 (ATF6), a member of the ATF/CREB family transcription factors, has two isoforms of 90 kDa (ATF6α) and 110 kDa (ATF6β): the latter contains five evolutionarily conserved N-linked glycosylation

sites and is a key transcriptional repressor of ATF6α, which regulates the strength and duration of ATF6-dependent ER stress response (ERSR) gene induction. It has been recently shown that unglycosylated ATF6β cannot be proteolytically cleaved to p60ATF6β, that actually inhibits ATF6α-mediated ERSR gene activation: thus, unglycosylated ATF6β may directly facilitate the expression of ERSR genes by losing its repressor function to ATF6α. One possible explanation is that different states of ATF6β N-glycosylation may affect protein topology or protein-protein interactions, required for its processing; in addition, loss of N-linked glycan chains from ATF6β could prevent its normal localization and such mutant could not be available for normal proteolysis [Guan et al., 2009].

To fully understand the role of glycosylation in REST repressor activity and correctly define the molecular mechanisms underlying this event, it will be necessary to identify the exact position of glycan chains in REST aminoacidic sequence by mass-spectrometry analysis. Moreover, to further confirm my results, the interaction among REST and co-repressors should be further investigated, possibly through co-immunoprecipitation assays, in order to correlate glycan chains presence to transcriptional repressor activity.

In conclusion, the present study aims to contribute to a more accurate comprehension of the processes responsible for REST activity modulation, which is important for a detailed knowledge of important events like neurogenesis and oncogenesis; moreover, considering the crucial role of glycosylation and transcription factors activity in the aetiology of many disease states, any further knowledge could find important and interesting pharmacological application.

7. REFERENCES

Apweiler R., Hermjakob H., & Sharon N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta-General Subjects* **1473,** 4-8.

Avezov E., Frenkel Z., Ehrlich M., Herscovics A., & Lederkremer G.Z. (2008) Endoplasmic reticulum (ER) mannosidase I is compartmentalized and required for N-glycan trimming to Man(5-6)GlcNAc(2) in glycoprotein ER-associated degradation. *Molecular Biology of the Cell* **19,** 216-225.

Ballas N., Battaglioli E., Atouf F., Andres M.E., Chenoweth J., Anderson M.E., Burger C., Moniwa M., Davie J.R., Bowers W.J., Federoff H.J., Rose D.W., Rosenfeld M.G., Brehm P., & Mandel G. (2001) Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31,** 353-365.

Ballas N. & Mandel G. (2005) The many faces of REST oversee epigenetic programming of neuronal genes. *Current Opinion in Neurobiology* **15,** 500-506.

Ballas N., Grunseich C., Lu D.D., Speh J.C., & Mandel G. (2005) REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121,** 645-657.

Bang A.G. & Goulding M.D. (1996) Regulation of vertebrate neural cell fate by transcription factors. *Current Opinion in Neurobiology* **6,** 25-32.

Bause E., (1983) Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *The Biochemical Journal* **209 (2)**, 331-336.

Bedini A, Baiula M, Spampinato S. (2008) Transcriptional activation of human mu-opioid receptor gene by insulin-like growth factor-I in neuronal cells is modulated by the transcription factor REST. *Journal of Neurochemistry* **105(6)**, 2166-2178.

Bedini A, Baiula M, Carbonari G, Spampinato S. (2010) Transcription factor REST negatively influences the protein kinase C-dependent up-regulation of human mu-opioid receptor gene transcription. *Neurochemistry International* **56 (2)**, 308-317.

Bergsland M., Werme M., Malewicz M., Perlmann T., & Muhr J. (2006) The establishment of neuronal properties is controlled by Sox4 and Sox11. Genes & Development **20,** 3475-3486.

Bruce A.W., Donaldson I.J., Wood I.C., Yerbury S.A., Sadowski M.I., Chapman M., Gottgens B., & Buckley N.J. (2004) Genome-wide analysis of repressor element 1 silencing

transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 10458-10463.

Bruce A.W., Lopez-Contreras A.J., Flicek P., Down T.A., Dhami P., Dillon S.C., Koch C.M., Langford C.F., Dunham I., Andrews R.M., & Vetrie D. (2009) Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level. *Genome Research* **19,** 994-1005.

Buse M.G. (2006) Hexosamines, insulin resistance, and the complications of diabetes: current status. *American Journal of Physiology, Endocrinology and Methabolism* **290(1)**, E1-E8.

Calderone A., Jover T., Noh K.M., Tanaka H., Yokota H., Lin Y., Grooms S.Y., Regis R., Bennett M.V.L., & Zukin R.S. (2003) Ischemic insults derepress the gene silencer REST in neurons destined to die. *Journal of Neuroscience* **23,** 2112-2121.

Cao X.W., Yeo G., Muotri A.R., Kuwabara T., & Gage F.H. (2006) Noncoding RNAs in the mammalian central nervous system. *Annual Review of Neuroscience* **29,** 77-103.

Caramelo J.J. & Parodi A.J. (2008) Getting in and out from calnexin/calreticulin cycles. *Journal of Biological Chemistry* **283,** 10221-10225.

Chen Z.F., Paquette A.J., & Anderson D.J. (1998) NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nature Genetics* **20,** 136-142.

Chong J.H.A., Tapiaramirez J., Kim S., Toledoaral J.J., Zheng Y.C., Boutros M.C., Altshuller Y.M., Frohman M.A., Kraner S.D., & Mandel G. (1995) Rest - A Mammalian Silencer Protein That Restricts Sodium-Channel Gene-Expression to Neurons. *Cell* **80,** 949-957.

Christlet T.H.T. & Veluraja K. (2001) Database analysis of O-glycosylation sites in proteins. *Biophysical Journal* **80,** 952-960.

Ciechanover A, Brundin P. (2003) The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron* **40(2)**, 427-446.

Clausen H. & Bennett E.P. (1996) A family of UDP-GalNAc: Polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology* **6,** 635-646.

Comer F.I. and Hart G.W. (2000) O-Glycosylation of nuclear and cytosolic proteins. Dynamic interplay between O-GlcNAc and O-phosphate. *The Journal of Biological Chemistry* **275(38),** 29179-29182.

Conaco C., Otto S., Han J.J., & Mandel G. (2006) Reciprocal actions of REST and a microRNA promote neuronal identity. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 2422-2427.

Coulson J.M. (2005) Transcriptional regulation: Cancer, neurons and the REST. *Current Biology* **15,** R665-R668.

Cowan J.M., Powers J., & Tischler A.S. (1996) Assignment of the REST gene to 4q12 by fluorescence in situ hybridization. *Genomics* **34,** 260-262.

De Virgilio M., Kitzmuller C., Schwaiger E., Klein M., Kreibich G., & Ivessa N.E. (1999) Degradation of a short-lived glycoprotein from the lumen of the endoplasmic reticulum: The role of N-linked glycans and the unfolded protein response. *Molecular Biology of the Cell* **10,** 4059-4073.

Dell A., Reason A.J., Khoo K.H., Panico M., Mcdowell R.A., Morris H.R. (1994) Mass-Spectrometry of carbohydrate-containing biopolymers. *Guide to Techniques in Glycobiology* **230,** 108-132.

Denic V., Quan E.M., & Weissman J.S. (2006) A luminal surveillance complex that selects misfolded glycoproteins for ER-associated degradation. *Cell* **126,** 349-359.

Dias W.B. & Hart G.W. (2007) O-GlcNAc modification in diabetes and Alzheimer's disease. *Molecular Biosystems* **3,** 766-772.

Elhammer A.P., Poorman R.A., Brown E., Maggiora L.L., Hoogerheide J.G., & Kezdy F.J. (1993) The Specificity of Udp-Galnac - Polypeptide N-Acetylgalactosaminyltransferase As Inferred from A Database of Invivo Substrates and from the Invitro Glycosylation of Proteins and Peptides. *Journal of Biological Chemistry* **268,** 10029-10038.

Finkel A., Iyer S.P., & Sutre G. (2003) Well-abstracted transition systems: application to FIFO automata. *Information and Computation* **181,** 1-31.

Francisco H., Kollins K., Varghis N., Vocadlo D., Vosseller K., & Gallo G. (2009) O-GlcNAc Post-Translational Modifications Regulate the Entry of Neurons Into an Axon Branching Program. *Developmental Neurobiology* **69,** 162-173.

Francisco H., Kollins K., Varghis N., Vocadlo D., Vosseller K., & Gallo G. (2009) O-GlcNAc Post-Translational Modifications Regulate the Entry of Neurons Into an Axon Branching Program. *Developmental Neurobiology* **69,** 162-173.

Fuller G.N., Su X.H., Price R.E., Cohen Z.R., Lang F.F., Sawaya R., & Majumder S. (2005) Many human medulloblastoma tumors overexpress repressor element-1 silencing

transcription (REST) neuron-restrictive silencer factor, which can be functionally countered by REST-VP16. *Molecular Cancer Therapeutics* **4,** 343-349.

Gao Z.L., Ure K., Ding P.G., Nashaat M., Yuan L.R., Ma J., Hammer R.E., & Hsieh J. (2011) The Master Negative Regulator REST/NRSF Controls Adult Neurogenesis by Restraining the Neurogenic Program in Quiescent Stem Cells. *Journal of Neuroscience* **31,** 9772-9786.

Grimes J.A., Nielsen S.J., Battaglioli E., Miska E.A., Speh J.C., Berry D.L., Atouf F., Holdener B.C., Mandel G., & Kouzarides T. (2000) The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex. *Journal of Biological Chemistry* **275,** 9461-9467.

Guan D., Wang H., Li V.E., Xu Y., Yang M., Shen Z. (2009) N-glycosylation of ATF6beta is essential for its proteolytic cleavage and transcriptional repressor function to ATF6alpha. *Journal of Cellular Biochemistry* **108(4)**, 825-831.

Guardavaccaro D., Frescas D., Dorrello N.V., Peschiaroli A., Multani A.S., Cardozo T., Lasorella A., Iavarone A., Chang S., Hernando E., & Pagano M. (2008) Control of chromosome stability by the beta-TrCP-REST-Mad2 axis. *Nature* **452,** 365-U10.

Guinez C., Mir A.M., Dehennaut V., Cacan R., Harduin-Lepers A., Michalski J.C., & Lefebvre T. (2008) Protein ubiquitination is modulated by O-GlcNAc glycosylation. *Faseb Journal* **22,** 2901-2911.

Guinez C., Mir A.M., Dehennaut V., Cacan R., Harduin-Lepers A., Michalski J.C., & Lefebvre T. (2008) Protein ubiquitination is modulated by O-GlcNAc glycosylation. *Faseb Journal* **22,** 2901-2911.

Hakomori S. (1996) Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Research* **56(23),** 5309-5318.

Han I. and Kudlow J.E. (1997) Reduced O-glycosylation of Sp1 is associated with increased proteasome susceptibility. *Molecular and Cellular Biology* **17(5),** 2550-2558.

Hanisch F.G., Jovanovic M., & Peter-Katalinic J. (2001) Glycoprotein identification and localization of O-glycosylation sites by mass spectrometric analysis of deglycosylated/alkylaminylated peptide fragments. *Analytical Biochemistry* **290,** 47-59.

Hanson S.R., Culyba E.K., Hsu T.L., Wong C.H., Kelly J.W., & Powers E.T. (2009) The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 3131-3136.

Hart G.W., Greis K.D., Dong L.Y., Blomberg M.A., Chou T.Y., Jiang M.S., Roquemore E.P., Snow D.M., Kreppel L.K., Cole R.N., et al. (1995) O-linked N-acetylglucosamine: the "yin-yang" of Ser/Thr phosphorylation? Nuclear and cytoplasmic glycosylation. *Advances in experimental medicine and biology* **376,** 115-123.

Hart G.W. (1997) Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annual Review of Biochemistry* **66,** 315-335.

Hart G.W., Housley M.P., & Slawson C. (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446,** 1017-1022.

Hart G.W., Slawson C., Ramirez-Correa G., & Lagerlof O. (2011) Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annual Review of Biochemistry, Vol 80* **80,** 825-858.

Hebert D.N., Garman S.C., & Molinari M. (2005) The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality-control tags. *Trends in Cell Biology* **15,** 364-370.

Hirschberg C.B., Robbins P.W., & Abeijon C. (1998) Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annual Review of Biochemistry* **67,** 49-69.

Holt G.D. & Hart G.W. (1986) The Subcellular-Distribution of Terminal N-Acetylglucosamine Moieties - Localization of A Novel Protein-Saccharide Linkage, O-Linked Glcnac. *Journal of Biological Chemistry* **261,** 8049-8057.

Holt G.D., Snow C.M., Senior A., Haltiwanger R.S., Gerace L., & Hart G.W. (1987) Nuclear-Pore Complex Glycoproteins Contain Cytoplasmically Disposed O-Linked N-Acetylglucosamine. *Journal of Cell Biology* **104,** 1157-1164.

Huang Z., Wu Q.L., Guryanova O.A., Cheng L., Shou W.N., Rich J.N., & Bao S.D. (2011) Deubiquitylase HAUSP stabilizes REST and promotes maintenance of neural progenitor cells. *Nature Cell Biology* **13,** 142-U91.

Jackson S.P. & Tjian R. (1988) O-Glycosylation of Eukaryotic Transcription Factors - Implications for Mechanisms of Transcriptional Regulation. *Cell* **55,** 125-133.

Johnson D.S., Mortazavi A., Myers R.M., & Wold B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316,** 1497-1502.

Jothi R., Cuddapah S., Barski A., Cui K., & Zhao K. (2008) Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Research* **36,** 5221-5231.

Julenius K., Mølgaard A., Gupta R., Brunak S. (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* **15(2)**, 153-164.

Juliandi B., Abematsu M., & Nakashima K. (2010) Chromatin remodeling in neural stem cell differentiation. *Current Opinion in Neurobiology* **20,** 408-415.

Kamemura K. & Hart G.W. (2003) Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: A new paradigm for metabolic control of signal transcluction and transcription. *Progress in Nucleic Acid Research and Molecular Biology, Vol 73* **73,** 107-136.

Kang P., Mechref Y., Klouckova I., Novotny M.V. (2005) Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Communications in Mass Spectrometry* **19(23),** 3421-3428.

Kato K. & Kamiya Y. (2007) Structural views of glycoprotein-fate determination in cells. *Glycobiology* **17,** 1031-1044.

Koenigsberger C., Chicca J.J., Amoureux M.C., Edelman G.M., & Jones F.S. (2000) Differential regulation by multiple promoters of the gene encoding the neuron-restrictive silencer factor. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 2291-2296.

Kreisler A., Strissel P.L., Strick R., Neumann S.B., Schumacher U., & Becker C.M. (2010) Regulation of the NRSF/REST gene by methylation and CREB affects the cellular phenotype of small-cell lung cancer. *Oncogene* **29,** 5828-5838.

Kuwabara T., Hsieh J., Nakashima K., Taira K., & Gage F.H. (2004) A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* **116,** 779-793.

Lamarre-Vincent N. and Hsieh-Wilson L.C. (2003) Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *Journal of the American Chemical Society* **125(22),** 6612-6613.

Laneve P., Gioia U., Andriotto A., Moretti F., Bozzoni I., & Caffarelli E. (2010) A minicircuitry involving REST and CREB controls miR-9-2 expression during human neuronal differentiation. *Nucleic Acids Research* **38,** 6895-6905.

Lazarus B.D., Love D.C., & Hanover J.A. (2009) O-GlcNAc cycling: Implications for neurodegenerative disorders. *International Journal of Biochemistry & Cell Biology* **41,** 2134-2146.

Lee J.H., Shimojo M., Chai Y.G., & Hersh L.B. (2000) Studies on the interaction of REST4 with the cholinergic repressor element-1/neuron restrictive silencer element. *Molecular Brain Research* **80,** 88-98.

Lemonde S., Rogaeva A., & Albert P.R. (2004) Cell type-dependent recruitment of trichostatin A-sensitive repression of the human 5-HT1A receptor gene. *Journal of Neurochemistry* **88,** 857-868.

Lim L.P., Lau N.C., Garrett-Engele P., Grimson A., Schelter J.M., Castle J., Bartel D.P., Linsley P.S., & Johnson J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433,** 769-773.

Link A.J., Eng J., Schieltz D.M., Carmack E., Mize G.J., Morris D.R., Garvik B.M., Yates J. (1999) Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology* **17 (7),** 676-682.

Liu F., Iqbal K., Grundke-Iqbal I., Hart G.W., & Gong C.X. (2004) O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 10804-10809.

Liu F., Shi J.H., Tanimukai H., Gu J.H., Gu J.L., Grundke-Iqbal I., Iqbal K., & Gong C.X. (2009) Reduced O-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer's disease. *Brain* **132,** 1820-1832.

Lowe JB and Marth JD (2003) A genetic approach to Mammalian glycan function. *Annual review of biochemistry* **72,** 643-691.

Lowry O.H., Rosebrough N.J., Farr A.L., Randall R.J (1951) Protein measurement with the Folin phenol reagent. *The Journal of Biological Chemistry* **193(1)**, 265-275.

Lunyak V.V., Prefontaine G.G., & Rosenfeld M.G. (2004) REST and peace for the neuronal-specific transcriptional program. *Gastroenteropancreatic Neuroendocrine Tumor Disease: Molecular and Cell Biological Aspects* **1014,** 110-120.

Lunyak V.V. & Rosenfeld M.G. (2005) No rest for REST: REST/NRSF regulation of neurogenesis. *Cell* **121,** 499-501.

Macdougall I.C. (2002) Optimizing the use of erythropoietic agents: pharmacokinetic and pharmacodynamic considerations. *Nephrol Dial Transplant* **17 Suppl 5,** 66-70.

Manning G., Whyte D.B., Martinez R., Hunter T., & Sudarsanam S. (2002) The protein kinase complement of the human genome. *Science* **298,** 1912-+.

Mitra N., Sinha S., Ramya T.N.C., & Surolia A. (2006) N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends in Biochemical Sciences* **31,** 156-163.

Moremen K.W. & Molinari M. (2006) N-linked glycan recognition and processing: the molecular basis of endoplasmic reticulum quality control. *Current Opinion in Structural Biology* **16,** 592-599.

Naruse Y., Aoki T., Kojima T., & Mori N. (1999) Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 13691-13696.

Ooi L., Belyaev N.D., Miyake K., Wood I.C., & Buckley N.J. (2006) BRG1 chromatin remodeling activity is required for efficient chromatin binding by repressor element 1-silencing transcription factor (REST) and facilitates REST-mediated repression. *Journal of Biological Chemistry* **281,** 38974-38980.

Opat AS, van Vliet C, Gleeson PA. (2001) Trafficking and localisation of resident Golgi glycosylation enzymes. *Biochimie.* **83(8),** 763-773.

Otto S.J., McCorkle S.R., Hover J., Conaco C., Han J.J., Impey S., Yochum G.S., Dunn J.J., Goodman R.H., & Mandel G. (2007) A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions. *Journal of Neuroscience* **27,** 6729-6739.

Palm K., Belluardo N., Metsis M., & Timmusk T. (1998) Neuronal expression of zinc finger transcription factor REST/NRSF/XRB gene. *Journal of Neuroscience* **18,** 1280-1296.

Palm K., Metsis M., & Timmusk T. (1999) Neuron-specific splicing of zinc finger transcription factor REST/NRSF/XBR is frequent in neuroblastomas and conserved in human, mouse and rat. *Molecular Brain Research* **72,** 30-39.

Pan G.J., Li J., Zhou Y.L., Zheng H., & Pei D.Q. (2006) A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *Faseb Journal* **20,** 1730.

Paquette A.J., Perez S.E., Anderson D.J. (2000) Constitutive expression of the neuron-restrictive silencer factor NRSF/REST in differentiating neurons disrupts neuronal gene expression and causes axon pathfinding errors in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **97(22),** 12318-12323.

Parodi A.J. (2000) Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochemical Journal* **348,** 1-13.

Petrescu A.J., Milac A.L., Petrescu S.M., Dwek R.A., & Wormald M.R. (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14,** 103-114.

Ravanpay A.C., Hansen S.J., & Olson J.M. (2010) Transcriptional inhibition of REST by NeuroD2 during neuronal differentiation. *Molecular and Cellular Neuroscience* **44,** 178-189.

Rexach JE, Clark PM, Hsieh-Wilson LC. (2008) Chemical approaches to understanding O-GlcNAc glycosylation in the brain. *Nature Chemical Biology* **4(2)**, 97-106.

Roopra A., Sharling L., Wood I.C., Briggs T., Bachfischer U., Paquette A.J., & Buckley N.J. (2000) Transcriptional repression by neuron-restrictive silencer factor is mediated via the SIN3-histone deacetylase complex. *Molecular and Cellular Biology* **20,** 2147-2157.

Roth J., Zuber C., Park S., Jang I., Lee Y., Kysela K.G., Le Fourn V., Santimaria R., Guhl B., & Cho J.W. (2010) Protein N-glycosylation, protein folding, and protein quality control. *Molecules and Cells* **30,** 497-506.

Sakabe K., Wang Z.H., & Hart G.W. (2010) beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 19915-19920.

Schoenherr C.J. & Anderson D.J. (1995) Silencing Is Golden - Negative Regulation in the Control of Neuronal Gene-Transcription. *Current Opinion in Neurobiology* **5,** 566-571.

Sethuraman N, Stadheim TA (2006) Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology* **17(4)**, 341-346.

Shental-Bechor D. & Levy Y. (2009) Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Current Opinion in Structural Biology* **19,** 524-533.

Shi Y.H., Zhao X.Y., Hsieh J., Wichterle H., Impey S., Banerjee S., Neveu P., & Kosik K.S. (2010) MicroRNA Regulation of Neural Stem Cells and Neurogenesis. *Journal of Neuroscience* **30,** 14931-14936.

Shimojo M., Paquette A.J., Anderson D.J., & Hersh L.B. (1999) Protein kinase A regulates cholinergic gene expression in PC12 cells: REST4 silences the silencing activity of neuron-restrictive silencer factor/REST. *Molecular and Cellular Biology* **19,** 6788-6795.

Shimojo M. & Hersh L.B. (2003) REST/NRSF-interacting LIM domain protein, a putative nuclear translocation receptor. *Molecular and Cellular Biology* **23,** 9025-9031.

Singh S.K., Kagalwala M.N., Parker-Thornburg J., Adams H., & Majumder S. (2008) REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453,** 223-U11.

Slawson C., Copeland R.J., & Hart G.W. (2010) O-GIcNAc signaling: a metabolic link between diabetes and cancer? *Trends in Biochemical Sciences* **35,** 547-555.

Spiro R.G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12,** 43R-56R.

Su X.H., Kameoka S., Lentz S., & Majumder S. (2004) Activation of REST/NRSF target genes in neural stem cells is sufficient to cause neuronal differentiation. *Molecular and Cellular Biology* **24,** 8018-8025.

Su X.H., Gopalakrishnan V., Stearns D., Aldape K., Lang F.F., Fuller G., Snyder E., Eberhart C.G., & Majumder S. (2006) Abnormal expression of REST/NRSF and Myc in neural stem/progenitor cells causes cerebellar tumors by blocking neuronal differentiation. *Molecular and Cellular Biology* **26,** 1666-1678.

Suzuki T. & Lennarz W.J. (2003) Hypothesis: a glycoprotein-degradation complex formed by protein-protein interaction involves cytoplasmic peptide:N-glycanase. *Biochemical and Biophysical Research Communications* **302,** 1-5.

Torres CR and Hart GW (1984) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *The Journal of Biological Chemistry* **259 (5),** 3308-3317.

Trombetta E.S. & Parodi A.J. (2003) Quality control and protein folding in the secretory pathway. *Annual Review of Cell and Developmental Biology* **19,** 649-676.

Vosseller K, Wells L, Hart GW. (2001) Nucleocytoplasmic O-glycosylation: O-GlcNAc and functional proteomics. *Biochimie* **83(7),** 575-81.

Walsh G. (2005) Biopharmaceuticals: recent approvals and likely directions. *Trends in Biotechnology* **23**, 553-558.

Wang C.Q., Eufemi M., Turano C., & Giartosio A. (1996) Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **35,** 7299-7307.

Weissman A.M. (2008) How much REST is enough? *Cancer Cell* **13,** 381-383.

Wells L, Whelan SA, Hart GW. (2003) O-GlcNAc: a regulatory post-translational modification. *Biochemical and Biophysical Research Communications* **302(3)**, 435-441.

Westbrook T.F., Martin E.S., Schlabach M.R., Leng Y.M., Liang A.C., Feng B., Zhao J.J., Roberts T.M., Mandel G., Hannon G.J., DePinho R.A., Chin L., & Elledge S.J. (2005) A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121,** 837-848.

Westbrook T.F., Hu G., Ang X.L.L., Mulligan P., Pavlova N.N., Liang A., Leng Y.M., Maehr R., Shi Y., Harper J.W., & Elledge S.J. (2008) SCF beta-TRCP controls oncogenic transformation and neural differentiation through REST degradation. *Nature* **452,** 370-U11.

Wopereis S, Lefeber DJ, Morava E, Wevers RA. (2006) Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clinical Chemistry* **52(4),** 574-600.

Yang X.Y., Zhang F.X., & Kudlow J.E. (2002) Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: Coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110,** 69-80.

Yang X.Y., Zhang F.X., & Kudlow J.E. (2002) Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: Coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110,** 69-80.

Yu H.B., Johnson R., Kunarso G., & Stanton L.W. (2011) Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells. *Genome Research* **21,** 1284-1293.

Zhang P.S., Lathia J.D., Flavahan W.A., Riche J.N., & Mattson M.P. (2009) Squelching glioblastoma stem cells by targeting REST for proteasomal degradation. *Trends in Neurosciences* **32,** 559-565.

Zheng L., Baumann U., Reymond J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acid Research* **32(14)**.

Zhou Q., Chipperfield H., Melton D.A., & Wong W.H. (2007) A gene regulatory network in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 16438-16443.

Zuccato C., Tartari M., Crotti A., Goffredo D., Valenza M., Conti L., Cataudella T., Leavitt B.R., Hayden M.R., Timmusk T., Rigamonti D., & Cattaneo E. (2003) Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature Genetics* **35,** 76-83.