MUCIN GENES (MUC1 AND MUC6) POLYMORPHISM AND GASTRIC CARCINOMA RISK

POLIMORFISMO DE GENES DAS MUCINAS (MUC1 E MUC6) E RISCO DE CARCINOMA GÁSTRICO

FILIPA ABREU GOMES DE CARVALHO

PORTO, 1999

DISSERTAÇÃO DE CANDIDATURA AO GRAU DE DOUTOR APRESENTADA À FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO.

Artigo 48º, § 3º - A Faculdade não responde pelas doutrinas expendidas na dissertação. (Regulamento da Faculdade de Medicina do Porto – Decreto de Lei nº19 337, de 29 de Janeiro de 1931).

Ao abrigo do Art 8º do Decreto-Lei nº 388/70 fazem parte integrante desta dissertação os seguintes trabalhos já publicados ou em publicação.

Carvalho F, Seruca R, David L, Amorim A, Seixas M, Bennett E, Clausen H, Sobrinho-Simões M: *MUC1* gene polymorphism and gastric cancer – an epidemiological study. *Glycoconjugate Journal* **14**: 107-111, 1997.

Garcia E, Carvalho F, Amorim A, David L: *MUC6* gene polymorphism in healthy individuals and in gastric cancer patients from Northern Portugal. *Cancer Epidemiology Biomarkers & Prevention* **6**: 1071-1074, 1997.

Carvalho F, Peixoto A, Steffensen R, Amorim A, David L, Sobrinho-Simões M: *MUC1* gene polymorphism does not contribute to explain the different incidences of gastric cancer in Portugal and Denmark. Submetido para publicação.

Carvalho F, David L, Aubert JP, Reis CA, Gärtner F, Peixoto A, Alves P, Sobrinho-Simões M: Mucins and mucin-associated carbohydrate antigens expression in gastric carcinoma cell lines. Submetido para publicação.

Em cumprimento do disposto no referido Decreto-Lei declara que participou activamente na recolha e estudo do material incluído em todos os trabalhos, tendo redigido os textos com a activa colaboração dos outros autores.
A dissertação inclui também resultados ainda não enviados para publicação.

AGRADECIMENTOS

À Professora Leonor David, orientadora deste trabalho, o meu mais sincero agradecimento por todo o estímulo, apoio e orientação científica e pedagógica.

Ao Professor Manuel Sobrinho-Simões, co-orientador deste trabalho, pelo incentivo e oportunidade que contribuíram de forma decisiva para a minha formação científica, bem como pela orientação e discussão dos trabalhos, o meu muito obrigada.

Agradeço ao Professor Victor Faria, Director do Serviço de Anatomia Patológica da Faculdade de Medicina da Universidade do Porto.

Aos co-autores das publicações incluídas nesta dissertação agradeço a sua colaboração.

I would like to thank Dr. Joyce-Taylor Papadimitriou, from Imperial Cancer Research Fund, and Prof. Nikolaus Blin, from University of Tübingen, for the oportunity of performing training periods in their labs.

Ao Professor António Amorim, à Professora Maria de S. José Nascimento e ao Dr. Filipe Sansonetty agradeço a sua colaboração em diversas fases deste trabalho.

A todos os amigos e colegas do IPATIMUP um sincero obrigada pelo apoio e colaboração.

À Fundação para a Ciência e a Tecnologia e ao Programa PRAXIS XXI o meu agradecimento pela Bolsa de Doutoramento bem como pelo financiamento de estágios e participação em Congressos.

Ao Armando, aos meus Pais e à Andreia, por tudo...

TABLE OF CONTENTS

Abbreviations 13

Introduction 14

General characteristics and importance of mucins 15

Human mucin genes and proteins 17

Non-human mucin genes and proteins 29

Mucins and infection: the case of *Helicobacter pylori* 32

Gastric mucins and gastric carcinoma risk 34

Aims 36

Materials and Methods 39

Results 47

I – Polymorphism of *MUC1* and *MUC6* genes and gastric carcinoma risk 48

Ia – *MUC1* VNTR polymorphism 48

Ib – *MUC1* CA repeat polymorphism 51

Ic – *MUC6* VNTR polymorphism 53

II – *MUC1* gene polymorphism and *Helicobacter pylori* infection 57

IIa – *MUC1* gene polymorphism in Portugal and Denmark: relation with *Helicobacter pylori* infection 57

IIb – *MUC1* gene polymorphism in dyspeptic patients from Northern Portugal: relation with *Helicobacter pylori* infection and other pathological parameters 63

III – Expression of mucins and *Helicobacter pylori* adherence to gastric carcinoma cell lines 67

IIIa – Characterization of the expression of mucins and carbohydrate side chains in gastric carcinoma cell lines 67

IIIb – *In vitro* adherence assay of *Helicobacter pylori* to gastric carcinoma cell lines 73

Discussion 75

References 89

Sumário e Conclusões 109

Papers I - IV

ABBREVIATIONS

bp – Base pair(s)

BSM – Bovine submaxillary mucin

Cys – Cystein

Da – Dalton(s)

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetracetic acid

EGF – Epidermal growth factor

FITC – Fluorescein isothiocyanate

HCl – Hydrochloric acid

Kbp – Kilobase pair(s)

KDa – Kilodalton(s)

MG1 – High molecular weight salivary mucin

MG2 – Low molecular weight salivary mucin

nm – Namometers

PBS – Phosphate buffered saline

PCR – Polymerase chain reaction

Pro – Proline

PSM – Porcine submaxillary mucin

RNA – Ribonucleic acid

rpm – Rotations per minute

RT – Room temperature

SDS – Sodium dodecyl sulphate

Ser – Serine

Thr – Threonine

VNTR – Variable number of tandem repeats

vWF – von Willebrand factor

INTRODUCTION

General characteristics and importance of mucins

Mucins are glycoproteins characterized by a high molecular mass (>$2x10^6$ Da), high carbohydrate content (50-80% of their mass) and high heterogeneity involving both the apomucin (also called protein backbone or mucin core protein) and the oligosaccharide chains. The majority of glycans on mucins are attached in O-linkage to serine and/or threonine via the linkage sugar N-acetylgalactosamine.

Mucin-type glycoproteins are widely distributed in nature, but our knowledge on this is limited by the scarcity of data on record in the literature for non-mammalian species.

Although mucins were originally defined as the highly glycosylated glycoproteins that form the major protein components of mucus and that are secreted by specialized epithelial cells, the term mucin is now rather loosely used to designate a heterogeneous family of O-linked glycoproteins.

Due to the integrated results of genetic, immunological and biochemical studies it is now possible to classify mucins into two groups: epithelial and non-epithelial mucins.

The epithelial mucins can be divided into two main subgroups: membrane-bound mucins, with a hydrophobic membrane-spanning domain, that do not form oligomeric structures and have specific biological functions, and secreted mucins. The secreted mucins can be divided into two further subgroups: gel-forming and non-gel-forming mucins. Secretory gel-forming mucins constitute the viscous mucus that covers most of the mucosal surfaces of respiratory, gastrointestinal and genitourinary tracts. This mucus has various and important physiological functions, including maintenance of viscosity and protection from pathogenic microorganisms, mechanical and chemical damage, proteolytic enzymes and pH extremes (For review see Taylor-Papadimitriou and Gendler, 1988; Hilkens, 1988; Ho and Kim, 1991; Devine and McKenzie, 1992; Lesuffleur *et al*, 1994; Kim and Gum, 1995; Gendler and Spicer, 1995; Kim *et al*, 1996).

Non-epithelial mucins are found in non-epithelial tissues serving as ligands for selectins and being involved in lymphocyte trafficking (Shimizu and Shaw, 1993). Non-epithelial mucins will not be discussed in the present thesis, and epithelial mucins will be "generically" designated as mucins.

At least nine human and several non-human mucin genes have been described in the literature, but to date only some of them have been completely sequenced and characterized.

As major components of the mucus coat, mucins are thought to play a mechanical protective role on mucosal surfaces. They may also have more specific functions, such

providing attachment sites for commensal or pathogenic microorganisms. In endothelial cells non-epithelial mucins may carry attachment sites for targeting leucocytes (For review see Taylor-Papadimitriou and Gendler, 1988; Hilkens *et al*, 1988; Ho and Kim, 1991; Devine and McKenzie, 1992; Shimizu and Shaw, 1993; Lesuffleur *et al*, 1994; Kim and Gum, 1995; Gendler and Spicer, 1995; Kim *et al*, 1996).

The ability of secreted mucins to form viscous gel layers depends primarily on the properties of mucin molecules: the hydrophilic carbohydrate side chains are able to bind water molecules and mucins form large polymers through disulphide bridges between the Cys residues in the non-repetitive regions of the molecules (Strous and Dekker, 1992).

Glycosylation is thought to be essencial for the role played by mucins. The same core protein can be glycosylated differently in different tissues; the same holds true when carcinomas are compared with their normal tissue counterparts (Ho and Kim, 1991; Devine and McKenzie, 1992). Many antigenic differences between carcinomas and adjacent normal epithelial tissues are due to either an increased synthesis/secretion of mucins and/or to altered glycosylation in the neoplastic cells. The most common mechanisms responsible for abnormal glycosylation are abnormal expression and/or abnormal activity of glycosyltransferases, and competition between different enzymes (Paulson and Colley, 1989; Brockhausen, 1993).

All human mucin genes have a similar structure and were named according to the Human Genome Mapping conventions by the gene symbol *MUC*, followed by a number reflecting the order in which the genes were cloned. Most of the *MUC* genes were identified using partial cDNA clones corresponding to the repetitive sequences (tandem repeat regions) and were distinguished on the basis of their chromosomal localization and pattern of tissue expression. Two general features of these human mucin genes (with exception of *MUC5B* gene) are the presence of tandem repeat sequences that encode Ser/Thr-rich domains, the potential O-glycosylation sites, and the genetic polymorphism reflecting the variable number of tandem repeats (VNTR) .

To date nine human mucin genes *(MUC1, MUC2, MUC3, MUC4, MUC5AC, MUC5B, MUC6, MUC7 and MUC8*) have been reported (Gendler *et al*, 1990a; Lan *et al*, 1990; Lancaster *et al*, 1990; Gum *et al*, 1994; Gum *et al*, 1990; Porchet *et al*, 1991; Aubert *et al*, 1991; Guyonnet-Duperat *et al*, 1995; Dufosse *et al*, 1993; Toribara *et al*, 1993; Bobek *et al*, 1993; Shankar *et al*, 1997). Some of the characteristics of these mucins are summarized in Table 1.

MUC1

Shimizu and Yamauchi (1982) originally purified a high molecular weight glycoprotein from human milk and found it had the properties of a mucin. Burchell *et al* (1983) and Griffiths *et al* (1987) identified the same glycoprotein as being the antigenic component recognized by a range of monoclonal antibodies (Taylor-Papadimitriou *et al*, 1981; Ashall *et al*, 1982; Cordell *et al*, 1985; Dekretser *et al*, 1985; Burchell *et al*, 1987).

cDNA clones for the mammary gland or milk mucin, now designated MUC1, have been obtained by several groups (Gendler *et al*, 1987; Siddiqui *et al*, 1988; Gendler *et al*, 1988; Abe *et al*, 1989; Ligtenberg *et al*, 1990; Gendler et al, 1990a; Wreschner *et al*, 1990). MUC1 has also been designated as polymorphic epithelial mucin (PEM) (Gendler *et al*, 1988; Gendler et al, 1990a), peanut-lectin binding urinary mucin (PUM) (Karlsson *et al*, 1983), episialin or MAM-6 (Ligtenberg *et al*, 1990), DF3 antigen (Siddiqui *et al*, 1988; Abe *et al*, 1989), H23 antigen (Wreschner *et al*, 1990), PAS-O (Shimizu and Yamauchi, 1982), epithelial membrane antigen (EMA) (Cordell *et al*, 1985), Ca1 (Bramwell *et al*, 1983), and NCRC11 antigen (Price *et al*, 1985).

The *MUC1* gene was the first human mucin gene for which the full length cDNA and genomic sequences were determined (Gendler *et al*, 1988; Gendler et al, 1990a; Lan *et*

Table 1. Some characteristics of human mucins.

Mucin	Source	cDNA sequence	Membrane anchor	Tandem repeat (a.a.)	Cystein residues	VNTR polymorphism	Chromosomal localization
MUC1	Mammary Pancreatic	Complete	Yes	20	C-terminal region (intracellular domain)	Yes	1q21-24
MUC2	Intestinal Tracheobronchial	Complete	No (gel-forming)	23	** C-terminal and N-terminal region	Yes	**11p15.5**
MUC3	Intestinal	Partial	?	17	*** C-terminal region	Yes	7q22
MUC4	Tracheobronchial	Partial	Yes	16	*** N-terminal region	Yes	3q29
MUC5AC	Tracheobronchial Gastric	Partial	No (gel-forming)	8	** C-terminal and N-terminal region	Yes	**11p15.5**
MUC5B	Tracheobronchial	Partial	No (gel-forming)	*29	** C-terminal region and within each super-repeat unit	No	**11p15.5**
MUC6	Gastric	Partial	No (gel-forming)	169	C-terminal region	Yes	**11p15.5**
MUC7	Salivary	Complete	No (non gel-forming)	23	N-terminal region	Yes	4q13-21
MUC8	Tracheobronchial	Partial	No (non gel-forming)	13 and 41	One Cys residue per repeat unit	?	12q24.3

* An irregular repeat of 29 a.a. is included within each super-repeat unit;
** Homology to D-domains from vWF;
*** Homology to EGF-like structural motifs.

al, 1990; Lancaster *et al*, 1990). The first cDNA clones were obtained by stripping the carbohydrate from purified milk mucin, raising antibodies to the core protein and using these antibodies to screen a λgt11 cDNA library prepared from breast cancer cell lines (Gendler *et al*, 1987). Sequencing of partial cDNA clones revealed that the core protein contained a large domain of a variable number of 20 amino acids repeat units (Gendler *et al*, 1988). The complete sequence of the *MUC1* gene has been described (Gendler *et al*, 1990a; Lan *et al*, 1990; Lancaster *et al*, 1990) and the deduced amino acid sequence has the features of a type I transmembrane protein with distinct domains: an amino terminal region with a putative signal peptide and degenerate tandem repeats, a large domain, which is the tandem repeat region, and a carboxyl terminal region with degenerate repeats, a unique sequence, a transmembrane region of 31 amino acids and a C-terminus cytoplasmatic tail of 69 amino acids. This structure results in a rod like molecule (Jentoff, 1990) because of extensive glycosylation, that extends far above the glycocalyx of the cell (Hilkens *et al*, 1992).

The amino acid composition of MUC1 is typical of a mucin, with serine, threonine, glycine and alanine accounting for more than 60% of the total amino acids. The five potential O-glycosylation sites present in the tandem repeat sequence of MUC1 are now possible to be all glycosylated *in vivo* (Muller *et al*, 1997). There are five potential N-glycosylation sites in the extracellular domain C-terminally to the tandem repeat region (Gendler et al, 1990a).

Although the MUC1 core protein expressed by different tissues is identical, the mature glycosylated forms differ from tissue to tissue (Shimizu and Yamauchi, 1982; Lan *et al*, 1987). This is probably dependent upon different glycosyltransferases and glycosidases expression or different availability of substrates from tissue to tissue (Clausen *et al*, 1992).

MUC1 is a transmembrane protein, although a soluble form exists in tissue culture supernatants and body fluids (Burchell *et al*, 1984; Pemberton *et al*, 1992; Boshell *et al*, 1992). The soluble form, which lacks the cytoplasmatic tail, can be released from the membrane by the action of a protease (Boshell *et al*, 1992). Recycling of MUC1 from the membrane to the *trans*-Golgi network followed by further sialylation and a cleavage event that removes the transmembrane and cytoplasmatic tail portions of the molecule has been reported (Ligtenberg *et al*, 1992a; Litvinov *et al*, 1993).

The *MUC1* gene is unusually compact, spanning approximately 4 to 7 kilobase pairs (Kbp) of genomic DNA (the size being variable due to the variable number of tandem repeat in exon 2), and contains seven exons and six short introns (Lancaster *et al*, 1990). Two splice variants of unknown significance have been described that result from the use of two alternative splice acceptor sites for exon 2 (Ligtenberg *et al*, 1990; Lancaster *et al*, 1990; Wreschner *et al*, 1990). These variants differed by 27 bp, which

result in an altered signal sequence and, hence, a different amino terminus. The splice recognition site is allele dependent and is based on a single adenine/guanine (A/G) nucleotide difference in exon 2, eight nucleotides downstream of the second splice acceptor site (Ligtenberg *et al*, 1990; Ligtenberg *et al*, 1991; Pratt *et al*, 1996).

The 60 bp tandem repeats found in exon 2 vary in number and are extremely homogeneous in sequence, making this region an expressed VNTR locus. The VNTR accounts for the extensive polymorphism observed at DNA, RNA and protein levels (Karlsson *et al*, 1983; Swallow *et al*, 1987a; Gendler *et al*, 1988; Gendler et al, 1990a; Ligtenberg *et al*, 1990; Carvalho *et al*, 1997).

Polymorphisms outside the tandem repeat region of *MUC1* are important to understand the evolution of the *MUC1* VNTR region. Besides the G/A substitution in exon 2 (Ligtenberg *et al*, 1990; Ligtenberg *et al*, 1991) an additional CA repeat polymorphism within intron 6 of *MUC1* gene has been reported (Pratt *et al*, 1996). The fact that the two markers flanking the *MUC1* tandem repeat are associated suggests that the VNTR polymorphism did not arise from events which caused the recombination of flanking markers (Pratt *et al*, 1996).

The *MUC1* gene was mapped to chromosome 1q21-24 (Swallow *et al*, 1987b) The regulatory sequences appear to be contained within a 10 kbp *Sac*II fragment that contains the *MUC1* gene, 1.6 Kbp of the 5' sequence and 1.9 Kbp of the 3' flanking sequence. A putative Sp1 binding site and an E box with homology to the E-cadherin palindromic sequence appear to play a role in the regulation of *MUC1* gene expression (Kovarik *et al*, 1993). In the same region, a putative binding site for retinoic acid receptors may also participate in transcriptional regulation (Kovarik *et al*, 1993). Abe and Kufe (1993) located also regulatory sequences 5' from position –618. In MCF-7 cells, a 45 KDa protein which recognizes a sequence between positions –505 and –485 was identified; three Sp1 binding sites (-582, -532 and -492) and an AP-2 like site (-587) were also located in this region (Abe and Kufe, 1993). Hormonal modulation of *MUC1* was observed in the human endometrium, with highest expression in the secretory phase (Rye *et al*, 1993; Hey *et al*, 1994), and during estrus and proestrus in the mice (Braga *et al*, 1993).

Specific roles have been ascribed to MUC1 mucin emphasising its membrane-bound characteristic. In normal conditions, MUC1 is present at the luminal side of most glandular epithelial cells. In carcinomas, MUC1 is upregulated, its luminal distribution is lost and, most importantly becomes antigenically distinct from the MUC1 expressed by normal mammary cells (Girling *et al*, 1989) due to aberrant glycosylation (Burchell *et al*, 1993; Brockhausen *et al*, 1995; Lloyd *et al*, 1996). This makes the molecule a potential target for immunotherapy (McKenzie *et al*, 1990; Taylor-Papadimitriou *et al*, 1993) and

there is evidence suggesting that MUC1 expressed by carcinomas can be recognized by the immune system. In agreement with this hypothesis, antibodies to the core protein of MUC1 were identified in the serum of patients with cancers of the breast, ovary, lung and pancreas (Burchell *et al*, 1984; Hilkens *et al*, 1986; Metzgar *et al*, 1984; McGuckin *et al*, 1992; Willsher *et al*, 1993; Rughetti *et al*, 1993).

A cell-mediated immune response against MUC1 was also reported in patients with breast, ovarian, pancreatic and colorectal carcinoma (Barnd *et al*, 1989; Jerome *et al*, 1991; Kotera *et al*, 1994). Free MUC1 protein inhibits T cell proliferation (Agrawal *et al*, 1998) as well as cytotoxic T cell lysis of target cells (Barnd *et al*, 1989). The available evidence suggests that cytotoxic T cell lysis is not MHC restricted and may arise from interactions of the multiple MUC1 epitopes with T cells (Barnd *et al*, 1989; Jerome *et al*, 1991; Taylor-Papadimitriou *et al*, 1993; Magarian-Blander *et al*, 1993).

Several approaches are currently being used in animal models, and in pre-clinical and clinical trials for cancer patients immunization (Acres *et al*, 1993; Bu *et al*, 1993; Goydos *et al*, 1996; Apostolopoulos *et al*, 1996; Apostolopoulos *et al*, 1998; Gong *et al*, 1998).

Furthermore, both adhesive and anti-adhesive properties have been ascribed to MUC1, directly dependent on the carbohydrate side chains carried by the molecule. The aberrant expression of sialylated Lewis antigens (Sialyl Le^a and Sialyl Le^x) which serve as ligands for the adhesion molecules of the selectin family, and are carried on MUC1 apomucin (Hanski *et al*, 1993) provided data on the putative role played by MUC1 in the metastization process (Bevilacqua and Nelson, 1993; Dejana *et al*, 1992; Majuri *et al*, 1992; Rice *et al*, 1989; Sawada *et al*, 1994). Similarly, sialylated components of the simple mucin-type may inhibit the action of natural killer cells and cytotoxic T cells (Ogata *et al*, 1992a; Sherblom *et al*, 1986; Van de Wiel-van Kemenade *et al*, 1993), facilitating immunologic escape from the host defense mechanisms.

Regarding the anti-adhesive properties of MUC1 it was reported that high levels of MUC1 can reduce cell-cell and cell-extracellular matrix adhesion in a nonspecific manner (Ligtenberg *et al*, 1992b; Wesseling *et al*, 1995). Recently, it was advanced that the length of MUC1 is the dominant factor that determines the inhibition of E-cadherin–mediated cell-cell interaction (Wesseling *et al*, 1996); it was shown that E-cadherin can not only be functionally suppressed by overexpression of MUC1 but also reactivated after suppression of MUC1 expression (Kondo *et al*, 1998).

All these roles have been attributed to the extracellular domain of MUC1 but it is known that the cytoplasmic tail of MUC1 can interact with the actin cytoskeleton and this interaction may be important for the membrane localisation of MUC1 (Parry *et al*, 1990). Yamamoto *et al* (1997) demonstrated that the cytoplasmic tail of MUC1 has a sequence

motif that is responsible for direct binding to β-catenin. Recent studies further showed that phosphorylation of MUC1 by glycogen synthase kinase 3β (GSK3β) decreases binding of MUC1 to β-catenin *in vitro* an *in vivo* and that overexpression of MUC1 in the absence of (GSK3β) activity inhibits formation of the E-cadherin-β-catenin complex (Li *et al*, 1998).

MUC2

The *MUC2* gene encodes a intestinal secretory gel-forming mucin. *MUC2* cDNA was isolated from a human small intestine cDNA library (Gum *et al*, 1989). Partial cDNA clones corresponding to *MUC2* have also been isolated from libraries of tracheobronchial tissue from patients with cystic fibrosis (Gerard *et al*, 1990) and chronic bronchitis (Jany *et al*, 1991). With the work of Gum *et al* (1994) *MUC2* became the second human mucin gene with a completely identified cDNA sequence of 15.563 bp.

Similarly to MUC1, MUC2 has a central VNTR domain rich in serine, threonine and proline. Each unit is composed of 23 amino acids with 14 potential O-glycosylation sites (Gum *et al*, 1989). As in *MUC1*, VNTR polymorphism has also been confirmed for *MUC2* (Griffiths *et al*, 1990; Toribara *et al*, 1991; Vinall *et al*, 1998; Debailleul *et al*, 1998). In addition to the central tandem repeat domain, there is another domain of 347 amino acids composed of irregular repeats with a high content of serine, threonine and proline (Toribara *et al*, 1991, Gum *et al*, 1992). The non-tandem repeat regions of MUC2 have a high Cys content (Gum *et al*, 1992; Gum *et al*, 1994). Interestingly, these Cys-rich domains exhibit a high degree of sequence similarity to the polymeric serum glycoprotein, vWF. vWF is important for the binding of platelets to damaged endothelium and platelet aggregation under high shear force conditions (Sadler, 1991). vWF contains four structurally similar domains of approximately 350 residues each, designated D-domains (Shelton-Inloes *et al*, 1986; Mancuzo *et al*, 1989; Dong *et al*, 1994). MUC2 also has four D-domains (three in the amino terminal region and one in the carboxy terminal region) (Gum *et al*, 1994). Moreover, the D-domains of MUC2 are positioned similarly to those of vWF relative to the ends of the protein. It is very likely that D-domains in MUC2 play the same role as vWF D-domains: Cys residues in the D-domains are involved in the formation of both intra- and intermolecular disulphide bridges, leading to the polymerisation of MUC2 monomers into large polymers characteristic of secretory mucins. In addition to the D-domains, there is considerable sequence similarity between MUC2 and pro-vWF in other regions of the molecule (Gum *et al*, 1994).

Asker *et al* (1995, 1998) demonstrated that O-glycosylation of MUC2 starts after dimerization and that dimerization of MUC2 in the endoplasmic reticulum is followed by a N-glycosylation-dependent transfer of mono- and dimers to the Golgi apparatus.

The *MUC2* gene is located at chromosome 11p15 locus (Griffiths *et al*, 1990) close to the genes encoding *MUC5AC*, *MUC5B* and *MUC6*. All these four genes are expressed in cells with histochemical detectable mucus. These findings suggest that the four mucin genes on chromosome 11p15.5 may be part of a multigene family whose members encode for secreted gel-forming mucins (Pigny *et al*, 1996). Recently, it was postulated that these genes are derived from the same ancestral gene (Desseyn *et al*, 1998).

Little is still known about the regulation of *MUC2* gene. Velcich *et al* (1997) detected two DNAseI-hypersensitive sites in the 5' region of *MUC2* gene suggesting the presence of DNA regulatory elements. The structure and activity of the human *MUC2* mucin gene promoter was also recently described (Gum *et al*, 1997a). This 5' region of *MUC2* gene is GC-rich and contains several repetitive elements (Gum *et al*, 1997a). It was also reported that methylation of the promoter region is one of the important factors regulating *MUC2* gene expression (Hanski *et al*, 1997).

Wesley *et al* (1985) reported that intestinal mucins could be fractionated into acidic and neutral components and proposed that MUC2 might be the major apomucin of the acidic fraction while MUC3 might be the major apomucin of the neutral fraction.

MUC3

A human small intestine λgt11 cDNA library was screened with antibodies to deglycosylated human small intestine mucin. Partial cDNA clones that define a second human intestinal mucin gene were isolated (Gum *et al*, 1990). The tandem repeat of 51 bp encodes for a 17 amino acids repetitive peptide (Gum *et al*, 1990). Southern analysis showed that *MUC3* gene is polymorphic like the majority of human mucin genes (Kim *et al*, 1991; Gum *et al*, 1997b; Debailleul *et al*, 1998).

The repeat units of both MUC2 and MUC3 are more heavily glycosylated than MUC1. More than half of the amino acids in the tandem repeats of MUC2 and MUC3 are serines and threonines (61% and 71%, respectively), and it was postulated that most of these amino acids carry carbohydrates (Gum *et al*, 1989), whereas MUC1 has only five potential glycosylation sites (25%) probably all glycosylated *in vivo* (Muller *et al*, 1997).

The *MUC3* gene maps to chromosome 7q22 (Fox *et al*, 1992). Recently, the gene structure, the carboxyl terminus and upstream repetitive region of *MUC3* were described

(Gum *et al*, 1997b). The *MUC3* carboxyl-terminal domain is 617 residues in length, including 511 residues of a non-repetitive mucin-like domain and a 106-residue Cys-rich domain with homology to the EGF-like structural motifs found in many proteins (Gum *et al*, 1997b). The region of *MUC3* located upstream of the aforementioned 51 bp tandem repeats, which encodes a major Ser- and Thr-rich domain, consists of a second type of repetitive structure with an imperfect periodicity of approximately 1125 bp (Gum *et al*, 1997b). *MUC3* is a large and complex gene that encodes a large glycoprotein with a structure very different from that of any mucin described to date. MUC3 may participate in membrane or intermolecular interactions and may not form gels by end-to-end polymerisation as suggested for the mucins encoded by the genes at 11p15.5 (Shekels *et al*, 1998).

Van KlinKen *et al* (1997) reported a novel large repetitive region of MUC3 located upstream of the 17 amino acids repeat region. The 177 nucleotide repeat region is at least 5 Kb in length and encodes 59 amino acids repetitive peptides (Van KlinKen *et al*, 1997), probably corresponding to the region previously described by Gum *et al* (1997b).

MUC4

A cDNA library constructed from human tracheobronchial mucosa was screened with polyclonal antiserum raised to chemically deglycosylated glycopeptides from human bronchial mucins. One partial cDNA clone was isolated and mapped a novel human tracheobronchial mucin gene (Porchet *et al*, 1991). This clone contains a tandem repeat of 48 bp, which encodes a protein containing about 50% of hydroxy amino acids (Porchet *et al*, 1991). The *MUC4* gene was mapped to chromosome 3q29 (Gross *et al*, 1992) and found to be polymorphic due to the VNTR (Porchet *et al*, 1991; Gross *et al*, 1992; Nollet *et al*, 1998; Debailleul *et al*, 1998).

Recently, it was demonstrated that the organization of *MUC4* gene is quite different from that of the 11p15 mucin genes (Nollet *et al*, 1998). Organisation and sequence of the 5'-region and its junction with the tandem repeat array of *MUC4* was described: exon 1 consists of a 5' untranslated region and a 82 bp fragment encoding the signal peptide; the second exon of *MUC4* encodes for a peptide rich in serine and threonine and can be divided into four distinct domains. These domains vary in length from 3000 to 7000 amino acid residues and contain only one Cys residue. This variation is due to the VNTR polymorphism (Nollet *et al*, 1998). The 3' exons encode a peptide containing many Cys residues, two EGF-like domains and a transmembrane domain (Moniaux *et al*, 1998, personal communication). Moreover, an intron downstream from the tandem repeat array consists mainly of a 15 bp tandem repeat that exhibits a polymorphism with a

variable number of tandem repeats. MUC4, after MUC1, is the second human mucin with a transmembrane-spanning domain.

MUC5AC

In addition to MUC4, three cDNA clones have been obtained from human tracheobronchial mucosa and mapped to 11p15 (Aubert *et al*, 1991; Van Cong *et al*, 1990). Because of their common provenance, these three genes have been indicated with the number 5: *MUC5A, MUC5B* and *MUC5C* (Audie *et al*, 1993). Further investigations have demonstrated that *MUC5A* and *MUC5C* are part of the same gene now called *MUC5AC* (Guyonnet-Duperat *et al*, 1995), and that this gene is distinct from *MUC5B*.

The *MUC5AC* gene encodes a tandem repeat domain, with the smallest repeat unit (8 amino acids) described for a human mucin gene. A consensus Cys-rich domain comprising ten Cys residues in a 130 amino acids sequence is repeated several times in the MUC5AC protein (Guyonnet-Duperat *et al*, 1995). The presence of such Cys–rich domains suggests that among the ten Cys residues some are likely to be involved in disulphide bridges linking mucin monomers into oligomers. This Cys-rich domain is inserted into the tandem repeat domain or into a non-repetitive threonine/serine rich domain (Rose *et al*, 1989). Thus, in the MUC5AC apomucin structure, it is possible to recognize a large and heavily glycosylated domain which contains some Cys-rich poorly glycosylated subdomains.

A cDNA clone, NP3a, isolated from a human nasal polyp cDNA library, was described as a candidate for the 3'-end of the *MUC5AC* gene (Meerzaman *et al*, 1994). The deduced amino acid sequence showed homology to the Cys-rich domains of *MUC2* (Gum *et al*, 1992; Gum *et al*, 1994) and vWF (Dong *et al*, 1994).

A 3310-bp mucin cDNA clone (L31) was isolated from an HT29-MTX expression library using a polyclonal serum specific for normal gastric mucosa. The nucleotide sequence of the L31 clone showed a very high level of identity to the clone NP3a (Lesuffleur *et al*, 1995). A striking conservation of the number and position of the Cys residues was found between the L31 clone and *MUC2* carboxyl terminal region. These observations provide strong evidence that the clone L31 forms the 3'-end of the *MUC5AC* gene.

Another *MUC5AC* cDNA of 2.5 Kb, designated HGM-1, was isolated from a stomach cDNA library (Klomp *et al*, 1995).

The 5'-end of *MUC5AC* was recently described and its sequence shows high homology with the 5'-end of *MUC2* and contains three domains homologous to the D-domains found in vWF (Van de Bovenkamp *et al*, 1998). Furthermore, the 5'-end of *MUC5AC* contains a putative leucine zipper motif not found in any other mucin identified so far (Van de Bovenkamp *et al*, 1998).

A genomic cosmid clone, designated ELO9, spanning the 3' region of *MUC5AC* and the 5' region of *MUC5B* allowed the conclusion that *MUC5AC* and *MUC5B* have the same transcriptional orientation (Buisine *et al*, 1998).

The amino terminus and 5'-flanking region of *MUC5AC* gene was cloned and it was demonstrated that within 4 Kbp of the transcription start site there are elements mediating transcriptional up-regulation in response to bacterial exoproducts (Li *et al*, 1998).

MUC5AC is one of the genes located in the MUC cluster at 11p15.5. The order of these genes in the cluster is the following: *MUC6-MUC2-MUC5AC-MUC5B*, with *MUC5B* being the most centromeric (Pigny *et al*, 1996). The MUC genes cluster spans a region of approximately 400 Kb and contains five putative CpG islands. The role of this Cys-subdomain type is not yet known, although the fact that the cystein residues are well conserved suggests that this region may be important. These sequence similarities prompted Desseyn *et al* (1998) to suggest that 11p15.5 mucin genes may be derived from a single ancestral gene.

Southern analysis showed that *MUC5AC* gene is highly polymorphic due to the VNTR region (Pigny *et al*, 1995; Vinall *et al*, 1998; Debailleul *et al*, 1998).

MUC5B

A human tracheobronchial cDNA library was screened using antiserum against deglycosylated human tracheobrochial mucins. Four partial cDNA clones with degenerated 87-bp tandem repeats were found to belong to the *MUC5B* gene (Dufosse *et al*, 1993), and to encode 29 amino acid degenerated tandem repeats.

The structural organization of the peptide deduced from the nucleotide sequence of the central region of *MUC5B* was published recently (Desseyn *et al*, 1997a). The single large exon of 10.713 bp, containing all the tandem repeat domain, is described as being the biggest for a vertebrate gene and encodes for a 3.570 amino acid peptide. Nineteen subdomains have been identified that showed sequence similarity to each other, creating four larger composite super-repeat units of 528 amino acids. Each super-repeat is made up of repeats consisting of an irregular repeat of 29 amino acids, one Cys-rich

subdomain, and one unique sequence of 111 amino acid residues also rich in serine and threonine (Desseyn *et al*, 1997a).

More recently, the complete genomic nucleotide sequence of the 3'-region, the exon-intron organization, and the full cDNA sequence for the carboxyl-terminal domain of the human *MUC5B* apomucin was described (Desseyn *et al*, 1997b). The last five Cys-rich subdomains exhibit extensive sequence similarity to *MUC2, MUC5AC*, and vWF (Gum *et al*, 1994; Lesuffleur *et al*, 1995; Sadler *et al*, 1994; Desseyn *et al*, 1997b). This evidence together with the comparative molecular analysis of *MUC5AC* and *MUC5B* that shows a remarkable similarity in the size and distribution of exons and in the type of splice sites, supports the concept that *MUC5AC* and *MUC5B* have evolved from a single common ancestral gene (Buisine *et al*, 1998; Desseyn *et al* 1998).

Keates *et al* (1997) described the complete C-terminal sequence and genomic organization of *MUC5B*, and found that MUC5B is the major human gallbladder mucin. Some discrepancies between the two 3'-terminal sequences and the genomic organization of *MUC5B* were recently shown, including differences in intron number and size as well as in sequence (Laine, 1998; Offner, 1998).

Sequence similarities between cDNA clones isolated from different cDNA libraries supported the conclusion that *MUC5B* gene encodes the high molecular weight salivary mucin (MG1) (Desseyn *et al*, 1997b; Troxler *et al*, 1997; Nielsen *et al*, 1997).

MUC5B contrasts with the other genes of the mucin family regarding the absence of VNTR polymorphism (Vinall *et al*, 1998).

MUC6

The *MUC6* gene has been isolated and charaterized using a human gastric cDNA library (Toribara *et al*, 1993). In addition to *MUC2, MUC5AC* and *MUC5B, MUC6* is the fourth gene to be localized in the 11p15 region (Toribara *et al*, 1993; Pigny *et al*, 1996). The cDNA sequence is characterized by the presence of a VNTR domain whose individual repeat unit is 507 nucleotides long (169 amino acids). The length of the individual repeat unit of *MUC6* is the largest among mucins.

The carboxyl-terminal sequence of *MUC6* contains 1083 bp of coding sequence (361 amino acids) followed by 632 bp of 3'-untranslated region (Toribara *et al*, 1997). The coding sequence consists of two distinct regions: region 1 containing the initial 270 amino acids (62% Ser-Thr-Pro with no Cys residues), and region 2 containing the COOH-terminal 91 amino acids (22% Ser-Thr-Pro with 12% Cys residues). Region 2 has approximately 25% amino acids homology to the COOH-terminal regions of human

mucins *MUC2, MUC5AC* and *MUC5B* and vWF (Gum *et al*, 1994; Lesuffleur *et al*, 1995; Sadler *et al*, 1994; Desseyn *et al*, 1997b). The conservation of the 11 cystein positions in region 2 suggests the importance of this short region to mucin polymerization (Toribara *et al*, 1997).

MUC6 is genetically polymorphic due to the VNTR (Toribara *et al*, 1993; Garcia *et al*, 1997; Vinall *et al*, 1998; Debailleul *et al*, 1998).

MUC7

Human saliva contains high and low molecular weight mucin glycoproteins (MG1 and MG2, respectively) that are structurally distinct.

The full-length sequence for *MUC7* has been published (Reddy *et al*, 1992; Bobek *et al*, 1993) and encodes MG2 glycoprotein. MG2 contains six tandem repeats of 23 amino acids rich in potential O-glycosylation sites.

In contrast to the *MUC2* gene, which encodes for a large secreted gel-forming mucin, the *MUC7* gene encodes a small secreted non-gel-forming mucin. The molecule is secreted as a monomer and therefore the two Cys residues, located upstream to the tandem repeat domain, are likely to be involved in intramolecular disulphide bonds (Bobek *et al*, 1993). Recently, the *MUC7* gene structure was determined and its chromosomal localization assigned to 4q13-q21 (Bobek *et al*, 1996).

It is now known that human submandibular/sublingual gland secretions contain a multimeric high molecular weight mucin (MG1) coded by *MUC5B* gene (Desseyn *et al*, 1997b; Troxler *et al*, 1997; Nielsen *et al*, 1997) and a monomeric low molecular weight mucin (MG2) encoded by the *MUC7* gene (Bobek *et al*, 1993; Bobek *et al*, 1996).

MUC8

A partial cDNA encoding a major airway mucin glycoprotein with a novel tandem repetitive sequence was cloned by Shankar *et al* (1994). The partial cDNA revealed the presence of almost perfect 41-bp tandem repeat sequence that encode two types of consensus peptide repeats of 41 and 13 amino acids, respectively. Recently, the extreme carboxyl terminus of *MUC8* was reported (Shankar *et al*, 1997).

MUC8 maps to chromosome 12q24.3 (Shankar *et al*, 1997). VNTR polymorphism for *MUC8* was not reported so far.

It is likely that in the future more human mucin genes will be identified in different tissues, in accordance to the known tissue- and cell-specific pattern of mucin expression.

Recently some putative new human mucin genes have been described and numbered according to their order of appearance as *MUC9* (human oviductin gene) and *MUC10* (otogelin) (Lapensée *et al*, 1997; Cohen-Salmon *et al*, 1997; Swallow, 1998, personal communication). Little is known about the structure of these genes; however, allelic polymorphism and chromosomal localization to 1p13 were reported for *MUC9* (Lapensée *et al*, 1997).

Non-human mucin genes and proteins

Several mammalian homologues of the human mucin genes have been described.

Full-length mouse and human *MUC1* clones have been obtained and, while their tandem repeats differ significantly, the human and mouse *MUC1* mucins share high sequence similarity in their non-repetitive and promotor regions; this finding is not surprising because the human and the mouse genes show similar patterns of expression (Spicer *et al*, 1991; Pemberton *et al*, 1992). The mouse *Muc1* gene encodes a protein that is smaller when compared to the human homologue because the mouse gene contains a fixed number of 16 repeats (Spicer *et al*, 1991; Vos *et al*, 1991).

The rat intestine was shown to express at least two different mucins (Khatri *et al*, 1993). The rat Muc2 mucin was cloned (Xu *et al*, 1992; Khatri *et al,* 1993; Ohmori *et al,* 1994; Hansson *et al*, 1994) and shares striking similarity to human *MUC2* in the non-repetitive regions. Homology with the human MUC2 both at C- and N- termini suggests that the rat gene is organised similarly to the human gene. Within the C-terminal domain Cys residues alignment in both mucins resemble the vWF. Bell *et al* (1998) demonstrated that the C-terminal domain of rat Muc2 can form disulphide-dependent dimers, and that N-glycosylation is required for dimerization and subsequent secretion, similarly to the human MUC2 (Asker *et al* 1995, 1998).

Two other partial rat intestinal mucin cDNA clones, rMUC176 and M2, were also isolated and are distinct from rat *Muc2* (Gum *et al*, 1991; Khatri *et al*, 1993).

Mouse caecal mucin has been pointed as being the mouse homologue of human *MUC3* gene. Despite the lack of homology between the human and mouse mucin tandem repeats, the mouse intestinal clone does share a significant sequence similarity with a portion of the recently published 3' non-repetitive region of human *MUC3* (Gum *et al*, 1997b). Mouse and human genes differ in their overall structural domains but share common Cys-rich regions, also present in rat Muc3 (Khatri *et al*, 1997). The arrangement of Cys within these regions corresponds to EGF-like domains. In mouse Muc3, a stretch of 19 amino acids may act as a transmembrane domain or allow for interaction with hydrophobic molecules (Shekels *et al*, 1998). Rat Muc3 appears to be also membrane-associated (Khatri *et al*, 1997; Khatri *et al*, 1998, personal communication). The spacing of the mouse and human MUC3 Cys is unlike that found in the Cys-rich regions of the MUC2, MUC5AC, MUC5B, MUC6 and vWF D domains. The differences in the Cys motifs may provide insight into the individual roles of the different mucins (Shekels *et al*, 1998).

A mouse homologue of *MUC5AC* (*mMuc5AC*) has been cloned by Shekels *et al* (1995). Guzman *et al* (1996) reported a partial sequence of a rat *Muc5AC* homologue. Rat and mouse have the same number of amino acids (16) in their Muc5AC repeat units and share 69% sequence similarity, whereas the human MUC5AC has only 8 amino acids in its tandem repeat (Inatomi *et al*, 1997). The C-terminal non-repetitive region of rat Muc5AC, which includes four Cys residues, shows similarity to the corresponding region in human MUC5AC and mouse Muc5AC (Inatomi *et al*, 1997).

Mucin molecules designated as rat ascites sialomucin glycoproteins, ASGP-1 and ASGP-2, with possible bifunctional capacities, were described on the cell surface of 13762 rat mammary adenocarcinoma cells. It is hypothesized that they protect against immune recognition and natural killer cell killing (Sherblom *et al*, 1986) and their expression correlates with metastatic potential (Steck *et al*, 1983). The sialomucin is synthetized as a large precursor that is cleaved during the passage through the endoplasmic reticulum into ASGP-1, the sialomucin portion, and ASGP-2, the membrane-associated N-glycosylated glycoprotein (Sheng *et al*, 1990). The full-length sequence of ASGP-1 was reported by Wu *et al* (1994). There are similarities between the heterodimeric cell surface glycoprotein complex (ASGP-1 and ASGP-2) and human MUC1. These similarities include: (a) the presence of three Cys residues in or near the transmembrane domain; (b) a cleavage event that occurs during the passage to the cell surface; and (c) endocytosis with further sialylation. The dual functionality of this protein complex raises the possibility that other mucins may also be multifunctional as proposed

for MUC1, with adhesion and anti-adhesion properties. This possibility will be discussed later.

Structural but not sequence homology is evident between the C-terminal regions of rat Muc3 and ASGP-2 (Khatri *et al*, 1997; Khatri *et al*, 1998 – personal communication).

Sequence similarity was also described between human MUC4 and ASGP-1 (Moniaux *et al*, 1998, personal communication).

Several other non-human mucins have been reported and some of them show sequence similarity to each other as is the case of porcine submaxillary mucin and bovine submaxillary mucin (Eckhardt *et al*, 1991; Bhargava *et al*, 1990). Ovine submaxillary mucin (Hill *et al*, 1997a, Hill *et al*, 1997b) and rat and mouse submandibular mucins (Albone *et al*, 1994; Denny *et al*, 1996) have been described but little is known with regard to the complete gene structures and possible homologies.

Partial cDNA clones that encode for airway mucins have been described in rat (rat airway mucin) (Tsuda *et al*, 1993) and in dog (canine tracheobronchial mucin) (Verma *et al*, 1993). cDNA clones encoding the pig gastric mucin were also published (Turner *et al*, 1995).

Three frog integumentary mucins (FIMs) produced by the merocrine mucous glands from the skin of *Xenopus laevis* have been reported in the literature. The first of these mucins detected by cDNA cloning was FIM-A.1 (Hoffmann *et al*, 1988), originally called spasmolysin. This 43 Kda mucin is a mosaic protein, consisting of an amino terminal signal sequence, 13 acidic Thr-rich repeats of 9 amino acids length, 7 repeats of the sequence ETTT, and four P-domains. The P-domain is a characteristic shuffled module (~50 amino acids in length) containing six invariant Cys residues that form three intramolecular disulphide bridges (or trefoil motif). It is a potential modulator of cell growth, by analogy to EGF-like repeats. In mucins, P-domains are thought to play a role in protein-protein interaction or lectin-like interaction (Hoffmann *et al*, 1993).

FIM-B.1, the second mucin described in *Xenopus laevis*, contains a Cys-rich carboxyl terminal domain that is homologous with vWF, MUC2, rat Muc2, MUC5AC, PSM, BSM, procollagen and thrombospondin (Probst *et al*, 1990; Probst *et al*, 1992). The gene is also polymorphic but differs from other mucin genes in that the repeat sequences are separated by introns (Probst *et al*, 1992). The function of FIM-B.1 is not known, although it could serve to protect against microbial infection (Probst *et al*, 1992).

Partial sequence for a third polymorphic frog integumentary mucin, FIM-C.1, was described by Hauser *et al* (1992). FIM-C.1 contains three semi repetitive Thr-rich clusters and six P-domains. The polymorphism observed among different animals appears to be due to the variable lengths of the Thr-rich domains.

Genes encoding mucin-type molecules in lower eucaryotes have been described in *Leishmania major* (Ilg *et al*, 1996) and *Trypanosoma cruzi* (Di Noia *et al,* 1995; Di Noia *et al,* 1998). Particularly in *T. cruzi*, the unicellular parasite agent of Chagas disease, mucin-like glycoproteins were demonstrated to play an important role in the interaction of the parasite with the surface of mammalian cells during the invasion process. The overall structure of the family of mucin-like genes in *T. cruzi* resembles that of mammalian mucin genes (Di Noia *et al*, 1995). It was demonstrated that the mucin gene family from *T. cruzi* is much larger and diverse than expected, with a minimal number of 484 mucin genes per haploid genome for a parasite clone (Di Noia *et al*, 1998). The regions encoding the N- and C-termini domains of these genes are highly conserved, whereas the central regions are highly variable (Di Noia *et al*, 1998).

A mucin-type glycoprotein with an apparent mass of 90 KDa was detected and partially characterized in embrionic cells of *Drosophila melanogaster* (Kramero *et al*, 1996). This glycoprotein was designated as mucin-D. In this species, mucin-D seems to be involved in the moulting events as it is biosynthetically regulated by the insect moulting hormone.

Mucins and infection: the case of *Helicobacter pylori*

Attachment is a prerequisite for bacterial colonization of epithelial cell surfaces. The initial host-microbial interaction is mediated by bacterial surface proteins named adhesins (Jones and Isaacson, 1983; Sharon, 1987) that bind to protein (Bliska *et al*, 1993) or carbohydrate (Karlsson, 1989; Hultgren *et al*, 1993) structures present on the cell surfaces. Many of these carbohydrates structures are carried by mucin proteins (Jentoff, 1990; Ho and Kim, 1991). Bacterial binding is restricted to a set of cell populations carrying the optimal receptor structures. This tropism partly determines the niche a bacterium is able to occupy.

The first well characterized carbohydrate epitope to be established as a microbial receptor was the ganglioside G_{M1}, used by cholera toxin for attachment to enterocytes (Holmgren *et al*, 1975). Since then, an overwhelming number of reports have

established the ability of microbial ligands to attach to carbohydrates and to host proteins (For review see Falk *et al*, 1994).

Many carbohydrate structures present on mucin glycan chains are recognized by surface adhesins or hemagglutinins of microorganisms, as is the case for pathogens which are to be eliminated from the lung and oral cavity, or symbiotic strains in the gastrointestinal tract that compete with pathogenic strains. Microorganisms were shown to recognize mucin-like carbohydrate structures, such as *Actinomyces naeslundii* and *Streptococcus sanguis* that colonize the oral cavity (Devine and McKenzie, 1992; Lamblin *et al*, 1991). There are also published examples for respiratory pathogens such as *Mycoplasma pneumoniae, Streptococcus pneunomiae, Pseudomonas aeroginosa,* Influenza viruses, and for a variety of *Escherichia coli* adhesins (Devine and McKenzie, 1992; Lamblin *et al*, 1991; Biesbrock *et al*, 1991; Sharfman *et al*, 1996). Adherence to intestinal mucin by an human pathogen, *Campylobacter upsaliensis,* was reported by Sylvester *et al* (1996).

Helicobacter pylori (*H. pylori*), a gram-negative, spiral-shaped, microaerophilic bacteria isolated from human gastric mucosa, was discovered in 1983 (Marshall and Warren, 1983). The *in vivo* tropism to gastric epithelium stimulates the need to characterize the epitopes mediating the cell lineage-specific attachment of *H. pylori* to human gastric surface mucous cells.

It was shown that *H. pylori* agglutinates erythrocytes (Evans *et al*, 1988) and binds to mouse adrenal gland (Y1) cells (Evans *et al,* 1989) in a sialic acid dependent manner. Furthermore, specific binding of *H. pylori* to acid glycosphingolipids extracted from human gastric mucosa, such as ganglioside G_{M3} and sulfatide (Saitoh *et al*, 1991), has been reported. Sialic acid-independent *H. pylori* adhesion to HeLa cells was also described (Fauchére and Blaser, 1990). It has also been reported that histo-blood group Lewis^b antigen is a candidate receptor for *H. pylori* (Borén *et al*, 1993; Falk *et al*, 1995). The Le^b-binding adhesin, BabA, was purified by receptor activity-directed affinity tagging (Ilver *et al*, 1998).

The apparent heterogeneity of these results probably reflects the known existence of several adhesins in *H. pylori* mediating adherence to various targets in gastric epithelium and/or the presence of multiple binding modes among different *H. pylori* strains (Moran, 1995). The multiplicity of adhesins described to date may participate at different steps in the process of *H. pylori* adherence, which involves different adhesion and recognition interactions at different stages of the colonizing process (Moran, 1995).

Gastric mucins and gastric carcinoma risk

In addition to sharing the general needs of mucosal surfaces for protection from pathogenic organisms, desiccation and mechanical damage, the gastrointestinal epithelium has specific needs that are different from other organ systems (Strous and Dekker, 1992). The mammalian stomach is unique in that its epithelial lining is nearly constantly exposed to pH 2 or less, as a consequence of HCl secretion. The mucin gel layer on the surface of the epithelium of the stomach has been proposed as a protective diffusion barrier against luminal HCl (Strous and Dekker, 1992).

There is consistent data indicating that mucin genes are expressed in a regulated cell- and tissue-specific manner. Stomach is a good example of such differential expression of mucin genes. MUC1 is widely expressed in all epithelia (Zotter *et al,* 1988; Ho and Kim, 1991; Xing *et al*, 1992a). In stomach, it can be detected in mucous cells of the surface epithelium and mucous neck cells of the antrum, as well as focally expressed in oxynthic glands of the body and pyloric glands of the antrum (Burchell *et al,* 1987; Girling *et al*, 1989; Ho *et al*, 1993; Lesuflleur *et al*, 1994; Ho *et al*, 1995; Reis *et al*, 1998a). MUC5AC is highly expressed in foveolar epithelium of both body and antrum (Byrd *et al*, 1997; DeBolós *et al*, 1995; Ho *et al*, 1995; Reis *et al*, 1997). In contrast, MUC6 protein expression is limited to mucous neck cells of the body and pyloric glands of the antrum (Byrd *et al*, 1997; DeBolós *et al*, 1995; Ho *et al*, 1995). Conflicting results regarding the expression of MUC2 have been reported, demonstrating either no expression of MUC2 (Ho *et al*, 1995; Xing *et al*, 1992b) or very low levels of expression (Audie *et al*, 1993; Carrato *et al*, 1994; Devine *et al*, 1993; Gambús *et al*, 1993; Ho *et al*, 1993; Reis *et al*, 1998b).

During gastric carcinogenesis, there are alterations in the mucin expression pattern. Frequently, there is expression of underglycosylated forms of MUC1 mucin, decreased expression of MUC5AC and MUC6 normally expressed in the stomach, and aberrant expression of mucins not found in the normal gastric mucosa (Baldús *et al*, 1998; Ho *et al*, 1995; Reis *et al*, 1997; Reis *et al*, 1998a; Reis *et al*, 1998c).

Gastric carcinoma is a major cause of cancer death worldwide (Muñoz, 1988). Portugal stands on the top of the list of European countries for the highest incidence and mortality from gastric cancer (Black *et al*, 1997). For unknown reasons the mortality rates have

not declined in Portugal as far as in most other countries (La Vecchia *et al*, 1992; Correa, 1992).

Human gastric carcinogenesis is a multistep, multifactorial process, and the relative contribution of environmental exposure and genetics for the risk of developing gastric carcinoma is far from being established. Diet and *H. pylori* infection have been identified as the major environmental exposure risk factors (Correa, 1988a; Correa, 1992; Solcia *et al*, 1996).

The contribution of host susceptibility factors has been difficult to elucidate. It has been known for a long time that there is a lower incidence of adenocarcinomas among blood group O versus A individuals within the ABO blood system (Mourant *et al*, 1978).

Mucins are attractive target molecules to study the relationship between genetics and environment because they play an important role in the protection of gastric mucosa and exhibit polymorphic variation in their length. Our group demonstrated that small *MUC1* alleles and genotypes are significantly more frequent in patients with gastric carcinoma than in control blood donors, suggesting that small *MUC1* genotypes are a risk factor for gastric carcinoma development (Carvalho *et al*, 1997).

Mucins are highly polymorphic glycoproteins that have in common the presence of extended arrays of tandemly repeated peptides. The variable number of tandem repeats accounts for the extensive polymorphism observed at DNA, RNA and protein levels (Karlsson *et al*, 1983; Swallow *et al*, 1987a; Gendler *et al*, 1988; Gendler et al, 1990a; Ligtenberg *et al*, 1990; Carvalho *et al*, 1997; Garcia *et al*, 1997; Debailleul *et al*, 1998). The biological significance of these polymorphisms had not been previously elucidated. We advanced the hypothesis that *MUC1* and *MUC6* genes polymorphism, resulting in individual size differences with respect to VNTR, could influence the function of the respective proteins, providing different protective effects against various environmental aggressions.

Most studies on incidence of gastric cancer have emphasized the role of environmental factors among which stands *H. pylori* infection. From the few studies on record, Portugal stands on the group of European countries with the highest prevalence of *H. pylori* infection (Soares *et al*, 1993; Estevens *et al*, 1993). In contrast, the Danish population is much less infected by *H. pylori* (Andersen *et al*, 1996). Fitting with these data and with the putative relationship between *H. pylori* infection and gastric carcinoma, Portugal has the highest and Denmark the lowest incidence rate of gastric carcinoma among the countries of the European Union (Black *et al*, 1997). We advanced the hypothesis that *MUC1* gene polymorphism could, at least partly, be used to explain the different incidence of gastric carcinoma between the two countries, probably acting together with *H. pylori* infection.

Taking into account the accumulated epidemiological data together with the development of our work we decided to study gastric carcinoma cell lines to see if they could be used as *in vitro* models to analyse the expression of mucins and carbohydrate side chains as well as to develop an *in vitro* adherence assay for *H. pylori*.

The aims of the present study can be summarized as follows:

I – Polymorphism of *MUC1* and *MUC6* genes and gastric carcinoma risk

In order to test whether or not mucin genes polymorphism can provide different protective effects against environmental aggressions, we performed case-control studies

using Southern blotting analysis to evaluate *MUC1* and *MUC6* genes polymorphism in a control group of healthy blood donors and in patients with gastric carcinoma.

II – *MUC1* gene polymorphism and *Helicobacter pylori* infection

In an attempt to gain insight into the reasons underlying the different incidences of gastric carcinoma in two european populations - Portugal and Denmark – we compared the *MUC1* gene polymorphism and the prevalence of *H. pylori* infection in the two populations.

III – Mucins expression and *Helicobacter pylori* adherence to gastric carcinoma cell lines

We aimed to characterize the expression of mucins and carbohydrate side chains in gastric carcinoma cell lines and to develop an *in vitro* adherence assay of *H. pylori* to gastric carcinoma cell lines.

MATERIALS AND METHODS

Materials

The biological material included in the different studies is described in Papers I-IV. Because some of the results have not been published yet, a summary of the biological material is presented:

- DNA and serum from Portuguese blood donors;
- DNA from Portuguese patients with gastric carcinoma;
- DNA, serum and biopsies from dyspeptic patients;
- Seven gastric cancer cell lines (GP202, GP220, MKN45, KATOIII, St23132; St2957, St3051), and one myeloid leukaemia cell line (HL60);
- *Helicobacter pylori* strain NCTC 11637, and *Escherichia coli* strain DH5α.

Methods

The methods used in the different studies are described in detail in Papers I-IV. Briefly, the following methods were used:

Extraction of DNA

DNA from blood donors and dyspeptic patients was isolated from blood samples (10 ml) pelleted and stored at –70ºC until DNA extraction. DNA from patients with gastric carcinoma was isolated from samples of non-neoplastic gastric mucosa collected immediately after surgery. The collected tissue samples were frozen in liquid nitrogen and stored at –70ºC until DNA extraction. High molecular weight DNA was isolated using a salt-chloroform extraction method (Mullenbach *et al*, 1989).

Southern blotting

DNA samples were digested with different restriction enzymes that recognize restriction sites in the flanking regions of the tandem repeats, separated on 0.7% agarose gel by electrophoresis and transferred to nylon membranes by alkaline blotting (Southern, 1975).

The MUC1 probe directed to the tandem repeat region was PCR-labelled with [α-^{32}P] dCTP. The MUC 6 probe directed to the tandem repeat region was labelled with [α-^{32}P] dCTP using the oligolabelling extension method (Feinberg and Vogelstein, 1982). Both probes were purified from non-incorporated nucleotides by passage through Sephadex-G50 columns.

The blots were prehybridized for 4 hours in a phosphate solution [0.5M $NaHPO_4$ (pH 7.2), 1mM EDTA, and 7% SDS] at 65ºC, followed by hybridization at 65ºC overnight with the specific probe for the VNTR region of the gene. Blots were washed under high stringency conditions (first with 40 mM $NaHPO_4$, 1mM EDTA, and 5% SDS, and then with 40 mM $NaHPO_4$, 1mM EDTA, and 1% SDS at 65ºC). Membranes were exposed to autoradiography at –70ºC with an intensifying screen. Autoradiography was developed after 1-3 days.

All bands from the autoradiographies were scored visually. The size of the alleles of *MUC1* and *MUC6* genes was determined by two independent observers by comparison with the size of the fragments of the marker λ *Hind*III. Reference alleles were included in every blot.

PCR for CA repeat of intron 6 of *MUC1* gene

Radioactive-PCR and visualization of the products in vertical acrylamide gels

Polymerase chain reaction was performed using standard techniques. The primer set was chosen from intron 6 of *MUC1* gene, flanking the CA repeat (Pratt *et al*, 1996). Upstream primer: 5'-AGAGAGTTTAGTTTTCTTGCTCC-3'; Downstream primer: 5'-CTTCTTGGCTCTAATCAGCCC-3'. Following an initial denaturation at 94ºC for 5 min, the PCR conditions were denaturation at 94ºC for 30 s, annealing at 58ºC for 30 s, and extension at 72ºC for 1.30 min. Thirty-three cycles were performed. Following the last

amplification cycle an additional extension period at 72ºC for 10 min was performed. PCR products were labelled with [α-^{32}P] dCTP during amplification, denatured in a solution containing 95% formamide for 5 min at 94ºC, separated by electrophoresis in 6% denaturing polyacrylamide gels and visualized through autoradiography.

PCR and visualization of the products in horizontal acrylamide gels followed by silver staining

The same PCR conditions were performed without using radioactivity. PCR products were run in ultrathin polyacrylamide gels (7%T, 5%C) and stained with silver. Briefly, the gel is placed in 10% ethanol solution for 5 min, oxidized in 1% nitric solution for 3 min and then rinsed with distilled water. Twenty min incubation in 0.012M silver nitrate solution was performed in dark with shaking. The gel is reduced with a solution containing 0.28M sodium carbonate (anhydrous) and 0.019% formalin. The redution process in stopped incubating the gel in 10% glacial acetic acid for 2 min and then the gel is placed in distilled water and dryed.

Enzyme-linked immunoassay determinations of anti-*H. pylori* IgG antibodies

Ten ml blood samples were centrifugated at 2500 rpm for 10 min and approximately 5 ml of serum was obtained, which was divided into aliquots and stored at –70ºC. The COBAS CORE Anti-*H. pylori* IgG EIA (Roche Diagnostic Systems) was performed according to the instructions of the manufacturer. Briefly, samples were diluted 1:400 with sample diluent and incubated with beads coated with highly purified *H. pylori* antigens (multi-component antigen preparation, including urease and free of cross-reacting flagella proteins). Specific antibodies were bound to the beads. After the removal of the unbound material by a washing step, the antigen-antibody complex on each bead was detected with peroxidase-conjugated goat anti-human IgG antibody. After the removal of unbound conjugate, the beads were incubated with a substrate solution containing tetramethylbenzidine (TMB) and hydrogen peroxide. A blue color develops, the intensity of which is proportional to the amount of *H. pylori*-specific IgG bound to the beads. The enzymatic reaction is stopped by the addition of acid, and absorvance values are determined at 450 nm. A standard curve is obtained by a dilution series of the positive control and by plotting each absorvance value *versus* the corresponding standard value. The concentrations of IgG antibody to *H. pylori* in the samples are determined by interpolation from this standard curve and expressed as

units per ml (U/ml). Each of the EIA tests was performed in duplicate for each sample with a serum specimen that had not been previously thawed.

Hematoxilin-eosin staining of gastric biopsies of dyspeptic patients

H. pylori infection was also evaluated in hematoxilin-eosin stained sections of gastric biopsies of dyspeptic patients. *Helicobacter pylori* infection was scored as – or +.

Culture of gastric carcinoma cell lines

Seven gastric carcinoma cell lines and one myeloid leukaemia cell line were analysed. Cell lines GP202 and GP220 were recently established in our lab (Gärtner *et al*, 1996), MKN45, KATOIII and HL60 were commercially available and cell lines St23132, St2957 and St3051 were a kind gift from Dr. Peter Vollmers (Vollmers *et al*, 1993). All cell lines were grown in RPMI 1640 with L-glutamax, supplemented with 10%FBS, 100U/ml penicilin, 100μg/ml streptomycin and 100U/ml nystatin. Cultures were maintained at 37ºC in a humidified 5% CO_2 atmosphere.

RT-PCR (Reverse Transcription-Polymerase Chain Reaction)

cDNAs were prepared using the 1st-Strand cDNA synthesis Kit (Clontech). PCRs were carried out in 50 μl of final solution (5μl Perkin Elmer buffer, 5μl MgCl2 25mM, 5μl dNTPs, 10 pmoles of each primer, 2U Taq Polymerase (Ampli Taq Gold), and 5μl cDNA). Cycling conditions were as follows: 1) denaturation at 96ºC, 30 seconds for one cycle; 2) denaturation at 96ºC, 30 seconds, annealing at 60ºC, 30 seconds, and extention at 72ºC, 60 seconds for 30 cycles; 3) extention at 72ºC, 15 minutes. The primers were chosen in the 3' part of the different mucin genes (unpublished data). The same PCR conditions were performed to study of all mucin genes (*MUC1-MUC7*). GADPH mRNAs were coamplified in each experiment with each primer set.

Northern blotting

Large RNAs were extracted from the different cell lines (~10^8 cells) at confluence using an improved method for the isolation and transfer of large RNAs as previously described (Debailleul *et al*, 1998). Briefly, RNA samples (10 μg) were denatured in loading buffer containing 50% formamide, 18% deionized formaldehyde, and 0.02M MOPS, heated at 68ºC for 10 min, and then fractionated through a 0.9% agarose denaturing gel (13 x 8 x 13 cm) containing 18% formaldehyde and 0.02M MOPS. The running gel buffer was 0.02M MOPS, pH 7.0. The electrophoresis was performed for 16 h at 30 V. Prior to transfer, the gel was soaked for 45 min in 20x SSC. The RNA was transferred onto Hybond-N^+ membrane (Amersham Corp.) via vacuum blotting for 30 min and crosslinked to the membrane by exposure to UV light for 4 min. All the probes used to *MUC1* (Gendler *et al*, 1990a), *MUC2* (Gum *et al*, 1989), *MUC3* (Gum *et al*, 1990), *MUC4* (Porchet *et al*, 1991), *MUC5AC* (Guyonnet-Duperat *et al*, 1995), *MUC5B* (Dufosse *et al*, 1993), *MUC6* (Toribara *et al*, 1993), *MUC7* (Vandenhaute *et al*, 1997) and actin (Clontech), used as a positive control, were labelled with [α-^{32}P] dCTP using a commercial random-primer labelling kit according to the manufacturer's protocol. These probes were used at 1 x 10^6 cpm/ml. Prehybridization and hybridization were performed in 10 ml of a solution containing 50% formamide, 5 x SSPE, 10 x Denhardt's solution, 2% (w/v) SDS, and 100 μg/ml of sheared salmon sperm DNA, for 2 and 16 h, respectively, at 42ºC. The membranes were washed twice in 0.1 x SSPE, 0.1% SDS buffer for 15 min at 65ºC. After a final wash with 6 x SSPE at room temperature, the membranes were exposed to Kodak X-Omat film at –80ºC.

Immunohistochemistry

Sections from formalin-fixed paraffin-embedded cytospins of the different cell lines were used for immunostaining. The avidin-biotin-peroxidase complex (ABC) method was applied. The paraffin sections were dewaxed and then treated with 0.3% hydrogen peroxide (H_2O_2) in methanol for 30 min, to block endogenous peroxidase. The sections were then incubated for 20 min with normal non-immune serum to eliminate non-specific staining. Excess of normal serum was removed and replaced by specific monoclonal antibodies in the appropriate dilution. Sections were incubated overnight (18 to 22 h) at 4ºC. After washing the slides the sections were incubated with a 1:200 dilution of biotin-labelled secondary antibody for 30 min and ABC for 60 min. Sections were stained for 7 min 0.05% 3'3 diaminobenzidinetetrahydrochloride (DAB) freshly prepared in 0.05M tris/hydroxymethylaminomethane buffer at pH 7.6, containing 0.01% H_2O_2, counterstained with hematoxylin, dehydrated and mounted.

Culture of *Helicobacter pylori* and *in vitro* adherence assay to gastric carcinoma cell lines

H. pylori strain NCTC 11637 was obtained from Public Health Laboratory Service (London), and grown in *Helicobacter pylori* agar plates containing 7% defibrinated horse blood (Becton Dickinson) in an atmosphere of 5% O_2 and 10% CO_2 for 5 days. Gram and Urease tests were performed to confirm the Gram- and Urease + characteristics of *H. pylori*. *E. coli* strain DH5α was used as a control.

Bacteria were harvested by rinsing the surface of each plate with 5 ml of PBS and removing growth by scraping with a sterile scraper. Quantitation of bacteria in suspension was determined by optical density measurement at 600 nm. After washing 3 times in PBS, bacteria was ressuspended in 0.04% FITC solution in sodium carbonate (0.05 mol/L) and sodium chloride (0.1 mol/L) (pH 9.2), freshly prepared, and labelled for 30 min at RT. After labelling, the bacteria were washed 3 times in PBS.

Different cell lines were trypsinized (except HL60 that grows in suspension) and counted in a Neubaeur chamber. Bacteria and cells were used at a ratio 300:1. Adhesion between bacteria and cells was performed at 37ºC for 1 hour with shaking (100 rpm – orbital shaker). After washing 2 times in PBS, 10000 gated events were analysed on a COULTER XL-MCL flow cytometer equipped with an argon laser tuned to 488 nm. FITC fluorescence was collected at 525 nm. Cells were gated to exclude dead cells. The threshold for positivity was set for each experiment by flow cytometric analysis of cells without adherent bacteria. *H. pylori* and *E. coli* binding indexes were calculated by the ratio of the mean fluorescence channel in the adherence experiment with the mean fluorescence channel in the control experiment. All the experiments were performed at least in duplicate.

Adherence assay was also performed in slides. Cell lines were grown in Biocoat culture slides from Becton Dickinson and infected with FITC-labelled *H. pylori*.

Statistical analysis

Allelic frequencies were estimated by gene counting method. Statistical analysis for Hardy-Weinberg equilibrium was performed using an exact test (Guo and Thompson, 1992), running the statistical software package GDA (Lewis and Zaykin, 1997) or GENEPOP (Raymond and Rousset, 1995). Expected heterozygosity was calculated according to the method of Nei (1978), and the difference between the expected and the

observed number of heterozygotes was used to calculate a chi-square with one degree of freedom. Heterogeneity analysis between populations was based on allelic contingency tables, and the corresponding chi-square was calculated using the statistical software package BIOSYS-1 (Swofford and Selander, 1981). The Monte Carlo test applied to the study of highly polymorphic genes, as described by Sham and Curtis (1995), was performed using the computer program CLUMP (ftp.diamond.gene.ucl.ac.uk [directory/pub/packages/dcurtis]). Odds ratio and 95% confidence limits were determined using the BMDP 4F computer program (Statistical Package Program BMDP; Los Angeles, CA, USA).

Statistical analysis of other results was performed by X^2 test, Student´s t test, unpaired t-test, ANOVA test and Mann-Whitney non-parametric test, using the Statview 4.02 Software.

Values of serum titres of IgG anti-*H. pylori* are expressed in mean ± standard error.

RESULTS

I – Polymorphism of *MUC1* and *MUC6* genes and gastric carcinoma risk

Ia – *MUC1* VNTR polymorphism (Paper I)

Fifteen alleles were identified in the 324 Portuguese blood donors using *Eco*RI as restriction enzyme. The allele size ranged from 8.9 Kbp to 12.6 Kbp – respectively alleles 15 and 1 in the ranking system (Figure 1). Thirteen alleles were identified in the 159 patients with gastric carcinoma, all of which were also present in the control population. Alleles 1 and 14 from the control population were not identified in patients with gastric carcinoma. Allelic frequencies are represented in Table 2 and Figure 2.

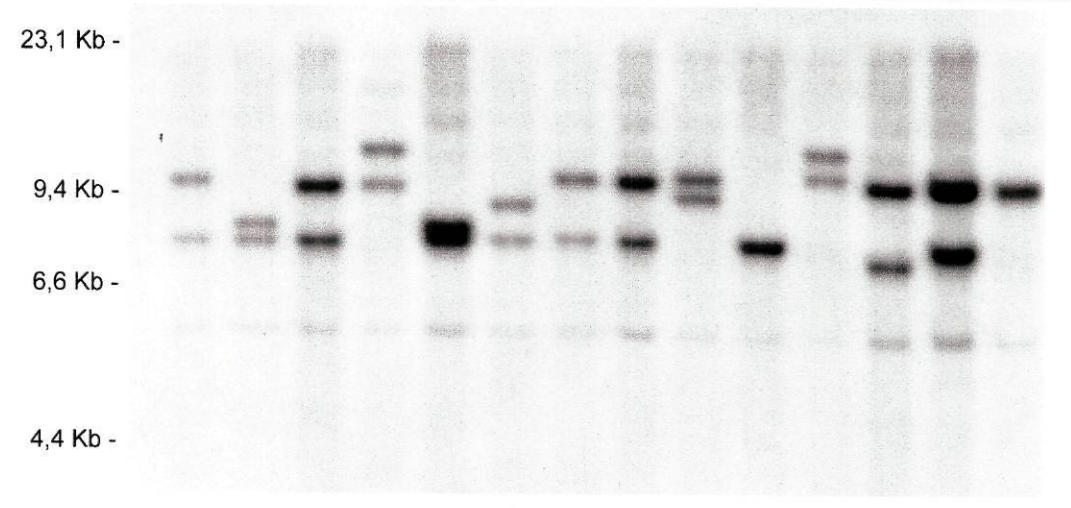


Figure 1. Detection of *MUC1* gene polymorphism. DNA samples from blood donors were digested with *EcoR*I, separated on 0.7% agarose gel, and transferred to nylon membranes by alkaline blotting. Membranes were hybridized with a MUC1 probe. Lanes 1-14 correspond to the following genotypes: 4-10; 7-10; 5-10; 1-4; 9-10; 6-10; 4-10; 4-10; 4-6; 10-10; 2-3; 4-12; 4-10; 4-4. Size marker is given in kilobase pairs.

Table 2. Distribution of *MUC1* allelic frequencies in blood donors and in patients with gastric carcinoma.

MUC1 alleles	Blood donors (n=324)	Patients with gastric carcinoma (n=159)
1	0.005	0.000
2	0.006	0.013
3	0.045	0.016
4	0.285	0.164
5	0.119	0.091
6	0.077	0.101
7	0.031	0.035
8	0.062	0.079
9	0.035	0.031
10	0.256	0.381
11	0.022	0.038
12	0.020	0.031
13	0.023	0.019
14	0.005	0.000
15	0.009	0.003

$p < 0.001$

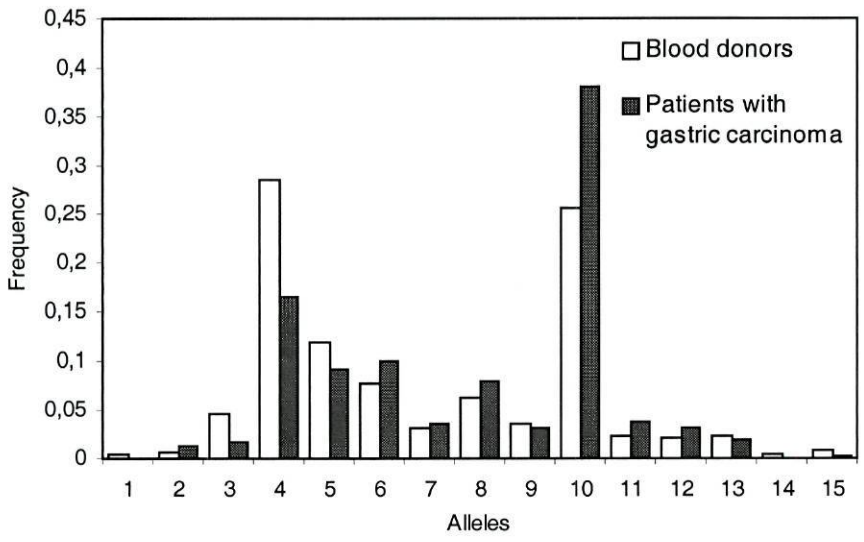


Figure 2. Distribution of allelic frequencies of *MUC1* gene in blood donors and in patients with gastric carcinoma. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the largest).

Alleles 4 and 10 were the most common alleles both in the control group and in patients with gastric carcinoma (Figure 2). Despite sharing almost the same alleles and having the same modal alleles, a significant difference in the distribution of allelic frequencies between blood donors and gastric carcinoma patients was observed. The majority of the alleles with a relatively high molecular weight (alleles 1 to 5, except allele 2) are more frequent in the control group whereas alleles with lower molecular weight (alleles 6 to 14, except alleles 9 and 13) are more frequent in patients with gastric carcinoma (X^2= 41.6; $p<0.001$). Due to the high number of alleles and the sparse representation of some of them we performed the Monte Carlo based test that confirmed a significantly ($p<0.001$) different distribution of the *MUC1* alleles between patients and controls.

We further evaluated the *MUC1* genotypes in the two populations. Observed heterozygosity was 72.2% in controls and 73.0% in patients with gastric carcinoma. The heterozygote 4-10 was the most frequent genotype in both populations: 44 of 324 controls (13.6%) and 28 of 159 patients (17.6%). The homozygous genotype for allele 4 was more frequent in controls than in patients with gastric carcinoma – 36 individuals (11.1%) and 3 patients (1.9%), respectively – and the homozygous genotype for allele 10 was less frequent in controls than in patients with gastric carcinoma – 28 individuals (8.6%) and 24 patients (15.1%), respectively.

Sixty six different genotypes were identified overall, most of them with a low number of individuals in both populations. Therefore, we grouped the genotypes dividing the alleles by the cut-point that corresponds to the median value of the blood donors population. The selected cut-point is also the one that maximizes the differences between the two populations: alleles 1-5 were grouped into a category of Large *MUC1* alleles (L) and alleles 6-15 were grouped into a category of Small *MUC1* alleles (S). Genotypes were thereafter recoded as having two Large *MUC1* alleles (LL), two Small *MUC1* alleles (SS), and one Large *MUC1* allele and one small *MUC1* allele (LS).

The comparison between patients and controls regarding the recoded MUC1 genotypes showed significant differences between patients and controls (Table 3). Genotypes for Large *MUC1* alleles (LL) are more frequent in controls (20.7%) than in patients with gastric carcinoma (7.5%) and genotypes for Small *MUC1* alleles (SS) are more frequent in patients with gastric carcinoma (50.9%) than in controls (29.3%) ($p<0.0001$). Genotypes with one Large *MUC1* allele and one Small *MUC1* allele (LS) have a similar frequency in both groups (50.0% for controls and 41.5% for patients with gastric carcinoma). The estimated relative risk (odds ratio), adjusted for age and gender, for gastric carcinoma development in individuals with SS *MUC1* genotype is 4.3 (95% confidence limits 1.8-10.5).

Table 3. Distribution of recoded *MUC1* genotypes, LL (two large *MUC1* alleles), LS (one large *MUC1* allele and one small *MUC1* allele) and SS (two small *MUC1* alleles), in blood donors and in patients with gastric carcinoma.

Recoded *MUC1* genotypes	Blood donors (n=324) n (%)	Patients with gastric carcinoma (n=159) n (%)
LL	67 (20.7%)	12 (7.5%)*
LS	162 (50.0%)	66 (41.5%)
SS	95 (29.3%)	81 (50.9%)*

* $p<0.0001$

Genotype distribution in gastric carcinoma patients was identical in gastric carcinomas with different morphologic types (p=0.611) (data not shown).

Loss of heterozygosity was not detected for *MUC1* gene when analysis of paired samples of gastric carcinoma cases (normal mucosa *versus* tumor) was performed. From the 159 paired samples of gastric carcinoma cases studied for *MUC1* gene polymorphism one case showed *MUC1* gene recombination.

Ib – *MUC1* CA repeat polymorphism

The results obtained with the two different methodologies (radioactive and non-radioactive PCR) were similar. Three different alleles 1, 2 and 3 with 176 bp $[(CA)_{11}]$, 178 bp $[(CA)_{12}]$ and 180 bp $[(CA)_{13}]$ respectively, were detected in the Portuguese blood donors (n= 207) and in patients with gastric carcinoma (n= 76) (Table 4, Figure 3). There are differences in the allelic frequencies between Portuguese blood donors and gastric carcinoma patients (Table 4). Allele 1 was more frequent in blood donors whereas allele 2 was more frequent in gastric carcinoma patients. A low representation of allele 3 was observed in both populations (Table 4). The heterozygous genotype 1-2 was the most frequent genotype in both populations (44.0% in blood donors and 39.5% in patients), followed by the genotypes 1-1 (29.0% in blood donors and 15.8% in patients), 2-2 (15.4% in blood donors and 30.3% in patients), 2-3 (6.8% in blood donors and 6.6% in patients) and 1-3 (4.8% in blood donors and 7.9% in patients). The homozygous genotype 3-3 was not detected in both populations. This locus is in Hardy-Weinberg equilibrium (p=0.16).

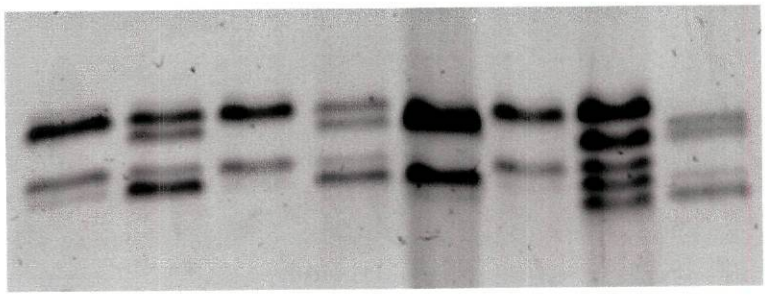


Figure 3. Silver stained gel of the CA repeat polymorphism in exon 6 of the *MUC1* gene. Lanes 1-8 correspond to the following genotypes: 1-1; 1-2; 1-1; 1-2; 2-2; 1-1; 1-3; 2-3.

Table 4. Distribution of *MUC1* CA repeat allelic frequencies in blood donors and in patients with gastric carcinoma.

MUC1 CA repeat alleles	Blood donors (n=207)	Patients with gastric carcinoma (n=76)
1 - $(CA)_{11}$	0.53	0.40
2 - $(CA)_{12}$	0.41	0.53
3 - $(CA)_{13}$	0.06	0.07

p= 0.0135

For association studies between VNTR polymorphism and CA repeat polymorphism we compared the recoded *MUC1* genotypes, as determined with *Eco*RI, with the genotypes and the alleles of the CA repeat polymorphism. Table 5 shows a statistical significant association between the genotype SS and the genotype 2-2 as well as with a higher frequency of allele 2. Genotype LL is associated with genotype 1-1 and with a higher frequency of allele 1 of the CA repeat. The same associations were observed for the patients with gastric carcinoma (data not shown).

Table 5. Distribution of recoded *MUC1* genotypes, LL (two large *MUC1* alleles), LS (one large *MUC1* allele and one small *MUC1* allele) and SS (two small *MUC1* alleles), in different CA repeat genotypes and alleles from blood donors of Northern Portugal.

VNTR genotypes	CA repeat genotypes					CA repeat alleles		
	1-1	1-2	1-3	2-2	2-3	1	2	3
LL	15	1	2	0	0	33	1	2
LS	12	40	3	0	3	67	43	6
SS	0	2	0	16	9	2	43	9
		p=0.0001				p=0.0001		

Because VNTR polymorphism as determined with *Eco*RI is not in Hardy-Weinberg equilibrium (Paper III) we decided to analyse a sub-sample of Portuguese blood donors with a different restriction enzyme, *Alu*I, that gives rise to smaller fragments. With this approach the *MUC1* locus is in Hardy-Weinberg equilibrium (Paper III). When we compared the *MUC1* genotypes as determined with *Alu*I with the genotypes of the CA repeat polymorphism we confirmed a significant association between the *MUC1* VNTR polymorphism, determined with *Alu*I, and the CA repeat polymorphism within intron 6, in the Portuguese blood donors. The statistical significance of the association was dependent on the statistical test applied and was stronger with Fisher (p=0) than with the X^2 (p=0.09) because of the large number of possible combinations of the different alleles in *Alu*I genotypes and consequent high number of degrees of freedom.

Ic – *MUC6* VNTR polymorphism (Paper II)

Ten different alleles with sizes ranging from 8.5 Kbp to 12.5 Kbp were identified in Portugueses blood donors (n= 376) (Figure 4, Table 6). The alleles were also numbered according to their molecular weight (allele 1 is the largest and allele 10 the smallest) and the most frequent alleles were: allele 4 (11.2 Kb; 37.5%), allele 7 (10.0 Kb; 20.9%) and allele 2 (12.0 Kb; 17.2%). Genotype distribution was in accordance with Hardy-Weinberg equilibrium expectations and the observed heterozygosity was 75.8%.

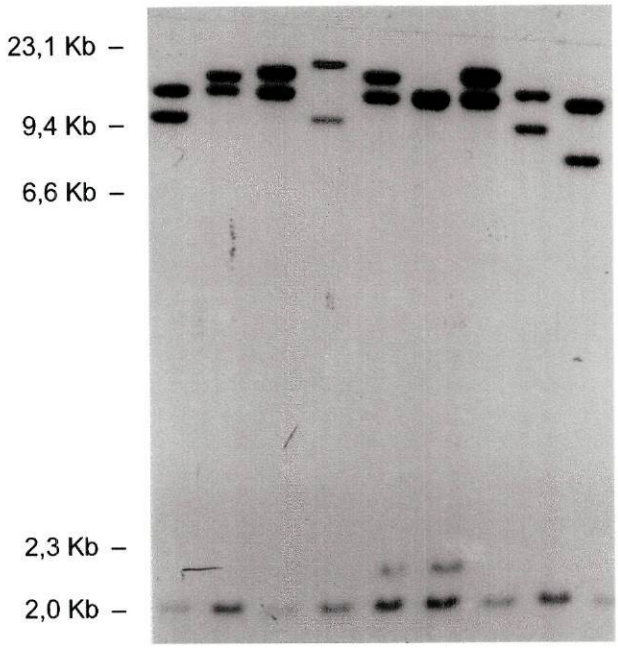


Figure 4. Detection of *MUC6* gene polymorphism. DNA samples from blood donors were digested with *Taq*I, separated on 0.7% agarose gel, and transferred to nylon membranes by alkaline blotting. Membranes were hybridized with a MUC6 probe. The constant band at 2.0Kbp represents the *Taq*I fragment incorporating the flanking DNA and the terminal end of the repeat. The polymorphism observed in two lanes, with an additional band at 2.1 kbp, represents sequence polymorphism in one of the two alleles. Size marker is given in kilobase pairs.

Table 6. Distribution of *MUC6* allelic frequencies in blood donors and in patients with gastric carcinoma.

MUC6 alleles	Blood donors (n=376)	Patients with gastric carcinoma (n=157)
1*	0.085	0.041
2	0.172	0.185
3	0.092	0.064
4	0.375	0.395
5*	0.012	0.045
6	0.021	0.019
7	0.209	0.188
8*	0.005	0.019
9	0.025	0.025
10*	0.004	0.019

p= 0.0004

*Alleles responsible for the differences between the two populations.

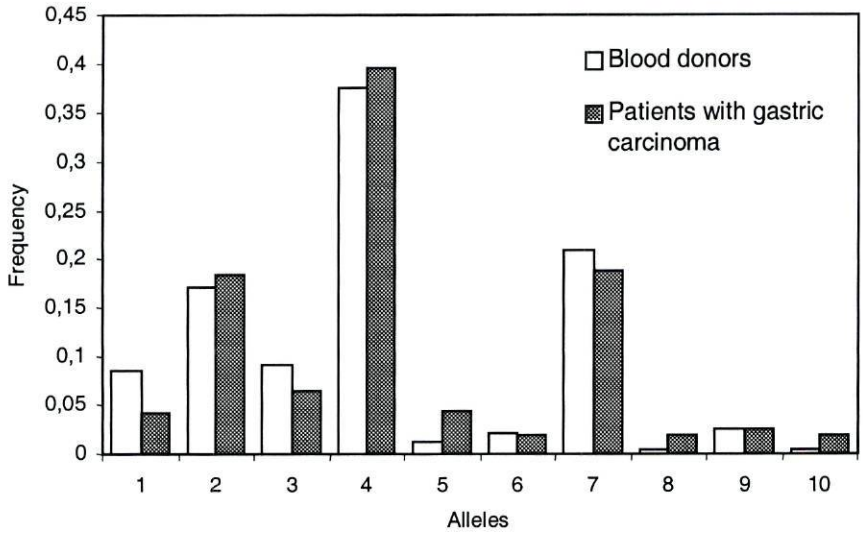


Figure 5. Distribution of allelic frequencies of *MUC6* gene in blood donors and in patients with gastric carcinoma. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the largest).

In gastric carcinoma patients the same 10 different alleles of the control group were identified and the modal alleles were also the same in both populations: allele 4, allele 7 and allele 2 (Table 6, Figure 5). Significant differences in allelic frequencies between blood donors and gastric carcinoma patients were observed for alleles 1, 5, 8 and 10. Allele 1, the largest allele, was more frequent in blood donors; alleles 5, 8 and 10, of smaller sizes, were more frequent in gastric carcinoma patients (Table 6, Figure 5). Genotype distribution in gastric cancer patients was as expected from Hardy-Weinberg equilibrium and the observed heterozygosity was 79.6%.

Allele distribution in gastric carcinoma patients was identical in gastric carcinomas with different morphologic types (p=0.628) (data not shown).

From the 157 paired samples of gastric carcinoma cases studied for *MUC6* gene polymorphism, 7 out of 125 informative cases (5.6%) showed allelic imbalance in comparison to the normal gastric mucosa. In a single case *MUC6* gene recombination was observed with the appearance of an additional band in the tumour sample.

Figure 6 shows *MUC1* alleles plotted against *MUC6* alleles. It is possible to observe that the distribution of *MUC1* alleles is independent from the distribution of *MUC6* alleles.

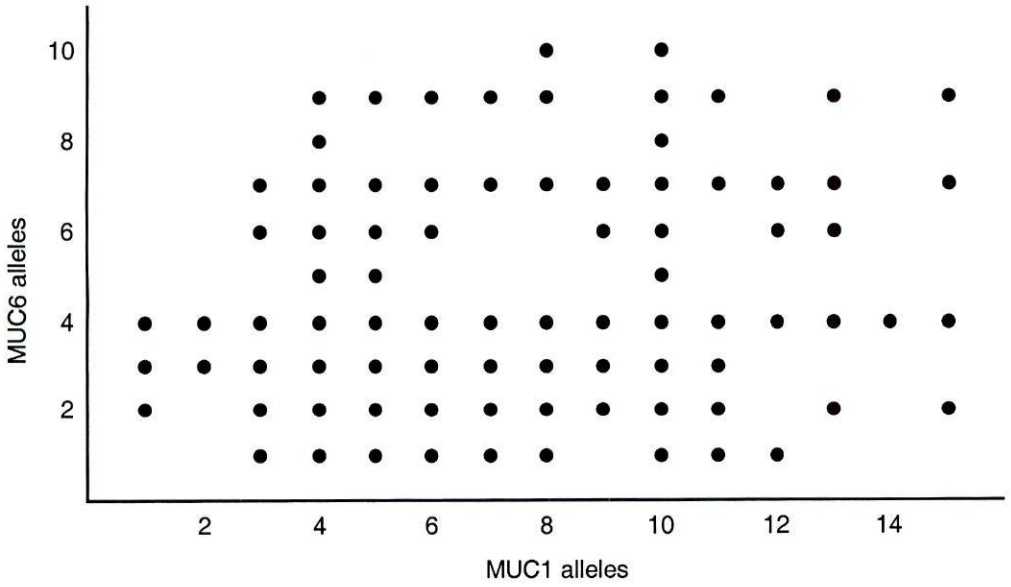


Figure 6. *MUC1* alleles were plotted against *MUC6* alleles.

II – *MUC1* gene polymorphism and *Helicobacter pylori* infection

IIa – *MUC1* gene polymorphism in Portugal and Denmark: relation with *Helicobacter pylori* infection (Part of these results are described in Paper III)

In the 324 blood donors from northern Portugal, with ages ranging from 19 to 64 years (38.0±0.9y) the male:female ratio was 4.6:1. In the 100 blood donors from Denmark, with ages ranging from 21 to 65 years (42.0±1.0y) the male:female ratio was 0.9:1. There were no differences between the two populations regarding ABO histo-blood groups distribution (p=0.25) (data not shown).

MUC1 gene polymorphism in the two populations using *Eco*RI

Allelic frequencies of *MUC1* gene estimated from the Danish samples differ from those previously described in the Portuguese blood donors (Carvalho *et al*, 1997 – Paper I) (Table 7, Figure 7) (p=0.0005), despite displaying the same modal peaks: allele 4 and allele 10, with allele 4 (one large *MUC1* allele) being the most frequent in Portuguese blood donors (28.5%) and allele 10 (one small *MUC1* allele) being the most frequent in Danish blood donors (30.1%).

The alleles responsible for the difference are alleles 2, 3 and 10. The smallest alleles (alleles 14 and 15) were not detected in Danish blood donors (Table 7, Figure 7).

With this restriction enzyme, allelic frequencies for *MUC1* gene are not in Hardy-Weinberg equilibrium in Portuguese blood donors (p=0.01) at variance with the findings in Danish blood donors (p=0.671). The heterozygosity was 72.2% and 77.0% for Portuguese and Danish blood donors, respectively. Observed and expected heterozygosities do not deviate significantly in the Danish sample. On the contrary, a large homozygote excess was observed in the Portuguese sample (X^2=23.58, p<0.005).

MUC1 gene polymorphism in the two populations using *Alu*I

Twenty different alleles were identified using *Alu*I as restriction enzyme (Table 7, Figure 8, Figure 9). The distribution of the alleles was bimodal as that obtained with *Eco*RI. There were differences in the allelic frequencies of Portuguese (n= 96) and Danish (n= 99) blood donors (p=0.01); however, the two modal alleles were the same in both populations - allele 5 and allele 16 - with allele 5 being the most frequent in the Portuguese blood donors (17.2%) and allele 16 being the most frequent in the Danish bloos donors (23.2%).

Allele 1 was not detected in the Portuguese blood donors. Six of the smallest alleles (alleles 14, 15, 16, 17, 19 and 20) were more frequent in Danish than in Portuguese blood donors.

Allelic frequencies determined with the use of *Alu*I are in Hardy-Weinberg equilibrium for both populations.

The heterozigosity was 86.5% and 92.9% for Portuguese and Danish blood donors, respectively. Observed and expected heterozygosities do not deviate significantly in both populations.

Table 7. Allelic frequencies of *MUC1* gene in Portuguese and Danish blood donors, using *Eco*RI and *Alu*I as restriction enzymes.

***Eco*RI**			***AluI***		
MUC1 Alleles	Portuguese blood donors (n=324)	Danish blood donors (n=100)	*MUC1* Alleles	Portuguese blood donors (n=96)	Danish blood donors (n=99)
1	3 (0.005)	1 (0.005)	1	0 (0)	5 (0.025)
2	4 (0.006)	5 (0.025)	2	1 (0.005)	1 (0.005)
3	29 (0.045)	2 (0.010)	3	1 (0.005)	2 (0.010)
4	185 (0.285)	46 (0.230)	4	2 (0.021)	4 (0.020)
5	77 (0.119)	25 (0.125)	5	33 (0.172)	24 (0.121)
6	50 (0.077)	8 (0.040)	6	13 (0.068)	14 (0.071)
7	20 (0.031)	5 (0.025)	7	26 (0.135)	20 (0.101)
8	40 (0.062)	5 (0.025)	8	9 (0.047)	12 (0.061)
9	23 (0.035)	12 (0.060)	9	11 (0.057)	2 (0.010)
10	166 (0.256)	80 (0.400)	10	1 (0.005)	4 (0.020)
11	14 (0.022)	1 (0.005)	11	12 (0.063)	3 (0.015)
12	13 (0.020)	6 (0.030)	12	7 (0.036)	1 (0.005)
13	15 (0.023)	4 (0.020)	13	11 (0.057)	9 (0.045)
14	3 (0.005)	0 (0)	14	7 (0.036)	8 (0.040)
15	6 (0.009)	0 (0)	15	15 (0.078)	23 (0.116)
-	-	-	16	27 (0.141)	46 (0.232)
-	-	-	17	7 (0.036)	8 (0.040)
-	-	-	18	3 (0.016)	1 (0.005)
-	-	-	19	3 (0.016)	7 (0.035)
-	-	-	20	3 (0.016)	4 (0.020)
	X^2= 38.1; p=0.0005			X^2= 35.6; p=0.01	

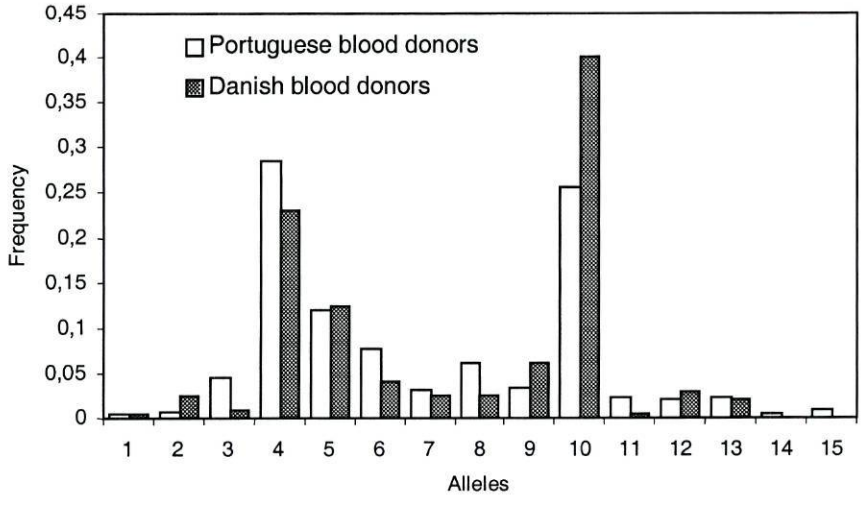


Figure 7. Allelic frequencies distribution of *MUC1* gene in blood donors from Portugal and Denmark, using *Eco*RI as restriction enzyme. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the largest).

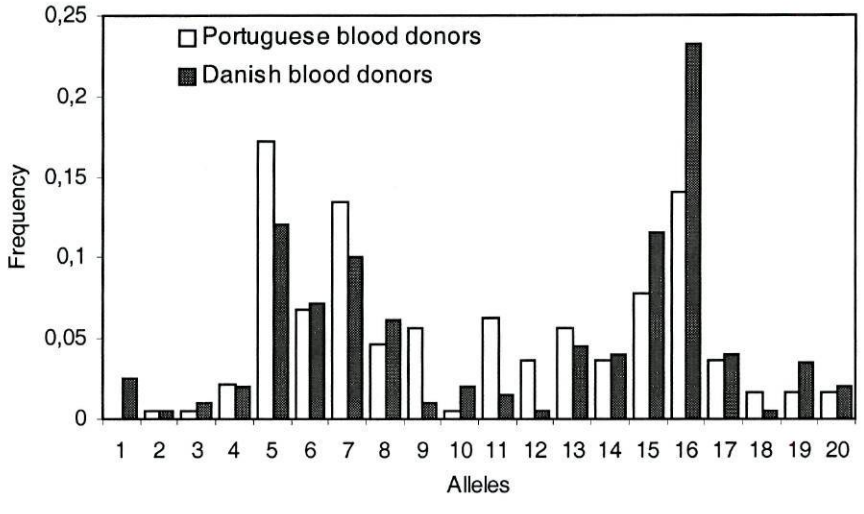


Figure 8. Allelic frequencies distribution of *MUC1* gene in blood donors from Portugal and Denmark, using *Alu*I as restriction enzyme. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the largest).

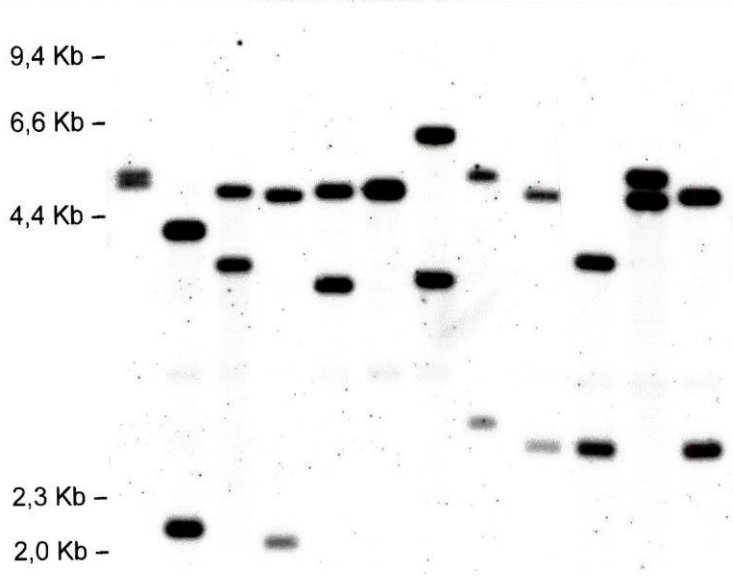


Figure 9. Detection of *MUC1* gene polymorphism. DNA samples from blood donors were digested with *Alu*I, separated on a 0.7% agarose gel, and transferred to nylon membranes by alkaline blotting. Membranes were hybridized with a MUC1 probe. Lanes 1-12 correspond to the following genotypes: 5-6, 8-19, 7-11, 7-20, 7-12, 7-7, 3-12, 5-14, 7-16, 11-16, 5-7, 7-16. Size marker is given in kilobase pairs.

Comparison of the serum levels of anti-*Helicobacter pylori* IgG antibodies in the two populations

The serum levels of anti-*H. pylori* IgG antibodies ranged from 0 to 1674 U/ml in donors from northern Portugal (n=151) and from 0 to 553 U/ml in Danish donors (n=100) (Figure 10). The serum levels of IgG anti-*H. pylori* were different between the two populations with a significantly higher mean titre in the Portuguese blood donors (154±20 U/ml) than in the Danish blood donors (32±9 U/ml) ($p<0.0001$).

Using the cut-off value , established for the Portuguese population and corroborated by the study of Raymond *et al* (1996), the prevalence of seropositivity was 82.1% and 18.0% in Portuguese and Danish blood donors, respectively.

The mean titre of antibodies in the 124 Portuguese men (171±23 U/ml) was suggestively higher ($p=0.06$) than that of the 27 women (74±23 U/ml). No differences were observed in the anti-*H. pylori* IgG distribution by gender in Danish blood donors. The mean titre of anti-*H. pylori* IgG antibodies did not differ with the age of the individuals nor with the blood group in both populations (data not shown).

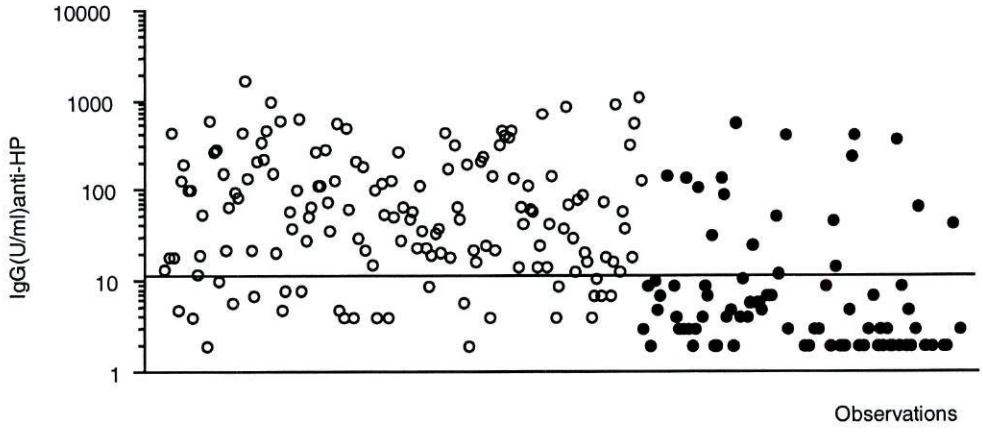


Figure 10. Scattergram (Log10 scale) of the titres of anti-*H.pylori* IgG antibodies in the 151 Portuguese blood donors (o) and in the 100 Danish blood donors (•). Horizontal line represents the cut-off value (11.5U/ml). The titres of most Portuguese donors are above the cut-off value at contrast to those of Danish donors.

Relationship between the levels of anti-H. pylori IgG antibodies and *MUC1* gene polymorphism

Considering the recoded *MUC1* genotypes in the *Eco*RI system, it was possible to observe that, in both populations, individuals with SS *MUC1* genotypes, having two small *MUC1* alleles, had higher levels of antibodies than individuals with LL *MUC1* genotypes (Table 8). The mean titres of antibodies were suggestively different (p=0.06) between genotypes SS and LL for the Portuguese population but did not attain the threshold of statistical significance in the Danish population (Table 8).

Table 8. Mean titres (± S.E.) of anti-*Helicobacter pylori* antibodies (U/ml) in individuals with LL, LS and SS *MUC1* genotypes, in Portuguese (n=151) and Danish (n=100) blood donors.

Recoded *MUC1* genotypes	Portuguese blood donors	Danish blood donors
LL	139±39*	7±4*
LS	140±28	45±16
SS	192±41*	20±12*
*p value	0.06	0.54

Relationship between the levels of anti-H. pylori IgG antibodies and *MUC6* gene polymorphism in Portuguese blood donors

The comparison of anti-*H. pylori* IgG antibodies and the polymorphism of *MUC6* gene showed that there was no significant association between different mean titres of anti-*H. pylori* IgG antibodies and different *MUC6* alleles (data not shown).

IIb – *MUC1* gene polymorphism in dyspeptic patients from Northern Portugal: relation with *Helicobacter pylori* infection and other pathological parameters

The group of 312 dyspeptic patients is composed by individuals selected at Hospital S. João (n= 120) and individuals recruited from a shipyard in northern Portugal (n=192). The mean age of these groups was 48.3±1.0 y and 43.0±0.4 y, respectively. The male:female ratio was 1.5:1 in the first group and 12.5:1 in the second group.

There were no differences in the ABO distribution (p=0.883) between the two populations (data not shown). Although there was a difference in the male:female ratio between the two groups, we decided to group all these individuals in a global group of 312 dyspeptic patients based on the presence of common clinicopathological diagnosis.

In the 312 dyspeptic patients the distribution of *MUC1* allelic frequencies, determined with *Eco*RI, was significantly different (p=0.0001) from that previously described for the Portuguese blood donors (Figure 11). However, the frequency of recoded *MUC1* genotypes did not differ significantly (p=0.304) from that of blood donors (Table 9).

Table 9. Distribution of recoded *MUC1* genotypes, LL (two large *MUC1* alleles), LS (one large *MUC1* allele and one small *MUC1* allele) and SS (two small *MUC1* alleles), in blood donors and in dyspeptic patients from Northern Portugal.

Recoded *MUC1* genotypes	Blood donors (n=324) n (%)	Dyspeptic patients (n=312) n (%)
LL	67 (20.7%)	62 (19.9%)
LS	162 (50.0%)	141 (45.2%)
SS	95 (29.3%)	109 (34.9%)

p=0.30

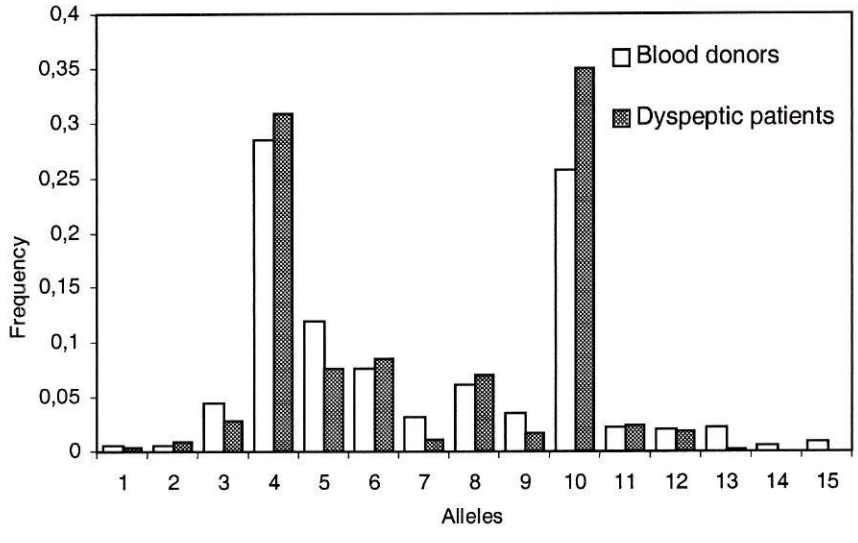


Figure 11. Distribution of allelic frequencies of *MUC1* gene, using *Eco*RI, in blood donors and dyspeptic patients from Northern Portugal.

H. pylori infection was evaluated using two different methodologies: serology, as previously described for the blood donors, and histology.

The serum levels of anti-*H. pylori* IgG antibodies ranged from 0 to 1622 U/ml, with a mean of 130±11 U/ml. Using the cut-off value of 11.5 U/ml the seroprevalence was 82.5 %, which is very similar to the one observed in the Portuguese blood donors. When we concentrate on the histology results the percentage of individuals with *H. pylori* identified in the gastric biopsies was 86.5%. The comparison of both methodologies, performed in 262 dyspeptic patients, demonstrated a high agreement between them: 91.1% of positive cases in serology had *H. pylori* in the biopsy and 48.0% of negative cases in serology had no *H. pylori* in the biopsy ($p<0.0001$) (Table 10). However, there are some discrepancies between both methods: 13 cases with negative serology had *H. pylori* in the biopsy and 21 cases had positive serology and no *H. pylori* in the biopsy.

Table 10. Comparison between the results obtained with serology and histology for the evaluation of *H. pylori* infection in dyspeptic patients.

		Serology	
		Positive	Negative
Histology	Positive	216	13
	Negative	21	12

$p<0.0001$

The comparison of the recoded *MUC1* genotypes with the titres of anti-*H-pylori* IgG antibodies, performed in 256 dyspeptic patients, showed no significant differences between the mean titres of IgG in the different groups (Table 11) although there is a trend for an higher IgG mean titre in the group of individuals with SS *MUC1* genotypes.

Table 11. Mean titres (± S.E.) of anti-*Helicobacter pylori* antibodies (U/ml) in 256 dyspeptic patients with LL, LS and SS *MUC1* genotypes.

Recoded *MUC1* genotypes	Dyspeptic patients
LL	141±32
LS	139±22
SS	153±24
p value	0.89

Similarly, there were no significant differences between the distribution of *MUC1* genotypes and the presence or absence of *H. pylori* in the gastric biopsies (data not shown).

Since we had previously described an association between the small *MUC1* genotypes and the risk for gastric carcinoma development we decided to find whether or not there is an association of *MUC1* genotypes with pre-neoplastic lesions of the gastric mucosa. Table 12 summarises the distribution of recoded *MUC1* genotypes according to the different morphologic aspects of the gastric mucosas observed in the biopsies of 281 dyspeptic patients. We observded a significant association between SS *MUC1* genotypes and glandular atrophy of the gastric mucosa ($p=0.03$): individuals with SS *MUC1* genotypes had a relatively higher frequency of glandular atrophy than individuals

with LL or LS *MUC1* genotypes (Table 12).No significant association was found between *MUC1* genotypes and intestinal metaplasia (p=0.53).

Table 12. Distribution of recoded *MUC1* genotypes according to the different morphologic aspects of the gastric mucosa of dyspeptic patients.

		Recoded *MUC1* genotypes			p value
		LL	LS	SS	
		n (%)	n (%)	n (%)	
Glandular Atrophy*	Absent	32 (60.4)	93 (71.5)	54 (55.1)	0.03
	Present	21 (39.6)	37 (28.5)	44 (44.9)	
Intestinal Metaplasia**	Absent	31 (59.6)	86 (70.5)	58 (59.8)	0.53
	Complete	10 (19.2)	12 (9.8)	13 (13.4)	
	Incomplete	11 (21.2)	24 (19.7)	26 (26.8)	

*281 biopsies were evaluated
**271 biopsies were evaluated

III – Mucins expression and *Helicobacter pylori* adherence to gastric carcinoma cell lines

IIIa – Characterization of expression of mucins and mucin-associated carbohydrate antigens in gastric carcinoma cell lines (Paper IV)

Northern blotting

The results of the Northern blotting study of different mucins and actin are summarised in Table 13. *MUC1* mRNA was detected in GP202, GP220, MKN45 and KATOIII cell lines (Figure 12). Low levels of *MUC2* mRNA were only detected in KATOIII. GP220 expresses *MUC3* mRNA. Low levels of *MUC3* and *MUC4* were detected in St23132 and KATOIII cell lines, respectively. *MUC5AC* mRNA was detected in St23132, St2957 and St3051 cell lines. *MUC5B, MUC6* and *MUC7* mRNAs were not detected in any of the cell lines.

Table 13. Northern blotting expression of mucins and actin in the different gastric carcinoma cell lines. The signal was scored as: - (no expression), +/- (low levels of expression), and + (high levels of expression).

Cell line	*MUC1*	*MUC2*	*MUC3*	*MUC4*	*MUC5AC*	*MUC5B*	*MUC6*	*MUC7*	Actin
GP202	+	-	-	-	-	-	-	-	+
GP220	+	-	+	-	-	-	-	-	+
MKN45	+	-	-	-	-	-	-	-	+
KATOIII	+	+/-	-	+/-	-	-	-	-	+
St23132	-	-	+/-	-	+	-	-	-	+/-
St2957	-	-	-	-	+	-	-	-	+/-
St3051	-	-	-	-	+	-	-	-	+/-

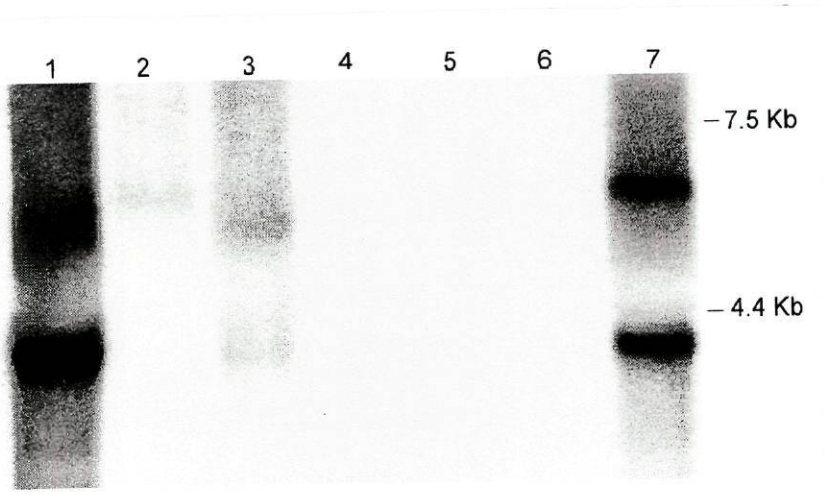


Figure 12. Northern blot analysis of *MUC1* mucin in gastric carcinoma cell lines. Lanes 1-7 correspond to: GP202, GP220, MKN45, St23132, St2957, St3051 and KATOIII. RNA size marker is given in kilobases.

RT-PCR (Reverse Transcription - Polymerase Chain Reaction)

Table 14 summarises the results from RT-PCR. All cell lines expressed *MUC1* (Figure 13). *MUC2* expression was detected in KATOIII and St2957 cell lines, and *MUC3* expression was detected in GP220 and St23132 cell lines. *MUC5AC* was detected in five cell lines and not in GP202 and KATOIII cell lines (Figure 14). *MUC5B* was detected in four cell lines: MKN45, St23132, St2957 and St3051. None of the cell lines expressed *MUC4*, *MUC6* or *MUC7* as detected by RT-PCR.

Table 14. RT-PCR results of mucins expression in the different gastric carcinoma cell lines. The signal was scored as : - (negative) or + (positive).

Cell line	*MUC1*	*MUC2*	*MUC3*	*MUC4*	*MUC5AC*	*MUC5B*	*MUC6*	*MUC7*
GP202	+	-	-	-	-	-	-	-
GP220	+	-	+	-	+	-	-	-
MKN45	+	-	-	-	+	+	-	-
KATOIII	+	+	-	-	-	-	-	-
St23132	+	-	+	-	+	+	-	-
St2957	+	+	-	-	+	+	-	-
St3051	+	-	-	-	+	+	-	-

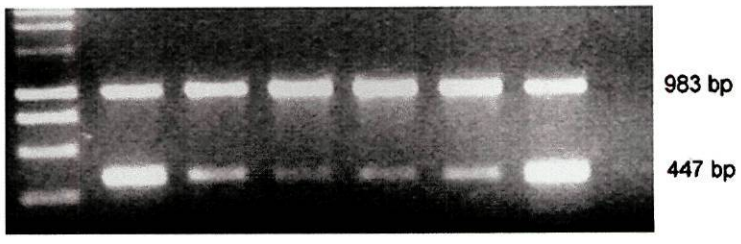


Figure 13. RT-PCR analysis of *MUC1* (447 bp product) and GAPDH (983 bp product) in gastric carcinoma cell lines. Lanes 1-8 correspond to: size marker, GP202, GP220, St2957, St3051, St23132, MNK45 and KATOIII cell lines.



Figure 14. RT-PCR analysis of *MUC5AC* (409 bp product) and GAPDH (983 bp product) in gastric carcinoma cell lines. Lanes 1-8 correspond to: GP202, GP220, St2957, St3051, St23132, MNK45 and KATOIII cell lines, and size marker.

Immunohistochemistry

The results from immunohistochemistry of the seven gastric carcinoma cell lines with the different monoclonal antibodies to the core proteins of mucins are summarised in Table 15. All cell lines expressed MUC1 and MUC5AC at different levels, except for GP220 that did not express MUC5AC. Expression of MUC1 was detected in all cell lines using three different monoclonal antibodies: HMFG-1 MAb stained more cells than HMFG-2 and HMFG-2 stained more cells than SM3. Four cell lines expressed MUC2: GP202 and St2957 at low levels and MKN45 and KATO III at high levels. None of the cell lines expressed MUC6. HL60, a myeloid leukaemia cell line, did not expressed any of the mucins studied.

The expression of mucin-associated carbohydrate antigens is summarised in Table 16. All the cell lines expressed high levels of type 1 and type 2 chain Lewis antigens. In most cell lines, the staining for the sialylated epitopes, S-Le^a and S-Le^x, was lower than that of the respective non-sialylated antigens (Le^a and Le^x, respectively). Exceptions are MKN45 and St2957 cell lines that show similar levels of Le^a and S-Le^a, KATOIII that shows higher levels of S-Le^x than Le^x, and St23132 that shows similar levels of Le^x and S-Le^x. The expression of simple mucin-type antigens is low or absent in every cell line. Expression of Tn antigen was observed in four cell lines: GP202, MKN45, St23132 and St2957. The sialylated form, S-Tn, was observed in MKN45, St23132, St2957, St3051

cell lines. T antigen was only detected in cell line GP202. HL60 cell line only expressed Le^x and S-Le^x antigens.

Table 15. Immunohistochemical expression of mucin proteins in the different cell lines. A semi-quantitative approach was used to score the staining: + (<25%), ++ (25-50%), +++ (50-75%) and ++++ (>75%).

	MUC1			MUC2	MUC5AC	MUC6
Cell line	HMFG-1	HMFG-2	SM3	PMH1	CLH2	CLH5
GP202	++++	++++	++	+	++++	-
GP220	+++	+	-	-	-	-
MKN45	++++	++++	++	++++	++++	-
KATO III	+++	+	-	++++	+	-
St23132	+++	+	-	-	++++	-
St2957	+++	-	-	+	+++	-
St3051	+++	+	-	-	+	-
HL60	-	-	-	-	-	-

Table 16. Immunohistochemical expression of mucin-associated carbohydrate antigens in the different cell lines. A semi-quantitative approach was used to score the staining: + (<25%), ++ (25-50%), +++ (50-75%) and ++++ (>75%).

	Simple mucin-type carbohydrate antigens			Lewis antigens					
				Type 1 chain			Type 2 chain		
Cell line	Tn	S-Tn	T	Le^a	S-Le^a	Le^b	Le^x	S-Le^x	Le^y
	5F4	3F1	3C9	Ca3F4	HB8059	BG.6	SH1	FH6	AH6
GP202	+	-	+	++	-	++++	++++	+	++++
GP220	-	-	-	++++	-	+++	++++	-	++++
MKN45	+	+	-	+	+	+	+++	++	-
KATOIII	-	-	-	++++	++	+	++	++++	++
St23132	+	+	-	++++	+++	++++	++	++	+++
St2957	+	+	-	++	++	++	++	+	+++
St3051	-	+	-	+++	++	++	+++	++	++++
HL60	-	-	-	-	-	-	+	+	-

Comparison of the results obtained by Northern blotting, RT-PCR and immunohistochemistry

MUC1 mRNA detected by RT-PCR as well as protein expression detected by immunohistochemistry were found in all gastric carcinoma cell lines. In St23132, St2957 and St3051 cell lines *MUC1* mRNA was not detected by Northern blotting (Table 17).

MUC2 protein expression was detected by immunohistochemistry in four cell lines: in KATOIII cell line mRNA was also detected by Northern blotting and RT-PCR, in St2957 cell line mRNA was detected by RT-PCR alone, and in GP202 and MKN45 cell lines mRNA detection was not achieved by Northern blotting or by RT-PCR (Table 17).

MUC3 mRNA was detected in GP220 and St23132 cell lines, both by Northern blotting and RT-PCR (Table 17).

MUC4 mRNA was detected in KATOIII cell line by Northern blotting alone (Table 17).

MUC5AC mRNA detected by RT-PCR and Northern blotting, as well as protein expression detected by immunohistochemistry, were found in St23132, St2957 and St3051 cell lines (Table 17). The four remaining cell lines showed discrepant results with the different methodologies: MKN45 cell line showed mRNA only by RT-PCR as well as protein expression; GP202 and KATOIII cell lines showed protein expression alone, and GP220 cell line showed mRNA expression detected by RT-PCR alone (Table 17).

MUC5B mRNA was detected in MKN45, St23132, St2957 and St3051 cell lines by RT-PCR without mRNA expression detected by Northern blotting (Table 17).

MUC6 and *MUC7* mucins were not detected in any of the cell lines (Table 17).

Table 17. Comparison between the results of mucin expression by Northern blotting, RT-PCR and immunohistochemistry. Results are expressed as positive + or negative -.

	MUC1			MUC2			MUC3		MUC4		MUC5AC			MUC5B		MUC6			MUC7	
Cell line	N	RT	I	N	RT	I	N	RT	N	RT	N	RT	I	N	RT	N	RT	I	N	RT
GP202	+	+	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-
GP220	+	+	+	-	-	-	+	+	-	-	-	+	-	-	-	-	-	-	-	-
MKN45	+	+	+	-	-	+	-	-	-	-	-	+	+	-	+	-	-	-	-	-
KATOIII	+	+	+	+	+	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-
St23132	-	+	+	-	-	-	+	+	-	-	+	+	+	-	+	-	-	-	-	-
St2957	-	+	+	-	+	+	-	-	-	-	+	+	+	-	+	-	-	-	-	-
St3051	-	+	+	-	-	-	-	-	-	-	+	+	+	-	+	-	-	-	-	-

N: Northern blotting; RT: RT-PCR; I: Immunohistochemistry.

IIIb –*In vitro* adherence assay of *Helicobacter pylori* to gastric carcinoma cell lines

Gastric carcinoma cell lines showed higher binding indexes to *H. pylori* than the myeloid leukaemia cell line HL60 (Table 18, Figure 15). The binding indexes for *E. coli* in all cell lines were similar (Table 18, Figure 15). Figure 16 shows one example of the adherence assay performed in slides for the cell line GP202.

Table 18. Binding indexes for *Helicobacter pylori* and *E. coli* in the different cell lines.

Cell line	GP202	GP220	MKN45	KATOIII	St23132	St2957	St3051	HL60
H.pylori binding Index	245.76	312.50	390.00	105.83	404.70	238.45	221.72	10.99
E. coli binding index	47.07	20.92	29.63	30.60	37.12	76.55	51.56	13.95

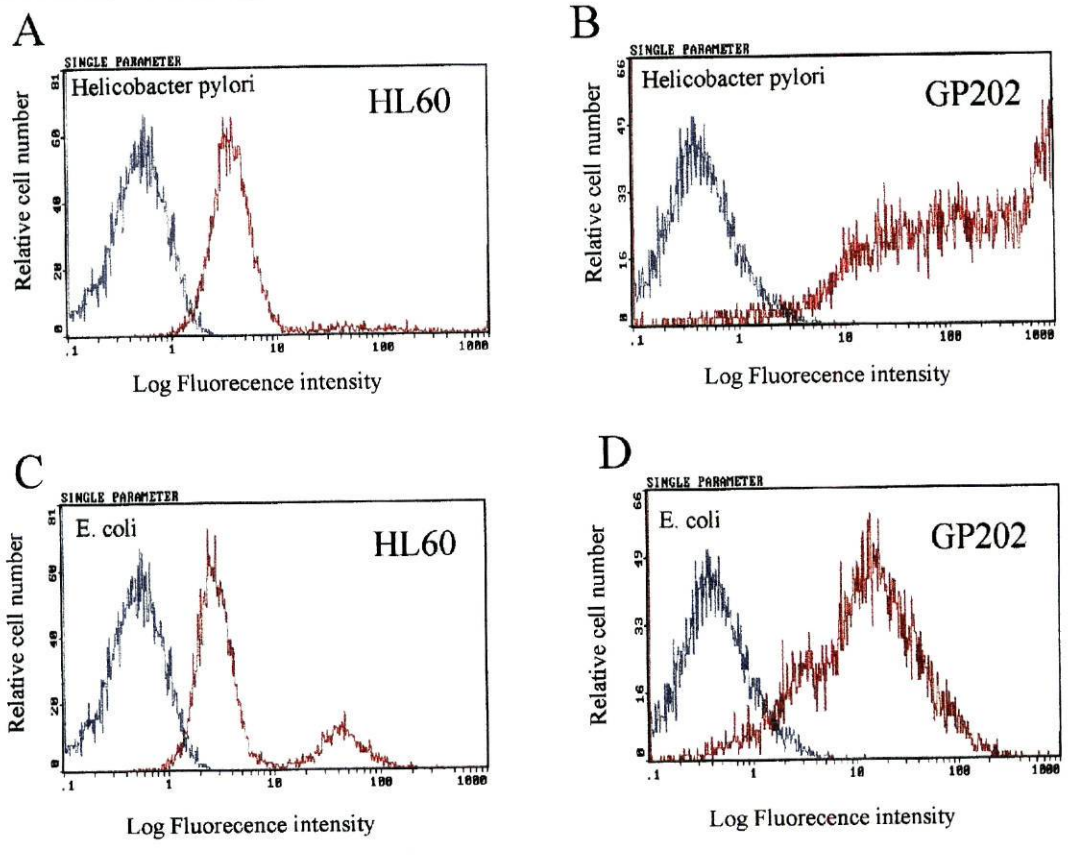


Figure 15. Quantitative analysis of *H. pylori* and *E. coli* adhering to HL60 (A, C) and GP202 (B, D) cell lines. The first peak (blue) represents cells which have not been exposed to *H. pylori* or *E.coli*. The second peak (red) represents cells which have been exposed to *H.pylori* (A, B) and *E. coli* (C, D).

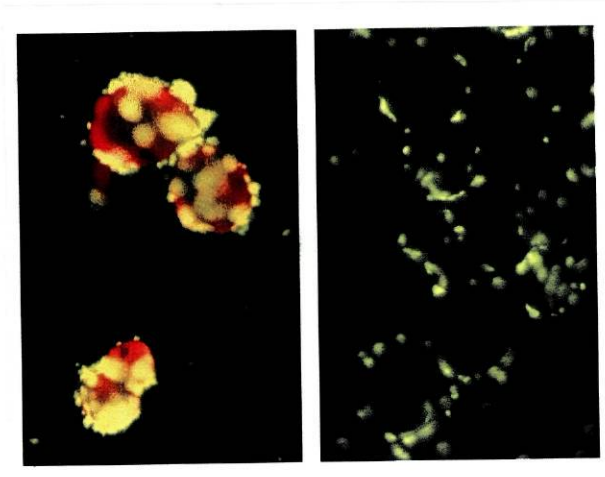


Figure 16. Left. Gastric carcinoma cell line GP202 (200x) stained in red by propidium iodide shows adherent bacteria stained in green by FITC. Right. Gastric carcinoma cell line GP202 (200x) (unstained) shows adherent bacteria stained in green by FITC.

DISCUSSION

Human gastric carcinogenesis is a multistep, multifactorial process, and the relative contribution of environmental exposure and genetics for the risk of developing carcinoma has not been established to date. Diet and *H. pylori* infection have been identified as the major environmental exposure risk factors (Correa, 1988a; Correa, 1992; Solcia *et al*, 1996). The contribution of host susceptibility factors is far from being elucidated. We undertook the present study in an attempt to see if mucin genes polymorphism, taken as a host susceptibility factor, might contribute to add useful information for the assessment of gastric carcinoma risk.

The *rationale* of the present thesis was based on previous studies showing that the *MUC1* gene polymorphism is reflected both at RNA and protein level (Karlsson *et al*, 1983; Swallow *et al*, 1987a; Gendler *et al*, 1988; Gendler et al, 1990a; Ligtenberg *et al*, 1990). The original studies demonstrating that the allelic variations in MUC1 were inherited in an autossomal codominant fashion were performed using lectins and antibodies to detect alleles of a mucin found in urine (Karlsson *et al*, 1983; Swallow *et al*, 1987). The evaluation of DNA polymorphism at the tandem repeat region of the *MUC1* gene was therefore used in the present study as a surrogate endpoint for the evaluation of protein polymorphism.

The putative biological significance of these polymorphisms has not been previously addressed. We advanced the hypothesis that individuals with *MUC1* alleles encompassing a lower number of tandem repeats, that encode for smaller proteins, might have a thinner mucus layer and be more susceptible to environmental aggressions namely those related to diet, toxins or microorganisms.

Fifteen *MUC1* alleles were identified in the 324 Portuguese blood donors using *EcoR*I as restriction enzyme and 13 alleles were identified in the 159 patients with gastric carcinoma, all of which were also present in the control population. We found a significantly different distribution of *MUC1* alleles between the control group of blood donors and the group of patients with gastric carcinoma. We also found that the prevalence of small sized alleles is higher in patients with gastric carcinoma than in controls. The higher prevalence of allele 10 and the lower prevalence of allele 4 in patients with gastric carcinoma are the most striking differences between patients and controls. These results fit with our hypothesis that individuals with small *MUC1* alleles are more susceptible to the development of gastric carcinoma. For the analysis of genotypes we classified the cases as described in the results section. We found that genotypes carrying two small *MUC1* alleles are more prevalent in patients with gastric carcinoma than in controls. Individuals with two small *MUC1* alleles carry an estimated relative risk for gastric carcinoma development of 4.3 (Carvalho *et al*, 1997).

To see whether or not *MUC1* genotypes are associated with pre-neoplastic lesions we analysed the distribution of *MUC1* genotypes according to the different morphologic aspects of the gastric mucosa in a series of 281 dyspeptic patients. A significant association was observed with glandular atrophy of the gastric mucosa, with a significantly higher prevalence of glandular atrophy in individuals with SS *MUC1* genotypes. We observed also a trend for a higher prevalence of incomplete type intestinal metaplasia in individuals with SS *MUC1* genotypes (26.8%) than in individuals with LL (19.7%) or LS (21.2%) *MUC1* genotypes but the differences do not attain the threshold of statistical significance. Chronic atrophic gastritis and intestinal metaplasia, namely incomplete type intestinal metaplasia, are considered as precursor lesions of gastric carcinoma (Correa, 1988a, 1988b; Filipe and Jass, 1986). The significantly higher prevalence of glandular atrophy in individuals with SS *MUC1* genotypes together with a higher prevalence of incomplete type intestinal metaplasia in individuals with SS *MUC1* genotypes, support that *MUC1* genotypes may play a role in the development of pre-neoplastic lesions of the gastric mucosa.

Recently, it was reported that individuals with a high number of tandem repeats in *MUC1* are at increased risk for development of severe acne in a japanese population (Ando *et al*, 1998). In another study, Denda *et al* (1998, personal communication) found a high frequency of small *MUC1* alleles in a japanese population, together with a high percentage of homozygous individuals for small *MUC1* alleles. This observation is conflicting with one of the characteristics of *MUC*1 – the high level of polymorphism. However, if the aforementioned results are confirmed by further studies they legitimate the speculation that the high prevalence of homozygous individuals for small *MUC1* alleles in the japanese population may contribute to the high prevalence of gastric carcinoma in Japan.

In order to see whether or not *MUC1* gene polymorphism could explain the geographic differences in the incidence of gastric carcinoma, we compared the *MUC1* gene polymorphism in blood donors from Portugal and Denmark, two countries with the highest and the lowest incidence of gastric carcinoma in the European Union, respectively (Black *et al*, 1997).

The analysis of *MUC1* gene polymorphism in the two populations using two different restriction enzymes indicates that *Alu*I is more suitable for genotyping than *Eco*RI. In fact, in the Portuguese sample we detected a significant deviation from the Hardy-Weinberg equilibrium, due to homozygote excess, when using *Eco*RI. This finding, which can not be attributed to technical errors since it was not found in the Danish sample, probably reflects a polymorphism at the *Eco*RI restriction site with a high frequency in Portugal. This assumption is further supported by the fact that the

differences observed in the distribution of the *MUC1* gene frequency between both Caucasian populations is higher with *Eco*RI than with *Alu*I.

We observed significant differences between Portuguese and Danish blood donors regarding the distribution of *MUC1* alleles with a tendency for a higher frequency of small sized *MUC1* alleles in the Danish population. The same tendency is detectable if one focuses on the two modal alleles in both systems. Alleles 4 (*Eco*RI) and 5 (*Alu*I), that are large alleles, are more frequent in Portuguese blood donors, and alleles 10 (*Eco*RI) and 16 (*Alu*I), that are small alleles, are more frequent in Danish blood donors. Our data indicate that the higher incidence of gastric carcinoma in Portugal than in Denmark (Black *et al*, 1997) can not be ascribed to *MUC1* gene polymorphism. The results of the present study show that *MUC1* gene polymorphism despite being a susceptibility trait for gastric carcinoma development, does not imply an isolated effect *per se*.

Further studies should be undertaken to see if the distribution of *MUC1* alleles differs between Danish gastric carcinoma patients and Danish blood donors, similarly to the differences we found in Portugal (Carvalho *et al*, 1997). If this comparison shows that Danish gastric carcinoma patients have a higher frequency of small *MUC1* alleles than Danish blood donors we can assume that, despite the inter-populational differences in the *MUC1* gene polymorphism, there is an increased risk for gastric carcinoma development associated with small *MUC1* alleles, probably acting together with the environmental conditions (notoriously different in Portugal and Denmark – see below). If this scenario is confirmed, the most likely explanation for the variable incidence of gastric carcinoma in Denmark and Portugal is that the aforementioned differences between both countries regarding environmental factors do play a major role in gastric carcinogenesis. If, in contrast, no differences in *MUC1* gene polymorphism are detected between gastric carcinoma patients and controls from Denmark we must assume that the differences observed in Portugal are restricted to the particular genetic/environmental circumstances of Northern Portugal (Carvalho *et al*, 1997).

The allele pattern distribution for *MUC1* gene in Portuguese blood donors is also different from that previously described by Gendler *et al* (1990a) in Northern European individuals, namely concerning the lower number of alleles in the population of our study. The differences between both studies are probably due to the use of different restriction enzymes producing fragments of different sizes. This methodological shortcome hampers the interpretation of the differences among populations. However, the clearly bimodal distribution of the 30 alleles described by Gendler *et al* (1990a) and the distribution reported by Pratt *et al* (1996) for the *MUC1* VNTR, are in keeping with our allelic distribution both with *Eco*RI and *Alu*I restriction enzymes. The bimodal

distribution of *MUC1* alleles may indicate that the polymorphism was in part generated by large-scale mutations, such as duplications of portions of the tandem repeat sequences (Vinall *et al*, 1998) or the fusion of two populations with different mutational demographic histories.

For enlarged studies on other populations, necessary to settle the putative relevance of the *MUC1* VNTR polymorphism, it is essential to find alternative techniques to the time consuming and expensive Southern blotting analysis. The CA repeat polymorphism within intron 6 may be a good alternative for the VNTR analysis. Our results suggest that the CA repeat polymorphism within intron 6 can be used as a pre-screening strategy. The association between the VNTR polymorphism and the CA repeat polymorphism was previously described by Pratt *et al* (1996). Furthermore, these authors demonstrated that there is an association of two polymorphisms in *MUC1*, located on both sides of the major tandem repeat region, the previously described CA repeat polymorphism within intron 6 and the G/A substitution in exon 2. These findings suggest that unequal crossing over during meiosis may not be a major cause of VNTR variation in *MUC1* (Pratt *et al*, 1996). Processes such as replication slippage, unequal sister chromatid exchange and gene conversion are likely to be responsible for the generation and maintenance of the VNTR polymorphism as postulated for non-coding DNA (Wolff *et al*, 1988, 1989, 1991; Martinson *et al*, 1993; Jeffreys *et al*, 1994; Ellsworth *et al*, 1995).

We observed that the genotype distribution in gastric carcinoma patients was identical in cases with different morphologic types, indicating that the susceptibility trait attributed to the *MUC1* gene polymorphism for the development of gastric carcinoma is independent of the gastric carcinoma histotype. Since gastric carcinoma is a non-homogeneous entity, with diffuse carcinomas being generally accepted as a morphologic sub-type associated with an underlying genetic predisposition, we envisioned the possibility that the genetic susceptibility conferred by small *MUC1* genotypes might be more relevant for the development of the diffuse/signet-ring cell type carcinoma. Our data, showing no difference in the *MUC1* genotype distribution among patients with carcinomas of the two main types suggest that the susceptibility for gastric carcinoma development mediated by *MUC1* gene polymorphim is independent from the gastric carcinoma histotype.

Loss of heterozygosity was not detected for *MUC1* gene when analysis of paired samples from normal tissues and gastric carcinomas was performed and just one case displayed *MUC1* gene recombination. Our results are corroborated indirectly by the findings obtained by our group using a microsatellite marker in chromosome 1q (D1S158); in this study Seruca *et al* (1995) observed a low frequency of allele loss in the aforementioned region of chromosome 1q (9% in a series of 23 cases). Our results are

at variance with previous data from breast and gastrointestinal tissues. Analysis of DNA prepared from breast cancer and blood cells from the same patients demonstrated that 30% of 70 informative patients showed loss of heterozygosity and 1% exhibited a new band indicating gene recombination (Gendler *et al*, 1990b). Waltz *et al* (1998) reported that *MUC1* gene was altered in 34 of 54 informative breast cancer patients. The 1q21-24 region, where *MUC1* gene maps, has been shown to be also involved in gastrointestinal cancers: 5 of 10 gastric carcinomas and 3 of 27 colorectal carcinomas showed loss of a *MUC1* allele (Fey *et al*, 1989). Sano *et al* (1991) reported that four of six well-differentiated advanced gastric carcinomas showed allele loss of *MUC1* gene. The gastric carcinoma of our series with *MUC1* recombination had the appearance of a mutant allele that probably arose by unequal mitotic recombination (Gendler *et al*, 1990b). *MUC1* gene recombination has not been previously described in gastric carcinomas (Fey *et al*, 1989; Armour *et al*, 1989; Sano *et al*, 1991) although it has been reported in breast carcinomas (Gendler *et al*, 1990b).

Taking into account the results obtained in the study of *MUC1* gene polymorphism we decided to evaluate *MUC6* gene polymorphism to see if it was also involved in individual susceptibility to gastric carcinoma development. The characterisation of *MUC6* gene polymorphism at DNA level as a surrogate endpoint for the evaluation of protein polymorphism is supported by the studies of Debailleul *et al* (1998), showing that allelic variations in length of mucin transcripts, including *MUC6*, are directly related to the variable number of tandem repeats seen at the DNA level.

Ten different *MUC6* alleles with sizes ranging from 8.5 to 12.5 Kbp were identified in the group of 376 blood donors. Observed genotype distribution was in accordance to Hardy-Weinberg equilibrium expectations, and the observed heterozygosity was 75.8% (Garcia *et al*, 1997). These results are in agreement with those reported previously by Toribara *et al* (1993) and Vinall *et al* (1998), demonstrating a high degree of polymorphism for *MUC6* gene.

Our patient and control samples were significantly different regarding sex and age distributions, but no differences were observed in allelic frequencies of *MUC6* gene regarding these parameters. Both populations were at Hardy-Weinberg equilibrium concerning ABO histo-blood system and therefore were considered appropriate samples for a case-control study of *MUC6* gene polymorphism similar to our study on *MUC1* gene polymorphism.

Significant differences in allelic frequencies between blood donors and patients with gastric carcinoma were observed for alleles 1, 5, 8 and 10, suggesting that *MUC6* gene polymorphism is involved in the susceptibility to gastric carcinoma development. The differences of *MUC6* allele frequencies between controls and patients are in agreement

with those observed previously for *MUC1* gene (Carvalho *et al*, 1997). Allele 1, the largest allele, was more frequent in blood donors; alleles 5, 8 and 10, which are smaller alleles, were more frequent in patients with gastric carcinoma.

Like for *MUC1* gene, the distribution of *MUC6* alleles in patients with gastric carcinoma with different morphologic types was identical, suggesting that *MUC6* gene polymorphism is involved in the predisposition to gastric carcinoma development regardless of the histotype.

Analysis of paired samples (tumor vs normal mucosa) demonstrated that 5.6% of the cases had allelic imbalance and one case displayed *MUC6* gene recombination. Our results are in accordance with previous studies showing absence or low levels of loss of heterozigosity in 11p15 region (Wada *et al*, 1988; Motomura *et al*, 1988; Sano *et al*, 1991; Uchino *et al*, 1992). In contrast to this, Rodriguez *et al* (1990) demonstrated 89% of chromosomal rearrangements involving the 11p13-15 region. Ranzani *et al* (1993) and Baffa *et al* (1996) detected 37% and 41% of loss of heterozygosity in 11p15.5, respectively. The aforementioned discrepancies may be due to the use of different probes and/or microsatellite markers and need further investigation.

We observed that the distribution of *MUC1* alleles is independent from the distribution of *MUC6* alleles. Similar results were previously reported by Vinall *et al* (1998) in the study of *MUC* gene cluster in 11p15, suggesting that an association at different loci is unlikely and that each gene should therefore be considered separately in disease-association studies.

For better understanding the role played by mucin genes polymorphism in the final mucin composition and its relevance for the protection of gastric mucosa some issues should be elucidated. MUC5AC gastric mucin, expressed in normal stomach together with MUC1 and MUC6 mucins, should also be taken into account. The fact that MUC1 mucin is a transmembrane protein and MUC6 is a secreted protein should be stressed since different biological functions are attributed to them. The results observed regarding cancer risk and *MUC1* gene polymorphism are clearer from a etiophatogenic standpoint because they fit within the system of different alleles ordered by size. At variance with this, the differences for *MUC6* gene are restricted to some alleles (1, 5, 8 and 10). This finding reinforces that, mostly in the case of secreted proteins (MUC5AC and MUC6), other factors should be analysed such as the amount of protein produced, the ability to polymerize into large structures and the interaction with other proteins. Most certainly, all these factors act in concert for the production and maintenance of the mucus layer.

Length variations of mucins are likely to have an impact on the properties of the mucus gel layer but it is still unknown to what extent the variations in length affect the variations in the protein product and in the mucus gel structure. Some issues remain unresolved: Are larger and shorter alleles equally expressed? What impact has heterozygosity in the final expression of cases depicting short and large alleles?

Considering the importance of environmental factors in gastric carcinoma development we evaluated the *H. pylori* infection in Portugal and Denmark in an attempt to find to what extend it may be involved in the different gastric carcinoma incidences in the two countries.

The available information indicates that more than 50% of the world population is colonized by *H. pylori* (Dunn *et al*, 1997). The majority of individuals acquire *H. pylori* infection early in life. An inverse relationship between socioeconomic status and prevalence of *H. pylori* has been observed (for review see Dunn *et al*, 1997). Host and environmental factors such as gender, smoking, diet, genetics, acid secretory profile and Lewis blood group antigens, may modulate the clinical response to *H. pylori* infection (Go, 1997). In 1991, four reports demonstrated for the first time associations between *H. pylori* infection and the presence (Talley *et al*, 1991) or the development (Forman *et al*, 1991; Nomura *et al*, 1991; Parsonnet *et al*, 1991) of gastric carcinoma. In 1994, the International Agency for Research on Cancer reviewed the available evidence and declared *H. pylori* a type I carcinogen for humans (International Agency for Research on Cancer, 1994).

Supporting the putative role played by *H. pylori* infection in gastric cancer development, we demonstrated, using a serological method, that Portuguese blood donors are highly infected (82.1%) by *H. pylori* when compared to Danish blood donors (18.0%). The prevalence of *H. pylori* infection in dyspeptic patients from Northern Portugal was very similar to the prevalence of infection in the blood donors. Our results fit with those previously reported on the *H. pylori* infection in Portugal (Estevens *et a*, 1993; Soares *et al*, 1993) and in several countries with high rates of infection by *H. pylori*. Mégraud *et al* (1989) showed that 80% to 90% of healthy adult individuals from developing countries are infected by *H. pylori*. Asaka *et al* (1994) detected 75% of infected asymptomatic individuals in the japanese population. Chang-Claude *et al* (1995), using histology, reported a prevalence of infection of 86% in a chinese population.

When we compared the prevalence of infection in the dyspeptic patients from our series using two different methodologies (serology and histology) some discrepancies were observed.

Infection of the gastric mucosa with *H. pylori* results in systemic as well as local immune responses, including elevation of specific IgG and IgA levels in serum and elevated levels of secretory IgA and IgM in the stomach, thus allowing the development of serologic tests for the detection of *H. pylori* infection (Perez-Perez *et al*, 1988; Kreuning *et al*, 1994). In contrast to endoscopic methods, serologic testing offers a quick, non-invasive and relatively inexpensive approach to the diagnosis of *H. pylori* infection. Serological testing circunvents the limitations of the histological diagnosis which is based upon biopsies that assess only small areas of the stomach and may therefore lead to false negative results particularly when dealing with mild and/or patchy forms of gastritis. Serological testing has nevertheless a relatively important drawback: unlike histology which diagnosis an existing infection, the quantification of anti-*H. pylori* antibodies can not distinguish between a present and a past infection. The 21 cases of dyspeptic patients from our series with a positive serology and no detectable *H. pylori* in the biopsies can either represent false negative results from histology due to sample bias or false positive results from serology due to a past infection. The 13 cases that were negative for anti-*H. pylori* IgG antibodies and showed *H. pylori* in the biopsy can represent false negatives in the serological testing. We can not disregard that the immune response against *H. pylori* is dependent upon various factors, including the infecting bacteria, the duration of infection and the host immune system itself; all these factors may contribute in the appropriate settings for the negative serology tests in the presence of *H. pylori* in the biopsies.

The very high rates of *H. pylori* infection in the Portuguese population fit well with the high incidence of gastric carcinoma in Portugal – the highest in western Europe (Estevens *et al*, 1993; Soares *et al*, 1993; Black *et al*, 1997). Similar associations are found in countries with high rates of *H. pylori* infection like Japan and China (Mégraud *et al*, 1989; Asaka *et al*, 1994; Chang-Claude *et al*, 1995). The low rates of *H. pylori* infection in the Danish population fit with those of a previous report of Andersen *et al* (1996) and with the very low incidence of gastric carcinoma in Denmark (Black *et al*, 1997).

Apart from the higher prevalence of infection we demonstrated also that the levels of anti-*H. pylori* IgG antibodies are significantly higher in the Portuguese population. Taking in consideration previous studies showing that the levels of anti-*H. pylori* antibodies are related to the severity of antral gastritis and density of *H. pylori* colonization (Kreuning *et al*, 1994), our results suggest that the Portuguese population further than being more frequently infected is also more severely infected by *H. pylori* than the Danish population.

The fact that *H. pylori* strains are highly diverse in genetic terms can be used to explain our results. Many studies have shown that infection of CagA (Cytotoxin Associated

Gene A) positive strains increases the risk of both atrophic gastritis and gastric cancer (Blaser *et al*, 1995; Beales *et al*, 1996; Shimoyama *et al*, 1998). The vacuolating cytotoxin (VacA) of *H. pylori* shows variation in 2 regions of the gene, with s1a/m1 strains being associated with increased risk of peptic ulcer disease and enhanced inflammation (Atherton *et al*, 1995, 1997). The high prevalence and more severe *H. pylori* infection in Portuguese blood donors may be explained by the high prevalence of s1b strains in individuals from Northern Portugal as well as by the high prevalence of infections by multiple *H. pylori* strains (Van Doorn *et al*, 1998).

Finally, our results suggest the existence of an association between the SS *MUC1* genotypes and the *H. pylori* infection. We observed that individuals with SS *MUC1* genotypes have a suggestively higher (p=0.06) mean titre of anti-*H. pylori* antibodies that individuals with LL *MUC1* genotypes, in the series of Portuguese blood donors. The same tendency, though not reaching the threshold of statistical significance, is observed in dyspeptic patients from Portugal and in Danish blood donors. Individuals with two small *MUC1* alleles apparently have a more severe *H. pylori* infection indirectly evaluated by the presence of higher titres of anti-*H. pylori* IgG antibodies.

The higher mean titres of anti-*H.pylori* IgG antibodies in the Portuguese blood donors of male gender is in agreement to that previously reported in the series of Replogle *et al* (1995), but was not confirmed in other epidemiological studies (Go, 1997). The reasons underlying this epidemiological evidence remain unclarified.

We analysed several gastric carcinoma cell lines to see if they could be used as *in vitro* models for adherence assays of *H. pylori,* in an attempt to gain insight into the mucins-bacterial interactions.

Our data showed that all gastric carcinoma cell lines evaluated express *MUC1* and most of the cell lines expressed *MUC5AC*. These two mucins are expressed in normal gastric mucosa (Burchell *et al,* 1987; Girling *et al*, 1989; Ho *et al*, 1993; Lesuffleur *et al*, 1994; Ho *et al*, 1995; Reis *et al*, 1998a; Byrd *et al*, 1997; DeBolós *et al*, 1995; Reis *et al*, 1997); MUC1 mucin is expressed in every gastric carcinoma of a series of 73 cases (Reis *et al*, 1998a) and MUC5AC is expressed in a large proportion of gastric carcinoma cases (63.6% from a series of 66 cases) (Reis *et al*, 1997). Our results thus fit with the *in vivo* data and suggest that gastric type differentiation, evaluated by mucin expression (MUC1 and MUC5AC), is maintained in all gastric carcinoma cell lines.

MUC6 mRNA and MUC6 protein were absent in every cell line of the present study. Expression of MUC6 protein is limited to mucopeptic cells of the normal stomach (Byrd *et al*, 1997; DeBolós *et al*, 1995; Ho *et al*, 1995) and in a series of 104 gastric carcinoma cases only 29.8% showed MUC6 protein expression (Reis *et al*, unpublished data). The

absence of expression of MUC6 reflects the low frequency of mucopeptic differentiation in gastric carcinoma cell lines as well as in gastric carcinomas *in vivo* (Solcia *et al*, 1996).

MUC 2 protein expression with or without detectable amounts of mRNA by RT-PCR and/or Northern blotting was observed in four cell lines. Conflicting results regarding the expression of MUC2 in normal gastric mucosa have been reported when different antibodies are used, demonstrating either no expression of MUC2 (Ho *et al*, 1995; Xing *et al*, 1992b) or very low levels of expression (Carrato *et al*, 1994; Devine *et al*, 1993; Gambús *et al*, 1993; Ho *et al*, 1993; Reis *et al*, 1998b). On the other hand, MUC2 is overexpressed or aberrantly expressed in intestinal metaplasia (Ho *et al*, 1995; Filipe *et al*, 1996; Reis *et al*, 1998c), a precursor lesion of gastric carcinoma, and in 38% of gastric carcinoma cases (Baldús *et al*, 1998). The present study, showing MUC2 expression in four cell lines, demonstrates that both *in vivo* and *in vitro* MUC2 expression reflects a non-universal aberrant differentiation pattern of gastric carcinoma cells.

Studies on the intestinal mucin *MUC3* in human stomach are scarce but low levels of *MUC3* mRNA have been demonstrated in normal gastric tissues (Audie *et al*, 1993; Van Klinken *et al*, 1996). Expression of *MUC4* in normal stomach has also been reported (Audie *et al*, 1993, Van Klinken *et al*, 1996) and the expression of *MUC3* and *MUC4* mRNA was previously demonstrated in gastric carcinomas (Ho *et al*, 1995). In accordance with the *in vivo* expression of *MUC3* and *MUC4* in gastric tissues, consistent results were obtained in the present study regarding Northern blotting and RT-PCR expression of *MUC3* in GP220 and St23132 cell lines and of *MUC4* in KATOIII cell line by Northern blotting.

To the best of our knowledge, expression of *MUC5B* has not been reported to date in normal stomach. Four cell lines expressed *MUC5B* mRNA as detected by RT-PCR. Low levels of expression of *MUC5B* mRNA can justify the absence of signal in Northern blotting. *In vitro* expression of *MUC5B* in cell lines may indicate an aberrant differentiation program of these gastric carcinoma cell lines in culture.

Expression of the low molecular weight salivary mucin *MUC7* has not been reported to date in normal gastric tissues and was not detected in any of the cell lines.

Expression of mucins by the different cell lines was not dependent on the type of carcinomas from which they were derived. This contrasts with previous results from human tumours showing preferential expression of MUC5AC in isolated cell-type gastric carcinomas (Reis *et al*, 1997) and preferential expression of MUC2 in glandular type gastric carcinomas (Baldús *et al*, 1998).

All cell lines displayed high levels of expression of type 1 and type 2 chain Lewis antigens. The most striking changes of glycosylation occurring in gastric carcinomas are the increased and/or *de novo* expression of Le^a antigen and the decreased or absent expression of type 2 Lewis antigens in comparison with normal gastric mucosa (Murata *et al*, 1992). The gastric carcinoma cell lines we have studied apparently do not reproduce the results described for gastric carcinomas since Le^a, S-Le^a and Le^b and all type 2 Lewis antigens (Le^x, S-Le^x and Le^y) are widely expressed in most cell lines. The relationship between mucin expression and Lewis antigens expression reported in the human stomach (DeBolós *et al*, 1995), namely the association between expression of MUC5AC and type 1 Lewis structures, and between MUC6 and type 2 Lewis structures, was not found in the cell lines. In fact, type 1 Lewis antigens are expressed in all gastric carcinoma cell lines and type 2 Lewis antigens are also expressed in all cell lines despite the absence of MUC6 expression. These findings suggest that different carrier mucins are used for building type 1 and type 2 Lewis chain antigens in gastric carcinoma cell lines.

The expression of simple mucin-type antigens is low or absent in the gastric carcinoma cell lines. Expression of Tn and S-Tn antigens was observed in four cell lines and T antigen was only detected in GP202 cell line. It has been reported that Tn and S-Tn antigens are expressed in the majority of gastric carcinomas and their metastases (Hirohashi *et al*, 1985; Kjeldsen *et al*, 1988; Thor *et al*, 1986; David *et al*, 1992; Iwata *et al*, 1993; Kushima *et al*, 1993; Carneiro *et al*, 1994a, 1994b; Werther *et al*, 1994), whereas T antigen is detected in a low proportion of cases (David *et al*, 1992). S-Tn expression is associated with the glandular type of gastric carcinoma (Iwata *et al*, 1993; Kushima *et al*, 1993; Carneiro *et al*, 1994a and 1994b), suggesting that mucin glycosylation may also be related to cell differentiation. The gastric carcinoma cell lines parallel the *in vivo* model in that the four cell lines expressing S-Tn are the ones derived from glandular type gastric carcinomas. The expression of T antigen by a single gastric carcinoma cell line is in keeping with the low levels of expression of T antigen in gastric carcinomas (David *et al*, 1992).

Our data also demonstrate the absence of relationship between mucin expression profile and simple mucin-type carbohydrate antigens expression, in accordance with previous results reported by Ogata *et al* (1992b) in colon cancer cell lines.

Summing up, our results show that the profile of mucin expression in gastric carcinoma cell lines: a) follows partially the mucin expression profile of normal gastric mucosa and gastric carcinomas with wide expression of MUC1 and MUC5AC; b) parallels the aberrant pattern of mucin expression observed in human gastric carcinomas with occasional expression of MUC2, MUC3, MUC4 and MUC5B; c) does not include, at least in our series, the expression of the gastric mucin MUC6; d) follows partially the

differentiation pattern of the carcinomas from which the cell lines have been originated keeping S-Tn expression in the cell lines derived from glandular type gastric carcinomas. Our results further demonstrate that there is no apparent relationship between the mucin core proteins and the simple mucin-type or Lewis carbohydrate antigens that are expressed in each cell line, suggesting that at variance to normal gastric tissues, the glycosylation is not related with the expressed mucins in gastric carcinoma cell lines.

The study of cell lines in the evaluation of the adherence of *H. pylori* showed that gastric carcinoma cell lines have higher binding indexes to *H. pylori* than HL60, a myeloid leukaemia cell line. HL60 is a non mucus-secreting cell line that displayed the lowest levels of expression of type 2 and no expression of type 1 Lewis carbohydrate chains, including Le^b. We demonstrated furthermore, that the binding of *H. pylori* to gastric carcinoma cell lines was not influenced by the presence and/or the levels of expression of different mucin cores. Similarly, the levels of expression of type 1 and type 2 Lewis carbohydrate chains, namely Le^b did not influence the binding of *H. pylori*. Our results fit with those reported by Clyne *et al* (1997) using the same *H. pylori* strain of our study; Clyne *et al* (1997) showed absence of effect of Le^a and Le^b expression on adherence of *H. pylori* to human gastric cells. At variance with these results are those reported by Bóren *et al* (1993), corroborated later on by Alkout *et al* (1997), showing that histo-blood group $Lewis^b$ antigen is a candidate receptor for *H. pylori*.

Our data support the conclusion that *in vitro* adherence of *H. pylori* to gastric carcinoma cell lines is not apparently influenced by the mucin profile and/or carbohydrate antigens expression on these cells. However, we can not disregard that *in vitro* conditions do not mimick *in vivo* conditions. In the *in vitro* adherence assays the amount of available bacteria and the exposure of potential binding structures by epithelial cells are far from representing the situation *in vivo*, where the mucus layer and the tissue organisation may influence the accessibility of bacteria to the epithelial cells.

Moreover, binding of *H. pylori* to different structures reflects the existence of several adhesins in *H. pylori* mediating adherence to various targets in gastric epithelium and/or the presence of multiple binding modes among different *H. pylori* strains (Moran, 1995).

All these factors contribute to shadow the putative relationship between *H. pylori* and mucins. Inhibition assays using purified synthetic carbohydrate structures in controlled *in vitro* models will hopefully help to elucidate the interaction between *H. pylori* and the host mucins. It will be also important to study different *H. pylori* strains to see whether or not they use different adherence mechanisms to human mucins. Finally, primary culture of gastric biopsies will probably provide a better model than the cell culture systems.

The main conclusions of the present work are the following:

- There is an increased risk of 4.3x for gastric carcinoma development in individuals with SS *MUC1* genotypes.

The role of *MUC1* gene polymorphism in gastric carcinoma development is supported by the association between SS *MUC1* genotypes and pre-neoplastic lesions, namely glandular atrophy of the gastric mucosa. The higher titres of anti-*H. pylori* IgG antibodies in individuals with SS *MUC1* genotypes suggest that in these individuals the *H. pylori* infection is more severe, thus proving a putative link between SS *MUC1* genotypes and increased risk for gastric carcinogenesis.

- There are also significant differences of allelic frequencies of *MUC6* gene in gastric carcinoma patients and blood donors but their relevance for gastric carcinoma development was not clarified in the present study.

- *In vitro* adherence of *H. pylori* to gastric carcinoma cell lines was higher than the adherence to a non-mucinous myeloid leukaemia cell line; despite this hint of relationship between mucins and *H. pylori* adherence no significant differences were observed regarding the pattern of expression of mucins and carbohydrates.

From the start, when we designed the present project aiming to clarify some of the gene/environment interactions in gastric carcinogenesis, we were aware that it would be impossible to get a comprehensive evaluation of more than a few of the numerous aspects involved in the process. Among such aspects it is worthwhile stressing that we never aimed to study the protective or the aggressive role of diet (Correa, 1992), the different *H. pylori* strains, which have variable cytotoxic potential (Blaser *et al*, 1995; Atherton *et al*, 1995, 1997), or other host characteristics, namely the expression of histo-blood group antigens that may be relevant for *H. pylori* colonisation of the gastric mucosa (Borén *et al*, 1993; Carneiro *et al*, 1996).

Despite these drawbacks, the line of work we have chosen allowed the identification of a host susceptibility factor that plays a role in gastric carcinoma development within the complex context of gene/environment interactions.

The attempt to address the role played by mucins in the interaction with *H. pylori*, using an *in vitro* system, has the advantage of limiting the number of involved factors but has the disadvantage of being far from the *in vivo* situation – i. e. the *in vitro* system represents a totally different and oversimplified microenvironment. The results we obtained with this model, showing that adherence of *H. pylori* to gastric carcinoma cell lines is not influenced by the pattern of expression of mucins and carbohydrates, are therefore hardly transposable to the *in vivo* situation.

REFERENCES

Abe M, Kufe D: Characterization of cis-acting elements regulating transcription of the human DF3 breast carcinoma-associated antigen (MUC1) gene. Proc Natl Acad Sci USA 90: 282-286, 1993.

Abe M, Siddiqui J, Kufe D: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated antigen gene. Biochem Biophys Res Commun 165: 644-649, 1989.

Acres B, Hareuveni M, Balloul J-M, Kieny M-P: Vaccinia virus MUC1 immunization of mice: immune response and protection against the growth of murine tumors bearing the MUC1 antigen. J Immunother 14: 136-143, 1993.

Agrawal B, krantz MJ, Reddish MA, Longenecker BM: cancer-associated MUC1 mucin inhibits human T-cell proliferation, which is reversible by IL-2. Nature Med 4: 43-49, 1998.

Albone EF, Hagen FK, Van Wuyckhuyse BC, Tabak LA: Molecular cloning of a rat submandibular gland apomucin. J Biol Chem 269: 16845-16852, 1994.

Alkout AM, Blackwell CC, Weir DM, Poxton IR, Elton RA, Luman W, Palmer K: Isolation of a cell surface component of *Helicobacter pylori* that binds H type 2, Lewis a, and Lewis b antigens. Gastroenterol 112: 179-1187, 1997.

Andersen LP, Rosenstock SJ, Bonnevie O, Jorgensen T: Seroprevalence of immunoglobulin G, M and A antibodies to *Helicobacter pylori* in an unselected Danish population. Am J Epidemiol 143: 1157-1164, 1996.

Ando I, Kukita A, Soma G, Hino H: A large number of tandem repeats in the polymorphic epithelial mucin gene is associated with severe acne. J Dermatol 25: 150-152, 1998.

Apostolopoulos V, Mckenzie IFC, Pietersz GA: Breast cancer immunotherapy – current status and future prospects. Immunol Cell Biol 74: 457-464, 1996.

Apostolopoulos V, Osinski C, McKenzie IFC: MUC1 cross-reactivity Galα(1,3)Gal antibodies in humans switch immune responses from cellular to humoral. Nature Med 4: 315-320, 1998.

Armour JAL, Patel I, Thein SL, Fey MF, Jeffreys: Analysis of somatic mutations at human minisatellite loci in tumors and cell lines. Genomics 4: 328-334, 1989.

Asaka M, Kimura T, Kato M, Kudo M, Miki K, Ogoshi K, Kato T, Tatsuta M, Graham D: Possible role of *Helicobacter pylori* infection in early gastric cancer development. Cancer 73: 2691-2694, 1994.

Ashall F, Bramwell ME, Harris H: A new marker for human cancer cells. 1. The Ca antigen and the Ca1 antibody. Lancet ii: 1-6, 1982.

Asker N, Baeckstrom D, Axelsson M, Carlstedt I, Hansson G: The human MUC2 mucin apoprotein appears to dimerize before O-glycosylation and shares epitopes with the insoluble mucin of rat small intestine. Biochem J 308: 873-880, 1995.

Asker N, Axelsson MA, Olofsson SO, Hansson GC: Dimerization of the human MUC2 mucin in the endoplasmic reticulum is followed by a N-glycosylation-dependent transfer of the mono- and dimers to the Golgi apparatus. J Biol Chem 273: 18857-18863, 1998.

Atherton JC; Cao P, Peek RM, Tummuru MK, Blaser MJ, Cover TL: Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific vacA types with cytotoxin production and peptic ulceration. J Biol Chem 270: 17771-17777, 1995.

Atherton JC, Peek RM, Tham KT, Cover TL, Blaser MJ: Clinical and pathological importance of heterogeneity in vacA, the vacuolating cytotoxin gene of *Helicobacter pylori*. Gastroenterol 112: 92-99, 1997.

Aubert J-P, Porchet N, Crepin N, Duterque-Coquillaud M, Vergnes G, Mazzuca M, Debuire B, Petitprez D, Degand P: Evidence for different human tracheobronchial mucin peptides deduced from nucleotide cDNA sequences. Am J Resp Cell Mol Biol 5: 178-185, 1991.

Audie JP, Janin A, Porchet N, Copin MC, Gosselin B, Aubert JP: Expression of human mucin genes in respiratory, digestive, and reprodutive tracts ascertained by *in situ* hybridization. J Histochem Cytochem 41: 1479-1485, 1993.

Baffa R, Negrini M, Mandes B, Rugge M, Ranzani GN, Hirohashi S, Croce CM: Loss of heterozygosity for chromosome 11 in adenocarcinoma of the stomach. Cancer Res 56: 268-272, 1996.

Baldús SE, Zirbes TK, Engel S, Hanish FG, Monig SP, Lorenzen J, Glossmann J, Fromm S, Thiele J, Pichmaier H, Dienes HP: Correlation of the immunohistochemical reactivity of mucin cores MUC1 and MUC2 with histopathological subtype and prognosis of gastric carcinoma. Int J Cancer 79: 133-138, 1998.

Barnd DL, Lan MS, Metzgar RS, Finn OJ: Specific, major histocompatibility complex-unrestricted recognition of tumor-associated mucins by human cytotoxic T cells. Proc Natl Acad Sci USA 86: 7159-7163, 1989.

Beales IL, Crabtree JE, Scunes D, Covacci A, Calam J: Antibodies to CagA protein are associated with gastric atrophy in *Helicobacter pylori* infection. Eur J Gastroenterol Hepatol 8: 645-649, 1996.

Bell S, Khatri I, Xu G, Forstner J: Evidence that a peptide corresponding to the rat Muc2 C-terminus undergoes disulphide-mediated dimerization. Eur J Biochem 253: 123-131, 1998.

Bevilacqua MP, Nelson RM: Selectins. J Clin Invest 91: 379-387, 1993.

Bhargava AK, Woitach JT, Davidson EA, Bhavanandan VP: Cloning and cDNA sequence of a bovine submaxillary gland mucin-like protein containing 2 distinct domains. Proc Natl Acad Sci USA 87: 6798-6802, 1990.

Biesbrock AR, Reddy MS, Levine MJ: Interaction of a salivary mucin-secretory immunocomplex with mucosal pathogens. Infect Immun 59: 3492-3497, 1991.

Black RJ, Bray F, Ferlay J, Parkin DM: Cancer incidence and mortality in the European Union: cancer registry data and estimates of national incidence for 1990. Eur J Cancer 33: 1075-1107, 1997.

Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, Stemmermann GN, Nomura A: Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res 55: 2111-2115, 1995.

Bliska JB, Galán JE, Falkow S: Signal transduction in the mammalian cell during bacterial attachment and entry. Cell (Cambridge, Mass.) 73: 903-920, 1993.

Bobek L, Tsai H, Biesbrock A, Levine M: Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (MUC7). J Biol Chem 268: 29563-20569, 1993.

Bobek L, Liu J, Sait S, Shows T, Bobek Y, Levine M: Structure and chromosomal localization of the human salivary mucin gene, MUC7. Genomics 31: 277-282, 1996.

Borén T, Falk P, Roth KA, Larson G, Normark S: Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science 262: 1892-1895, 1993.

Boshell M, Lalani E-N, Pemberton L, Burchell J, Gendler S, Taylor-Papadimitriou J: The product of the human *MUC1* gene when secreted by mouse cells transfected with the full-length cDNA lacks the cytoplasmic tail. Biochem Biophys Res Commun 185: 1-8, 1992.

Braga VMM, Gendler S: Modulation of Muc1 mucin expression in the mouse uterus during estrus cycle, early pregnancy and placentation. J Cell Sci 105: 397-405, 1993.

Bramwell ME, Bhavanandan VP, Wiseman G, Harris H: Structure and function of the Ca antigen. Br J Cancer 48: 177-183, 1983.

Brockhausen I, Yang J-M, Burchell J, Whitehouse C, Taylor-Papadimitriou J: Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. Eur J Biochem 233: 607-617, 1995.

Brockhausen I: Clinical aspects of glycoprotein biosynthesis. Crit Rev Clin Lab Sci 30: 65-151, 1993.

Bu D, Domenech N, Lewis J, Taylor-Papadimitriou J, Finn OJ: Recombinant vaccinia mucin vector: *in vitro* analysis of expression of tumor-associated epitopes for antibody and human cytotoxic T-cell recognition. J Immunother 14: 127-135, 1993.

Buisine MP, Desseyn JL, Porchet N, Degand P, Laine A, Aubert J-P: Genomic organization of the human MUC5AC mucin gene: additional evidence for a common ancestral gene for the 11p15.5 mucin gene family. Biochem J 332: 729-738, 1998.

Burchell J, Durbin H, Taylor-Papadimitriou J: Complexity of expression of antigenic determinants, recognized by monoclonal antibodies HMFG-1 and HMFG-2, in normal and malignant human mammary epithelial cells. J Immunol 131: 508-513, 1983.

Burchell J, Wang D, Taylor-Papadimitriou J: Detection of tumour associated antigens recognised by the monoclonal antibodies HMFG-1 and –2 in serum from patients with breast cancer. Int J Cancer 34: 763-768, 1984.

Burchell J, Gendler S, Taylor-Papadimitriou J, Girling A, Lewis A, Millis R, Lamport D: Development and characterization of breast cancer reactive monoclonal antibodies directed to the core protein of the human milk mucin. Cancer Res 47: 5476-5482, 1987.

Burchell J, Taylor-Papadimitriou J: Effect of modification of carbohydrate side chains on the reactivity of antibodies with core-protein epitopes of the MUC1 gene product. Epith Cell Biol 2: 155-162, 1993.

Byrd JC, Yan P, Sternberg L, Yunker CK, Scheiman JM, Bresalier RS: Aberrant expression of gland-type gastric mucin in the surface epithelium of *Helicobacter pylori*-infected patients. Gastroenterol 113: 455-464, 1997.

Carneiro F, Amado M, David L, Clausen H, Sobrinho-Simões M: Glycosylation features of gastric carcinoma initiation and progression. A review with emphasis on simple mucin-type and histo-blood group antigens of the Lewis system. Eur J Cancer Prev 3 (Suppl 2): 39-46, 1994a.

Carneiro F, David L, Sobrinho-Simões M: Sialyl-Tn expression in gastric carcinoma. Eur J Cancer 30A: 1398-1399, 1994b.

Carneiro F, Amado M, Lago P, Taveira-Gomes A, Amil M, Barreira R, Soares J, Pinho C: *Helicobacter pylori* infection and blood groups. Am J Gastroenterol 91: 2646, 1996.

Carrato C, Balagué C, DeBolós C, Gonzalez E, Gambus G, Planas J, Perini JM, Andreu D, Real FX: Differential apomucin expression in normal and neoplastic human gastrointestinal tissues. Gastroenterol 107: 160-172, 1994.

Carvalho F, Seruca R, David L, Amorim A, Seixas M, Bennett E, Clausen H, Sobrinho-Simões M: MUC1 gene polymorphism and gastric cancer – an epidemiological study. Glycoconj J 14: 107-111, 1997.

Chang-Claude J, Raedsch R, Waldherr R, von Wulffen H, Crespi M, Yang GR, Qui SL, Munoz N, Correa P, Wahrendorf J: Prevalence of *Helicobacter pylori* infection and gastritis among young adults in China. Eur J Cancer Prev 4: 73-769, 1995.

Clausen H, Bennett E, Dabelsteen E: Carbohydrates of the cell surface: Molecular aspects of glycosyltransferases and their genes. APMIS 100 (Suppl 27): 9-17, 1992.

Clyne M, Drumm B: Absence of effect of Lewis a and Lewis b expression on adherence of *Helicobacter pylori* to human gastric cells. Gastroenterol 113: 72-80, 1997,

Cohen-Salmon M, El-Amraqui A, Leibovici M, Petit C: Otogelin: a glycoprotein specific to the acellular membranes of the inner ear. Proc Natl Acad Sci USA 94: 14450-14455, 1997.

Cordell J, Richardson TC, Pulford KAF, Ghosh AK, Gatter KC, Heyderman E, Mason DY: Production of monoclonal antibodies against human epithelial membrane antigen for use in diagnostic immunocytochemistry. Br J Cancer 52: 347-354, 1985.

Correa P: A human model of gastric carcinogenesis. Cancer Res 48: 3554-3560, 1988a.

Correa P: Chronic gastritis: a clinico-pathological classification. Am J Gastroenterol 83: 504-509, 1988b.

Correa P: Human gastric carcinogenesis: a multistep and multifactorial process – first american cancer society award lecture on cancer epidemiology and prevention. Cancer Res 52: 6735-6740, 1992.

David L, Nesland JM, Clausen H, Carneiro F, Sobrinho-Simões M: Expression of simple mucin type carbohydrate antigens (Tn, sialosyl Tn, T and sialosyl T) in gastric carcinomas and their metastases. APMIS 100 (Suppl 27): 162-172, 1992.

Debailleul V, Laine A, Huet G, Mathon P, d'Hooghe M, Aubert J-P, Porchet N: Human mucin genes *MUC2, MUC3, MUC4, MUC5AC, MUC5B,* and *MUC6* express stable and extremely large mRNAs and exhibit a variable length polymorphism. J Biol Chem 273: 881-890, 1998.

DeBolós C, Garrido M, Real FX: MUC6 apomucin shows a distinct normal tissue distribution that correlates with Lewis antigen expression in the human stomach. Gastroenterol 109: 723-734, 1995.

Dejana E, Martin-Padura I, Lauri D, Bernasconi S, Bani MR, Garofalo A, Giavazzi R, Magnani J, Menard S: Endothelial leucocyte adhesion molecule-1-dependent adhesion of colon carcinoma cells to vascular endothelium is inhibited by an antibody to Lewis fucosylated type I carbohydrate chain. Lab Invest 66: 324-330, 1992.

Dekretser TA, Thorne HJ, Jacobs DJ, Jose DG: The sebaceous gland antigen defined by OM-1 monoclonal antibody is expressed at high density on the surface of ovarian carcinoma cells. Eur J Cancer Clin Oncol 21: 1019-1035, 1985.

Denda K, Fujita K, Fujime M, Irimura T: High frequency of short *MUC1* mucin alleles in Japanese population. 5th International Workshop on Carcinoma-Associated Mucins, 18-23 July 1998, Cambridge.

Denny PA, Mirels L, Denny PA: Mouse submandibular gland salivary apomucin contains repeated N-glycosylation sites. Glycobiol 6: 43-50, 1996.

Desseyn JL, Guyonnet-Duperat V, Porchet N, Aubert J-P, Laine A: Human mucin gene *MUC5B*, the 10.7-kb large central exon encodes various alternate subdomains resulting in a super-repeat. J Biol Chem 272: 3168-3178, 1997a.

Desseyn JL, Aubert J-P, Van-Seuningen I, Porchet N, Laine A: Genomic organization of the 3' region of the human mucin gene *MUC5B*. J Biol Chem 272: 16873-16883, 1997b.

Desseyn JL, Buisine M-P, Porchet N, Aubert J-P, Degand P, Laine A: Evolutionary history of the 11p15 human mucin gene family. J Mol Evol 46: 102-106, 1998.

Devine PL, McGuckin MA, Birrell GW, Whitehead RH, Sachdev GP, Shield P, Ward BG: Monoclonal antibodies reacting with the human MUC2 mucin core protein. Br J Cancer 67: 1182-1188, 1993.

Devine PL, McKenzie IFC: Mucins: structure, function and associations with maligancy. BioEssays 14: 619-625, 1992.

Di Noia J, Sanchez D, Frasch A: The protozoan *Trypanosoma cruzi* has a family of genes resembling mucin genes of mammalian cells. J Biol Chem 270: 24146-24149, 1995.

Di Noia J, D'Orso I, Aslund L, Sanchez D, Frasch: The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions. J Biol Chem 273: 10843-10850, 1998.

Dong Z, Thoma RS, Crimmins DL, McCourt DW, Tuley EA, Sadler JE: Disulfide bonds required to assemble functional von Willebrand factor. J Biol Chem 269: 6753-6758, 1994.

Dufosse J, Porchet N, Audie JP, Gyuonnet-Duperat V, Laine A, Van-Seuningen I, Marrakchi S, Degand P, Aubert J-P: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternating domains in human mucin peptides mapped to 11p15. Biochem J 293: 329-337, 1993.

Dunn BE, Cohen H, Blaser MJ: *Helicobacter pylori*. Clin Microbiol Rev 10: 720-741, 1997.

Eckhardt AE, Timpte CS, Abernethy JL, Zhao Y, Hill RL: Porcine submaxillary mucin contains a cystein-rich, carboxyl-terminal domain in addition to a highly repetitive, glycosylated domain. J Biol Chem 266: 9678-9686, 1991.

Ellsworth DL, Shirver MD, Boerwinkle E: Nucleotide sequence analysis of the apopoliprotein B 3' VNTR. Hum Mol Genet 4: 937-944, 1995.

Estevens J, Fidalgo P, Tendeiro T, Chagas C, Ferra A, Nobre Leitão C, Costa Mira F: Anti-*Helicobacter pylori* antibodies prevalence and gastric adenocarcinoma in Portugal: report of a case-control study. Eur J Cancer Prev 2: 377-380, 1993.

Evans DG, Evans DJ Jr, Moulds JJ, Graham DY: N-acetylneuraminyllactose-binding fibrillar hemagglutinin of *Campylobacter pylori*: a putative colonization factor antigen. Infect Immun 56: 2896-2906, 1988.

Evans DG, Evans DJ Jr, Graham DY: Receptor-mediated adherence of *Campylobacter pylori* to mouse Y-1 adrenal cell monolayers. Infect Immun 57: 2272-2278, 1989.

Falk P, Borén T, Normark S: Characterization of microbial host receptors. Methods Enzymol 236: 353-374, 1994.

Falk PG, Bry L, Holgersson J, Gordon JI: Expression of a human α-1,3/4-fucosyltransferase in the pit cell lineage of FVB/N mouse stomach results in production of Le^b-containing glycoconjugates. A potentail transgenic mouse model for studying *Helicobacter pylori* infection. Proc Natl Acad Sci USA 92: 1515-1519, 1995.

Fauchére JL, Blaser MJ: Adherence of *Helicobacter pylori* cells and their surface components to HeLa cell membrabes. Microb Pathog 9: 427-439, 1990.

Feinberg AP, Volgelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6-13, 1982.

Fey MF, Hesketh C, Wainscoat JS, Gendler S, Thein SL: Clonal allele loss in gastrointestinal cancers. Br J Cancer 59: 750-754, 1989.

Filipe MI, Jass JR: Intestinal metaplasia subtypes and cancer risk. In Filipe MI, Jass JR (eds): Gastric Carcinoma. Edinburgh, Churchill Livingstone, 87-115, 1986.

Filipe MI, Linehan JM, Durrant LG, Price MR, Smeeton NC, Pathak S, Swallow DM: Expression of a peptide epitope of the colonic mucin MUC2 in precursor lesions to gastric carcinoma. Eur J Cancer Prev 5: 287-295, 1996.

Forman D, Newel DG, Fullerton F, Yarnell JW, Stacey AR, Wald N, Sitas F: Association between infection with *Helicobacter pylori* and the risk of gastric cancer: evidence from a prospective investigation. Br Med J 302: 1302-1305, 1991.

Fox MF, Lahbib F, Pratt WS, Atwood J, Gum JR, Kim YS, Swallow DM: Regional localization of the intestinal mucin gene MUC3 to chromosome 7q22. Ann Hum Genet 56: 281-287, 1992.

Gambús G, DeBolós C, Andreu D, Francí C, Egea G, Real FX: Detection of the MUC2 apomucin tandem repeat with a monoclonal antibody. Gastroenterol 104: 93-102, 1993.

Garcia E, Carvalho F, Amorim A, David L: *MUC6* gene polymorphism in healthy individuals and in gastric cancer patients from Northern Portugal. Cancer Epidemiol Biomarkers Prev 6: 1071-1074, 1997.

Gärtner F, David L, Seruca R, Machado JC, Sobrinho-Simões M: Establishment and characterization of two cell lines derived from human diffuse gastric carcinomas xenografted in nude mice. Virchows Arch 428: 91-98, 1996.

Gendler S, Burchell J, Duhig T, Lamport D, White R, Parker M, Taylor-Papadimitriou J: Cloning of partial cDNA encoding differentiation and tumor-associated mucin glycoproteins expressed by human mammary epithelium. Proc Natl Acad Sci USA 84: 6060-6064, 1987.

Gendler S, Taylor-Papadimitriou J, Duhig T, Rothbard J, Burchell J: A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats. J Biol Chem 263: 12820-12823, 1988.

Gendler S, Lancaster CA, Taylor-Papadimitriou J, Duhig T, Peat N, Burchell J, Pemberton L, Lalani E-N, Wilson D: Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. J Biol Chem 265: 15286-15293, 1990a.

Gendler S, Cohen EP, Craston A, Duhig T, Johnstone G, Barnes D: The locus of the polymorphic epithelial mucin (PEM) tumour antigen on chromosome 1q21 shows a high frequency of alteration in primary human breast tumours. Int J Cancer 45: 431-435, 1990b.

Gendler SJ, Spicer AP: Epithelial mucin genes. Annu Rev Phisiol 57: 607-634, 1995.

Gerard C, Eddy RL, Shows TS: The core polypeptide of cystic fribosis tracheal mucin contains a tandem repeat structure. Evidence for a common mucin in airway and gastrointestinal tissue. J Clin Invest 86: 1921-1927, 1990.

Girling A, Bartkova J, Burchell J, Gendler S, Gliiet C, Taylor-Papadimitriou J: A core protein epitope of the polymorphic epithelial mucin detected by monoclonal antibody SM-3 is selectively exposed in a range of primary carcinomas. Int J Cancer 43: 1072-1076, 1989.

Go MF: What are the host factors that place an individual at risk for *Helicobacter pylori*-associated disease. Gastroenterol 113: S15-S20, 1997.

Gong J, Chen D, Kashiwaba M, li Y, Chen L, Takeuchi H, Rowse GJ, Gendler S, Kufe D: Reversal of tolerance to human MUC1 antigen in MUC1 transgenic mice immunized with fusions of dendritic and carcinoma cells. Proc Natl Acad Sci USA 95: 6279-6283, 1998.

Goydos JS, Elder E, Whiteside T, Finn OJ, Lotze MT: A Phase I trial of a synthetic mucin peptide vaccine. J Surg Res 63: 298-304, 1996.

Griffiths AB, Burchell J, Gendler S, Lewis A, Blight K, Tilly R, Taylor-Papadimitriou J: Immmunological analysis of mucin molecules expressed by normal and malignant mammary epithelial cells. Int J Cancer 40: 319-327, 1987.

Griffiths D, Matthews DJ, West L, Attwood J, Povey S, Swallow DM, Gum JR, Kim YS: Assignment of the polymorphic intestinal mucin gene (MUC2) to chromosome 11p15. Ann Hum Genet 54: 277-285, 1990.

Gross MS, Guyonnet-Duperat V, Porchet N, Bernheim A, Aubert J-P, Van cong N: Mucin 4 (MUC4) gene: regional assignment (3q29) and RFLP analysis. Ann Génét 35: 21-26, 1992.

Gum JR, Byrd JC, Hicks JW, Toribara NW, Lamport DTA, Kim YS: Molecular cloning of human intestinal cDNAs. J Biol Chem 264: 6480-6487, 1989.

Gum JR, Hicks JW, Swallow DM, Lagace RL, Byrd JC, Lamport DTA, Siddiki B, Kim YS: Molecular cloning of cDNAs derived from a novel human intestinal mucin gene. Biochem Biophys Res Commun 171: 407-415, 1990.

Gum JR, Hicks JW, Lagace RE, Byrd JC, Toribara NW, Siddiki B, Fearney FJ, Lamport DT, Kim YS: Molecular cloning of rat intestinal mucin. Lack of conservation between mammalian species. J Biol Chem 266: 22733-22738, 1991.

Gum JR, Hicks JW, Toribara NW, Roteh E-M, Lagace RE, Kim YS: The human MUC2 intestinal mucin has cystein-rich subdomains located both upstream and downstream of its central repetitive region. J Biol Chem 267: 21375-21383, 1992.

Gum JR, Hicks JW, Toribara NW, Siddiki B. Kim YS: Molecular cloning of human intestinal mucin (MUC2) cDNA. J Biol Chem 269: 2440-2446, 1994.

Gum JR, Hicks JW, Kim YS: Identification and characterization of the *MUC2* (human intestinal mucin) gene 5'-flanking region: promoter activity in cultures cells. Biochem J 325: 259-267, 1997a.

Gum JR, Ho JJL, Pratt WS, Hicks JW, Hill As, Vinall LE, Roberton AM, Swallow DM, Kim YS: MUC3 human intestinal mucin: analysis of gene structure, the carboxyl terminus, and a novel upstream repetitive region. J Biol Chem 272: 26678-26686, 1997b.

Guo SW, Thompson EA: Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics 48: 361-372, 1992.

Guyonnet-Duperat V, Audie JP, Debailleul V, Laine A, Buisisne M, Galiegue-Zouitina S, Pigny P, Degand P, Aubert J-P, Porchet N: Characterization of the human mucin gene MUC5AC: a consensus cystein-rich domain for 11p15 mucin genes? Biochem J 305: 211-219, 1995.

Guzman K, Bader T, Nettesheim P: Regulation of MUC5 and MUC1 genes expression: correlation with airway mucuos differentiation. Am J Physiol 270: L846-L853, 1996.

Hanski C, Drechsler K, Hanish FG, Sheehan J, Manske M, Ogorek D, Klussmann E, Hanski ML, Blank M, Xing PX, McKenzie IFC, Devine PL, Riecken EO: Altered glycosylation of the MUC1 protein core contributes to the colon carcinoma-associated increase of mucin-bound sialyl-Lewisx expression. Cancer Res 53: 4082-4088, 1993.

Hanski C, Riede E, Gratchev A, Foss HD, Bohm C, Klumann E, Hummel M, Mann B, Buhr HJ, Stein H, Kim YS, Gum J, Riecken EO: MUC 2 gene suppression in human colorectal carcinomas and their metastases: in vitro evidence of the modulatory role of DNA methylation. Lab Invest 77: 685-695, 1997.

Hansson G, Baeckstrom D, Carlstedt I, Klinga-Levan K: Molecular cloning for a region of an apoprotein from the insoluble mucin complex of rat small intestine. Biochem Biophys Res Commun 198: 181-190, 1994.

Hauser F, Hoffman W: P-domains as shuffled cystein-rich modules in integumentary mucin C.1 (FIM-C.1) from *Xenopus laevis*. Polydispersity and genetic polymorphism. J Biol Chem 267: 24620-24624, 1992.

Hey NA, Graham RA, Seif MW, Aplin JD: The polymorphic epithelial mucin MUC1 in human endometrium is regulated with maximal expression in the implantation phase. J Clin Endocrinol Metab 78: 337-342, 1994.

Hilkens J, Kroezen V, Bronfrer JM, De J, Bakker M, Bruning PF: MAM-6 antigen, a new serum marker for breast cancer monitoring. Cancer Res 46: 2586-2587, 1986.

Hilkens J: Biochemistry and function of mucins in malignant disease. Cancer Rev 11-12: 25-53, 1988.

Hilkens J, Ligtenberg MJ, Vos H, Litivinov SV: Cell membrane-associated mucins and their adhesion-modulating property. Trends Biochem Sci 17: 359-363, 1992.

Hill HD, Schwyzer M, Steinman HM, Hill RL: Ovine submaxillary mucin. Primary structure and peptide substrates of UDP-N-acetylgalactosamine:mucin transferase. J Biol Chem 252: 3799-3804, 1997b.

Hill HD; Reynolds JA, Hill RL: Purification, composition, molecular weight, and subunit structure of ovine submaxillary mucin. J Biol Chem 252: 3791-3798, 1997a.

Hirohashi S, Clausen H, Yamada T, Shimosato Y, Hakomori S: Blood group A cross-reacting epitope defined by monoclonal antibodies NCC-LU-35 and –81 expressed in cancer of blood group O or B individuals: its identification as Tn antigens. Proc Natl Acad Sci USA 82: 7039-7043, 1985.

Ho SB, Kim YS: Carbohydrates antigens on cancer-associated mucin-like molecules. Sem Cancer Biol 2: 389-400, 1991.

Ho SB, Niehans GA, Lyftogt C, Yan PS, Cherwitz DL, Gum ET, Dahiya R, Kim YS: Heterogeneity of mucin gene expression in normal and neoplastic tissues. Cancer Res 53: 641-651, 1993.

Ho SB, Shekels L, Toribara NW, Kim YS, Lyftogt C, Cherwitz DL, Niehans GA: Mucin gene expression in normal, preneoplastic, and neoplastic human gastric epithelium. Cancer Res 55: 2681-2690, 1995.

Hoffmann W: A new repetitive protein from *Xenopus laevis* skin highly homologous to pancreatic spasmolytic polypeptide. J Biol Chem 263: 7686-7690, 1988.

Hoffmann W, Hauser F: The P-domain or trefoil motif: a role in renewal and pathology of mucous epithelial? Trends Biochem Sci 18: 239-243, 1993.

Holmgren J, Lonnroth I, Masson J, Svennerholm L: Interaction of cholera toxin and membrane GM1 ganglioside of small intestine. Proc Natl Acad Sci USA 72: 2520-2524, 1975.

Hultgren SJ, Abraham S, Caparon M, Falk P, St Geme JW III, Normark S: Pilus and non-pilus bacterial adhesins: assembly and function in cell recognition. Cell (Cambridge, Mass.) 73: 887-901, 1993.

Ilg T, Stierhof YD, Craik D, Simpson R, Handman E, Bacic A: Purification and structural characterization of a filamentous, mucin like proteophosphoglycan secreted by *Leishmania* parasites. J Biol Chem 271: 21583-21596, 1996.

Ilver D, Arnqvist A, Ogren J, Frick IM, Kersulyte D, Incecik E, Berg D, Covacci A, Emgstrand L, Borén T: *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. Science 279: 373-377, 1998.

Inatomi T, Tisdale A, Zhan Q, Spurr-Michaud S, Gipson I: Cloning of rat Muc5AC mucin gene: comparison of its structure and tissue distribution to that of human and mouse homologues. Biochem Biophys Res Commun 236: 789-797, 1997.

International Agency for Research on Cancer. Schistosomes, liver flukes and *Helicobacter pylori*. Working Group on the Evaluation of Carcinogenic Risks to Humans. IARC Monogr Eval Carcinog Risks Hum 61: 177-241, 1994.

Iwata H, Itzkowitz SH, Werther JL, Hayashi K, Nakamura H, Ichinose M, Miki K, Tatematsu M: Expression of sialosyl-Tn in intestinal type cancer cells of human gastric cancers. Acta Pathol Jpn 43: 646-653, 1993.

Jany BH, Gallup MW, Yan P-S, Gum JR, Kim YS, Basbaum CB: Human bronchus and intestine express the same mucin gene. J Clin Invest 87: 77-82, 1991.

Jeffreys AJ, Tamaki K, MacLeod A, Monckon DG, Neil DL, Armour JAL: Complex gene conversion events in germline mutation at human minisatellites. Nat Genet 6: 136-145, 1994.

Jentoff N: Why are proteins O-glycosylated? Trends Biochem 15: 291-294, 1990.

Jerome KR, Barnd DL, Bendt KM, Boyer CM, Taylor-Papadimitriou J, McKenzie IFC, Bast RC, Finn OJ: Cytotoxic T-lymphocytes derived from patients with breast adenocarcinoma recognize an epitope present on the protein core of a mucin molecule preferentially expressed by malignant cells. Cancer Res 51: 2908-2916, 1991.

Jones GW, Isaacson RE: Proteinaceous bacterial adhesins and their receptors. CRC Crit Rev Microbiol 10: 229-260, 1983.

Karlsson KA: Animal glycosphingolipids as membrane attachment sites for bacteria. Annu Rev Biochem 58: 309-350, 1989.

Karlsson S, Swallow D, Griffiths B, Corney G, Hopkinson DA: A genetic polymorphism of a human urinary mucin. Ann Hum Genet 47: 263-269, 1983.

Keates A, Nunes D, Afdhal N, Troxler R, Offner G: Molecular cloning of a major human gall bladder mucin: complete C-terminal sequence and genomic organization of MUC5B. Biochem J 324: 295-303, 1997.

Khatri IA, Forstner GG, Forstner JF: Suggestive evidence for two different mucin genes in rat intestine. Biochem J 294: 391-399, 1993.

Khatri IA, Forstner GG, Forstner JF: The carboxyl-terminal sequence of rat intestinal mucin rMuc3 contains a putative transmembrane region and two EGF-like motifs. Biochim Biophys Acta 1326: 7-11, 1997.

Khatri IA, Specian RD, Forstner JF: Rat Muc3, a membrane-associated mucin of rat intestine: mRNA localization, peptide expression and separation on a CsCl gradient. 5th International Workshop on Carcinoma-Associated Mucins, 18-23 July 1998, Cambridge.

Kim YS, Gum JR, Byrd JC, Toribara NW: The structure of human intestinal apomucins. Am Rev Respir Dis 144: S10-S14, 1991.

Kim YS, Gum JR: Diversity of mucin genes, structure, function, and expression. Gastroenterol 109: 999-1001, 1995.

Kim YS, Gum J, Brockhausen I: Mucin glycoproteins in neoplasia. Glycoconj J 13: 693-707, 1996.

Kjeldsen T, Clausen H, Hirohashi S, Ogawa T, Iijima H, Hakomori S: Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl- 2-6α-N-acetylgalactosaminyl (sialosyl-Tn) epitope. Cancer Res 4: 2214-2220, 1988.

Klomp L, Van Rens L, Strous G: Cloning and analysis of human gastric mucin cDNA reveals two types of conserved cystein-rich domains. Biochem J 308: 831-838, 1995.

Kondo K, Kohno N, Yokoyama A, Hiwada K: Decreased MUC1 expression induces E-cadherin-mediated cell adhesion of breast cancer cell lines. Cancer Res 58: 2014-2019, 1998.

Kotera Y, Fontenot JD, Pecher G, Metzgar RS, Finn OJ: Humoral immunity against a tandem repeat epitope of human mucin MUC1 in sera from breast, pancreatic, and colon cancer patients. Cancer Res 54: 2856-2860, 1994.

Kovarik A, Peat N, Wilson D, Gendler S, Taylor-Papadimitriou J: Analysis of the tissue-specific promoter of the *MUC1* gene. J Biol Chem 268: 9917-9926, 1993.

Kramero AA, Arbatsky NP, Rozovsky YM, Mikhaleva EA, Polesskaya OO; Gvozdev VA, Shibaev VN: Mucin type glycoprotein from *Drosophila melanogaster* embrionic cells: characterization of carbohydrate component. FEBS Letters 378: 213-218, 1996.

Kreuning J, Lindermen J, Biemond I, Lamers CBHW: Relation between IgG and IgA antibodies titres against *Helicobacter pylori* in serum and severity of gastritis in asymptomatic subjects. J Clin Pathol 47: 227-231, 1994.

Kushima R, Jancic S, Hattori T: Association between expression of sialosyl-Tn antigen and intestinalization of gastric carcinomas. Int J Cancer 55: 904-908, 1993.

La Vecchia C, Lucchini F, Negri E, Boyle P, Maisonneuve P, Levi F: Trends of cancer mortality in Europe, 1955-1989: I, Digestive sites. Eur J Cancer 28: 132-235, 1992.

Laine A: Complete C-terminal sequence and genomic organization of MUC5B? Biochem J 331: 349-350, 1998.

Lamblin G, Lhermitte M, Klein A, Houdret N, Scharfman A, Ramphal R, Poussel P: The carbohydrate diversity of human respiratory mucins: a protection of the underlying mucosa? Am Rev Resp Dis 144: S19-S24, 1991.

Lan MS, Khorrami A, Kaufman B, Metzgar RS: Molecular characterization of a mucin-type antigen associated with human pancreatic cancer. J Biol Chem 262: 12863-12870, 1987.

Lan MS, Batra SK, Qi W-N, Metzgar RS, Hollingsworth MA: Cloning and sequencing a pancreatic tumor mucin cDNA. J Biol Chem 265: 15294-15299, 1990.

Lancaster CA, Peat N, Duhig T, Wilson D, Taylor-Papadimitriou J, Gendler S: Structure and expression of the human polymorphic epithelial mucin gene: an expressed VNTR unit. Biochem Biophys Res Commun 173: 1019-1029, 1990.

Lapensée L, Paquette Y, Bleau G: Allelic polymorphism and chromosomal localization of the human oviductin gene (*MUC9*). Fertil Steril 68: 702-708, 1997.

Lesuffleur T, Zweibaum A, Real FX: Mucins in normal and neoplastic human gastrointestinal tissues. Critical Rev Oncol/Hematol 17: 153-180, 1994.

Lesuffluer T, Roche F, Hill A, Lacasa M, Fox M, Swallow D, Zweibaum A, Real FX: Characterization of a mucin cDNA clone isolated from HT-29 mucus-secreting cells. J Biol Chem 270: 13665-13673, 1995.

Lewis PO, Zaykin D: Genetic Data Analysis: Software for the analysis of discrete genetic data. Version 1.0.1997; Free program distributed by the authors over internet from GDA Home page at http://chee.unm.edu/gda/.

Li D, Gallup M, Fan N, Szymkiwksi D, Basbaum C: Cloning of the amino-terminal and 5'-flanking region of the *MUC5AC* mucin gene and transcriptional up-regulation by bacterial exproducts. J Biol Chem 273: 6812-6820, 1998.

Li Y, Bharti A, Chen D, Gong J, Kufe D: Interaction of glycogen synthase kinase 3beta with the DF3/MUC1 carcinoma-associated antigen and beta-catenin. Mol Cell Biol 18: 7216-7224, 1998.

Ligtenberg MJL, Vos HL, Gennissen AMC, Hilkens J: Episialin, a carcinoma-asociated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. J Biol Chem 265:5573-5578, 1990.

Ligtenberg MJL, Gennissen AMC, Vos HL, Hilkens J: A single nucleotide polymorphism in an exon dictates allele dependent differential splicing of episialin mRNA. Nucleic Acids Res 19: 297-301, 1991.

Ligtenberg MJL, Kruijshaar L, Buijs F, van Meijer M, Litvinov SV, Hilkens J: Cell-associated episialin is a complex containing two proteins derived from a common precursor. J Biol Chem 267: 6171-6177, 1992a.

Ligtenberg MJL, Buijs F, Vos H, Hilkens J: Suppression of cellular aggregation by high levels of episialin. Cancer Res 52: 2318-2324, 1992b.

Litvinov SV, Hilkens J: The epithelial sialomucin, episialin, is sialylated during recycling. J Biol Chem 268: 21364-21371, 1993.

Lloyd K, Burchell J, Kudryashov V, Yin B, Taylor-Papadimitriou J: Comparison of O-linked carbohydrate chains in MUC1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. J Biol Chem 271: 33325-33334, 1996.

Magarian-Blander J, Domenech N, Finn OJ: Specific and effective T-cell recognition of cells transfected with truncated human mucin cDNA. Ann N Y Acad Sci 690: 231-242, 1993.

Majuri ML, Mattila P, Renkonen R: Recombiant E-selectin-protein mediates tumor cell adhesion via sialyl-Le(a) and sialyl-Le(x). Biochem Biophys Res Commun 182: 1376-1382, 1992.

Mancuzo DJ, Tuley EA, Westfield LA, Worral NK, Shelton-Inloes BB, Sorace JM, Alevy YG, Sadler JE: Structure of the gene for human von Willebrand factor : J Biol Chem 264: 19514-19527, 1989.

Marshall BJ, Warren JR: Unidentified curved bacilli on gastric epithelium in chronic active gastritis. Lancet *I*: 1273-1274, 1983.

Martinson JJ, Harding RM, Philippon G, Flye Sainte-Marie F, Roux J, Boyce AJ, Clegg JB: Demographic reductions and genetic bottlenecks in humans: minisatellite allelic distributions in Oceania. Hum Genet 91: 445-450, 1993.

McGuckin MA, Ramm LE, Joy GJ, Free KE, Ward BG: Preoperative discrimitation between ovarian carcinoma, non-ovarian gynecological malignancy and benign adnexal masses using serum levels of CA125 and the polymorphic epithelial mucin antigens CASA, OSA and MSA. Int J Gynecol Cancer 2: 119-128, 1992.

McKenzie IFC, Xing PX: Mucins in breast cancer: recent immunological advances. Cancer Cells 2: 75-78, 1990.

Meerzaman D, Charles P, Daskal E, Polymeropoulos M, Martin B, Rose M: Cloning and analysis of cDNA encoding a major airway glycoprotein, human tracheobronchial mucin (MUC5). J Biol Chem 269: 12932-12939, 1994.

Mégraud F, Brassens-Rabbé MP, Denis F, Belbouri A, Hoa DQ: Seroepidemiology of *campylobacter pylori* infection in various populations. J Clin Microbiol 27: 1870-1873, 1989.

Metzgar RS, Rodriguez N, Finn OJ, Lan MS, Daasch VN, Fernsten PD, Meyers NC, Sindelar WF, Sandler RS, Seigler HF: Detection of a pancreatic cancer-associated antigen (DU-PAN-2 antigen) in serum and ascites of patients with adenocarcinoma. Proc Natl Acad Sci USA 81: 5242-5246, 1984.

Moniaux N, Nollet S, Desseyn JL, Buisine MP, Laine A, Porchet N, Aubert J-P: Genomic organization of the human mucin genes. MUC4: an organization which differs from the 11p15 genes. 5th International Workshop on Carcinoma-Associated Mucins, 18-23 July 1998, Cambridge.

Moran AP: Cell surface characteristics of *Helicobacter pylori* . FEMSIM 10: 271-280, 1995.

Motomura K, Nishisho I, Takai S, Tateishi H, Okazaki M, Yamamoto M, Miki T, Honjo T, Mori T: Loss of alleles at loci on chromosome 13 in human primary gastric cancers. Genomics 2: 180-184, 1988.

Mourant AE, Kopec AC, Domaniewska-Sobczak K: Blood group and diseases. A study of associations of diseases with blood groups and other polymorphisms. Oxford: Oxford Medical Publications, 1978.

Mullenbach R, Lagoda PJL, Welter C: An efficient salt-chloroform extraction of DNA from blood and tissues. Trends Genet 5: 391, 1989.

Muller S, Goletz S, Packer N, Gooley A, Lawson AM, Hanish F-G: Localization of O-glycosylation sites on glycopeptide fragments from lactation-associated MUC1. J Biol Chem 272: 24780-24793, 1997.

Muñoz N: Descriptive epidemiology of stomach cancer. In: PI Reed and MJ Hill (eds.), Gastric carcinogenesis, pp 51-69, Elsevier, Amsterdam, 1988.

Murata K, Egami H, Shibata Y, Sakamoto K, Misumi A, Ogawa M: Expression of bloos group- related antigens, ABH, Lewis a, Lewis b, Lewis x, Lewis y, CA 19.9, and CSLEX1 in early cancer, intestinal metaplasia, and uninvolved mucosa of the stomach. Am J Clin Pathol 98: 67-75, 1992.

Nei M: Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590, 1978.

Nielsen PA, Bennett EP, Wandal HH, Therkildson MH, hannibal J, Clausen H: Identification of a major high molecular weight salivary mucin (MG1) as tracheobronchial mucin MUC5B. Glycobiol 7: 413-419, 1997.

Nollet S, Moniaux N, Maury J, Petitprez D, Degand P, Laine A, porchet N, Aubert J-P: Human mucin gene MUC4: organization of its 5'-region and polymorphism of its central tandem repeat array. Biochem J 332: 739-748, 1998.

Nomura A, Stemmermann GN, Chyou PH, Kato I, Perez-Perez GI, Blaser MJ: *Helicobacter pylori* infection in gastric carcinoma among Japanese Americans in Hawaii. N Engl J Med 325: 1132-1136, 1991.

Offner G: Complete C-terminal sequence and genomic organization of MUC5B? a reply. Biochem J 331: 350-352, 1998.

Ogata S, Maimonis PJ, Itzkovitz SH: Mucins bearing the cancer-associated sialosyl-Tn antigen mediate inhibition of natural killer cell cytotoxicity. Cancer Res 52: 4741-4746, 1992a.

Ogata S, Uehara H, Chen A, Itzkowitz SH: Mucin gene expression in colonic tissues and cell lines. Cancer Res 52: 5971-5978, 1992b.

Ohmori H, Dohrman AF, Gallup M, Tsuda T, Kai H, Gum JR, Kim YS, Basbaum CB: Molecular cloning of amino terminal region of a rat Muc2 mucin gene homologue: evidence for expression in both intestine and airway. J Biol Chem 269: 17833-17840, 1994.

Parry G, Beck JC, Moss L, Bartley J, Ojakian GK: Determination of apical membrane polarity in mammary epithelial cell cultures: the role of cell-cell, cell-substratum- and membrane-cytoskeleton interactions. Exp Cell Res 188: 302-311, 1990.

Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman N, Orentreich N, Sibley RK: *Helicobacter pylori* infection and the risk of gastric carcinoma. N Engl J Med 325: 1127-1131, 1991.

Paulson JC, Colley KJ: Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. J Biol Chem 264: 17615-17618, 1989.

Pemberton L, Taylor-Papadimitriou J, Gendler S: Antibodies to the cytoplasmatic domain of MUC1 mucin show conservation throughout mammals. Biochem Biophys Res Commun 185: 167-175, 1992.

Perez-Perez GI, Dworkin BM, Chodos JE, Blaser MJ: *Campylobacter pylori* antibodies in humans. Ann Intern Med 109: 11-17, 1988.

Pigny P, Pratt WS, Laine A, Leclercq A, Swallow DM, Nguyen VC, Aubert J-P, Porchet N: The MUC5AC gene: RFLP analysis with the Jer58 probe. Hum Genet 96: 367-368, 1995.

Pigny P, Guyonnet-Duperat V, Hill As, Pratt WS, Galiegue-Zouitina S, Hooge MC, Laine A, Van-Seuningen I, Degand P, Gum JR, Kim YS, Swallow DM, Aubert J-P, Porchet N: Human mucin genes assigned to 11p15.5: identification and organization of a cluster of genes. Genomics 38: 340-352, 1996.

Porchet N, Van Cong N, Dufosse J, Audie JP, Guyonnet-Duperat V, Gross MS, Denis C, Degand P, Bernheim A, Aubert J-P: Molecular cloning and chromosomal localization of a novel human tracheobronchial mucin cDNA containing tandemly repeated sequences of 48-bp. Biochem Biophys Res Commun 175: 414-422, 1991.

Pratt W, Islam I, Swallow DM: Two additional polymorphisms within the hypervariable MUC1 gene: association of alleles either side of the VNTR region. Ann Hum Genet 60: 21-28, 1996.

Price MR, Edwards S, Owainati A, Bullock JE, Ferry B, Robins RA, Baldwin RW: Multiple epitopes on human breast carcinoma-associated antigen. Int J Cancer 36: 567-572, 1985.

Probst JC, Gertzen EM, Hoffmann W: An integumentary mucin (FIM-B.1) from *Xenopus laevis* homologous with the von Willebrand factor. Biochemistry 29: 6240-6244, 1990.

Probst JC, Hauser F, Joba W, Hoffmann W: The polymorphic integumentary mucin B.1 from *Xenopus laevis* contains the short consensus repeat. J Biol Chem 267: 6310-6316, 1992.

Ranzani GN, Renault B, Pellegata NS, Fattorini P, Magni E, Bacci F, Amadori D: Loss of heterozygosity and K-ras gene mutations in gastric cancer. Hum Genet 92: 244-249, 1993.

Raymond J, Kalach N, Bergeret M, Barbet JP, Benhamou PH, Gendrel D, Dupont C: Evaluation of a serological test for diagnosis of *Helicobacter pylori* infection in children. Eur J Clin Microbiol Infect Dis 15: 415-417, 1996.

Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86: 248-249, 1995.

Reddy M, Bobek L, Haraszthy G, Biesbrock A, Levine M: Structural features of the low-molecular-mass human salivary mucin. Biochem J 287: 639-643, 1992.

Reis CA, David L, Nielsen PA, Clausen H, Mirgorodskaya E, Roepstorff P, Sobrinho-Simões M: Immunohistochemical study of MUC5AC expression in human gastric carcinomas using a novel monoclonal antibody. Int J Cancer 74: 112-121, 1997.

Reis CA, David L, Seixas M, Burchell J, Sobrinho-Simões M: Expression of fully and under-glycosylated forms of MUC1 mucin in gastric carcinoma. Int J Cancer 79: 402-410, 1998a.

Reis CA, Sørensen T, Mandel U, David L, Mirgorodskaya E, Roepstorff P, Kihlberg J, Stig-Hansen JE, Clausen H: Development and characterization of an antibody directed to an α-N-acetyl-D-galactosamine glycosylated MUC2 peptide. Glycoconj J 15: 51-62, 1998b.

Reia CA, David L, Correa P, Carneiro F, DeBolós C, Garcia E, Mandel U, Clausen H, Sobrinho-Simões M: Intestinal metaplasia of human stomach displays distinct patterns of mucin (MUC1, MUC2, MUC5AC, MUC6) expression. Cancer Res (accepted for publication).

Replogle ML, Glaser SL, Hiatt RA, Parsonnet J: Biologic sex as a risk factor for *Helicobacter pylori* infection in healthy young adults. Am J Epidemiol 142: 856-863, 1995.

Rice GE, Bevilacqua MP: An inducible endothelial cell surface glycoprotein mediates malanoma adhesion. Science 246: 1303-1306, 1989.

Rodriguez E, Rao PH, Ladanyi M, Altorki N, Albino AP, Kelsen DP, Jhanwar SC, Chaganti RSK: 11p13-15 is a specific region of chromosomal rearragment in gastric and esophageal adenocarcinomas. Cancer Res 50: 6410-6416, 1990.

Rose MC, Kaufman B, Martin BM: Proteolytic fragmentation and peptide mapping of human carboxyamidomethylated tracheobronchial mucin. J Biol Chem 264: 8193-8199.

Rughetti A, Turchi V, Ghetti CA, Scambia G, Panici PB, Roncucci G, Mancoso S, Frati L, Nuti M: Human B-cell immune response to the polymorphic epithelial mucin. Cancer Res 53: 2457-2459, 1993.

Rye PD, Bell SC, Walker RA: Immunohistochemical expression of tumour-associated glycoprotein and polymorphic epithelial mucin in the human endometrium during menstrual cycle. J Reprod Fertil 97: 551-556, 1993.

Sadler JE: von Willebrand factor. J Biol Chem 266: 22777-22780, 1991.

Saitoh T, Natomi H, Zhao WL, Okuzumi K, Sugano K, Iwamori M, Nagai Y: Identification of glycolipid receptors for *Helicobacter pylori* by TLC-immunostaining. FEBS Lett 282: 385-387, 1991.

Sano T, Tsujino T, Yoshida K, Nakayama H, Haruma K, Ito H, Nakamura Y, Kajiyama G, Tahara E: Frequent loss of heterozigosity on chromosome 1q, 5q, and 17p in human gastric carcinomas. Cancer Res 51: 2926-2931, 1991.

Sawada T, Ho J, Chung Y-S, Sowa M, Kim YS: E-selectin binding by pancreatic tumor cells is inhibited by cancer sera. Int J Cancer 57: 901-907, 1994.

Scharfman A, Kroczynski H, Carnoy C, Van Brussel E, Lamblin G, Ramphal R, Roussel P: Adhesion of *Pseudomonas aeruginosa* to respiratory mucins and expression of mucin-binding proteins are increased by limiting iron during growth. Infect Immun 64: 5417-5420, 1996.

Seruca R, Constância M, Dos Santos N, David L, Queimado L, Carvalho F, Carneiro F: Allele loss in human gastric carcinomas: relation to tumor progression and differentiation. Int J Oncol 7: 1159-1166, 1995.

Sham PC, Curtis D: Monte Carlo tests for association between disease and alleles at highly polymorphic loci. Ann Hum Genet 59: 97-105, 1995.

Shankar V, Gilmore M, Elkins R, Sachdev G: A novel human airway mucin cDNA encodes a protein with unique tandem-repeat organization. Biochem J 300: 295-298, 1994.

Shankar V, Pichan P, Eddy R, Tonk V, Nowak N, Sait S, Shows T, Schultz R, Gotway G, Elkins R, Gilmore M, Sachdev G: Chromosomal localization of a human mucin gene (MUC8) and cloning of the cDNA corresponding to the carboxy terminus. Am J Resp Cell Mol Biol 16: 232-241, 1997.

Sharon N: Bacterial lectins, cell-cell recognition and infectious disease. FEBS Lett. 217: 145-157, 1987.

Shekels L, Lyftogt C, Kieliszewski M, Filie JD, Kozac CA, Ho S: Mouse gastric mucin: Cloning and chromosomal localization. Biochem J 311: 775-785, 1995.

Shekels L, Hunninghake D, Tisdale A, Gipson I, Kieliszewski M, Kozac C, Ho S: Cloning and characterization of mouse intestinal Muc3 mucin: 3' sequence contains epidermal-growth-factor-like domains. Biochem J 330: 1301-1308, 1998.

Shelton-Inloes BB, Titani K, Sadler JE: cDNA sequences for human von Willebrand factor reveal five types of repetitive domains and five possible protein sequence polymorphism. Biochemistry 25: 3164-3171, 1986.

Sheng Z, Hull SR, Carraway KL: Biosynthesis of the cell surface sialo-mucin complex of ascites 13762 rat mammary adenocarcinoma cells from a high Mr precursor. J Biol Chem 265: 8505-8510, 1990.

Sherblom AP, Moody CE: Cell surface sialomucin and resistance to natural cell-mediated cytotoxicity of rat mammary tumor ascites cells. Cancer Res 9: 4543-4546, 1986.

Shimizu M, Yamauchi K: Isolation and characterization of mucin-like glycoproteins in human milk fat globule membranes. J Biochem 91: 515-524, 1982.

Shimizu Y, Shaw S: Mucins in mainstream. Nature 366: 630-631, 1993.

Shimoyama T, Fukuda S, Tanaka M, Mikami T, Munakata A, Crabtree JE: CagA seropositivity associated with development of gastric cancer in a Japanese population. J Clin Pathol 51: 225-228, 1998.

Siddiqui J, Abe M, Hayes D, Shani E, Yunis E, Kufe D: Isolation and sequencing of a cDNA coding for the human DF3 breast carcinoma-associated antigen. Proc Natl Acad Sci USA 85: 2320-2323, 1988.

Soares J, Carneiro F, Cotter J, Perreira F, Fonseca E, Pinho C: Prevalência da infecção pelo *Helicobacter pyloril* e características da mucosa gástrica em doentes dispépticos sujeitos a endoscopia no norte de Portugal. Rev Gastroenterol 10: 119-131, 1993.

Solcia E, Fiocca R, Luinetti O, Villani L, Padivan L, Calistri D, Ranzani G, Chiaravalli A, Capella C: Intestinal and diffuse gastric cancer arise in a different background of *Helicobacter pylori* gastritis through different gene involvement. Am J Surg Path 20 (Suppl. 1): S8-S22, 1996.

Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503-517, 1975.

Spicer AP, Parry G, Patton S, Gendler S: Molecular cloning and analysis of the mouse homologue of the tumour-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of a minisatellite-like polymorphism. J Biol Chem 266: 15099.15109, 1991.

Steck PA, Nicolson GL: Cell surface glycoprotein of 13762NF mammary adenocarcinoma clones of differing metastasis potentials. Exp Cell Res 147: 255-267, 1983.

Strous GJ, Dekker J: Mucin-type glycoproteins. Crit Rev Biochem Mol Biol 27: 57-92, 1992.

Swallow DM, Gendler S, Griffiths B, Corney G, Taylor-Papadimitriou J, Bramwell ME: The human tumour-associated epithelial mucins are coded by an expressed hypervariable gene locus PUM. Nature 328: 82-84, 1987a.

Swallow DM, Gendler S, Griffiths B, Kearney A, Povey S, Sheer D, Palmer RW, Taylor-Papadimitriou J: The hypervariable gene locus PUM, which codes for the tumour associated epithelial mucins is located on chromosome 1, within the region 1q21-24. Ann Hum Genet 51: 289-294, 1987b.

Swallow DM: Introduction to epithelial mucins. 5th International Workshop on Carcinoma-Associated Mucins, 18-23 July 1998, Cambridge.

Swofford DL, Selander RB: BIOSYS-1: A Fortran program for the comprehensive analysis for electrophoretic data in population genetics and systematics. J Hered 72: 281-283, 1981.

Sylvester FA, Philpott D, Gold B, Lastovica A, Forstner JF: Adherence to lipids and intestinal mucin by a recently recognized human pathogen, *Campylobacter upsaliensis*. Infect Immun 64: 4060-4066, 1996.

Talley NJ, Zinsmeister AR, Weaver A, DiMagno EP, Carpenter HA, Perez-Perez GI, Blaser MJ: Gastric adenocarcinoma and *Helicobacter pylori* infection. J Natl Cancer Inst 83: 1734-1739, 1991.

Taylor-Papadimitriou J, Peterson JA, Arklie J, Burchell J, ceriani RL, Bodmer WF: Monoclonal antibodies to epithelium-specific components of the human milk fat globule membrane: production and reaction with cells in culture. Int J Cancer 28: 17-21, 1981.

Taylor-Papadimitriou J, Gendler SJ: Molecular aspects of mucins. Cancer Rev 11-12: 11-24, 1988.

Taylor-Papadimitriou J, Stewart L, Burchell J, Beverley P: The polymorphic epithelial mucin as a target for immunotherapy. Ann N Y Acad Sci 690: 69-79, 1993.

Thor A, Ohuchi N, Szpak CA, Johnston WW, Schlom J: Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. Cancer Res 46: 3118-3124, 1986.

Toribara NW, Gum JR, Culhane PJ, Lagace RE, Hicks JW, Peterson GM, Kim YS: MUC2 human small intestine mucin gene structure. J Clin Invest 88: 1005-1013, 1991.

Toribara NW, Roberton A, Ho S, Kuo W-L, Gum E, Hicks J, Gum J, Byrd J, Siddiki B, Kim YS: Human gastric mucin. Identification of a unique species by expression cloning. J Biol Chem 268: 5879-5885, 1993.

Toribara NW, Ho S, Gum E, Gum J, Lau P, Kim YS: The carboxyl-terminal sequence of the human secretory mucin, MUC6. J Biol Chem 272: 16398-16403, 1997.

Troxler R, Iontcheva I, Oppenheim F, Nunes D, Offner G: Molecular characterization of a major high molecular weight mucin from human sublingual gland. Glicobiol 7: 965-973, 1997.

Tsuda T, Gallup M, Jany B, Gum J, Kim YS, Basbaum C: Characterization of a rat airway cDNA encoding a mucin-like protein. Biochem Biophys Res Commun 195: 363-373, 1993.

Turner B, Bhaskar R, Hadzopoulou-Claradas M, Specian R, Lamont T: Isolation and characterization of cDNA clones encoding pig gastric mucin. Biochem J 308: 89-96, 1995.

Uchino S, Tsuda H, Noguchi M, Yokota J, Terada M, Saito T, Kobayashi M, Sugimura T, Hirohashi S: Frequent loss of heterozygosity at the DCC locus in gastric cancer. Cancer Res 52: 3099-3102, 1992.

Van Cong N, Aubert J-P, Gross M, Porchet N, Degand P, Frézal J: Assignment of human tracheobrochial mucin genes(s) to 11p15 and tracheobronchial mucin-related sequence to chromosome 13. Hum Genet 86: 167-172, 1990.

Van de Bovenkamp J, Hau C, Strous G, Buller H, Deckker J, Einerhand A: Molecular cloning of human gastric mucin MUC5AC reveals conserved cystein-rich D-domains and a putative leucine zipper motif. Biochem Biophys Res Commun 245: 853-859, 1998.

Van de Wiel-Van Kemenade E, Ligtenberg MJ, deBoer AJ, Buijs F, Vos HL, Melief CJ, Hilkens J, Figdor CG: Episialin (MUC1) inhibits cytotoxic lymphocyte-target cell interaction. J Immunol 151: 767-776, 1993.

Van Doorn LJ, Figueiredo C, Rossau R, Jannes G, van Asbroeck M, Sousa JC, Carneiro F, Quint WGV: Typing of *Helicobacter pylori vacA* gene and detection of *cagA* gene by PCR and reverse hybridization. J Clin Microbiol 36: 1271-1276, 1998.

Van Klinken BJW, Oussoren E, Weenink JJ, Strous G, Buller HA, Dekker J, Einerhand AWC: The human intestinal cell lines Caco-2 and LS174T as models to study cell-type specific mucin expression. Glycoconj J 13: 757-768, 1996.

Van Klinken BJW, Van Dijken TC, Oussoren E. Buller HA, Dekker J, Einerhand AWC: Molecular cloning of human MUC3 cDNA reveals a novel 59 amino acid tandem repeat region. Biochem Biophys Res Commun 238: 143-148, 1997.

Vandenhaute B, Buisine MP, Debailleul V, Clement B, Moniaux N, Dieu MC, Degand P, Porchet N, Aubert JP: Mucin gene expression in biliary epithelial cells. J Hepatol 27: 1057-1066, 1997.

Velcich A, Palumbo L, Selleri L, Evans G, Augenlich L: Organization and regulatory aspects of the human intestinal mucin gene (*MUC2*) locus. J Biol Chem 272: 7968-7976, 1997.

Verma M, Davidson EA: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA containing a cystein-rich domain. Proc Natl Acad Sci USA 90: 7144-7148, 1993.

Vinall LE, Hill AS, Pigny P, Pratt WS, Toribara N, Gum JR, Kim YS, Porchet N, Aubert J-P, Swallow DM: Variable number tandem repeat polymorphism of the mucin genes located in the complex 11p15.5. Hum Genet 102: 357-366, 1998.

Vollmers HP, Stulle K, Dammrich J, Pfaff M, Papadopoulos T, Betz C, Saal K, Muller-Hermelink HK: Characterization of four new cell lines. Virchows Arch 63: 335-343, 1993.

Vos HL, Devries Y, Hilkens J: The mouse episialin (Muc1) gene and its promoter-rapid evolution of the repetitive domain in the protein. Biochem Biophys Res Commun 181: 121-130, 1991.

Wada M, Yokota J, Mizoguchi H, Sugimura T, Terada M: Infrequent loss of chromosomal heterozygosity in human cancer. Cancer Res 48: 2988-2992, 1988.

Waltz MR, Pandelidis SM, Pratt W, Barnes D, Swallow DM, Gendler SJ, Cohen EP: A microsatellite within the MUC1 locus at 1q21 is altered in the neoplastic cells of breast cancer patients. Cancer Genet Cytogenet 100: 63-67, 1998.

Ward BG, Mcguckin MA, Ramm LE, Coglan M, sanderson B, Tripcony L, Free KE: The management of ovarian carcinoma is improved by the use of cancer associated serum antigen (CASA) and CA125 assays. Cancer 71: 430-438, 1993.

Wesley A, Mantle M, Man D, Qureshi R, Forstner J: Neutral and acidic species of human intestinal mucin. J Biol Chem 260: 7955-7959, 1985.

Wesseling J, Van der Valk S, Vos H, Sonnenberg A, Hilkens J: Episialin (MUC1) overexpression inhibits integrin-mediated cell adhesion to extracellular matrix components. J Cell Biol 129: 255-265, 1995.

Wesseling J, van der Valk S, Hilkens J: A mechanism for inhibition of E-cadherin-mediated cell-cell adhesion by membrane-associated mucin episialin/MUC1. Mol Biol Cell 7: 565-577, 1996.

Wherter JL, Rivera-MacMurray S, Bruckner H, Tatematsu M, Itzkowitz SH: Mucin-associated sialosyl-Tn antigen in gastric cancer correlates with adverse outcome. Br J Cancer 69: 613-616, 1994.

Willsher PC, Xing P-X, Clarke CP, Ho DWM, McKenzie IFC: Mucin 1 antigens in the serum and bronchial lavage fluid of patients with lung cancer. Cancer 72: 2936-2942, 1993.

Wolff RK, Nakamura Y, White R: Molecular characterization of spontaneously generated new allele at a VNTR locus: no exchange of flanking DNA sequence. Genomics 3: 347-351, 1988.

Wolff RK, Plaetke R, Jeffreys AJ, White R: Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics 5: 382-384, 1989.

Wolff RK, Nakamura Y, Odelberg S, Shiang R, White R: Generation of variability at VNTR loci in human DNA. EXS 58: 20-38, 1991.

Wreschner DH, Hareuveni M, Tsarfaty I, Smorodinsky N, Horey J, Zaretsky J, kotkes P, Weiss M, Lathe R, Dion A, Keydar I: Human epithelial tumor antigen cDNA sequences. Eur J Biochem 189: 463-473, 1990.

Wu K, Fregein N, Carraway KL: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifunctional cell surface glycoprotein complex of ascites rat mamamry adenocarcinoma cells. J Biol chem 269: 11950-11955, 1994.

Xing P-X, Prenzoska J, Quelch K, McKenzie IFC: Second-generation anti-MUC1 peptide monoclonal antibodies. Cancer Res 52: 2310-2317, 1992a.

Xing P-X, Prenzoska J, Layton GT, Devine PL, McKenzie IFC: Second-generation monoclonal antibodies to intestinal MUC2 peptide reactive with colon cancer. J Natl Cancer Inst 84: 699-703, 1992b.

Xu G, Haun L-J, Khatri IA, Wang D, Bennick A, Fahim RE, Forstner GG, Forstner JF: cDNA for the carboxyl-terminal region of a rat intestinal mucin like peptide. J Biol Chem 267: 5401-5407, 1992.

Yamamoto M, Bharti A, Li Y, Kufe D: Interaction of the DF3/MUC1 breast carcinoma-associated antigen and β-catetin in cell adhesion. J Biol Chem 272: 12492-12494, 1997.

Zotter S, Hageman PC, Loonitzer A, Mooi WJ, Hilgers J: Tissue and tumor distribution of human polymorphic epithelial mucin. Cancer Rev 11-12: 55-101, 1988.

SUMÁRIO E CONCLUSÕES

Neste trabalho procurou esclarecer-se a contribuição de factores de risco do hospedeiro para o desenvolvimento do carcinoma gástrico, no complexo contexto das interacções gene/ambiente.

I – Polimorfismo dos genes *MUC1 e MUC6* e risco de carcinoma gástrico

Utilizando a técnica de Southern blotting, realizaram-se estudos de caso-controlo para avaliar o polimorfismo dos genes *MUC1* e *MUC6*, em dadores de sangue e em doentes com carcinoma gástrico.

Verificou-se que a distribuição alélica e genotípica do *MUC1* é significativamente diferente no grupo de dadores de sangue e nos doentes com carcinoma gástrico. Os indivíduos com o genótipo SS de *MUC1* (dois alelos leves, de baixo peso molecular) possuem um risco relativo estimado 4.3x superior para o desenvolvimento de carcinoma gástrico.

O polimorfismo do gene *MUC1* foi também estudado num grupo de doentes dispépticos do Norte de Portugal. A frequência dos genótipos de *MUC1* recodificados observada nos doentes dispépticos não diferiu significativamente da frequência obtida para os dadores de sangue.

Observou-se uma prevalência elevada de atrofia glandular e metaplasia intestinal do tipo incompleta em indivíduos com genótipo SS de *MUC1* relativamente aos outros grupos de genótipos, suportando o papel do *MUC1* no desenvolvimento de lesões pré-neoplásicas da mucosa gástrica.

A análise do polimorfismo do *CA repeat* do intrão 6 do gene *MUC1* foi efectuada em dadores de sangue e em doentes com carcinoma gástrico, utilizando a técnica de PCR. Este estudo permitiu verificar que em ambas as populações os polimorfismos do *VNTR* e do *CA repeat* estão associados. Sugere-se, por isso, a utilização do polimorfismo do *CA repeat* como estratégia de pré-screening para o polimorfismo do *VNTR*.

Diferenças estatisticamente significativas foram observadas nas frequências alélicas do gene *MUC6* entre os dadores de sangue e os doentes com carcinoma gástrico, mas a sua relevância para o desenvolvimento de carcinoma gástrico não foi esclarecida neste trabalho.

II – Polimorfismo do gene *MUC1* e infecção por *Helicobacter pylori*

Comparou-se o polimorfismo do gene *MUC1* em dadores de sangue Portugueses e Dinamarqueses com o propósito de verificar se este polimorfismo poderia contribuir para explicar a diferente incidência de carcinoma gástrico em Portugal e na Dinamarca. Utilizando as enzimas de restrição *Eco*RI e *Alu*I identificam-se respectivamente quinze e

vinte alelos de *MUC1*. A distribuição alélica é significativamente diferente entre as duas populações, independentemente da enzima utilizada, com uma maior frequência de alelos leves de *MUC1* na população da Dinamarca. A elevada frequência de alelos leves de *MUC1* na população Dinamarquesa sugere que o polimorfismo de *MUC1*, apesar de ser um factor de susceptibilidade, não desempenha um papel isolado no complexo processo da carcinogénese gástrica. A diferente incidência de carcinoma gástrico em Portugal e na Dinamarca pode ser explicada por condições ambientais claramente variáveis nos dois países.

Verficou-se, utilizando um método serológico, que os dadores de sangue Portugueses têm uma prevalência de infecção por *H.pylori* (82.1%) superior à observada nos dadores de sangue Dinamarqueses (18.0%). A prevalência de infecção nos doentes dispépticos, do Norte de Portugal, é semelhante à observada nos dadores de sangue.

Demonstrou-se, na população de dadores de sangue Portugueses, uma associação entre a infecção por *H. pylori* e os genótipos de alelos leves de *MUC1*, apresentando estes indivíduos títulos serológicos mais altos do que os indivíduos com genótipos de alelos pesados de *MUC1*. A mesma tendência foi observada nos doentes dispépticos e nos dadores de sangue dinamarqueses.

III – Expressão de mucinas e adesão do *Helicobacter pylori* a linhas celulares de carcinoma gástrico

Analisou-se, em sete linhas celulares de carcinoma gástrico, a expressão de mucinas (por imunocitoquímica, Northern blotting e RT-PCR) e carbohidratos (por imunocitoquímica), no sentido de verificar se as linhas celulares poderiam ser usadas como modelos de adesão *in vitro*, com o objectivo de esclarecer a possível interacção entre as mucinas e o *H. pylori*.

Demonstrou-se que o padrão de expressão de mucinas e carbohidratos nas linhas celulares de carcinoma gástrico: a) reflecte, em parte, o padrão de expressão de mucinas da mucosa gástrica e dos carcinomas gástricos, com elevada expressão de MUC1 e MUC5AC; b) está de acordo com o padrão aberrante de expressão de mucinas apresentado por alguns carcinomas gástricos, com expressão ocasional de MUC2, MUC3, MUC4, MUC5B; c) não inclui, na série analisada, a expressão da mucina gástrica MUC6; d) reflecte, em parte, o padrão de diferenciação dos carcinomas gástricos, mantendo-se a expressão de S-Tn nas linhas celulares derivadas de carcinomas gástricos do tipo glandular. Demonstrou-se ainda que não existe relação aparente entre a expressão de mucinas e de carbohidratos, sugerindo que, ao contrário do que está descrito para a mucosa gástrica normal, a glicosilação nas linhas celulares não é dependente do tipo(s) de core(s) proteíco(s) de mucinas expressos.

A adesão *in vitro* do *H. pylori* a linhas celulares de carcinoma gástrico foi superior à observada para uma linha de leucemia mielóide. Demonstrou-se que adesão do *H. pylori*, neste sistema *in vitro*, não era influenciada pelo padrão de expressão de mucinas e carbohidratos, conclusão dificilmente extrapolável para a realidade *in vivo*, uma vez que as condições *in vitro* e *in vivo* não são comparáveis.

PAPER I

Glycoconjugate Journal (1997) **14:** 107–111

MUC1 gene polymorphism and gastric cancer – an epidemiological study

F. Carvalho[1], R. Seruca[1], L. David[1], A. Amorim[1,2], M. Seixas[1], E. Bennett[3], H. Clausen[3] and M. Sobrinho-Simões[1]*

[1]*Institute of Molecular Pathology and Immunology of the University of Oporto, IPATIMUP, Dept. of Pathology, Hospital de S. João, 4200 Porto, Portugal*
[2]*Institute of Anthropology, Faculty of Sciences of the University of Oporto, 4000 Porto, Portugal*
[3]*Department of Oral Diagnostics, School of Dentistry, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark*

Gastric carcinoma is a major cause of cancer death worldwide and, like most human cancers, probably develops after environmental insults acting on normal individuals and/or individuals with increased genetic susceptibility. Mucins are attractive molecules to study the relationship between genetics and environment because they play an important role in the protection of gastric mucosa against environmental insults and exhibit a highly polymorphic genetic variation. We performed a case-control study using Southern blot analysis to evaluate the *MUC1* gene polymorphism in a series of blood donors (n=324) and in patients with gastric carcinoma (n=159). We found that the distribution of *MUC1* alleles is significantly different in the two populations and that small *MUC1* alleles and small *MUC1* genotypes are significantly more frequent in patients with gastric carcinoma than in controls. Individuals with small *MUC1* genotypes are at increased risk for gastric carcinoma development.

***Keywords*: *MUC1* gene polymorphism, gastric cancer**

Introduction

Gastric carcinoma is a major cause of cancer death worldwide [1] and, like most human cancers, probably develops after environmental insults acting on normal individuals and/or individuals with increased genetic susceptibility [2, 3]. The relative contribution of environmental exposure and genetics for the risk of developing gastric carcinoma is far from being established: Diet and infections, with particular emphasis on *Helicobacter pylori* infection, have been identified as exposure risk factors [4, 5]. In contrast to most Western countries, the mortality rates from gastric carcinoma have not declined in Portugal [6] raising the possibility that the Portuguese population may have some particular genetic susceptibility for gastric carcinoma development.

Mucins are attractive molecules to study the relationship between genetics and environment because they play an important role in the protection of gastric mucosa and exhibit a highly polymorphic genetic variation in their length. In fact, mucins are the major structural components of the mucus viscous gel covering the gastric mucosa, and represent the first line defence barrier against environmental aggressions. All mucins have in common the presence of extended arrays of tandemly repeated peptides rich in serine and threonine residues that are potential O-glycosylation sites. The variable number of tandem repeats accounts for the extensive polymorphism observed at DNA, RNA and protein levels [7–14]. *MUC1* [10, 15, 16] and *MUC2* [17] are the only two fully sequenced human mucin genes. Several other human mucin genes have been cloned and partially sequenced [12–14, 18–21]. The protein product of the *MUC1* gene (mammary/pancreatic mucin) has a molecular weight that correlates to the size of the mRNA and to the size of the DNA restriction fragments, suggesting an important gene effect on the final structure of the mucin [22]. All known mucins are secreted products except *MUC1* which has a transmembrane anchorage domain leading to a membrane-bound mucin.

MUC1 mucin is highly expressed on the gastric mucosa [23] and therefore we advanced the hypothesis that *MUC1* polymorphism, resulting in individual size differences with respect to VNTR (variable number of tandem repeats), might influence the function of *MUC1* mucins, i.e. could provide different protective effects against various environmental aggressions.

*To whom correspondence should be addressed. Tel: + 351-2-590591; Fax: + 351-2-5503940.

0282–0080 © 1997 Chapman & Hall

In order to verify our hypothesis we performed a case-control study using Southern blot analysis to evaluate *MUC1* gene polymorphism in patients with gastric carcinoma and in a control group of blood donors.

Materials and methods

We have evaluated, in a case-control study, the *MUC1* gene polymorphism in a Caucasian control group of blood donors ($n = 324$) and in a series of patients with gastric carcinoma ($n = 159$). DNA from the control population was isolated from blood samples (10 ml), collected after obtaining informed consent. Blood samples were pelleted and stored at −70 °C until DNA extraction. DNA from patients with gastric carcinoma was isolated from samples of non-neoplastic gastric mucosa collected immediately after surgery. The collected tissue samples were frozen in liquid nitrogen and stored at −70 °C until DNA extraction.

High-molecular weight DNA was isolated using a salt-chloroform extraction method, as previously described by Mullenbach *et al.* [24]. DNA samples from blood donors and gastric cancer patients were digested with *Eco*RI that recognizes restriction sites in the flanking regions of the tandem repeats of the *MUC1* gene, separated on 0.7% agarose gel by electrophoresis, for 17–19 h at approximately 53 V, and transferred to nylon membranes (Hybond-N, Amersham) by alkaline blotting [25].

The *MUC1* probe was constructed using primers to the flanking regions of the tandem repeats (Primer sense – 5'TGGGCTGGGGGGGCGGTGG3': Primer anti-sense – 5'CGGTGTCACCTCGGCCCCGGACA3'). The PCR product of 94 bp (base pairs) was subcloned into *Eco* RI site of pT7T3 plasmid vector (Pharmacia). The total plasmid was PCR-labelled with [α-^{32}P] dCTP (Amersham) and used as a probe for the *MUC1* tandem repeat region. Following an initial denaturation of 94 °C for 3 min, the amplification was performed for 30 cycles with denaturation at 94 °C for 1 min, annealing at 50 °C for 2 min and extension at 72 °C for 3 min. An additional extension period of 10 min at 72 °C was performed in the end of the cycles. The labelled probe was purified from non-incorporated nucleotide by passage through Sephadex-G50 columns.

Prehybridization was performed for 3–4 h in a phosphate solution (0.5 M $NaHPO_4$, pH = 7.2, 1 mM EDTA, 7% SDS) at 65 °C. Hybridization was performed overnight at 65 °C. Washing of the blots was performed under high stringency conditions (first with 40 mM $NaHPO_4$, 1 mM EDTA, 5% SDS, and secondly with 40 mM $NaHPO_4$, 1mM EDTA, 1% SDS, at 65 °C). Membranes were exposed for autoradiography at −70 °C with an intensifying screen. Autoradiography was developed after 7 days.

All bands from the autoradiograms were scored visually. The size of the different alleles of the *MUC1* gene, both in the blood donors and in the gastric carcinoma patients, was determined by comparison with the size of the fragments of the marker λ*Hind*III (Amersham). All the alleles identified in controls and patients were ranked and numbered from 1 to 15 according to their molecular weight. Allele 1 represents the allele with higher molecular weight and allele 15 represents the allele with lower molecular weight.

The comparison between cases and controls was performed using χ^2 statistics and Student's t test. The Monte Carlo test applied to the study of highly polymorphic genes, as described by Sham and Curtis [26], was performed using the computer program CLUMP (ftp.diamond.gene.ulc.ac.uk [directory/pub/packages/dcurtis]). Odds ratio and 95% confidence limits were determined using the BMDP 4F computer program (Statistical Package program BMDP; Los Angeles, CA, USA).

Results

Comparison between controls and cases regarding age, gender and ABO histo-blood groups

The age of the control population ($n = 324$) was 39.7 ± 11.0 and the age of the gastric carcinoma patients ($n = 159$) was 61.0 ± 12.3 ($p < 0.0001$). The male:female sex-ratio was 3.6:1 in the control population and 1.6:1 in the patients with gastric carcinoma ($p < 0.0001$). The blood group was known in all the donors and in 113 patients with gastric carcinoma. Both populations were in Hardy-Weinberg equilibrium regarding the ABO histo-blood groups. The differences between the two populations were not significant ($p > 0.05$), regarding the ABO system.

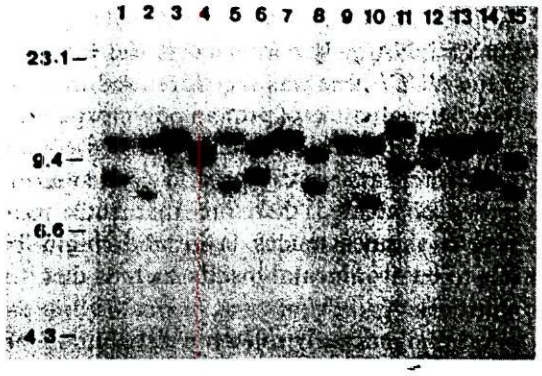
Both populations are Caucasian and originate from the same districts in the Northwest part of Portugal.

We considered the samples appropriate for a case-control study of the *MUC1* gene polymorphism based on the presence of Hardy-Weinberg equilibrium in the ABO system. The differences in age and gender of controls and patients with gastric carcinoma will be further analysed when comparing the two populations for the *MUC1* gene polymorphism.

Comparison between controls and cases regarding *MUC1* gene polymorphism

Fifteen alleles were identified in the 324 blood donors. The allele size ranged from 8.9 kb (kilobase pairs) to 12.6 kb – alleles 15 and 1 in the ranking system (Fig. 1). Thirteen alleles were identified in the 159 patients with gastric carcinoma, all of which were also present in the control population. Alleles 1 and 14 from the control population were not identified in patients with gastric carcinoma. Allelic frequencies are represented in Table 1 and Fig. 2.

Alleles 4 and 10 were the most common alleles both in the control group and in patients with gastric carcinoma (Fig. 2). Most of the alleles were represented in both populations and the modal alleles were the same in both populations. Despite this, a difference in the distribution of allelic frequencies between the two groups is clearly observed



re 1. Detection of *MUC1* gene polymorphism. Human DNA sam- rom blood donors were digested with *Eco* RI, separated on 0.7% se gel, transferred to nylon membranes by alkaline blotting, and dized with the *MUC1* probe. Lanes numbered 1–15 correspond to llowing genotypes: 4–8; 4–11; 4–4; 5–6; 3–10; 5–8; 3–3; 6–10; 4–12; 2–7; 4–6; 4–4; 3–9; 6–10. Size marker is given in kilobases.

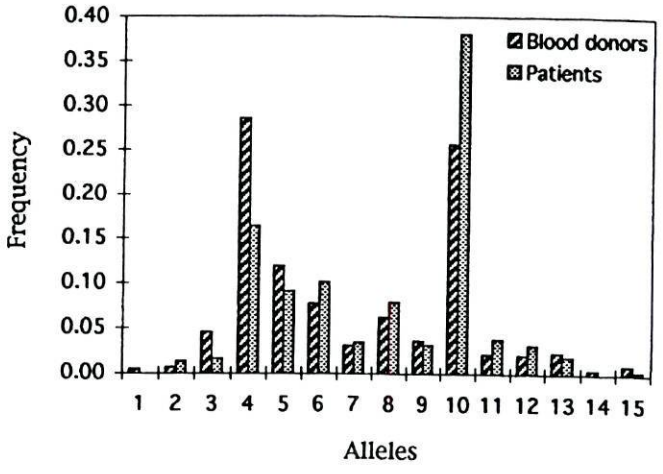


Figure 2. Allelic frequencies distribution of *MUC1* gene in blood donors and in patients with gastric carcinoma. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the heaviest).

e 1. Distribution of *MUC1* allelic frequencies in blood rs and in patients with gastric carcinoma.

1 alleles	*Blood donors* (n = *324*)	*Patients with gastric carcinoma* (n = *159*)
	0.005	0.000
	0.006	0.013
	0.045	0.016
	0.285	0.164
	0.119	0.091
	0.077	0.101
	0.031	0.035
	0.062	0.079
	0.035	0.031
	0.256	0.381
	0.022	0.038
	0.020	0.031
	0.023	0.019
	0.005	0.000
	0.009	0.003

2). In fact the majority of the alleles with a relatively molecular weight (alleles 1–5, except allele 2) were e frequent in the control group and alleles with a lower ecular weight (alleles 6–14, except alleles 9 and 13) were e frequent in patients with gastric carcinoma ($\chi^2 = 41.6$; 0.001). Due to the high number of alleles and the sparse esentation of some of them we performed a Monte o based test that confirmed a significantly ($p < 0.001$) rent distribution of the *MUC1* alleles between patients controls.

e further evaluated the *MUC1* genotypes distribution in wo populations. Observed heterozygosity was 72.2% in rols and 73.0% in patients with gastric carcinoma. The heterozygote 4–10 was the most frequent genotype in both populations: 44 of 324 controls (13.6%) and 28 of 159 patients with gastric carcinoma (17.6%). The homozygous genotype for allele 4 was more frequent in controls than in patients with gastric carcinoma – 36 individuals (11.1%) and three patients (1.9%), respectively – and the homozygous genotype for allele 10 was less frequent in controls than in patients with gastric carcinoma – 28 individuals (8.6%) and 24 patients (15.1%), respectively.

Sixty-six different genotypes were identified overall, most of them with a low number of individuals in both populations, which prevented the use of all genotypes for some statistical analyses. To group the genotypes we divided the alleles using the cut-point that corresponds to the median value of the blood donors population. The selected cut-point is also the one that maximizes the differences between the two populations: alleles 1–5 were grouped into a category of Large *MUC1* alleles (*L*) and alleles 6–15 were grouped into a category of Small *MUC1* alleles (*S*). Genotypes were thereafter recoded as having two large *MUC1* alleles (LL), two small *MUC1* alleles (SS) and one large *MUC1* allele and one small *MUC1* allele (LS).

In the control group of blood donors no differences were observed regarding the mean age ($p > 0.1$) and the male:female sex-ratio ($p < 0.1$) between the three different recoded genotypes: LL, SS, LS (data not shown). We assumed, based on these observations, that no major gender or age dependent selection of the different genotypes has occurred in the control population.

The comparison between patients and controls regarding the recoded *MUC1* genotypes is shown in Table 2. Significant differences are observed in the genotypic frequencies between patients and controls. Genotypes for large *MUC1* alleles (LL) are more frequent in controls (50.9%) than in patients with gastric carcinoma (29.3%) and genotypes for

Table 2. Distribution of recoded *MUC1* genotypes, LL (two large *MUC1* alleles), LS (one large *MUC1* allele and one small *MUC1* allele) and SS (two small *MUC1* alleles), in blood donors and in patients with gastric carcinoma.

Genotypes	*Blood donors* (n = *324*) n (%)	*Patients with gastric carcinoma* (n = *159*) n (%)
LL	67 (50.9%)	12 (29.3%)*
LS	162 (41.5%)	66 (50.0%)
SS	95 (7.5%)	81 (20.7%)*

The estimated relative risk (odds ratio) adjusted for age and gender, for gastric carcinoma development between individuals with SS genotype and individuals with LL genotype is 4.3 (95% confidence limits 1.8–10.5). *$p < 0.0001$.

small *MUC1* alleles (SS) are more frequent in patients with gastric carcinoma (20.7%) than in controls (7.5%) ($p < 0.0001$). Genotypes with one large *MUC1* allele and one small *MUC1* allele (LS) have a similar frequency in both groups (Table 2). The estimated relative risk (odds ratio), adjusted for age and gender, for gastric carcinoma development in individuals with SS genotype is 4.3 (95% confidence limits 1.8–10.5).

Figure 3 depicts the results obtained after grouping the genotypes LS together with the genotypes LL (Fig. 3A) or with the genotypes SS (Fig. 3B). In both scenaria the differences between controls and patients with gastric carcinoma are highly significant ($p < 0.0001$). The results obtained suggest that the presence of a small *MUC1* genotype is a risk factor for gastric carcinoma development, either in a LL or in a LS combination.

Discussion

The *rationale* of the present work was based on previous studies showing that the *MUC1* gene polymorphism is reflected both at the RNA and protein levels [7–14]. The evaluation of DNA polymorphism at the tandem repeat region of the *MUC1* gene was therefore used as a surrogate endpoint to the evaluation of protein polymorphism. In fact, individuals with *MUC1* alleles encompassing a lower number of tandem repeats are expected to code smaller *MUC1* proteins. We postulated therefore that such individuals might have a thinner mucus layer and might be more susceptible to environmental insults such as diet, toxins or microorganisms.

We found a signficantly different distribution of *MUC1* alleles between the control group of blood donors and the group of patients with gastric carcinoma. We have also found that the prevalence of small sized alleles is higher in patients with gastric carcinoma than in controls. The higher prevalence of allele 10 and the lower prevalence of allele 4 in patients with gastric carcinoma are the most striking differences between patients and controls. A similar tendency for a higher prevalence of smaller alleles in carcinomas was observed in respect to alleles 6, 7, 8, 11 and 12.

These results are in accordance with our hypothesis that individuals with small *MUC1* alleles are more susceptible to the development of gastric carcinoma.

The second approach we have undertaken was aimed at determining if the allele combination per individual, i.e. the genotype, was in keeping with this interpretation. In fact the expression of *MUC1* alleles is codominant [8] and the net effect of *MUC1* expression will be dependent on the polymorphism of both alleles.

The results observed with homozygous genotypes for the two most common alleles (alleles 4 and 10) followed the same tendency of allele distribution: patients with gastric carcinoma showed a significantly higher prevalence of homozygous genotype for the small sized allele (allele 10).

For the analysis of all genotypes we classified the cases as described in the Results section: LL, two large *MUC1* alleles; SS, two small *MUC1* alleles; and LS, one large *MUC1* allele and one small *MUC1* allele. We found that genotypes

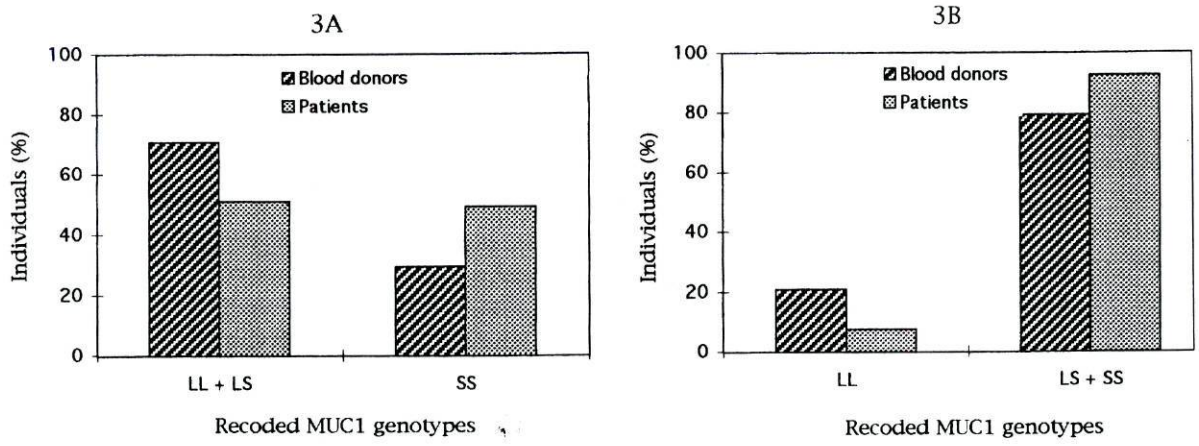


Figure 3. Distribution of recoded *MUC1* genotypes in blood donors and patients; (3A) Genotypes LS grouped with genotypes LL; (3B) Genotypes LS grouped with genotypes SS.

ying small *MUC1* alleles, either combined with another l *MUC1* allele (SS) or with a large *MUC1* allele (SL) more prevalent in patients with gastric carcinoma in controls. The results suggest that the presence of ngle small *MUC1* allele is sufficient to confer genetic eptibility for the development of gastric carcinoma. The viduals with two small *MUC1* alleles (SS) carry an esti-d relative risk for gastric carcinoma development of 4.3. few issues have to be addressed before we can recom-d the study of mucin genes polymorphism for the identi-ion of individuals at risk for the development of gastric noma: (1) The young age of our control population have introduced a bias in the statistical analysis, in sense that we do not know how many individuals his population will develop a gastric carcinoma in future; (2) The DNA restriction fragments obtained *Eco*RI digestion are of a quite large size contributing to poor resolution of allele sizes. This may be responsible n underestimation of the total number of alleles and for high number of individuals classified as homozygous in populations; (3) Finally, we need to get more informa-regarding the basis of our study. Namely, we need to irm the assumption that the genetic polymorphism is ctly reflected at the protein level and the genetic/protein morphism is causally involved in *in vivo* and *in vitro* els of gastric carcinogenesis.

these shortcomings are resolved we will be able to n that the newly proposed risk factor might be of or importance in the future. All the other reported risk ors for gastric carcinoma rely upon the identification of ursor conditions evaluated on histological slides, as is case of chronic atrophic gastritis, intestinal metaplasia dysplasia [27]. The detection of precursor conditions in ptomatic individuals forms the present basis for enter-in follow-up protocols intended to achieve early diag-s of gastric carcinoma. This strategy implicates invasive edures with frequent endoscopies and biopsies. We pt that the molecular epidemiology of mucin genes morphism will allow the establishment of non-invasive ening strategies for identification of individuals at risk for ric carcinoma development in the general population.

knowledgements

s work was supported by Junta Nacional de Investigação ntífica e Tecnológica (JNICT), PECS/SAU/255/95 and AXIS XX1 2/2.1/SAU/1356/95.

ferences

Munoz N (1988) In *Gastric carcinogenesis* (PI Reed and MJ Hill, eds) pp. 51–69. Amsterdam: Elsevier.
Knudson AG (1985) *Cancer Res* **45**: 1437–43.
3 Venitt S (1994) *Clin Chem* **40**: 1421–25.
4 La Vecchia C, Negri E, Decarli A, D'Avanzo B, Franceschi S (1987) *Int J Cancer* **40**: 484–89.
5 Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, Sibley RK (1991) *N Engl J Med* **325**: 1127–31.
6 La Vecchia C, Lucchini F, Negri E, Boyle P, Maisonneuve P, Levi F (1992) *Eur J Cancer* **28**: 132–235.
7 Karlson S, Swallow DM, Griffiths B, Corney G, Hopkinson DA (1983) *Ann Hum Genet* **47**: 263–69.
8 Swallow DM, Gendler S, Griffiths B, Corney G, Taylor-Papadimitrou J, Bramwell ME (1987) *Nature* **328**: 82–84.
9 Gendler S, Taylor-Papadimitrou J, Duhig T, Rothbard J, Burchell J (1988) *J Biol Chem* **263**: 12820–23.
10 Gendler SJ, Lancaster CA, Taylor-Papadimitrou J, Duhig T, Peat N, Burchell J (1990) *J Biol Chem* **265**: 15286–93.
11 Toribara NW, Gum JR, Culhane PJ, Lagace RE, Hicks JW, Peterson GM, Kim YS (1991) *J Clin Invest* **88**: 1005–13.
12 Porchet N, Van Cong N, Dufosse J, Audie JP, Guyonnet-Duperat V, Gross MS, Denis C, Degand P, Bernheim A, Aubert JP (1991) *Biochem Biophys Res Commun* **175**: 414–22.
13 Toribara NW, Roberton AM, Ho SB, Kuo W-L, Gum E, Hicks JW, Gum JR, Byrd JC, Siddiki B, Kim YS (1993) *J Biol Chem* **268**: 5879–85.
14 Bobek LA, Tsai H, Biesbrock AR, Levine MJ (1993) *J Biol Chem* **268**: 20563–69.
15 Lan MS, Batra SK, Qi W-N, Metzgar RS, Hollinsworth MA (1990) *J Biol Chem* **265**: 15294–99.
16 Lancaster CA, Peat N, Duhig T, Wilson D, Taylor-Papadimitrou J, Gendler SL (1990) *Biochem Biophys Res Commun.* **173**: 1019–29.
17 Gum JR, Hicks JW, Toribara NW, Siddiki B, Kim YS (1994) *J Biol Chem* **269**: 2440–46.
18 Gum JR, Hicks JW, Swallow DM, Lagace RL, Byrd JC, Lamport DTA, Siddiki B, Kim YS (1990) *Biochem Biophys Res Commun* **171**: 407–15.
19 Van Cong N, Aubert JP, Gross MS, Porchet N, Degand P, Frézal J (1990) *Hum Genet* **86**: 167–72.
20 Aubert JP, Porchet N, Crepin M, Duterque-Coquillaud M, Vergnes G, Mazzuca M, Debuire B, Petitprez D, Degand P (1991) *Am J Resp Cell Mol Biol* **5**: 178–85.
21 Meerzaman D, Charles P, Daskal E, Polymeropolus MH, Martin BM, Rose MC (1994) *J Biol Chem* **269**: 12932–39.
22 Ligtenberg MJL, Vos HL, Gennissen AMC, Hilkens J (1990) *J Biol Chem* **265**: 5573–78.
23 Ho SB, Niehans GA, Lyftogt C, Yan PS, Cherwitz DL, Gum ET, Dahiya R, Kim YS (1993) *Cancer Res* **53**: 641–51.
24 Mullenbach R, Lagoda PJL, Welter C (1989) *Trends Genet* **5**: 391.
25 Southern EM (1975) *J Mol Biol* **98**: 503–17.
26 Sham PC, Curtis D (1995) *Ann Hum Genet* **59**: 97–105.
27 Correa P (1992) *Cancer Res* **52**: 6735–40.

Received 8 September 1995, revised 28 December 1995

PAPER II

MUC6 Gene Polymorphism in Healthy Individuals and in Gastric Cancer Patients from Northern Portugal[1]

Elsa Garcia, Filipa Carvalho, António Amorim, and Leonor David[2]

Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP, 4200 Porto, Portugal

Abstract

Mucins exhibit a high degree of genetic polymorphism because of the presence of a variable number of tandem repeats. The aims of this work were to describe the *MUC6* gene polymorphism in the Portuguese population and to evaluate whether *MUC6* gene polymorphism was involved in individual susceptibility to gastric cancer development, as observed previously for the *MUC1* gene. We found that the 10 alleles identified in the population of blood donors ($n = 376$), by Southern blot analysis, were also found in gastric cancer patients ($n = 157$). However, significant differences in allelic frequencies between the two populations were observed for 4 of the 10 alleles, in agreement with those described previously for the *MUC1* gene; the largest allele was more frequent in blood donors, and smaller alleles were more frequent in gastric cancer patients. Our results suggest that *MUC6* gene polymorphism is involved in the predisposition to gastric carcinoma development.

Introduction

Mucins are high molecular weight glycoproteins with core proteins (apomucins) rich in serine and threonine residues to which *O*-linked carbohydrate chains are attached. A constant feature of the eight mucins identified thus far is the presence of a VNTR[3] in the coding regions of the genes that are responsible for the highly polymorphic structure of the genes (1–9).

Previous studies of our group showed that *MUC1* gene polymorphism was involved in the individual susceptibility to gastric cancer development (10). It was demonstrated that individuals with *MUC1* genotypes displaying a small number of tandem repeats, and therefore with small glycoprotein products, were at increased risk for gastric cancer development. It was postulated that such individuals might have a thinner mucus layer, probably leading to a less efficient protection against environmental insults (10).

Another mucin gene, *MUC6* gene, is highly expressed in the gastric mucosa (1, 11–13). *MUC6* was identified by expression cloning using antibodies raised against deglycosylated mucin isolated from human gastric tissue (1, 11). The individual repeat unit of the *MUC6* gene is composed by 507 bp that code for a peptide stretch with 31 and 18% of threonine and serine residues, respectively (1). The aims of this work were: to describe the *MUC6* gene polymorphism in the Portuguese population, characterized by a high incidence of gastric cancer; and to evaluate whether *MUC6* gene polymorphism was involved in individual susceptibility to gastric cancer development, as demonstrated previously for *MUC1* gene.

Materials and Methods

The polymorphism of the *MUC6* gene was studied in a control group of 376 blood donors (recruited from the blood bank during 1 month in 1993 and 1 month in 1995) and in 157 patients with gastric cancer (consecutive cases from which frozen material was available, collected from 1988 to 1995). Both groups are from a population of Northern Portugal. Gastric cancer cases were representative from the general distribution of gastric carcinoma in Northern Portugal regarding age and sex as well as localization (antrum, 86; body, 41; cardia, 30), morphology (intestinal carcinomas, 82; diffuse carcinomas, 38; atypical carcinomas, 37), and staging. DNA from blood donors and cancer patients was isolated by a method described previously (14) from pelleted blood samples (10 ml) and nonneoplastic gastric tissue, respectively. Pelleted blood samples and collected tissue samples (frozen in liquid nitrogen immediately after surgical removal) were stored at −70°C until DNA extraction.

Isolated high molecular weight DNA was digested with *Taq*I, which cuts the DNA outside the tandem repeat region of the *MUC6* gene (1), separated on 0.7% agarose gel by electrophoresis for 17–19 h at approximately 53 V, and transferred to nylon membranes (Hybond-N; Amersham) by alkaline blotting (15).

The blots were prehybridized for 3–4 h in a phosphate solution [0.5 M $NaHPO_4$ (pH 7.2), 1 mM EDTA, and 7% SDS] at 65°C, followed by hybridization at 65°C overnight with a *MUC6* probe specific for the VNTR region of the gene. Blots were washed under high stringency conditions (first with 40 mM $NaHPO_4$, 1 mM EDTA, 5% SDS, and next with 40 mM $NaHPO_4$, 1 mM EDTA, and 1% SDS at 65°C). Membranes were exposed for autoradiography at −70°C. Autoradiography was developed after 1 day.

All bands from the autoradiograms were scored visually. The size of the different alleles of the *MUC6* gene was determined by two independent observers by comparison with the size of the fragments of the marker λ *Hin*dIII (Amersham). All

Received 6/25/97; revised 8/18/97; accepted 8/19/97.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by Junta Nacional de Investigação Científica e Tecnológica - JNICT (Project PECS/C/SAU/255/95).

[2] To whom requests for reprints should be addressed, at Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP, Rua Dr. Roberto Frias s/n, 4200 Porto, Portugal. Phone: 351-2-5570700; Fax: 351-2-5570799; E-mail: leonor.david@ipatimup.up.pt.

[3] The abbreviation used is: VNTR, variable number of tandem repeats.

of the cases were analyzed twice, and reference alleles were included in every blot.

MUC6 probe was constructed using primers to the tandem repeat region of the gene (primer sense, 5′-GCAGGCTAACCACACCCTTCA-3′; primer antisense, 5′-ATGTTGCAGTCATAGGACCCTG-3′). Thirty cycles of amplification were undertaken at 95°C for 45 s (denaturing) and at 61°C for 1 s (annealing) as well as at 72°C for 1 min (extension). The purified PCR product of 603 bp (Prep-A-Gene DNA purification kit; Bio-Rad) was digested with *Eco*RI and subcloned in the vector pT7T3 (Pharmacia) digested previously with *Eco*RI. The ligation product was then used to transform Epicurian Coli Sure Competent Cells (Stratagene). Selection of recombinant clones was made on plates supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (galactose) and isopropylthio-β-D-galactosidase. The total plasmid was extracted using the Qiagen Plasmid Mini kit (Qiagen, Inc., Chatsworth, CA), labeled with [α-^{32}P] dCTP (Amersham) using the oligo-labeling extension method (16) and purified from non-incorporated nucleotides by passage through Sephadex-G50 columns.

Allelic frequencies were estimated by gene counting, and population conformity with Hardy-Weinberg expectations was tested according to Guo and Thompson (17), using the statistical software package GENEPOP (18).

The statistical analysis of the results was made by Student's t and χ^2 tests. The results were considered significantly different when $P < 0.05$. Comparison of gene frequencies between patients and controls was made separately for each allele using 2×2 contingency χ^2 analysis with 1 df.

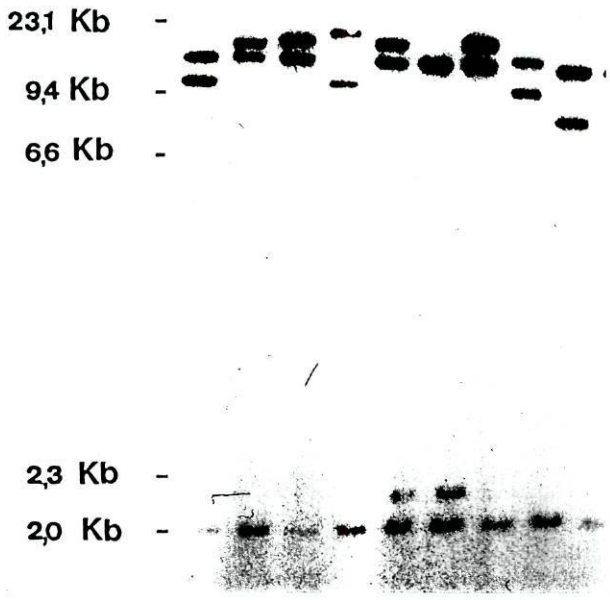


Fig. 1. Detection of *MUC6* gene polymorphism. DNA samples from blood donors were digested with *Taq*I, separated on 0.7% agarose gel, and transferred to nylon membranes by alkaline blotting. Membranes were hybridized with a MUC6 probe. The constant band at 2.0 Kb possibly represents the *Taq*I fragment incorporating the flanking DNA and the terminal end of the repeat. The polymorphism observed in two lanes, with an additional band at 2.1 Kb, represents sequence polymorphism in one of the two alleles (1). The size marker is given in kilobases.

Results

***MUC6* Gene Polymorphism in the Portuguese Population (Blood Donors).** We have studied a Portuguese population of 376 blood donors 39.7 ± 11.2 years of age with a male:female sex ratio of 3.8:1. Ten different alleles with sizes ranging from 8.5 to 12.5 Kb were identified (Fig. 1). The alleles were numbered according to their molecular weight (allele 1 is the heaviest, and allele 10 the lightest), and the most frequent were: allele 4, 11.2 Kb (37.5%); allele 7, 10.0 Kb (20.9%); and allele 2, 12.0 Kb (17.2%).

Genotype distribution was in accordance with Hardy-Weinberg equilibrium expectations, and the observed heterozygosity was 75.8%. No differences were found in allelic pattern distributions regarding sex, age, and histo-blood group (data not shown).

Comparison between Controls and Gastric Cancer Patients. Blood donors and gastric cancer patients were significantly different regarding sex ($P = 0.0001$) and age ($P < 0.0001$) distributions. The control population was younger than the group of gastric cancer patients (61.3 ± 12.3); the male:female sex:ratio was lower in the group of gastric cancer patients (1.6:1). In both populations, *MUC6* allele distribution was independent of age and sex (Table 1). The ABO histo-blood group was found in all blood donors and in 111 gastric cancer patients, and the distributions were not significantly different. We considered the samples appropriate for a case-control study of the *MUC6* gene polymorphism based also on the same origin of both populations.

In gastric cancer patients, the same 10 different alleles of the control group were identified, and the modal alleles were also the same in both populations: allele 4, allele 7, and allele 2 (Table 2 and Fig. 2). Significant differences in allelic frequencies between blood donors and gastric cancer patients were observed for alleles 1, 5, 8, and 10 (Table 2). Allele 1, the largest allele, was more frequent in blood donors; alleles 5, 8, and 10, of smaller sizes, were more frequent in gastric cancer patients (Table 2 and Fig. 2). Allele distribution in gastric cancer patients was identical in cases with different localizations ($P = 0.229$) and different morphological types ($P = 0.628$; data not shown).

Genotype distribution in gastric cancer patients was as expected from Hardy-Weinberg equilibrium, and the observed heterozygosity was 79.6%.

Discussion

The first aim of this work was to describe the *MUC6* gene polymorphism in the Portuguese population, characterized by a high incidence of gastric cancer. Ten different alleles with sizes ranging from 8.5 to 12.5 Kb were identified in a group of 376 healthy individuals. The modal alleles had 11.2 Kb (allele 4), 10.0 Kb (allele 7), and 12.0 Kb (allele 2). Observed genotype distribution was in accordance with Hardy-Weinberg equilibrium expectations, and the observed heterozygosity was 75.8%. No associations were found between the *MUC6* gene polymorphism and sex, age, and ABO histo-blood group.

Our results are in agreement with those reported previously by Toribara *et al.* (1) in a population in the United States. They studied the *MUC6* gene polymorphism in DNA samples of nine individuals, digested with *Taq*I, and observed a high degree of polymorphism.

The second aim of our work was to evaluate whether *MUC6* gene polymorphism was involved in individual susceptibility to gastric cancer development. The rationale of this approach was based on a previous study of our group showing that the polymorphism of *MUC1* gene was involved in the

Table 1 Allelic frequencies[a] in blood donors and gastric cancer patients according to age and sex distributions

Alleles	Donors					Carcinomas				
	Age			Sex		Age			Sex	
	<30 $n = 78$	30–60 $n = 281$	>60 $n = 17$	Male $n = 297$	Female $n = 79$	<30 $n = 3$	30–60 $n = 66$	>60 $n = 88$	Male $n = 96$	Female $n = 61$
1	0.038	0.100	0.059	0.089	0.070	0	0.068	0.023	0.047	0.033
2	0.160	0.165	0.324	0.172	0.171	0	0.174	0.199	0.193	0.172
3	0.135	0.080	0.088	0.098	0.070	0	0.083	0.051	0.073	0.049
4	0.442	0.365	0.235	0.372	0.386	0.500	0.348	0.426	0.375	0.426
5	0	0.016	0	0.010	0.019	0	0.045	0.045	0.036	0.057
6	0.032	0.018	0.029	0.024	0.013	0.167	0.008	0.023	0.026	0.008
7	0.167	0.221	0.206	0.209	0.209	0.333	0.205	0.170	0.193	0.180
8	0.006	0.005	0	0.003	0.013	0	0.030	0.011	0.016	0.025
9	0.013	0.027	0.059	0.022	0.038	0	0.023	0.028	0.031	0.016
10	0.006	0.004	0	0.002	0.013	0	0.015	0.023	0.010	0.033

[a] P for all the comparisons between allele frequencies within age and sex categories was >0.05.

Table 2 Allelic distributions in blood donors and gastric cancer patients

Alleles	Blood donors Frequency (n)	Cancer patients Frequency (n)	P
1	0.085 (64)	0.041 (13)	0.01198[a]
2	0.172 (129)	0.185 (58)	0.60623
3	0.092 (69)	0.064 (20)	0.13110
4	0.375 (282)	0.395 (124)	0.54181
5	0.012 (9)	0.045 (14)	0.00083[a]
6	0.021 (16)	0.019 (6)	0.82042
7	0.209 (157)	0.188 (59)	0.43950
8	0.005 (4)	0.019 (6)	0.03326[a]
9	0.025 (19)	0.025 (8)	0.98347
10	0.004 (3)	0.019 (6)	0.01392[a]

[a] $P < 0.05$.

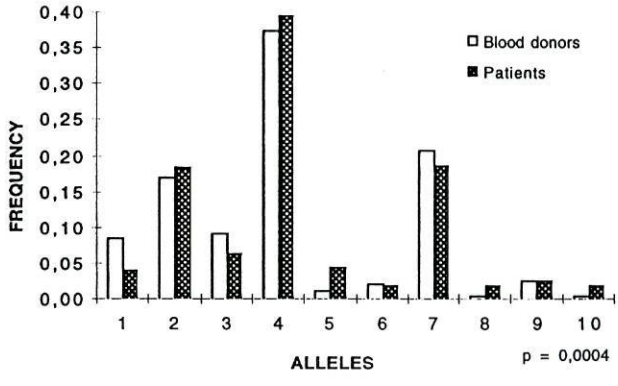


Fig. 2. Distribution of allelic frequencies of *MUC6* gene in blood donors and in gastric cancer patients. Alleles are numbered according to their size by decreasing order of molecular weight.

individual susceptibility to gastric cancer development (10). It was demonstrated that individuals with *MUC1* genotypes displaying a small number of tandem repeats, and therefore with small glycoprotein products, were at increased risk for gastric cancer development. It was postulated that such individuals might have a thinner mucus layer, probably due to a less efficient protection against environmental insults. It was demonstrated previously that the protein product of the *MUC1* gene has a molecular weight that correlates to the size of mRNA and to the size of the polymorphic region, VNTR, of the gene (2, 3, 9, 19). Regarding the *MUC6* gene, it has not been demonstrated that alleles showing a small VNTR region originate comparably small gene products, *i.e.*, there is no evidence that the polymorphism at the DNA level (VNTR) accounts for a polymorphism at both mRNA and protein levels. In fact, *MUC6* like other secretory mucins typically demonstrates a polydisperse mRNA pattern on Northern blot analysis. Polydispersity of mucin mRNA is a characteristic of most mucins, with *MUC1* being a notable exception to this rule. The reasons for this polydispersity are unknown but may be due to either rapid turnover of mucin mRNA or increased susceptibility of the long mucin mRNAs to degradation (20). It is also possible that the repetitive structure of mucin mRNAs renders them physically unstable or more susceptible to RNase degradation (20).

Although our patient and control samples were significantly different regarding sex and age, no differences were observed in allelic frequencies of *MUC6* polymorphism regarding these parameters. Both populations were originally from Northern Portugal and were at Hardy-Weinberg equilibrium in regard to the ABO system. We, therefore, assumed that the samples were appropriate for a case-control study of the *MUC6* gene polymorphism.

Significant differences in allelic frequencies of blood donors and gastric cancer patients were observed for alleles 1, 5, 8, and 10, suggesting that *MUC6* gene polymorphism is involved in the susceptibility to gastric cancer development. The differences of *MUC6* allele frequencies between controls and patients are in agreement with those observed previously for *MUC1* gene (10). Allele 1, the largest allele, was more frequent in blood donors; alleles 5, 8, and 10, which are smaller alleles, were more frequent in gastric cancer patients. Interestingly, the distribution of *MUC6* alleles in patients with gastric carcinomas from different localizations and with different morphological types was identical.

Altogether, our data suggest that *MUC6* gene polymorphism is involved in the predisposition to gastric carcinoma development, regardless of the histotype. For a full understanding of the role of mucins in the protection of gastric mucosa MUC5AC mucin, widely expressed in the gastric mucosa, should also be taken into consideration (21, 22). Finally, one should not disregard that the amount of protein produced and the final glycoprotein composition of secreted mucins, like MUC5AC and MUC6, may be more important than the gene structure *per se*.

References

1. Toribara, N. W., Roberton, A. M., Ho, S. B., Kuo, W-L, Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B., and Kim, Y. S. Human gastric mucin: identification of a unique species by expression cloning. J. Biol. Chem., *268:* 5879–5885, 1993.

2. Swallow, D. M., Gendler, S., Griffiths, B., Corney, G., Taylor-Papadimitrou, J., and Bramwell, M. E. The human tumor-associated epithelial mucins are coded by an expressed hypervariable gene locus *PUM*. Nature (Lond.), *328:* 82–84, 1987.

3. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitrou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E-N., and Wilson, D. Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. J. Biol. Chem., *265:* 15286–15293, 1990.

4. Toribara, N. W., Gum, J. R., Culhane, P. J., Lagace, R. E., Hicks, J. W., Peterson, G. M., and Kim, Y. S. MUC-2 human small intestinal mucin gene structure: repeated arrays and polymorphism. J. Clin. Invest., *88:* 1005–1013, 1991.

5. Bobek, L. A., Tsai, H., Biesbrok, A. R., and Levine, M. J. Molecular cloning, sequence, and specificity of expression of the gene coding the low molecular weight human salivary mucin (MUC7). J. Biol. Chem., *268:* 20563–20569, 1993.

6. Porchet, N., Van Cong, N., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A., and Aubert, J. P. Molecular cloning and chromosomal localization of a novel human tracheo-bronchial mucin cDNA containing tandemly repeated sequences of 48 base pairs. Biophys. Res. Commun., *175:* 414–422, 1991.

7. Kim, Y. S., Gum, J. R., Byrd, J. C., and Toribara, N. W. The structure of human intestinal apomucins. Am. Rev. Respir. Dis., *144:* S10–S14, 1991.

8. Shankar, V., Gilmore, M. S., Elkins, R. C., and Sachdev, G. P. A novel human airway mucin cDNA encodes a protein with unique tandem-repeat organization. Biochem. J., *300:* 295–298, 1994.

9. Karlson, S., Swallow, D. M., Griffiths, B., Corney, G., and Hopkinson, D. A. A genetic polymorphism of a human urinary mucin. Ann. Hum. Genet., *47:* 263–269, 1983.

10. Carvalho, F., Seruca, R., David, L., Amorim, A., Seixas, M., Bennett, E., Clausen, H., and Sobrinho-Simões, M. *MUC1* gene polymorphism and gastric cancer-an epidemiological study. Glycoconj. J., *14:* 107–111, 1997.

11. Ho, S. B., Roberton, A. M., Shekels, L. L., Lyftogt, C. T., Niehans, G. A., and Toribara, N. W. Expression cloning of gastric mucin complementary DNA and localization of mucin gene expression. Gastroenterology, *109:* 735–747, 1995.

12. De Bolós, C., Garrido, M., and Real, F. X. MUC6 apomucin shows a distinct normal tissue distribution that correlates with Lewis antigen expression in the human stomach. Gastroenterology, *109:* 723–734, 1995.

13. Ho, S. B., Shekels, L. L., Toribara, N. W., Kim, Y. S., Lyftog, C., Cherwitz, D. L., and Niehans, G. A. Mucin gene expression in normal, preneoplastic and neoplastic human gastric epithelium. Cancer Res., *55:* 2681–2690, 1995.

14. Müllenbach, R., Lagoda, P. J. L., and Welter, C. An efficient salt-chloroform extraction of DNA from blood and tissues. Trends Genet., *5:* 391, 1989.

15. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., *98:* 503–517, 1975.

16. Feinberg, A. P., and Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Annal. Biochem., *132:* 6–13, 1983.

17. Guo, S. W., and Thompson, E. A. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics, *48:* 361–372, 1992.

18. Raymond, M., and Rousset, F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Hered., *86:* 248–249, 1995.

19. Ligtenberg, M. J. L., Vos, H. L., Gennissen, A. M. C., and Hilkens, J. Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. J. Biol. Chem., *265:* 5573–5578, 1990.

20. Gum, J. R. Mucin genes and the proteins they encode: structure, diversity and regulation. Am. J. Respir. Cell Mol. Biol., *7:* 557–564, 1992.

21. Guyonnet Duperat, V., Audie, J. P., Debailleul, V., Laine, A., Buisine, M. P., Galiegue-Zouitina, S., Pigny, P., Degand, P., Aubert, J. P., and Porchet, N. Characterization of the human mucin gene *MUC5AC*: a consensus cysteine-rich domain for 11p15 mucin genes? Biochem. J., *305:* 211–219, 1995.

22. Reis, C. A., David, L., Nielsen, P. A., Clausen, H., Mirgorodskaya, K., Roepstorff, P., and Sobrinho-Simões, M. Immunohistochemical study of MUC5AC expression in human gastric carcinomas using a novel monoclonal antibody. Int. J. Cancer, *74:* 112–121, 1997.

PAPER III

MUC1 gene polymorphism does not contribute to explain the different incidences of gastric cancer in Portugal and Denmark

Filipa Carvalho[1], António Peixoto[1], Rudi Steffensen[2], António Amorim[1,3], Leonor David[1,4], Manuel Sobrinho-Simões[1,4]

[1]*Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP, Rua Dr. Roberto Frias s/n, 4200 Porto, Portugal, and*
[2]*Department of Blood Transfusion and Clinical Immunology, Aalborg Hospital, PO Box 561, DK-9100, Aalborg, Denmark, and*
[3]*Faculty of Sciences, University of Porto, 4050 Porto, Portugal, and*
[4]*Medical Faculty of University of Porto, 4200 Porto, Portugal*

Running title: *MUC1* gene polymorphism in Portugal and Denmark

Keywords: *MUC1*, mucins, polymorphism, gastric cancer

Correspondence:

Filipa Carvalho
Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP, Rua Dr. Roberto Frias s/n, 4200 Porto, Portugal
Tel: +351-2-5570700
Fax: +351-2-5570799
E-mail: filipa.carvalho@ipatimup.pt

Abstract

Mucins are highly polymorphic glycoproteins that constitute the major structural component of the mucus gel that protects epithelia from environmental aggressions. We demonstrated previously that individuals from northern Portugal with small *MUC1* genotypes are at increased risk for gastric carcinoma development. In this study we compared the *MUC1* gene polymorphism in blood donors from northern Portugal and Denmark, using Southern blotting with two restriction enzymes (*EcoRI* and *AluI*), to see whether or not *MUC1* gene polymorphism may contribute to explain the differences in the incidence of gastric carcinoma between the two populations. We observed a different distribution of *MUC1* alleles in the blood donors from Portugal and Denmark using both *EcoRI* and *AluI* as restriction enzymes, with smaller alleles being more frequent in danish than in portuguese donors. We conclude that the much higher incidence of gastric cancer in Portugal can not be ascribed to *MUC1* gene polymorphism *per se*.

Introduction

Mucins are highly polymorphic glycoproteins (for Review, see [1]) that constitute the major structural component of the mucus gel that protects epithelia from environmental aggressions. The product of *MUC1* gene is the polymorphic epithelial mucin (PEM)[2] which is expressed in the majority of epithelial tissues[3-5] including gastric mucosa[6,7]. In a previous study we demonstrated that the distribution of *MUC1* alleles is significantly different in blood donors and in gastric carcinoma patients from northern Portugal, and that small *MUC1* alleles and small *MUC1* genotypes are significantly more frequent in patients with gastric carcinoma than in controls, thus indicating that individuals with small *MUC1* genotypes are at increased risk for gastric carcinoma development[8].

In order to see if *MUC1* gene polymorphism may contribute to explain the geographic differences in the incidence of gastric carcinoma, we compared *MUC1* gene polymorphism in blood donors from two the countries with the highest (Portugal) and lowest (Denmark) incidences of gastric carcinoma in the European Union[9].

Materials and Methods

Southern blotting was performed as previously described[8] using DNA extracted from ten ml blood samples, digested with *EcoRI*, from 324 portuguese blood donors and 100 danish blood donors after obtaining informed consent. A sub sample of each population (n=96 portuguese blood donors and n= 99 danish blood donors) was digested with *AluI*, a restriction enzyme cutting closer to the tandem repeat of *MUC1* gene, producing smaller fragments. Alleles were numbered according to their molecular weight, allele 1 being the largest in both systems.

Allele frequencies were estimated by the gene counting method. Statistical analysis for Hardy-Weinberg equilibrium was performed using an exact test[10], running the statistical software package GDA[11]. Expected

heterozygosity was calculated according to the method of Nei[12], and the difference between the expected and the observed number of heterozygotes was used to calculate a chi-square with one degree of freedom. Heterogeneity analysis between populations was based on allelic contigency tables, and the corresponding chi-square was calculated using the statistical software package BYOSYS-1[13].

Results

Characterization of the two populations

In the 324 blood donors from northern Portugal, with ages ranging from 19 to 64 years (38.0±0.9y) the male:female ratio was 4.6:1. In the 100 blood donors from Denmark, with ages ranging from 21 to 65 years (42.0±1.0y) the male:female ratio was 0.9:1. There were no differences between the two populations regarding ABO histo-blood groups distribution (data not shown).

MUC1 gene polymorphism in the two populations using EcoRI

Allelic frequencies of *MUC1* gene estimated from the danish samples differ from those previously described in the portuguese blood donors[8] (Table 1, Figure 1) (p=0.0005), despite displaying the same modal peaks: allele 4 and allele 10, with allele 4 (one large *MUC1* allele) being the most frequent in portuguese blood donors (28.5%) and allele 10 (one small *MUC1* allele) being the most frequent in danish blood donors (30.1%).

The alleles mainly responsible for the difference are alleles 2, 3 and 10. The smallest alleles (alleles 14 and 15) were not detected in danish blood donors (Figure 1, Table 1).

With this restriction enzyme, allelic frequencies for *MUC1* gene are not in Hardy-Weinberg equilibrium in portuguese blood donors (p=0.01) at variance with the findings in danish blood donors (p=0.671). The heterozigosity was 72.2% and 77.0% for portuguese and danish blood donors,

respectively. Observed and expected heterozygosities do not deviate significantly in the danish sample. On the contrary, a large homozygote excess was observed in the portuguese sample ($\chi 2$ =23.58, $p<0.005$).

MUC1 gene polymorphism in the two populations using AluI

Twenty different alleles were identified using *AluI* as restriction enzyme (Table 1, Figure 2). The distribution of the alleles was bimodal as the one obtained with *EcoRI*. There were differences in the allelic frequencies of portuguese and danish blood donors (p=0.01); however, the two modal alleles were the same in both populations - allele 5 and allele 16 - with allele 5 being the most frequent in the portuguese blood donors (17.2%) and allele 16 being the most frequent in the danish bloos donors (23.2%).

Allele 1 was not detected in the portuguese blood donors. Six of the smallest alleles (alleles 14, 15, 16, 17, 19 and 20) were more frequent in danish than in portuguese blood donors.

Allelic frequencies determined with the use of *AluI* are in Hardy-Weinberg equilibrium for both populations.

The heterozigosity was 86.5% and 92.9% for portuguese and danish blood donors, respectively. Observed and expected heterozygosities do not deviate significantly in both populations.

Discussion

The analysis of *MUC1* polymorphism using two different restriction enzymes indicates that *AluI* appears to be more suitable for genotyping than *EcoRI*. Indeed, in the portuguese sample we have detected a significant deviation from the Hardy-Weinberg equilibrium, due to homozygote excess, when using *EcoRI*. This finding, which cannot be attributed to technical errors since it is not found in the danish sample, probably reflects a polymorphism at the *EcoRI* restriction site with high frequency in Portugal. This assumption is supported by the fact that the difference observed in the

distribution of the *MUC1* gene frequency in both Caucasian populations is higher with *EcoRI* than with *AluI*.

With both restriction enzymes we observed a distribution of *MUC1* alleles in the blood donors from Portugal and Denmark that does not meet the hypothesis put forward as the starting point of the present study: the higher frequency of small *MUC1* alleles/genotypes in the portuguese population might contribute to explain the higher incidence of gastric carcinoma in Portugal than in Denmark. As a matter of fact we observed a tendency for a higher frequency of small sized *MUC1* alleles in the danish than in the portuguese population. The same tendency is detectable if one focus on the two modal alleles observed in both systems. Alleles 4 (*EcoRI*) and 5 (*AluI*), that are larger alleles, are more frequent in portuguese blood donors, and alleles 10 (*EcoRI*) and 16 (*AluI*), that are smaller alleles, are more frequent in danish blood donors.

In conclusion the much higher incidence of gastric cancer in Portugal in comparison to Denmark[9] can not be ascribed to *MUC1* gene polymorphism *per se*. The results of the present study show, furthermore, that if *MUC1* polymorphism is a susceptibility trait for gastric cancer, it does not imply an isolated major effect on gastric cancer development. For the time being, we will have to continue relying upon the influence of environmental factors such as *Helicobacter pylori* infection to explain the aforementioned differences[14].

Acknowledgments

This work was supported by PRAXIS XXI (Grant 2/2.1/SAU/1356/95). The authors thank the staff of the Immunohemotherapy Department of Hospital S. João.

References

1 Gendler SJ, Spicer AP. Epithelial mucin genes. *Annu Rev Physiol* 1995; **57**: 607-634.

2 Gendler SJ, Lancaster CA, Taylor-Papadimitrou J *et al.* Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J Biol Chem* 1990; **265**: 15286-15293.

3 Zotter S, Hageman PC, Lossnitzer A, Mooi WJ, Hilgers J. Tissue and tumor distribution of human polymorphic epithelial mucin. *Cancer Rev* 1988; **11-12:** 55-101.

4 Ho SB, Kim YS. Carbohydrates antigens on cancer-associated mucin-like molecules. *Sem Cancer Biol* 1991; **2**: 389-400.

5 Devin PL, McKenzie IFC. Mucins: structure, function and associations with malignancy. *BioEssays* 1992; **14**: 619-625.

6 Ho SB, Niehans GA, Lyftogt C *et al.* Heterogeneity of mucin expression in normal and neoplastic tissues. *Cancer Res* 1993; **53**: 641-651.

7 Reis CA, David L, Seixas M, Burchell J, Sobrinho-Simões M. Expression of fully and under-glycosylated forms of *MUC1* mucin in gastric carcinoma. *Int J Cancer* 1998; **79**: 402-410.

8 Carvalho F, Seruca R, David L *et al. MUC1* gene polymorphism and gastric cancer - an epidemiological study. *Glycoconj J* 1994; **14**: 107-111.

9 Black RJ, Bray F, Ferlay J, Parkin DM. Cancer incidence and mortality in the European Union: cancer registry data and estimates of national incidence for 1990. *Eur J Cancer* 1997; **33:** 1075-1107.

10 Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 1992; **48**: 361-372.

11 Lewis PO, Zaykin D. Genetic Data Analysis: Software for the analysis of discrete genetic data. Version 1.0. 1997; Free program distributed by authors over the internet from GDA Home Page at http://chee.unm.edu/gda/.

12 Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 1978; **89**: 583-590.

13 Swofford DL, Selander RB. BIOSYS-1: A Fortran program for the comprehensive analysis for electrophoretic data in population genetics and systematics. *J Hered* 1981; **72**: 281-283.

14 Forman D. Review article: Is there significant variation in the risk of gastric cancer associated with *Helicobacter pylori* infection? *Aliment Pharmacol Ther* 1998; **12 Suppl 1**: 3-7.

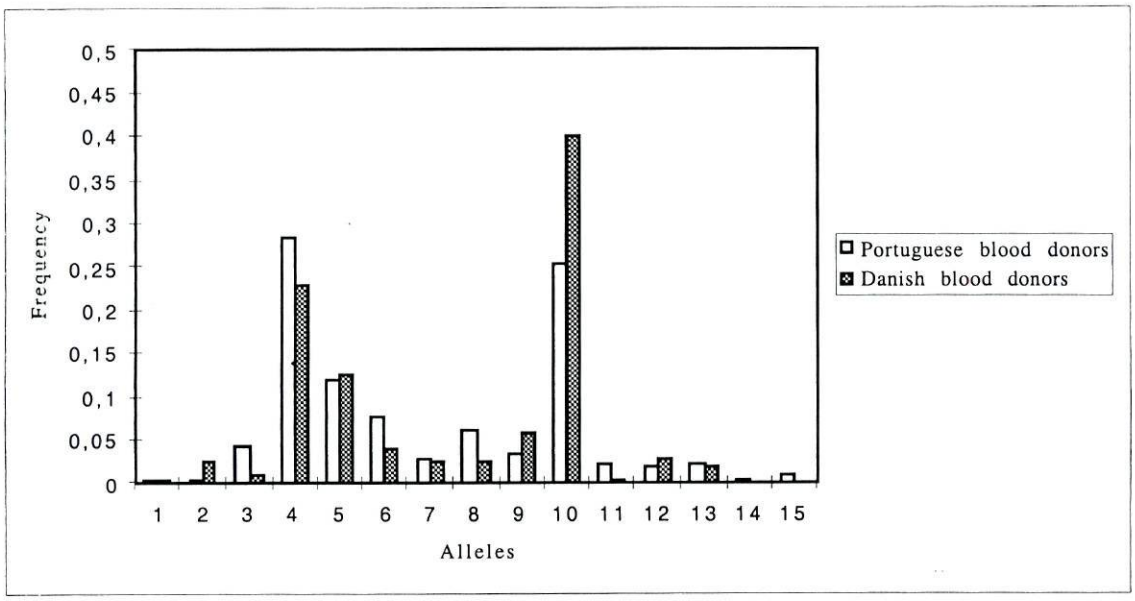
Table 1 Allelic frequencies of *MUC1* gene in portuguese and danish blood donors, using *EcoRI* and *AluI* as restriction enzymes.

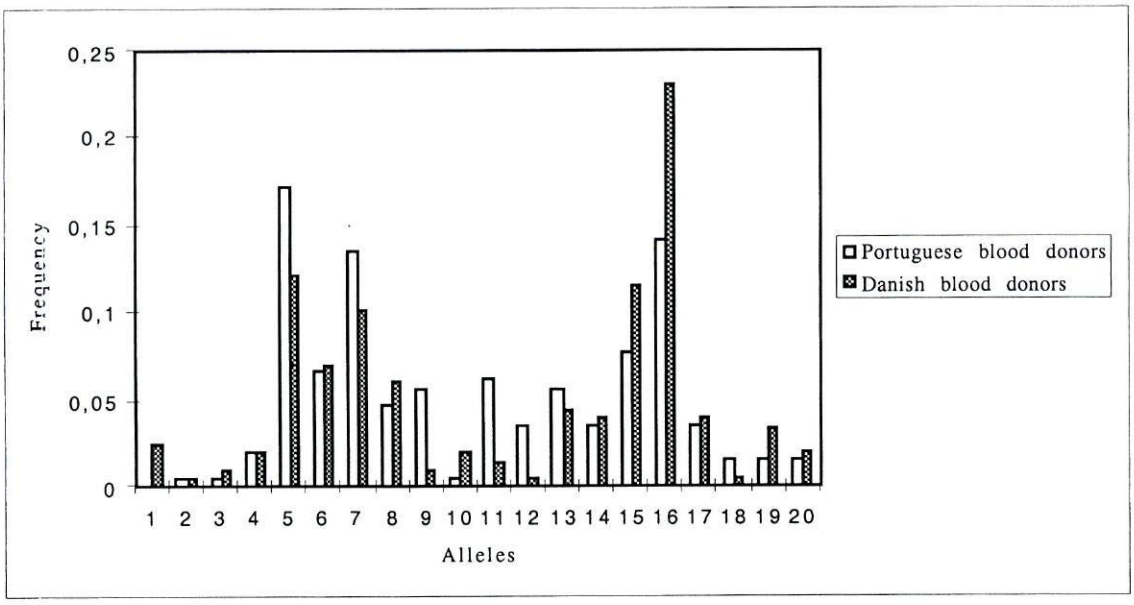
EcoRI			*AluI*		
MUC1 Alleles	Portuguese blood donors (n=324)	Danish blood donors (n=100)	*MUC1* Alleles	Portuguese blood donors (n=96)	Danish blood donors (n=99)
1	3 (0.005)	1 (0.005)	1	0 (0)	5 (0.025)
2	4 (0.006)	5 (0.025)	2	1 (0.005)	1 (0.005)
3	29 (0.045)	2 (0.010)	3	1 (0.005)	2 (0.010)
4	185 (0.285)	46 (0.230)	4	2 (0.021)	4 (0.020)
5	77 (0.119)	25 (0.125)	5	33 (0.172)	24 (0.121)
6	50 (0.077)	8 (0.040)	6	13 (0.068)	14 (0.071)
7	20 (0.031)	5 (0.025)	7	26 (0.135)	20 (0.101)
8	40 (0.062)	5 (0.025)	8	9 (0.047)	12 (0.061)
9	23 (0.035)	12 (0.060)	9	11 (0.057)	2 (0.010)
10	166 (0.256)	80 (0.400)	10	1 (0.005)	4 (0.020)
11	14 (0.022)	1 (0.005)	11	12 (0.063)	3 (0.015)
12	13 (0.020)	6 (0.030)	12	7 (0.036)	1 (0.005)
13	15 (0.023)	4 (0.020)	13	11 (0.057)	9 (0.045)
14	3 (0.005)	0 (0)	14	7 (0.036)	8 (0.040)
15	6 (0.009)	0 (0)	15	15 (0.078)	23 (0.116)
-	-	-	16	27 (0.141)	46 (0.232)
-	-	-	17	7 (0.036)	8 (0.040)
-	-	-	18	3 (0.016)	1 (0.005)
-	-	-	19	3 (0.016)	7 (0.035)
-	-	-	20	3 (0.016)	4 (0.020)
	$\chi 2 = 38.1$; p=0.0005			$\chi 2 = 35.6$; p=0.01	

Legends to Figures

Figure 1 *Allelic frequencies distribution of MUC1 gene in blood donors from Portugal and Denmark, using EcoRI as restriction enzyme. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the largest).*

Figure 2 *Allelic frequencies distribution of MUC1 gene in blood donors from Portugal and Denmark, using AluI as restriction enzyme. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the largest).*


0,5
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
Frequency
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Alleles
Portuguese blood donors
Danish blood donors


0,25
0,2
0,15
0,1
0,05
0
Frequency
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Alleles
Portuguese blood donors
Danish blood donors

PAPER IV

Mucins and mucin-associated carbohydrate antigens expression in gastric carcinoma cell lines

Filipa CARVALHO*, Leonor DAVID*†, Jean-Pierre AUBERT‡§, Celso Albuquerque REIS*, Fátima GÄRTNER*|, António PEIXOTO*, Pedro ALVES*, Manuel SOBRINHO-SIMÕES*†

*Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP, Rua Dr. Roberto Frias s/n, 4200 Porto, Portugal, and

†Medical Faculty of the University of Porto, 4200 Porto, and

‡Unité INSERM U 377, Place de Verdun , 59045 Lille Cedex, France, and

§Laboratory of Biochemistry and Molecular Biology, Hospital C. Huriez, CHRU 59037 Lille Cedex, and

|Institute of Biomedical Sciences Abel Salazar, Largo Prof. Abel Salazar, 4099/003

Porto, Portugal

Page-heading title: Mucins in gastric carcinoma cell lines

Keywords: mucins, MUC1, MUC5AC, MUC6, gastric carcinoma cell lines

Correspondence:

Filipa Carvalho

Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP, Rua Dr. Roberto Frias s/n, 4200 Porto, Portugal

Tel: +351-2-5570700

Fax: +351-2-5570799

E-mail: filipa.carvalho@ipatimup.pt

SYNOPSIS

Mucins are high molecular weight glycoproteins characterised by high molecular mass, high carbohydrate content (50-80% of their mass) and marked heterogeneity involving both the apoprotein and the oligosaccharide side chains. There is consistent data indicating that mucin genes are expressed in a regulated cell- and tissue-specific manner. Stomach is a good example of such differential expression of mucin genes. The first aim of the present study was to characterise the expression of mucins at mRNA and protein level, and of mucin-associated carbohydrate antigens, at protein level, in a panel of seven gastric carcinoma cell lines, and to compare the expression profile in the cell lines with the results previously described in normal gastric tissues and human gastric carcinomas. Secondly, we aimed to see whether or not there is an association between the expression of mucins and of mucin-associated carbohydrayte antigens in order to evaluate if, like in normal conditions, the glycosylation is related with the expressed mucin proteins. Our results show that the profile of mucin expression in gastric carcinoma cell lines: a) follows partially the mucin expression profile of normal gastric mucosa and gastric carcinomas with wide expression of MUC1 and MUC5AC; b) parallels the aberrant pattern of mucin expression observed in human gastric carcinomas with occasional expression of MUC2, MUC3, MUC4 and MUC5B; c) does not include, at least in our series, the expression of the gastric mucin MUC6; d) follows partially the differentiation pattern of the carcinomas from which the cell lines have been originated keeping S-Tn expression in the cell lines derived from glandular type gastric carcinomas. Our results further demonstrate that there is no apparent relationship between the mucin core proteins and the simple mucin-type or Lewis carbohydrate antigens that are expressed in each cell

line, suggesting that in contrast to normal gastric tissues, the glycosylation is not related with the mucins expressed in gastric carcinoma cell lines.

INTRODUCTION

Mucins are high molecular weight glycoproteins characterized by high molecular mass, high carbohydrate content (50-80% of their mass) and marked heterogeneity involving both the apoprotein and the oligosaccharide side chains [1](for review see Gendler and Spicer, 1995). To date nine human mucin genes have been partially or totally sequenced and characterized [2-13] (Gendler *et al*, 1990; Lan *et al*, 1990; Lancaster *et al*, 1990; Gum *et al*, 1994; Gum *et al*, 1990; Porchet *et al*, 1991; Aubert *et al*, 1991; Guyonnet-Duperat *et al*, 1995; Dufosse *et al*, 1993; Toribara *et al*, 1993; Bobek *et al*, 1993; Shankar *et al*, 1997).

There is consistent data indicating that mucin genes are expressed in a regulated cell- and tissue-specific manner. Stomach is a good example of such differential expression of mucin genes. MUC1 is detected in mucous cells of the surface epithelium and neck region of the gastric antrum, as well as in pyloric glands and in oxynthic glands of the body region [14-19] (Burchell *et al,* 1987; Girling *et al*, 1989; Ho *et al*, 1993; Lesuffleur *et al*, 1994; Ho *et al*, 1995; Reis *et al*, 1998a). MUC5AC is highly expressed in foveolar epithelium of both body and antrum [18, 20-23] (Audié *et al*, 1993; Byrd *et al*, 1997; DeBolós *et al*, 1995; Ho *et al*, 1995; Reis *et al*, 1997), whereas MUC6 protein expression is limited to mucous neck cells of the body and pyloric glands of the antrum [18, 21, 22] (Byrd *et al*, 1997; DeBolós *et al*, 1995; Ho *et al*, 1995). Finally, the

expression of MUC2 in normal stomach has been reported either as absent [18, 24] (Ho *et al*, 1995; Xing *et al*, 1992b) or as very weak [16, 20, 24-28] (Audie *et al*, 1993; Carrato *et al*, 1994; Devine *et al*, 1993; Gambús *et al*, 1993; Ho *et al*, 1993; Reis *et al*, 1998b).

Glycosylation is thought to be essential for the biological functions of mucins and it has been postulated that the cell type-specific mucin expression in gastric cells may be related to the type of glycosylation. De Bolós *et al* (1995) [22] demonstrated namely, that there is an association between the expression of MUC5AC and those of Lewis^{b}, Lewis^{a}, and sialyl-Lewis^{a} and between the expression of MUC6 and that of Lewis^{y}.

During gastric carcinogenesis, there is often expression of underglycosylated forms of MUC1 mucin, decreased expression of MUC5AC and MUC6, and aberrant expression of mucins not found in the normal gastric mucosa, including MUC2 [18, 19, 23, 29] (Baldús *et al*, 1998; Ho *et al*, 1995; Reis *et al*, 1997; Reis *et al*, 1998a). It has been shown that aberrant glycosylation occurs in cancer development of the stomach like elsewhere [30] (Hakomori, 1989), leading to the accumulation of simple mucin-type carbohydrates such as Tn, Sialyl-Tn and T (Thomsen-Friedenreich) antigens [31-33] (Hirohashi, 1985; Kjeldsen, 1988, Springer, 1984).

The first aim of the present study was to characterize the expression of mucins at mRNA and protein level, and of mucin-associated carbohydrate antigens, at protein level, in a panel of seven gastric carcinoma cell lines, and to compare the expression profile in the cell lines with the results previously described in normal gastric tissues and human gastric carcinomas taking in consideration that three of the cell lines (GP202, GP220 and KATOIII) were developed from isolated cell carcinomas and the remaining four cell lines (MKN45, St23132, St2957 and St3051) were originated from glandular type gastric

carcinomas. Secondly, we aimed to see whether or not there is an association between the expression of mucins and of mucin-associated carbohydrayte antigens in order to evaluate if, like in normal conditions, the glycosylation is related with the expressed mucin proteins.

EXPERIMENTAL

Materials

Culture materials and standard RNA marker were obtained from Life Technologies, Inc. Nylon membranes Hybond-N^+ and [α-^{32}P]dCTP were purchased from Amersham. Mucin probes were obtained from different sources: MUC1 [2], MUC2 [34], MUC3 [6], MUC4 [7], MUC5AC [9], MUC5B [10], MUC6 [11], MUC7 [35]. Actin probe and 1st –Strand cDNA synthesis Kit were obtained from Clontech (Clontech, Ozyme, France). Smart ladder used as a standard for RT-PCR was purchased from Eurogentec. Random-primer labelling kit was obtained from Boehringer Mannheim. Ampli Taq Gold was purchased from Perkin Elmer. The primers were synthesized by MWG-BIOTECH. Anti-MUC1 monoclonal antibodies HMFG-1, HMFG-2 and SM3 were a generous gift from Dr. Joyce Taylor-Papadimitriou and Dr. Joy Burchell (Imperial Cancer Research Fund, London). Monoclonal antibodies PMH1 [28], CLH2 [23] and CLH5 (unpublished) were produced by our group. Anti-Le^b monoclonal antibody BG.6 was purchased from Signet Laboratories Inc.. All the other monoclonal antibodies to the carbohydrate structures were a kind gift from Prof. Henrik Clausen (Faculty of Health

Sciences, Copenhagen). Rabbit normal non-immune serum, biotinilated rabbit anti-mouse secondary antibody, and avidin-biotin-peroxidase complex were obtained from DAKO.

Cell culture

Seven gastric cancer cell lines were analysed. GP202 and GP220 cell lines were recently established in our lab [36], MKN45 and KATOIII are commercially available cell lines (American Type Cell Collection) and St23132, St2957, St3051 cell lines were a kind gift from Dr. Peter Vollmers [37]. All cells lines were grown in RPMI 1640 with L-glutamax, supplemented with 10% inactivated (30 min, 56°C) fetal bovine serum, 100U/ml penicilin, 100 mg/ml streptomycin and 100U/ml nystatin. Cultures were maintained at 37°C in a humidified 5% CO_2 atmosphere.

Northern blotting

Large RNAs were extracted from the different cell lines (~10^8 cells) at confluence using an improved method for the isolation and transfer of large RNAs as previously described [38] (Debailleul *et al*, 1998). All the probes used to *MUC1, MUC2, MUC3, MUC4, MUC5AC, MUC5B, MUC6, MUC7* and actin (positive control) were labelled with [α–^{32}P]dCTP using a commercial random-primer labelling kit according to the manufacturer's protocol.

RT-PCR (Reverse Transcription - Polymerase Chain Reaction)

cDNAs were prepared using the 1st-Strand cDNA synthesis Kit. PCRs were carried out in 50 µl of final solution (5µl Perkin Elmer buffer, 5µl MgCl2 25mM, 5µl

dNTPs, 10 pmoles of each primer, 2U Taq Polymerase, and 5µl cDNA). Cycling conditions were as follows: 1) denaturation at 96°C, 30 seconds for one cycle; 2) denaturation at 96°C, 30 seconds, annealing at 60°C, 30 seconds, and extention at 72°C, 60 seconds for 30 cycles; 3) extention at 72°C, 15 minutes. The primers were chosen in the 3' part of the different mucin genes (unpublished data). The same PCR conditions were performed to study all mucin genes. GAPDH mRNAs were coamplified in each experiment with each primer set.

Immunohistochemistry

Sections from formalin-fixed paraffin-embedded cytospins were used for immunostaining. The avidin-biotin-peroxidase complex (ABC) method was applied. The paraffin sections were dewaxed and then treated with 0.3% hydrogen peroxide (H_2O_2) in methanol for 30 min, to block endogenous peroxidase. The sections were incubated for 20 min with normal non-immune serum to eliminate non-specific staining. Excess normal serum was removed and replaced by specific primary monoclonal antibodies in the appropriate dilution (Table 1). Sections were incubated overnight (18 to 22 h) at 4°C. After washing the slides, sections were incubated with biotin-labelled secondary antibody diluted 1:200 in PBS with 5% BSA for 30 min and with ABC for 60 min. Sections were stained for 5 min with 0.05% 3'3 diaminobenzidinetetrahydrochloride (DAB) freshly prepared in 0.05M tris/hydroxymethylaminomethane (Tris) buffer at pH 7.6, containing 0.01% H_2O_2, counterstained with hematoxylin, dehydrated and mounted.

RESULTS

Northern blotting

The results of the Northern blotting study of the different mucins and actin are summarised in Table 2. *MUC1* mRNA was detected in GP202, GP220, MKN45 and KATOIII cell lines (Figure 1). Low levels of *MUC2* mRNA were only detected in KATOIII. GP220 expresses *MUC3* mRNA. Low levels of MUC3 and MUC4 were detected in St23132 and KATOIII cell lines, respectively. *MUC5AC* mRNA was detected in St23132, St2957 and St3051 cell lines. *MUC5B, MUC6* and *MUC7* mRNAs were not detected in any of the cell lines.

RT-PCR (Reverse Transcription - Polymerase Chain Reaction)

Table 3 summarises the results from RT-PCR. All cell lines expressed *MUC1* (Figure 2). *MUC2* expression was detected in KATOIII and St2957 cell lines, and *MUC3* expression was detected in GP220 and St23132 cell lines. *MUC5AC* was detected in five cell lines and not in GP202 and KATOIII cell lines (Figure 3). *MUC5B* was detected in four cell lines: MKN45, St23132, St2957 and St3051. None of the cell lines expressed *MUC4*, *MUC6* or *MUC7* as detected by RT-PCR.

Immunohistochemistry

The results from immunohistochemistry of the seven cell lines with the different monoclonal antibodies to the core proteins of mucins are summarised in Table 4. All cell lines expressed MUC1 and MUC5AC at different levels, except for GP220 that did not

express MUC5AC. Expression of MUC1 was detected in all cell lines using three different monoclonal antibodies: HMFG-1 MAb stained more cells than HMFG-2 and HMFG-2 stained more cells than SM3. Four cell lines expressed MUC2: GP202 and St2957 at low levels and MKN45 and KATO III at high levels. None of the cell lines expressed MUC6.

The expression of mucin-associated carbohydrate antigens is summarised in Table 5. All the cell lines expressed high levels of type 1 and type 2 chain Lewis antigens. In most cell lines, the staining for the sialylated epitopes, S-Le^a and S-Le^x, was lower than that of the respective non-sialylated antigens (Le^a and Le^x, respectively). Exceptions are MKN45 and St2957 cell lines that show similar levels of Le^a and S-Le^a, KATOIII that shows higher levels of S-Le^x than Le^x and St23132 that shows similar levels of Le^x and S-Le^x. The expression of simple mucin-type antigens is low or absent in every cell line. Expression of Tn antigen was observed in four cell lines: GP202, MKN45, St23132 and St2957. The sialylated form, S-Tn, was observed in MKN45, St23132, St2957, St3051 cell lines. T antigen was only detected in cell line GP202.

Comparison of the results obtained by Northern blotting, RT-PCR and immunohistochemistry

MUC1 mRNA detected by RT-PCR as well as protein expression detected by immunohistochemistry were found in all gastric carcinoma cell lines. In St23132, St2957 and St3051 cell lines *MUC1* mRNA was not detected by Northern blotting (Table 6).

MUC2 protein expression was detected by immunohistochemistry in four cell lines: in KATOIII cell line, mRNA was also detected by Northern blotting and RT-PCR;

in St2957 cell line mRNA was detected by RT-PCR alone; and in GP202 and MKN45 cell lines mRNA detection was not achieved by Northern blotting nor by RT-PCR (Table 6).

MUC3 mRNA was detected in GP220 and St23132 cell lines, both by Northern blotting and RT-PCR (Table 6).

MUC4 mRNA was detected in KATOIII cell line by Northern blotting alone (Table 6).

MUC5AC mRNA detected by RT-PCR and Northern blotting, as well as protein expression detected by immunohistochemistry, were found in St23132, St2957 and St3051 cell lines (Table 6). The four remaining cell lines showed discrepant results with the different methodologies: MKN45 cell line showed mRNA only by RT-PCR as well as protein expression; GP202 and KATOIII cell lines showed protein expression alone, and GP220 cell line showed mRNA expression detected by RT-PCR alone (Table 6).

MUC5B mRNA was detected in MKN45, St23132, St2957 and St3051 cell lines by RT-PCR without mRNA expression detected by Northern blotting (Table 6).

MUC6 and MUC7 mucins were not detected in any of the cell lines (Table 6).

DISCUSSION

To our knowledge this is the first comprehensive study on mucin expression in a panel of gastric carcinoma cell lines. Mucin expression was evaluated at mRNA and protein level using different methodologies.

Our data show that all cell lines evaluated in this study express MUC1 and most of the cell lines express MUC5AC. These two mucins are expressed in normal gastric mucosa [14-19, 21-23] (Burchell *et al,* 1987; Girling *et al*, 1989; Ho *et al*, 1993; Lesuffleur *et al*, 1994; Ho *et al*, 1995; Reis *et al*, 1998a; Byrd *et al*, 1997; DeBolós *et al*, 1995; Reis *et al*, 1997); MUC1 mucin is expressed in all gastric carcinomas from a series of 73 cases [19] and MUC5AC is expressed in a large proportion of gastric carcinoma cases (63.6% from a series of 66 cases) [23]. Our results thus fit with the *in vivo* data and suggest that gastric type differentiation, evaluated by MUC1 and MUC5AC expression, is maintained in all gastric carcinoma cell lines.

The absence of *MUC1* mRNA detected by Northern blotting in three of the gastric carcinoma cell lines, despite their positivity by RT-PCR and immunohistochemistry, may be explained by the presence of lower amounts of *MUC1* mRNA, under the detection threshold for Northern blotting. This result is in keeping with low levels of actin expression in the St23132, St2957, St3051 cell lines (data not shown).

A similar interpretation is tenable for the understanding of MUC5AC protein expression in MKN45 cell line without detectable amounts of *MUC5AC* mRNA by Northern blotting. In GP220 cell line, *MUC5AC* expression was detected by RT-PCR alone, suggesting that very low levels of gene expression are responsible for undetectable

levels of *MUC5AC* mRNA by Northern blotting and of MUC5AC protein by immunohistochemistry. Finally, the immunohistochemical expression of MUC5AC in GP202 and KATOIII cell lines, which are negative by Northern blotting and RT-PCR, may represent a false positive result of immunohistochemistry despite the high specificity of the antibody we have used [23].

In the present study three antibodies for MUC1 mucin were used and higher reactivities were observed with HMFG-1 and HMFG-2 than with SM3, reflecting different specificities of the MAbs for the epitopes they recognize. In fact, HMFG-1 reacts with fully glycosylated and deglycosylated MUC1 whereas HMFG-2 reacts with deglycosylated MUC1 and more weakly with the fully glycosylated MUC1 [39, 40] (Taylor-Papadimtriou, 1981; Burchel *et al*, 1983). SM3 detects a cancer-associated epitope on the core peptide of MUC1 which is not exposed in fully glycosylated molecules [14] (Burchell *et al*, 1987). Underglycosylated forms of MUC1, detected by SM3, were only observed in GP202 and MKN45 cell lines. These results are in accordance with previous results from our group showing that HMFG-1 is positive in every gastric carcinoma whereas SM3 epitope is expressed in 80% of the cases [19] (Reis et al, 1998). Gastric carcinoma cell lines parallel gastric carcinoma tissues regarding the close association between the immunoreactivity for SM3 epitope and the level of expression of HMFG-1 epitope: the two cell lines staining for SM3 depicted high levels of immunoreactivity for HMFG-1. As proposed for gastric carcinomas *in vivo* it is tempting to speculate that high levels of MUC1 expression lead to a relative insufficiency of the glycosylation system with exposure of the underglycosylated MUC1 epitope

recognized by SM3. Alternatively, increased amounts of MUC1 synthesis may lead to the cytoplasmic accumulation of immature protein detected by SM3 [19] (Reis *et al,* 1998).

MUC6 mRNA and MUC6 protein were absent in every cell line of the present study. Expression of MUC6 protein is limited to mucopeptic cells of the normal stomach [18, 21, 22] (Byrd *et al*, 1997; DeBolós *et al*, 1995; Ho *et al*, 1995) and in a series of 104 gastric carcinoma cases only 29.8% showed MUC6 protein expression (Reis *et al*, unpublished data). The absence of expression of MUC6 probably reflects the low frequency of mucopeptic differentiation in gastric carcinoma cell lines as well as in gastric carcinomas *in vivo* (Solcia *et al*, 1996) [41].

MUC 2 protein expression with or without detectable amounts of mRNA by RT-PCR and/or Northern blotting was observed in four cell lines. Expression of *MUC2* mRNA detected by Northern blotting was observed in KATOIII cell line in accordance with the RT-PCR and immunohistochemistry results. *MUC2* mRNA and the respective protein were also detected in St2957 cell line by RT-PCR and immunohistochemistry, respectively. MUC2 protein showed further staining in GP202 and MKN45 cell lines where *MUC2* mRNA was not detected by Northern blotting nor by RT-PCR. One can not exclude that some cross-reactivity of monoclonal antibody PMH1 may be responsible for these discrepant results although it seems unlikely since the detection of MUC2 protein was confirmed by another anti-MUC2 monoclonal antibody (data not shown). In fact, conflicting results regarding the expression of MUC2 in normal gastric mucosa have been reported when different antibodies were used, demonstrating either no expression of MUC2 [18, 24] (Ho *et al*, 1995; Xing *et al*, 1992b) or very low levels of expression [16, 25-28] (Carrato *et al*, 1994; Devine *et al*, 1993; Gambús *et al*, 1993; Ho *et al*, 1993; Reis

et al, 1998b). On the other hand, MUC2 is overexpressed or aberrantly expressed in intestinal metaplasia [18, 28, 42] (Ho, 1995; Filipe, 1996; Reis, 1998), a precursor lesion of gastric carcinoma, and in 38% of gastric carcinoma cases [29] (Baldus *et al*,1998). The present study, showing MUC2 expression in four cell lines, demonstrates that both *in vivo* and *in vitro* MUC2 expression reflects a non-universal aberrant differentiation pattern of gastric carcinoma cells.

Studies on the intestinal mucin MUC3 in human stomach are scarce but low levels of *MUC3* mRNA have been demonstrated in normal gastric tissues [20, 43] (Audie et al, 1993; van Klinken, 1996). Expression of *MUC4* in normal stomach has also been reported in the literature [20, 43] (Audie et al, 1993, van Klinken et al, 1996) and the expression of *MUC3* and *MUC4* mRNA was previously demonstrated in gastric carcinomas [18] (Ho, 1995). In accordance with the *in vivo* expression of MUC3 and MUC4 in gastric tissues, consistent results were obtained in the present study regarding Northern blotting and RT-PCR expression of MUC3 in GP220 and St23132 cell lines and of MUC4 in KATOIII cell line by Northern blotting.

To the best of our knowledge expression of MUC5B has not been reported in normal stomach. Four cell lines expressed *MUC5B* mRNA as detected by RT-PCR. Low levels of expression of *MUC5B* mRNA can justify the absence of signal in Northern blotting. *In vitro* expression of MUC5B in cell lines may indicate an aberrant differentiation program of these gastric carcinoma cell lines in culture.

Expression of the low molecular weight salivary mucin, MUC7 has not been reported in the literature for normal gastric tissues and was not detected in any of the cell lines.

In the present study we evaluated mucin protein and mRNA expression in seven gastric carcinoma cell lines using cells at a confluence stage. Time course experiments should be performed in order to see if the levels of mucin mRNAs and proteins may be influenced by the relative differentiation status of the cell cultures at the time of harvest. In fact, it has been reported for LS174T, a colon cancer cell line, that MUC1-6 mRNAs reach maximal expression levels at preconfluence and remain at stable levels thereafter [43] (van Klinken, 1996).

Expression of mucins by the different cell lines was not correlated with the characteristics of the carcinomas from which they derived. This contrasts with previous results from human tumours showing preferential expression of MUC5AC in isolated cell-type gastric carcinomas [23] (Reis , 1997) and preferential expression of MUC2 in glandular type gastric carcinomas [29] (Baldus, 1998).

All cell lines displayed high levels of expression of type 1 and type 2 chain Lewis antigens. The most striking changes of glycosylation occurring in gastric carcinomas are the increased and/or *de novo* expression of Le^a antigen and the decreased or absent expression of type 2 Lewis antigens in comparison with normal gastric mucosa [44] (Murata, 1992). The gastric carcinoma cell lines we have studied apparently do not reproduce the results described for gastric carcinomas since Le^a, S-Le^a and Le^b and all type 2 Lewis antigens (Le^x, S-Le^x and Le^y) are widely expressed in most cell lines. The relationship between mucin expression and Lewis antigens expression reported for the human stomach [22] (De Bolos, 1995), namely the association between expression of MUC5AC and type 1 Lewis structures, and between MUC6 and type 2 Lewis structures, was not found in the cell lines. In fact, type 1 Lewis antigens are expressed in all gastric

carcinoma cell lines and type 2 Lewis antigens are also expressed in all cell lines despite the absence of MUC6 expression. These findings suggest that different carrier mucins are used for building type 1 and type 2 Lewis chain antigens in gastric carcinoma cell lines. Several studies have established that sialyl-Le^a and sialyl-Le^x are carried on the MUC1 apomucin in colon cancer tissues [45] (Hanski) and colon cancer cell lines [46] (Baecstrom), and recent data indicate that sialyl-Le^a is carried on MUC1 and MUC3 in bile [47] (Baecstrom). Biochemical studies will be necessary to elucidate the possible association between apomucin(s) and carbohydrate expression in malignant stomach cells.

The expression of simple mucin-type antigens is low or absent in the gastric carcinoma cell lines. Expression of Tn and S-Tn antigens was observed in four cell lines and T antigen was only detected in GP202 cell line. It has been reported that Tn and S-Tn antigens are expressed in the majority of gastric carcinomas and their metastases [31, 32, 48-54] (Hirohashi, 1985; Kjeldsen, 1988; Thor, 1986; David, 1992; Iwata, 1993; Kushima, 1993; Carneiro, 1994a e b, Werther, 1994), whereas T antigen is detected in a low proportion of cases [49] (David, 1992). S-Tn expression is associated with the glandular type of gastric carcinoma [50-53] (Iwata, 1993; Kushima, 1993; Carneiro, 1994a e b), suggesting that mucin glycosylation may also be related to cell differentiation. The gastric carcinoma cell lines parallel the *in vivo* model in that the four cell lines expressing S-Tn are the ones derived from glandular type gastric carcinomas. The expression of T antigen by a single gastric carcinoma cell line is in keeping with the low levels of expression of T antigen in gastric carcinomas [49] (David, 1992).

Similarly to the absence of an association between mucin expression and Lewis-type antigens expression our data also demonstrate absence of relationship between mucin expression profile and simple mucin-type carbohydrate antigens expression in accordance with previous results reported by Ogata *et al* (1992) [55] in colon cancer cell lines.

Summing up, our results show that the profile of mucin expression in gastric carcinoma cell lines: a) follows partially the mucin expression profile of normal gastric mucosa and gastric carcinomas with wide expression of MUC1 and MUC5AC; b) parallels the aberrant pattern of mucin expression observed in human gastric carcinomas with occasional expression of MUC2, MUC3, MUC4 and MUC5B; c) does not include, at least in our series, the expression of the gastric mucin MUC6; d) follows partially the differentiation pattern of the carcinomas from which the cell lines have been originated keeping S-Tn expression in the cell lines derived from glandular type gastric carcinomas Our results further demonstrate that there is no apparent relationship between the mucin core proteins and the simple mucin-type or Lewis carbohydrate antigens that are expressed in each cell line, suggesting that in contrast to normal gastric tissues, the glycosylation is not related with the expressed mucins in gastric carcinoma cell lines.

Acknowledgements

The authors thank Mrs D. Leitão and M.P. Delescaut for technical assistance. This work was supported by the Fundação para a Ciência e a Tecnologia, PRAXIS XXI (Project nº PECS/P/SAU/255/95) and by BIOMED 2-PL963222.

REFERENCES

1 Gendler, S. J. and Spicer, A. P. (1995) Annu. Rev. Phisiol. **57**, 607-634

2 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.-N. and Wilson, D.(1990) J. Biol. Chem. **265**, 15286-15293

3 Lan, M. S., Batra, S.K., Qi, W.-N., Metzgar, R. S. and Hollingswort, M. A. (1990) J. Biol. Chem. **265**, 15294-15299

4 Lancaster, C. A., Peat, N., Duhig, T., Wilson, D., Taylor-Papadimitriou, J. and Gendler, S. (1990) Biochem. Biophys. Res. Commun. **173**, 1019-1029

5 Gum, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) J. Biol. Chem. **269**, 2440-2446

6 Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R.L., Byrd, J. C., Lamport, D. T. A., Siddiki, B. and Kim, Y. S. (1990) Biochem. Biophys. Res. Commun. **171**, 407-415

7 Porchet, N., Van Cong, N., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A. and Aubert, J.-P. (1991) Biochem. Biophys. Res. Commun. **175**, 414-422

8 Aubert, J.-P., Porchet, N., Crepin, N., Duterque-Coquillaud, M., Vergnes, G., Mazzuca, M., Debuire, B., Petitprez, D. and Degand, P. (1991) Am. J. Resp. Cell. Mol. Biol. **5**, 178-185

9 Guyonnet-Duperat, V., Audie, J. P., Debailleul, V., Laine, A., Buisisne, M., Galiegue-Zouitina, S., Pigny, P., Degand, P., Aubert, J.-P. and Porchet, N. (1995) Biochem. J. **305**, 211-219

10 Dufosse, J., Porchet, N., Audie, J. P., Gyuonnet-Duperat, V., Laine, A., Van-Seuningen, I., Marrakchi, S., Degand, P. and Aubert, J.-P. (1993) Biochem J **293**, 329-337

11 Toribara, N. W., Roberton, A., Ho, S., Kuo, W.-L., Gum, E., Hicks, J., Gum, J., Byrd, J., Siddiki, B. and Kim, Y. S. (1993) J. Biol. Chem. **268**, 5879-5885

12 Bobek, L., Tsai, H., Biesbrock, A. and Levine, M. (1993) J. Biol. Chem. **268**, 29563-20569

13 Shankar, V., Pichan, P., Eddy, R., Tonk, V., Nowak, N., Sait, S., Shows, T., Schultz, R., Gotway, G., Elkins, R., Gilmore, M. and Sachdev, G. (1997) Am. J. Resp. Cell. Mol. Biol. **16**, 232-241

14 Burchell, J., Gendler, S., Taylor-Papadimitriou, J., Girling, A., Lewis, A., Millis, R. and Lamport, D. (1987) Cancer Res. **47**, 5476-5482

15 Girling, A., Bartkova, J., Burchell, J., Gendler, S., Gliiet, C. and Taylor-Papadimitriou, J. (1989) Int. J. Cancer **43**, 1072-1076

16 Ho, S. B., Niehans, G. A., Lyftogt, C., Yan, P. S., Cherwitz, D. L., Gum, E. T., Dahiya, R. and Kim, Y. S. (1993) Cancer Res. **53**, 641-651

17 Lesuffleur, T., Zweibaum, A. and Real, F. X. (1994) Critical Rev. Oncol./Hematol. **17**, 153-180

18 Ho, S. B., Shekels, L. L., Toribara, N. W., Kim, Y. S., Lyftogt, C., Cherwitz, D. L. and Niehans, G. A. (1995) Cancer Res. **55**, 2681-2690

19 Reis, C. A., David, L., Seixas, M., Burchell, J. and Sobrinho-Simões, M. (1998) Int. J. Cancer **79**, 402-410

20 Audie, J. P., Janin, A., Porchet, N., Copin, M. C., Gosselin, B. and Aubert J. P. (1993) J. Histochem. Cytochem. **41**, 1479-1485

21 Byrd, J. C., Yan, P., Sternberg, L., Yunker, C. K., Scheiman, J. M. and Bresalier, R. S. (1997) Gastroenterology **113**, 455-464

22 DeBolós, C., Garrido, M. and Real, F. X. (1995) Gastroenterology **109**, 723-734

23 Reis, C. A., David, L., Nielsen, P. A., Clausen, H., Mirgorodskaya, E., Roepstorff, P. and Sobrinho-Simões, M. (1997) Int. J. Cancer **74**, 112-121

24 Xing, P.-X., Prendoska, J., Layton, G. T., Devine, P. L. and McKenzie, I. F. C. (1992) J. Natl. Cancer Inst. **84**, 699-703

25 Carrato, C., Balagué, C., DeBolós, C., Gonzalez, E., Gambús, G., Planas, J., Perini, J. M., Andreu, D. and Real, F. X. (1994) Gastroenterology **107**, 160-172

26 Devine, P. L., McGuckin, M. A., Birrell, G. W., Whitehead, R. H., Sachdev, G. P., Shield, P. and Ward, B. G. (1993) Br. J. Cancer. **67**, 1182-1188

27 Gambús, G., DeBolós, C., Andreu, D., Francí, C., Egea, G. and Real, F. X.(1993) Gastroenterology **104**, 93-102

28 Reis, C.A., Sørensen, T., Mandel, U., David, L., Mirgorodskaya, E., Roepstorff, P., Kihlberg, J., Stig-Hansen, J. E. and Clausen, H. (1998) Glycoconj. J **15**, 51-62

29 Baldús, S. E., Zirbes. T. K., Engel, S., Hanish, F. G., Monig, S. P., Lorenzen, J., Glossmann, J., Fromm, S., Thiele, J., Pichmaier, H. and Dienes, H. P. (1998) Int. J. Cancer **79**, 133-138

30 Hakomori, S. (1989) Adv. Cancer Res. **52**, 257-331

31 Hirohashi, S., Clausen, H,, Yamadam T,, Shimosatom Y. and Hakomori, S. (1985) Proc. Natl .Acad. Sci. USA **82**, 7039-7043

32 Kjeldsen, T., Clausen, H,, Hirohashi, S., Ogawa, T., Iijima, H. and Hakomori S. (1988) Cancer Res. **4**, 2214-2220

33 Springer, G. F. (1984) Science **224**, 1198-1206

34 Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A. and Kim, Y. S. (1989) J Biol Chem **264**, 6480-6487

35 Vandenhaute, B., Buisine, M. P., Debailleul, V., Clement, B., Moniaux, N., Dieu, M. C., Degand, P., Porchet, N, and Aubert, J. P. (1997) J Hepatol **27**, 1057-1066

36 Gärtner, F., David, L., Seruca, R., Machado, J. C. and Sobrinho-Simões, M. (1996) Virchows Arch. **428**, 91-98

37 Vollmers, H. P., Stulle, K., Dammrich, J., Pfaff, M., Papadopoulos, T., Betz, C., Saal, K. and Muller-Hermelink, H. K. (1993) Virchows Arch **63**, 335-343

38 Debailleul, V., Laine, A., Huet, G., Mathon, P., d'Hooghe, M., Aubert, J. P. and Porchet, N. (1998) J. Biol. Chem. **273**, 881-890

39 Taylor-Papadimitriou, J., Peterson, J. A., Arklie, J., Burchell, J., Ceriani, R. L. and Bodmer, W. F. (1981) Int. J. Cancer **28**, 17-21

40 Burchell, J., Durbin, H. and Taylor-Papadimitriou, J. (1983) J. Immunol. **131**, 508-513

41 Solcia, E., Fiocca, R., Luinetti, O., Villani, L., Padovan, L., Calistri, D., Ranzani, G., N., Chiaravalli, A. and Capella, C. (1996) Am J Surg Pathol **20**, S8-S22

42 Filipe, I., Linehan, J., M., Durrant, L., G., Price, M., R., Smeeton, N., C., Pathak, S. and Swallow, D., M. (1996) Eur J Cancer Prev **5**, 287-295

43 Van Klinken, B. J. W., Oussoren, E., Weenink, J. J., Strous, G., Buller, H. A., Dekker, J., Einerhand, A. W. C. (1996) Glycoconj. J. **13**, 757-768

44 Murata, K., Egami, H., Shibata, Y., Sakamoto, K., Misumi, A. and Ogawa, M. (1992) Am. J. Clin. Pathol. **98**, 67-75

45 Hanski, C., Drechesler, K., Hanish, F. G., Sheehan, J., Manske, M., Ogorek., D., Klussmann, E., Hanski, M., L., Blank, M., Xing, P. X., McKenzie, I. F. C., Devine, P. L. and Riecken, E. O. (1993) Cancer Res **53**, 4082-4088

46 Baeckstrom, D., Nilsson, O., Price, M. R., Lindholm, L. and Hansson, G. C. (1993) Cancer Res **53**, 755-761

47 Baeckstrom, D., Karlsson, N. and Hansson, G. C. (1994) J Biol Chem **269**, 14430-14437

48 Thor, A., Ohuchi, N., Szpak CA, Johnston, W. W. and Scholm, J. (1986) Cancer Res. **46**, 3118-3124

49 David, L., Nesland, J. M., Clausen, H., Carneiro, F. and Sobrinho-Simões, M. (1992) APMIS **100 Suppl 27**, 162-172

50 Iwata, H., Itzkowitz, S. H. and Werther, J. (1993) Acta Pathol. Jpn. **43** 646-653

51 Kushima, R., Jancic, S. and Hattori, T. (1993) Int. J Cancer. **55**, 904-908

52 Carneiro, F., Amado, M., David, L., Clausen, H. and Sobrinho-Simões, M. (1994) Eur. J. Cancer Prev. **3 Suppl 2**, 39-46

53 Carneiro, F., David, L. and Sobrinho-Simões. (1994) Eur. J. Cancer **30A**, 1398-1399

54 Werther, J. L., Rivera-MacMurray, S., Bruckner, H., Tatematsu, M. and Itzkowitz, S. H. (1994) Br. J. Cancer **69**, 613-616

55 Ogata, S., Uehara, H., Chen, A. and Itzkowitz, S.H. (1992) Cancer Res. **52**, 5971-5978

Table 1 Monoclonal antibodies used for immunohistochemistry and respective dilutions.

Antigen	Antibody	Dilution
MUC1	HMFG-1	*
	HMFG-2	*
	SM3	*
MUC2	PMH1	*
MUC5AC	CLH2	*
MUC6	CLH5	*
Tn	5F4	1:15
S-Tn	3F1	1:8
T	3C9	1:10
Le^a	CA3F4	1:5
S-Le^a	CA19.9	1:5
Le^b	BG.6	1:60
Le^x	SH1	1:5
S-Le^x	FH6	1:5
Le^y	AH6	1:5

* Hybridoma supernatant

Table 2 Northern blotting expression of mucins and actin in the different gastric carcinoma cell lines. The signal obtained was scored as: - (no expression), +/- (low levels of expression), and + (high levels of expression).

Cell line	*MUC1*	*MUC2*	*MUC3*	*MUC4*	*MUC5AC*	*MUC5B*	*MUC6*	*MUC7*	Actin
GP202	+	-	-	-	-	-	-	-	+
GP220	+	-	+	-	-	-	-	-	+
MKN45	+	-	-	-	-	-	-	-	+
KATOIII	+	+/-	-	+/-	-	-	-	-	+
St23132	-	-	+/-	-	+	-	-	-	+/-
St2957	-	-	-	-	+	-	-	-	+/-
St3051	-	-	-	-	+	-	-	-	+/-

Table 3 RT-PCR results of mucins expression in the different gastric carcinoma cell lines. The signal obtained was scored as: - (negative) or + (positive).

Cell line	*MUC1*	*MUC2*	*MUC3*	*MUC4*	*MUC5AC*	*MUC5B*	*MUC6*	*MUC7*
GP202	+	-	-	-	-	-	-	-
GP220	+	-	+	-	+	-	-	-
MKN45	+	-	-	-	+	+	-	-
KATOIII	+	+	-	-	-	-	-	-
St23132	+	-	+	-	+	+	-	-
St2957	+	+	-	-	+	+	-	-
St3051	+	-	-	-	+	+	-	-

Table 4 Immunohistochemical expression of mucin proteins in the different gastric carcinoma cell lines. A semi-quantitative approach was used to score the staining: + (<25%), ++ (25-50%), +++ (50-75%) and ++++ (>75%).

	MUC1			MUC2	MUC5AC	MUC6
Cell line	HMFG-1	HMFG-2	SM3	PMH1	CLH2	CLH5
GP202	++++	++++	++	+	++++	-
GP220	+++	+	-	-	-	-
MKN45	++++	++++	++	++++	++++	-
KATO III	+++	+	-	++++	+	-
St23132	+++	+	-	-	++++	-
St2957	+++	-	-	+	+++	-
St3051	+++	+	-	-	+	-

Table 5 Immunohistochemical expression of mucin-associated carbohydrate antigens in the different gastric carcinoma cell lines. A semi-quantitative approach was used to score the staining: + (<25%), ++ (25-50%), +++ (50-75%) and ++++ (>75%).

	Simple mucin-type carbohydrate antigens			Lewis antigens					
				Type 1 chain			Type 2 chain		
Cell line	Tn	S-Tn	T	Le^a	S-Le^a	Le^b	Le^x	S-Le^x	Le^y
	5F4	3F1	3C9	Ca3F4	HB8059	BG.6	SH1	FH6	AH6
GP202	+	-	+	++	-	++++	++++	+	++++
GP220	-	-	-	++++	-	+++	++++	-	++++
MKN45	+	+	-	+	+	+	+++	++	-
KATOIII	-	-	-	++++	++	+	++	++++	++
St23132	+	+	-	++++	+++	++++	++	++	+++
St2957	+	+	-	++	++	++	++	+	+++
St3051	-	+	-	+++	++	++	+++	++	++++

Table 6 Comparison between the results from mucin expression by Northern blotting, RT-PCR and immunohistochemistry. Results are expressed as positive + or negative - .

	MUC1			MUC2			MUC3		MUC4		MUC5AC			MUC5B		MUC6			MUC7	
Cell line	N	RT	I	N	RT	I	N	RT	N	RT	N	RT	I	N	RT	N	RT	I	N	RT
GP202	+	+	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-
GP220	+	+	+	-	-	-	+	+	-	-	-	+	-	-	-	-	-	-	-	-
MKN45	+	+	+	-	-	+	-	-	-	-	-	+	+	-	+	-	-	-	-	-
KATOIII	+	+	+	+	+	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-
St23132	-	+	+	-	-	-	+	+	-	-	+	+	+	-	+	-	-	-	-	-
St2957	-	+	+	-	+	+	-	-	-	-	+	+	+	-	+	-	-	-	-	-
St3051	-	+	+	-	-	-	-	-	-	-	+	+	+	-	+	-	-	-	-	-

N: Northern blotting; RT: RT-PCR; I: Immunohistochemistry.

Figures

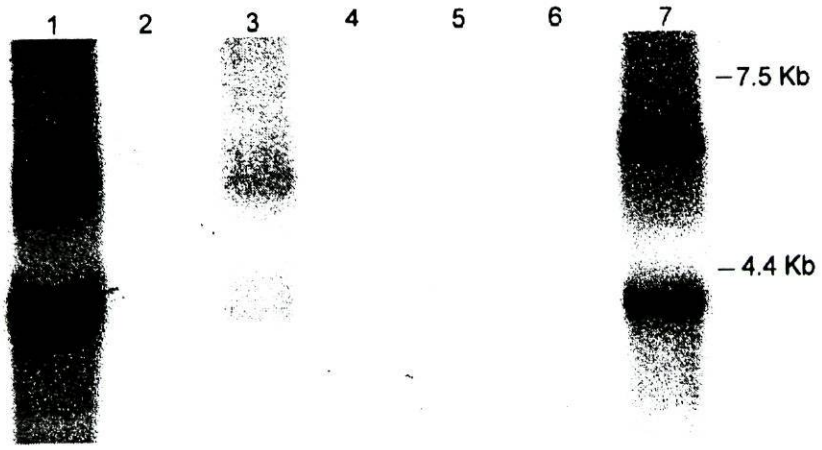


Figure 1. Northern blot analysis of *MUC1* mucin in gastric carcinoma cell lines. Lanes 1-7 correspond to: GP202, GP220, MKN45, St23132, St2957, St3051 and KATOIII. RNA size marker is given in kilobases.

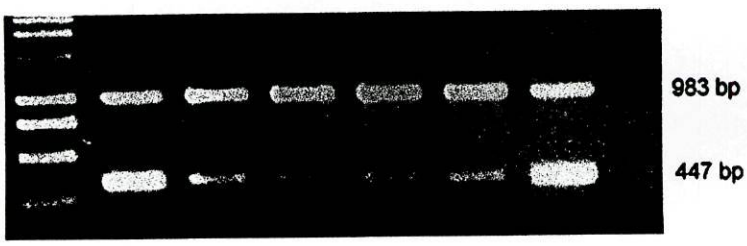


Figure 2. RT-PCR analysis of *MUC1* (447 bp product) and GAPDH (983 bp product) in gastric carcinoma cell lines. Lanes 1-8 correspond to: size marker, GP202, GP220, St2957, St3051, St23132, MKN45 and KATOIII cell lines.

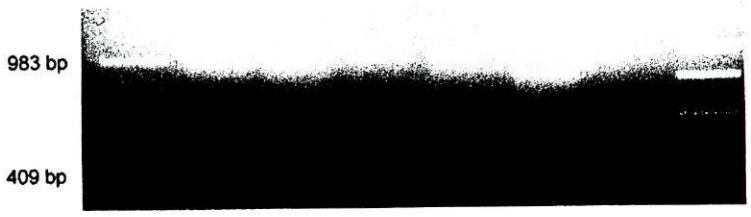


Figure 3. RT-PCR analysis of *MUC5AC* (409 bp product) and GAPDH (983 bp product) in gastric carcinoma cell lines. Lanes 1-8 correspond to: GP202, GP220, St2957, St3051, St23132, MKN45 and KATOIII cell lines, and size marker.