Celiac disease : from basic insight to therapy development

Stepniak, D.T.

Am J Physiol Gastrointest Liver Physiol.
2005 Sep;289(3):G495-500.

5

No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population.

Begoña Diosdado[1]*, Dariusz Stepniak[2]*, Alienke Monsuur[1]*, Lude Franke[1], Martin Wapenaar[1], Manrique L. Mearin[3], Frits Koning[2] and Cisca Wijmenga[1]

[1] Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Center, Utrecht, the Netherlands;
[2] Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, the Netherlands;
[3] Department of Paediatrics, Unit of Paediatric Gastroenterology, Leiden University Medical Center, Leiden, the Netherlands;

* These authors have all contributed equally to this paper.

No genetic association of the human prolyl endopeptidase gene found in the Dutch celiac disease population

Begoña Diosdado, Dariusz Stepniak, Alienke J. Monsuur, Lude Franke, Martin C. Wapenaar, Maria Luisa Mearin, Frits Koning and Cisca Wijmenga

ABSTRACT

Celiac disease (CD) is a complex genetic disorder of the small intestine. The DQ2/DQ8 HLA genes explain approximately 40% of the genetic component of the disease, but the remaining non-HLA genes have not yet been identified. The key environmental factor known to be involved in the disease is gluten, a major protein present in wheat, barley and rye. Integrating microarray data and linkage data from chromosome 6q21-22 revealed the prolyl endopeptidase (*PREP*) gene as a potential CD candidate in the Dutch population. Interestingly, this gene encodes for the only enzyme that is able to cleave the proline rich gluten peptides. To investigate the role of the human *PREP* gene as a primary genetic factor in CD, we conducted gene expression, sequence analysis and genetic association studies of the *PREP* gene and determined PREP enzyme activity in biopsies from CD patients and controls. Sequence analysis of the coding region of the *PREP* gene revealed two novel polymorphisms. Genetic association studies using two novel polymorphisms and three known *PREP* variants excluded a genetic association between *PREP* and CD. Determination of PREP activity revealed weak but significant differences between treated and untreated CD biopsies ($p < 0.05$). Our results from the association study indicate that *PREP* is not a causative gene for CD in the Dutch population. These are further supported by the activity determinations in which we observed no differences in PREP activity between CD patients and controls.

INTRODUCTION

Celiac disease (CD) is a chronic autoimmune disorder caused by the ingestion of dietary gluten. Gluten toxicity in CD patients is, in part, determined by the proline- and glutamine-rich gliadins, secalins and hordeins present in wheat, rye and barley, respectively. This toxicity results from the presence of a repertoire of T cells in the lamina propria of the intestines of CD individuals that are able to recognize many different gluten peptides and provoke an erroneous immune response in the small intestine. This leads to specific tissue damage characterized by lymphocytic infiltration of the mucosa (Marsh I), a Marsh II stage presenting crypt hyperplasia together with the Marsh I features, and Marsh III (MIII) stage in which - in addition to Marsh II - villous atrophy develops [1–3].

So far, the only treatment for CD patients is a strict gluten-free diet, but new alternatives have been recently proposed based on an improved understanding of the disease ethiopathogenesis [4–6]. One of the most attractive new approaches consists of an enzymatic therapy using the bacterial prolyl endopeptidase from *Flavobacterium meningosepticum*, an enzyme that can remove gluten toxicity by cleaving it into small fragments that lack T cell stimulatory properties [4]. This bacterial enzyme has a well-conserved evolutionary homologue in humans (EC 3.4.21.26) [7] which encodes for a cytosolic enzyme that also hydrolyzes amide bonds of very proline-rich peptides shorter than 30 amino acids [8]. It is tempting to speculate that an impaired function of PREP would result in the accumulation of long, immunostimulatory gluten peptides in the lumen or lamina propria, and that this could play a role in breaking down an individual's tolerance to gluten.

Interestingly, the human *PREP* gene is located in the chromosomal region 6q21-22 that showed suggestive linkage (lod score 3.10, p=1.3 x10-4) to CD in the Dutch population [9]. In addition, microarray experiments performed in the same population showed an approximately two-fold up-regulation of *PREP* in seven untreated CD patients compared to four treated CD patients, all eleven of whom still showed villous atrophy ($p<0.005$) [10].

As these results suggested a role for the human *PREP* gene as a primary candidate for CD in the Dutch population, we performed a detailed analysis of PREP activity and follow-up expression in biopsies of patients and controls, sequenced the *PREP* gene in a large group of patients, and carried out genetic association studies.

MATERIALS AND METHODS

Subjects

Seven CD patients from seven independent sibpairs who contributed to the linkage peak on chromosome 6p21-22 and showed two alleles identical-by-descent for this region were selected for re-sequencing the *PREP* gene in order to define new variants in exon and exon-intron boundaries.

We collected 47 biopsies for the enzyme activity studies (Table 1) from 24 CD patients with a MIII biopsy proven lesion, and 23 controls that had a biopsy examination for other reasons such as abdominal pain or failure to thrive. The diagnosis of the CD patients was done according to the ESPGHAN criteria [11]. DNA material was available for 37 of these samples (18 CD patients (24-41, table 1) and 19 controls (1-19, Table 1)), which allowed us to assess both genotype and activity data.

The genetic study comprised a group of 311 independent CD cases and 180 independent age- and sex-matched random hospital controls, all of Dutch Caucasian origin. Only CD patients with a biopsy proven MIII lesion were included in this study. We collected blood samples and isolated DNA according to standard laboratory procedures [9].

Initially 16 biopsies from eight MIII CD patients and eight Mo CD patients, and a pool of 16 RNA samples from control individuals, were used to validate the microarray results for the *PREP* gene using real-time RT-PCR (see Table 1 of supplementary data; http://humgen.med.uu.nl/publications/CD/Diosdado 2005_ 2/). These samples were not used in the further studies.

The study was approved by the Medical Ethics Committees of the University Medical Centers in Utrecht and Leiden, and informed consent was obtained from all individuals.

Determination of PREP enzyme activity.

To measure the PREP activity, we modified the method described by Goosens [12]. The duodenal biopsies were washed with PBS, frozen and stored in –80°C for no longer than 18 months. The biopsies were thawed on ice and ground with an Ultra Turrax homogeniser (Ika Labortechnik, Staufen, Germany) at 22,000 rpm in the presence of 500 µl of lysis buffer (20mM Tris/HCl pH 7.4,137mM NaCl, 2mM EDTA, 10% glycerol, 1%Triton X-100). The lysates were centrifuged (14,000 rpm, 15 min, 4°C) and the assay was performed in 96-well black plates with a clear bottom (Corning Inc., NY, USA). Every measurement was performed four times. 20 µl of lysates were pre-incubated with 75 µl of incubation buffer (100mM K_3PO_4 pH 7.5, 1mM EDTA, 1mM DTT) for 5 minutes at 37°C. The reaction was started by adding 5 µl of substrate solution (4 mM Z-Gly-Pro-AMC in 60% methanol). After one hour of

Table 1. Data on individuals (CD patients and controls) included in the study

Patient	Age	Gender	Status	Biopsy Number	Histological Stage	Diet
1	4	M	Control		Control	None
2	17	F	Control		Control	None
3	16	M	Control		Control	None
4	15	M	Control		Control	None
5	10	M	Control		Control	None
6	16	M	Control		Control	None
7	9	M	Control		Control	None
8	2	M	Control		Control	None
9	5	F	Control		Control	None
10	6	F	Control		Control	None
11	13	M	Control		Control	None
12	18	M	Control		Control	None
13	4	F	Control		Control	None
14	1	M	Control		Control	None
15	5	M	Control		Control	None
16	5	M	Control		Control	None
17	14	M	Control		Control	None
18	4	M	Control		Control	None
19	11	F	Control		Control	None
20	4	F	Control		Control	None
21	10	F	Control		Control	None
22	9	F	Control		Control	None
23	9	F	Control		Control	None
24	2	F	CD patient	1st (Diagnostic)	MIII	None
25	15	F	CD patient	1st (Diagnostic)	MIII	None
26	3	F	CD patient	1st (Diagnostic)	MIII	None
27	10	F	CD patient	1st (Diagnostic)	MIII	None
28	10	M	CD patient	1st (Diagnostic)	MIII	None
29	7	M	CD patient	1st (Diagnostic)	MIII	None
30	6	F	CD patient	1st (Diagnostic)	MIII	None
31	3	M	CD patient	1st (Diagnostic)	MIII	None
32	7	F	CD patient	3rd (Challenge)	MIII	Challenge
33	9	M	CD patient	2nd (Control)	MIII	GFD
34	4	F	CD patient	2nd (Control)	M0	GFD
35	6	F	CD patient	2nd (Control)	MI-II	GFD
36	17	F	CD patient	2nd (Control)	M0	GFD
37	9	F	CD patient	2nd (Control)	M0	GFD
38	6	F	CD patient	2nd (Control)	MI-II	GFD
39	8	F	CD patient	2nd (Control)	MI	GFD
40	2	F	CD patient	2nd (Control)	M0	GFD
41	12	F	CD patient	2nd (Control)	M0	GFD
42	5	F	CD patient	1st (Diagnostic)	MIII	None
43	15	F	CD patient	1st (Diagnostic)	MIII	None
44	3	F	CD patient	1st (Diagnostic)	MIII	None
45	3	F	CD patient	1st (Diagnostic)	MIII	None
46	15	F	CD patient	2nd (Control)	M0	GFD
47	4	M	CD patient	2nd (Control)	M0	GFD
48					M*	
49					M*	
50					M*	
51					M*	
52					M*	
53					M*	
54					M*	

M* - independent sibpairs from genome screen; M - male; F - female; EMA-IgA - Antibodies anti-endomysium; TGA-IgA - antibodies anti-gliadin; CD - celiac disease; GFD - gluten-free diet; M0 - celiac patients with complete remission; MIII - Marsch III stage; ND - not determined.

DQ2	DQ8	EMA-IgA	TGA-IgA	Clinical Symptoms
Pos	Neg	Neg	Neg	Lassitude
ND	ND	Neg	ND	Epigastric pain
ND	ND	ND	ND	Diarrhoea& abdominal pain
ND	ND	ND	ND	Diarrhoea
Pos	Neg	ND	Neg	Short stature
ND	ND	ND	ND	Chronic vomiting
Pos	Neg	ND	ND	Constipation
ND	ND	Neg	Neg	Failure to thrive
ND	ND	ND	ND	Suspected CD
Pos	Neg	Neg	Neg	Diarrhoea
ND	ND	ND	ND	Diarrhoea& anal fistels
Neg	Neg	Neg	Neg	Abdominal pain
ND	ND	ND	ND	Short stature, constipation
ND	ND	dubious	ND	Vomits& failure to thrive
Neg	Neg	Neg	Neg	Abdominal pain
Pos	Neg	Neg	Neg	Abdominal pain
ND	ND	ND	ND	Epigastric pain
Pos	Pos	Pos	ND	Vomits
ND	ND	ND	ND	Unknown
Pos	Neg	Neg	ND	Suspected CD
ND	ND	Neg	ND	Epigastric pain
ND	ND	ND	ND	Suspected CD
ND	ND	ND	ND	Suspected CD
Pos	Neg	Pos	Pos	Asymptomatic
ND	ND	Pos	ND	Unknown
Pos	Neg	Pos	Pos	Chronic diarrhoea& lassitude
ND	ND	Pos	ND	Unknown
Pos	Neg	Pos	ND	Abdominal pain
Pos	Pos	Pos	Pos	Chronic diarrhoea& lassitude
UN	UN	ND	ND	None
ND	ND	ND	ND	Unknown
Pos	Neg	Pos	Pos	Asymptomatic
Pos	Neg	ND	ND	Unknown
Pos	Neg	ND	ND	None
Pos	Neg	ND	ND	Unknown
Pos	Neg	ND	ND	None
Pos	Neg	Neg	Neg	None
yes	Neg	ND	ND	Unknown
Pos	Neg	Neg	Neg	None
Pos	Neg	Neg	Neg	None
Pos	Neg	Neg	Neg	None
ND	ND	ND	ND	Failure to thrive
ND	ND	Pos	Pos	Chronic diarrhoea
ND	ND	ND	ND	Unknown
ND	ND	ND	ND	Unknown
ND	ND	Neg	ND	None
Pos	Neg	Neg	Neg	None

incubation at 37°C, the reaction was stopped with 50 µl of 1M acetic acid. The concentration of the released AMC was measured fluorimetrically at λ_{ex} 360 nm and λ_{em} 460 nm using a CytoFluor multi-well plate reader (PerSeptive Biosciences). One unit of the enzyme was defined as the catalytic activity that releases 1µmol of AMC per minute. Both Z-Gly-Pro-AMC substrate and standard AMC were purchased from Fluka Chemie AG (Buchs, Switzerland). Total protein concentration in lysates was determined using a Bradford protein assay (Bio Rad, Munchen, Germany) and a CBA protein assay (Pierce, Rockford, IL, USA), with BSA (Pierce) as the standard in both cases.

Quantitative real-time RT-PCR

Quantification of *PREP* transcriptional activity was performed by real-time RT-PCR on RNA from biopsies as previously described [13]. We used an Assay-on-Demand Gene Expression product for the *PREP* gene (ABI Hs.00267576), and the GUSB gene (detected by PARD 4326320E) as an endogenous reference to correct for expression-independent sample-to-sample variability (Applied Biosystems, Foster City, CA, USA). In order to quantify the relative expression by the $2^{-\Delta\Delta Ct}$ method [13], equimolar amounts of total RNA from 16 control individuals were pooled and used for normalisation of the expression data. Both genes were tested in duplicate for all the individual patient samples and the control pool on an ABI 7900 HT (Applied Biosystems, Foster City, CA, USA).

Sequence analysis

PCR amplification was performed on all 15 exons and exon-intron boundaries of the *PREP* gene. Details about the primer sequences and the PCR conditions can be found in table 2 of the supplementary data. The PCR products were examined on a 2% agarose gel and purified with the Millipore Vacuum Manifold, according to the manufacturer's protocol (Billerica, MA, USA). Samples were prepared with the ABI PRISM BigDye terminator cycle sequencing ready kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. PCR and sequencing amplification were performed on a GeneAmp PCR system 9700 (Perkin Elmer, Foster City, CA, USA). Sequencing was performed on a 3730 DNA sequencer (Applied Biosystems, Foster City, CA, USA). Analysis and alignment was carried out with the Sequence Navigator (Applied Biosystems, Foster City, CA, USA) and Vector NTI (InforMax Inc, MA, USA).

Genetic association studies and data analysis

Five of the selected SNPs were typed using assay-on-demand probes from the *PREP* gene: hCV1963751 (ABI no. C___1963751_10), rs9486069 (ABI no. C__1163 8424_10), rs1078725 (ABI no. C___8304693_10), rs2793389 (ABI no. C__1163

5753_10) and rs1051484 (ABI no. C___8304751_20). The sixth selected SNP, rs12192054 was typed by using an assay-by-design probe from Applied Biosystems (Applied Biosystems, Foster City, CA, USA). These SNPs were tested in a case-control study (311 cases and 180 controls) and analyzed on an ABI Prism 7900 HT system (Applied Biosystems, Foster City, CA, USA).

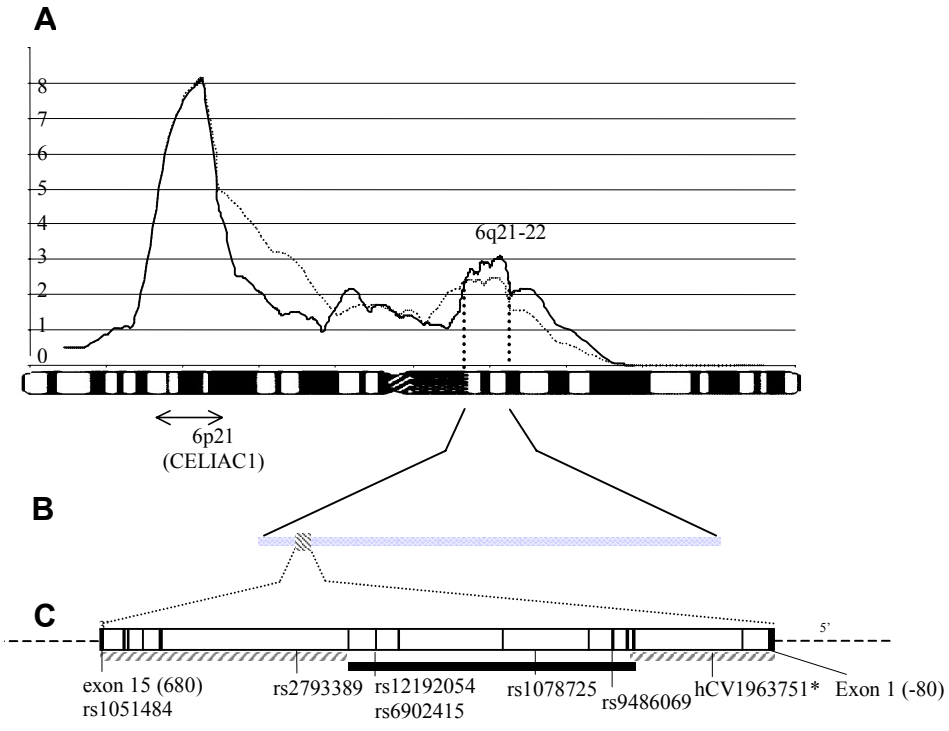
HWE was evaluated separately in cases and control, for all SNPs tested (data not shown). Differences in allele frequencies and genotype distributions were compared between cases and controls using the χ^2 test.

RESULTS

We compiled our earlier microarray [10] and linkage [9] data from Dutch CD patients using TEAM, a bioinformatics tool developed in-house [14], that allowed us to define the physical location of the differentially expressed genes under the genetic linkage peaks. Integrating and analyzing these two data sets revealed that *PREP* was one of the differentially expressed genes located under the linkage peak on chromosome 6q21-22 in the Dutch genome screen (Fig. 1 A and B). The 6q21-22 region encompasses 22 megabases and contains 111 genes. The relative risk in the Dutch CD population attributed to this locus is 2.3 [9]. Quantitative expression studies by real-time RT-PCR on a set of eight RNA samples from treated CD patients in complete remission (M0), eight untreated CD patients with total villus atrophy (MIII) and a pool of normal controls validated these findings. The experiments showed that *PREP* was significantly down-regulated in treated M0 patients compared to MIII patients ingesting gluten (1.3 fold, $p<0.05$; Table 1 of supplementary data), although to a lesser extent than previously described [10].

Sequence analysis

To investigate whether the enzymatic properties of this gene product or its expression levels were different in CD patients due to an underlying genetic variation, we performed sequence analysis on the entire coding region and exon-intron boundaries of the *PREP* gene to identify putative mutations or variants in the CD population. The human *PREP* gene is fully annotated in the public databases and consists of 2,905 nucleotides distributed over 15 exons that encode 710 amino acid residues [7]. Since the tertiary structure of the human PREP has not yet been described, we used the tertiary structure of its porcine homologue as reference in defining which exons were encoded by which domains. The human and porcine enzymes are 97% homologous at the amino acid level, and in the porcine PREP, exons 1-3 and 10-15 encode the catalytic domain and exons 3-10 the characteristic beta-propeller domain that regulates its proteolitic activity (Fig. 1C) [15].



D

SNP ID	Exon location	Base pair location	Amino acid position	Nucleotide change	Amino acid change	Allele frequency in public databases	Found in 44 sequenced individuals
exon1 (-80)*	exon 1	105896438	(-80)	**G**/**T**	UTR 5'	Not present	33 G/G, 11 G/T
hCV1963751*	intron 2	105891476	-	G/A	-	0.66G/0.34A	ND
rs9486069*	exon 5	105867066	130	TA**T**/TA**C**	Tyr/Tyr	0.66T/0.34C	17 T/T, 22 C/T, 1 C/C 4 ND
rs1078725*	intron 6	105858273	-	T/C	-	0.77C/0.23T	ND
rs12192054*	exon 9	105822511	351	**T**TA/**G**TA	Leu/Val	ND	29 T/T, 9 T/G, 2 G/G, 4 ND
rs6902415	exon 9	105822399	375	TT**C**/TT**T**	Phe/Phe	ND	43 C/C, 1C/T
rs2793389*	intron 10	105816736	-	C/A	-	0.85C/0.15A	ND
exon15 (680)	exon 15	105771761	680	CA**C**/CA**T**	His/His	Not present	43 C/C, 1T/T
rs1051484*	exon 15	105771702	706	**G**TC/**A**TC	Val/Ile	0.72G/0.28A	33 G/G, 11 G/A

Figure 1. A. Linkage data of 101 sibpairs (Dutch CD patients) on chromosome 6. The dashed line indicates the linkage graph before fine mapping, while the continous line is after fine mapping. **B**. 95% confidence interval (CI) containing 111 genes. The dashed square indicates the position of the *PREP* gene. **C**. Exonic-intronic view of *PREP*. The dashed line represents the catalytic domains of the protein and the continuous line the beta-propeller domain. **D**. The table includes the 6 exonic single nucleotide polymorphysms (SNPs) identified by sequencing in 44 individuals and the 3 intronic SNPs. The SNPs selected for the genetic studies are indicated with an asterisk. SNP ID - SNP number; UTR - untranslated region; ND - not determined.

Sequence analysis of all 15 exons and exon-intron boundaries in 44 individuals revealed 6 SNPs in the coding region of *PREP*. These SNPs were present in exon 1, exon 5, exon 9 (two SNPs) and exon 15 (two SNPs) (Fig. 1B). The SNP in exon 1 and one of the two SNPs in exon 15 have not yet been annotated in public databases. The published allele frequencies and the frequency of occurrence of these SNPs in the sequenced individuals are shown in Fig. 1D.

Only two of the identified SNPs lead to an amino acid change in the PREP protein. A SNP found in exon 9, 1050T→G, gives rise to a leucine to valine substitution at position 351 while a SNP in exon 15, 2118 G→A, gives rise to a valine to isoleucine substitution at position 706. This latter substitution is not expected to have any impact on the function of the PREP protein as the amino acid at position 706 is not conserved (valine in man, isoleucine in pigs, bovines, rats and mice). The leucine to valine substitution at position 351 is a conservative one and, therefore, we cannot rule out that this substitution may impact PREP function.

Genetic association studies

To further investigate whether genetic polymorphisms in PREP are associated to CD in the Dutch population, we performed genetic association studies. For our linkage peak on chromosome 6p22, with a relative risk of 2.3 and a SNP frequency in the range of 0.1-0.4, our sample size had 80% power to detect a CI of 95%.

Four exonic SNPs (exon 1 (-80), rs9486069, rs12192054 and rs1051484) were selected based on their high heterozygosity in our sequence samples and their possible influence on the protein. Unfortunately, the SNP in the 5' UTR could not be designed because of the extreme repetitiveness in the region. None of the three SNPs, however, showed a statistical difference between the cases and controls (Table 2).

To further exclude *PREP* as a causative gene, we selected three non-coding SNPs for further genetic association studies on the basis of a minor allele frequency of >10% (Table 2). These three SNPs also showed a lack of statistical difference between the cases and controls (Table 2). Haplotype analysis did not change these results (data not shown). Finally, we tested a microsatellite marker located in intron

2 of the *PREP* gene, which also showed no association with CD (data not shown). Overall, we found no association between any of our genetic markers and CD.

Activity of PREP in biopsy material from patients and controls

In order to further investigate whether an impaired enzymatic activity of PREP could be responsible for a decreased digestion of gluten peptides in the small intestine of CD patients and, hence, activation of an aberrant immune response, the catalytic activity of the enzyme was measured in 47 biopsies from CD patients and controls. The activity values lay in the range of 1.71 to 8.52 U/g protein with an average of 4.8 U/g protein (SD = 1.61, which is in agreement with the described PREP activities measured in other human tissues [12]. First, patients were grouped according their histological status and adhering to the treatment in treated CD (M0) and untreated CD (MIII), and independently of their genotypes. The average PREP activity levels measured in the untreated CD patients were lower than in the treated CD patients ($p < 0.05$). No significant differences were observed between the treated or untreated CD patients and the controls (Fig. 2). We were not able to correlate PREP activity levels with the age or gender of the studied individuals (data not shown).

Activity-genotype correlations

To further detect an influence of the tested genetic variants on the expression and activity results, we calculated whether there was any association between the different genotypes of the SNPs and the enzymatic activity of PREP. For activity-genotype correlation, the genotypes of four identified coding SNPs of the gene (Fig. 1D) and the activity measurements of 37 individuals were studied (Table 1, individuals 1-19 and 24-41). To do so, individuals were grouped according to their

Table 2. P-values obtained from testing the case-control cohort for 3 coding and 3 intronic SNPs

SNP name	SNP type	MAF*	Number of cases of MAF	Number of controls of MAF	p-value
hCV1963751	G/A	0.34 (A)	308 (34.7)	178 (36.5)	0.577
rs9486069	T/C	0.34 (C)	306 (28.3)	175 (31.4)	0.301
rs1078725	T/C	0.23 (T)	309 (22.0)	177 (20.9)	0.688
rs12192054	G/T	ND (G)	305 (16.4)	176 (12.2)	0.079
rs2793389	C/A	0.15 (A)	307 (21.5)	179 (18.2)	0.211
rs1051484	A/G	0.28 (A)	309 (15.4)	180 (14.7)	0.784

Values in parentheses are percentages. * Minor allele frequencies were obtained from the Celera or Ensembl databases. All SNPs were tested on 311 cases and 180 controls.

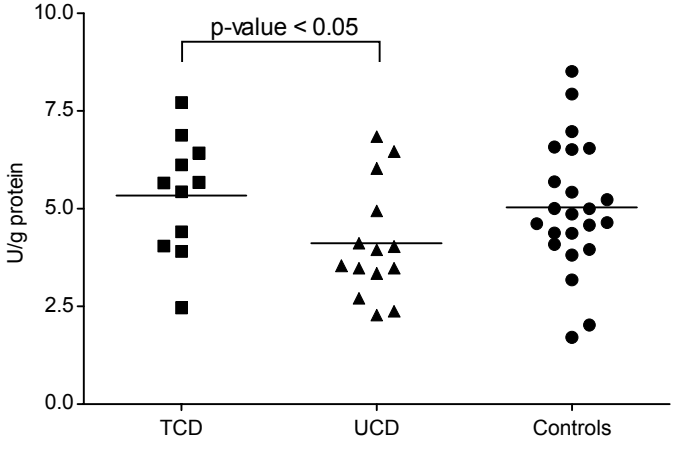


Figure 2. PREP activity in duodenal biopsies. The activity was measured with Z-Gly-Pro-AMC substrate and corrected for protein concentration determined with BCA assay. TCD – treated celiac disease (gluten-free diet); UCD – untreated celiac disease (normal diet); Controls – no celiac disease and normal diet. (TCD 5.35 ± 0.46; UCD 4.12 ± 0.39; $p<0.05$; controls 5.03 ± 0.033)

genotypes and the average of the activity for each group was calculated for each of the coding SNPs, (except SNP rs6902415 since all individuals were homozygote C/C, and exon 15 (680) since all individuals but one were C/C) (Table 3 of supplementary data). An association t-test was used to find genotype-activity correlations but revealed no significant association for any of the four SNPs (data not shown). We concluded that the activity is not modulated by the sequence of the gene, which further supports the findings of our genetic association studies.

Discussion

CD is a complex genetic trait in which genetic and environmental factors are the primary causative determinants for the disease. Although gluten has been identified as the major environmental factor [16], only the genetic contribution of the HLA region is well understood [2]. Recently, a genome-wide screen in our Dutch population has been successful in finding significant linkage to two non-HLA associated regions, one to chromosome 19 and another to chromosome 6q21-22 [9]. No causative gene has been identified yet for either of these regions.

By integrating a data set from our microarray experiments with the genetic information of the 6q21-22 region, we identified eight differentially expressed genes located under this linkage peak. As one of these differentially expressed genes was *PREP*, we hypothesized that an altered PREP activity in the intestinal mucosa could be responsible for the inefficient breakdown of gluten peptides, which could consequently facilitate the onset of CD. We therefore performed a comprehensive set of complementary studies to investigate the putative role of *PREP* in the pathogenesis of CD.

Since expression studies showed the existence of altered levels of *PREP* mRNA in the biopsies of CD patients, we hypothesized we might identify a DNA polymorphism or a variant that would slightly alter the activity of the enzyme, rather than a major mutation that would fully abolish its function. Sequence analysis did not reveal any major mutations in 25 CD patients, but six SNPs were found in the coding region of this gene. One of the SNPs is in one of the residues of the catalytic triad (His680) but it does not give rise to an amino acid change. A novel SNP was found in the 5' UTR of *PREP*. Since the promoter region of *PREP* is not known, *in silico* studies using Transfac TF professional v8.2 were used to define whether putative binding sites and regulatory sequences in the 5' UTR of *PREP* reside at the position of this SNP. No putative regulatory sequence was predicted at the site of the SNP (data not shown), suggesting that this SNP may not affect the transcriptional regulation of *PREP*. SNP rs9486069, located within the first 10 nucleotides of exon 5, was also of potential interest since it has been well established that sequences within the first or last 20 nucleotides of an exon can influence the splicing machinery by enhancing or silencing its effects [17]. We therefore looked for a possible influence of this SNP on the splicing machinery using Spring Harbor software [18], but found none (data not shown).

From the sequence and follow-up analysis we concluded that none of the SNPs would directly provoke a change in the structure of the protein. Neither did our later genetic studies support a role for *PREP* as a primary gene in CD. The microsatellite marker and the six SNPs inside *PREP* did not show any significant differences, nor any trend towards significance. Besides, since the promoter region of the *PREP* gene is unknown, SNPs in this region could not be totally excluded.

Finally, to further exclude any functional consequence of these coding polymorphisms in PREP activity that could implicate it in the pathogenesis of CD, we determined the catalytic activity of PREP in biopsies from 47 children. As expected from the genetic association studies, we found no significant differences between the treated or untreated CD children and the pediatric controls, or when individuals were grouped by their genotypes. Nevertheless, since the biopsies for normal controls came from individuals who might have had altered intestinal mucosa due to diarrhea or abdominal pain, these results could be an underestimation. However,

the PREP activity in untreated CD children was slightly decreased compared to treated CD pediatric patients, possibly as the result of intestinal tissue damage associated with the disease. These observations are perfectly in line with findings of Donlon and Stevens (J. Donlon – personal communication) but do not support results published by Matysiak-Budnik and colleagues, who described an increased *PREP* activity in the intestinal mucosa of eight treated (i.e. following a gluten-free diet) CD patients compared to seven controls [19]. It remains to be established why our results differ from those of Matysiak-Budnik.

In conclusion, these results clearly indicate that no genetic polymorphisms in the *PREP* gene can be linked to CD. This finding is further supported by the activity determinations, in which we found no differences in the enzyme activity between CD patients and controls. Thus, *PREP* does not seem to be implicated in the pathogenesis of CD.

Supplementary data for this article is available at:
http://humgen.med.uu.nl/publications/CD/Diosdado2005_2/

Acknowledgements

The authors would like to thank Ellen van Koppen and Remi Steens for their help in acquiring patient data, Jackie Senior for editing the text, and Alfons Bardoel, Daniel Chan and Alexandra Zhernakova for their practical work.

Grants

This study was supported by the Netherlands Organization for Scientific Research (ZonMW, grants 902-22-094 and 912-02-028), the Celiac Disease Consortium, an innovative cluster approved by the Netherlands Genomics Initiative and partially funded by a Dutch government grant (BSIK03009), and the Dutch Digestive Disease Foundation (MLDS, grant WS00-13).

REFERENCES

1. Dieterich, W., B. Esslinger, and D. Schuppan. 2003. Pathomechanisms in celiac disease. *Int.Arch.Allergy Immunol.* 132:98-108.
2. Sollid, L. M. 2002. Coeliac disease: dissecting a complex inflammatory disorder. *Nat.Rev.Immunol.* 2:647-655.
3. Vader, W., Y. Kooy, P. Van Veelen, A. De Ru, D. Harris, W. Benckhuijsen, S. Pena, L. Mearin, J. W. Drijfhout, and F. Koning. 2002. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 122:1729-1737.
4. Senger, S., D. Luongo, F. Maurano, M. F. Mazzeo, R. A. Siciliano, C. Gianfrani, C. David, R. Troncone, S. Auricchio, and M. Rossi. 2003. Intranasal administration of a recombinant alpha-gliadin down-regulates the immune response to wheat gliadin in DQ8 transgenic mice. *Immunol.Lett.* 88:127-134.
5. Shan, L., O. Molberg, I. Parrot, F. Hausch, F. Filiz, G. M. Gray, L. M. Sollid, and C. Khosla. 2002. Structural basis for gluten intolerance in celiac sprue. *Science* 297:2275-2279.
6. Vader, L. W., D. T. Stepniak, E. M. Bunnik, Y. M. Kooy, W. de Haan, J. W. Drijfhout, P. A. Van Veelen, and F. Koning. 2003. Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. *Gastroenterology* 125:1105-1113.
7. Vanhoof, G., F. Goossens, L. Hendriks, M. De, I, D. Hendriks, G. Vriend, C. Van Broeckhoven, and S. Scharpe. 1994. Cloning and sequence analysis of the gene encoding human lymphocyte prolyl endopeptidase. *Gene* 149:363-366.
8. Venalainen, J. I., R. O. Juvonen, and P. T. Mannisto. 2004. Evolutionary relationships of the prolyl oligopeptidase family enzymes. *Eur.J.Biochem.* 271:2705-2715.
9. Van Belzen, M. J., J. W. Meijer, L. A. Sandkuijl, A. F. Bardoel, C. J. Mulder, P. L. Pearson, R. H. Houwen, and C. Wijmenga. 2003. A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 125:1032-1041.
10. Diosdado, B., M. C. Wapenaar, L. Franke, K. J. Duran, M. J. Goerres, M. Hadithi, J. B. Crusius, J. W. Meijer, D. J. Duggan, C. J. Mulder, F. C. Holstege, and C. Wijmenga. 2004. A microarray screen for novel candidate genes in coeliac disease pathogenesis. *Gut* 53:944-951.
11. Revised criteria for diagnosis of coeliac disease. 1990. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch.Dis.Child* 65:909-911.
12. Goossens, F., M. De, I, G. Vanhoof, and S. Scharpe. 1996. Distribution of prolyl oligopeptidase in human peripheral tissues and body fluids. *Eur.J.Clin.Chem.Clin. Biochem.* 34:17-22.
13. Wapenaar, M. C., M. J. Van Belzen, J. H. Fransen, A. F. Sarasqueta, R. H. Houwen, J. W. Meijer, C. J. Mulder, and C. Wijmenga. 2004. The interferon gamma gene in celiac disease: augmented expression correlates with tissue damage but no evidence for genetic susceptibility. *J.Autoimmun.* 23:183-190.
14. Franke, L., H. van Bakel, B. Diosdado, M. van Belzen, M. Wapenaar, and C. Wijmenga. 2004. TEAM: a tool for the integration of expression, and linkage and association maps. *Eur.J.Hum.Genet.* 12:633-638.

15. Fulop, V., Z. Bocskei, and L. Polgar. 1998. Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell* 94:161-170.
16. Van de Kamer, J. H., H. A. Weijers, and W. K. Dicke. 1953. Coeliac disease. IV. An investigation into the injurious constituents of wheat in connection with their action on patients with coeliac disease. *Acta Paediatr.* 42:223-231.
17. Cartegni, L., S. L. Chew, and A. R. Krainer. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat.Rev.Genet.* 3:285-298.
18. Cartegni, L., J. Wang, Z. Zhu, M. Q. Zhang, and A. R. Krainer. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 31:3568-3571.
19. Matysiak-Budnik, T., C. Candalh, C. Dugave, A. Namane, C. Cellier, N. Cerf-Bensussan, and M. Heyman. 2003. Alterations of the intestinal transport and processing of gliadin peptides in celiac disease. *Gastroenterology* 125:696-707.