Investigating the mechanism between cardiovascular risk and haplotypes lacking the *C4B* gene

Doctoral thesis

Zsófia Bánlaki

Semmelweis University
Molecular Medicine Doctoral School

Supervisor:
Dr. György Füst, DSc

Official reviewers:
Dr. Zoltán Pós, PhD
Dr. Krisztián Papp, PhD

President of the examining committee:
Dr. Erzsébet Ligeti, DSc

Members of the examining committee:
Dr. Barna Vásárhelyi, DSc
Dr. József Prechl, PhD

Budapest
2012

Introduction

Cardiovascular diseases are considered as a major public health problem of our time; possible factors in their background are subject to intensive studies. Decades ago our workgroup observed that the so-called C4B*Q0 carrier chromosomes which lack the *C4B* gene (encoding isotype B of complement component 4) are present in a significantly lower proportion of healthy elderly than healthy young individuals both in Hungarian and Icelandic Caucasians. Later it has been revealed that individuals carrying *C4B* gene lacking haplotypes have an increased risk of myocardial infarction, stroke and coronary artery disease and their chance of short-term survival (concerning the time spent in hospital) following infarction is also lower. The above associations proved to be strong especially in the case of smokers.

However, a molecular explanation to the phenomenon is not self-evident. Even though C4B serum levels of C4B*Q0 carriers are relatively low due to their lower than average *C4B* gene dosage, this does not cause a decrease in the functional activity of the complement system. Theoretically, it is also possible that elimination of immune complexes is disturbed because of the relatively low C4 serum levels, inducing inflammatory processes and thus blood vessel damage. Several

indirect indications support this. However, no association between haplotypes lacking *C4A* genes and cardiovascular morbidity has been found, even though C4A isotype proteins presumably play a greater role in the clearance of immune complexes than C4B isotype proteins. Besides – with the only exception of C2 deficiency – not even homozygote complement deficiencies have been associated with increased cardiovascular mortality and morbidity.

Considering all this, our workgroup has assumed that instead of relatively low *C4B* gene dosage itself, it is some genetic variation(s) strongly linked to the lack of *C4B* genes that is in the background of the increased cardiovascular risk of C4B*Q0 carriers. As *C4* genes are located in the MHC region well known for its strong linkage disequilibrium, there is a chance that the lack of *C4B* genes is specifically characteristic to a certain frequent, conserved (ancestral) MHC haplotype. Should such a haplotype indeed exist, its strongly linked remote polymorphisms might jointly contribute to an increased vulnerability to cardiovascular diseases.

According to another alternate, the reason of the above outlined phenomenon is to be looked for in the *C4*-gene-containing RCCX modular structures. RCCX is constituted of four, protein-encoding genes that are always duplicated or deleted together, creating mono-, bi-, tri- or rare

quadrimodular structures. Normally, the monomodular structure contains only functional genes, while in any additional modules only the *C4* genes encode a protein product. *CYP21* genes, of which the active form (*CYP21A2*) encodes the 21-hydroxylase enzyme indispensable for cortisol and aldosterone production, are also located within the RCCX. In the most frequently occurring bimodular RCCX structure, the functional *CYP21A2* gene follows directly after a *C4B* gene. Three cryptic *CYP21* regulation elements have been identified within *C4B* genes, whose sequences in *C4A* genes have not yet been investigated. Besides, given the strong linkage between *C4* and *CYP21* genes it is conceivable that the lack of *C4B* genes is associated with some functional polymorphism of the *CYP21A2* gene. As optimal expression and activity of the 21-hydroxylase enzyme is essential for adequate mobilization of corticosteroid hormones thus for adaptive stress response, disturbances in the latter process are in turn known risk factors for the development of cardiovascular diseases.

Aim

In the course of my research I investigated whether the association between C4B*Q0 haplotypes and cardiovascular risk might be due to genetic linkage. The specific goals were as follows:

- Estimating the composition and frequency of ancestral MHC haplotypes occurring in the Caucasian population, and exploring whether the lack of *C4B* genes is characteristic to one single or several different ancestral, extended MHC haplotypes.
- Detailed characterizing of RCCX structures occurring in the Caucasian population and examining whether the lack of *C4B* genes is associated with a certain RCCX variant.
- Comparing *CYP21* regulatory elements in intron 35 of *C4A* and *C4B* genes, considering that the active *CYP21A2* gene is generally located after a *C4B* gene variant.
- Comparing baseline and stimulated corticosteroid hormone levels of C4B*Q0 carriers and non-carriers, proceeding from the fact that polymorphisms of the *C4* gene neighboring *CYP21A2* gene can have a direct influence on cortisol and aldosterone production, thus on stress response as well.

Methods

Study population

Examination of MHC haplotypes and RCCX structural variants was carried out by means of pedigree analysis. Children with leukemia awaiting bone-marrow transplantation as well as their healthy parents and siblings (n=203) were recruited. The results were compared to the data of a healthy Boston Caucasian population (2675 independent haplotypes).

For comparing intron 35 sequences of *C4A* and *C4B* genes, we analyzed samples from homozygote C4B*Q0 (n=12) and homozygote C4A*Q0 carriers (n=10) gathered from formerly investigated, clinically heterogenic study populations.

Investigation of the relationship between C4B*Q0 carrier state and steroid hormone levels was conducted on individuals (n=76) diagnosed with benign, non-functional adrenal incidentaloma.

Determination of genetic polymorphisms

Genomial DNA was extracted from peripheral blood mononuclear cells using the traditional salting out procedure

described by Miller *et al.* In the case of adrenal incidentaloma samples and most family study samples, already extracted DNA arrived to our laboratory from the collaboration partners.

Determination of *HLA-A*, *-B*, *-DRB1* and *-DQB1* alleles was performed by our collaboration partners using sequence-specific amplification (PCR-SSP) and reverse dot-blot technique.

Of the single nucleotide polymorphisms (SNPs) investigated, *LTA 252A>G* (rs909253), *HSPA1B 1267A>G* (rs1061581), *CFB S>F* (rs641153) and *AGER -374T>A* (rs1800624) were detected with restriction fragment length polymorphism (PCR-RFLP) technique, while the *AGER -429T>C* (rs1800625) and *TNF -308G>A* (rs1800629) variations were examined by allele discrimination based on TaqMan probes.

Copy numbers of different *C4* gene variants (*C4A*, *C4B*, *C4L*, *C4S*) as well as *CYP21A2* active- and *CYP21A1P* pseudogenes were determined using quantitative real-time PCR.

Haplotypic arrangement of long and short *C4A* and *C4B* genes was estimated using allele-specific long-range PCR reactions.

Sequencing of intron 35 of *C4* genes was performed in the Biological Research Centre of Szeged. Chromatograms were evaluated with the CLC DNA Workbench 5.6.1 software.

Analysis of steroid and ACTH hormonal levels

Baseline and stimulated hormone concentrations were determined by our collaboration partners following blood taking between 8:00 and 9:00 a.m. using electrochemiluminescent and radioimmunoassay methods.

ACTH stimulation was performed by the intramuscular administration of 1 mg synthetic ACTH analogue (Cortrosyn), inhibition of cortisol production was achieved by 1 mg dexamethasone taken orally.

Statistical analyses

Statistical analyses were performed by the aid of GraphPadPrism 4.0, SPSS 13.0 and Arlequin 3.1 softwares.

For comparing haplotype patterns, correlation coefficients were determined by the parametric Pearson correlation. Categorical values were compared by the χ^2 test. Hardy-Weinberg equilibrium was determined based on Fisher's exact test.

For investigating steroid hormone levels, comparisons between groups were done by the Mann-Whitney test, paired comparisons were performed with the Wilcoxon matched pair test. Optimal cut points between high and low corticosteroid reactivity values were determined by ROC analysis. Associations between *C4B* gene dosages and hormone levels were calculated by logistic regression analysis. All tests were two-tailed. Differences with a p value <0,05 were considered as significant.

Results

MHC haplotypes occurring in the healthy Caucasian population

In the course of our investigations we determined the polymorphisms *LTA 252A>G, HSPA1B 1267A>G, CFB S>F, AGER -429T>C, AGER -374T>A* and *TNF -308G>A* as well as *C4A* and *C4B* copy numbers in altogether 203 Caucasian individuals with known *HLA-A, -B, -DRB1* and *-DQB1* alleles. Tracing inheritance within families we have assembled 196 genetically independent haplotypes, of which 188 were evaluated after disclosing those affected by recombination.

Seventeen different, internationally recognized ancestral MHC haplotypes were identified in the study population and 13 additional MHC haplotypes identical in all markers examined were present in at least two independent copies. Considering all 188 chromosomes, 89 carried repeatedly occurring MHC haplotypes, while 99 contained individual allele combinations. On the whole, 47.3% of all MHC haplotypes studied were ancestral or repeatedly occurring.The observed haplotype pattern was compared to 2675 chromosomes from a healthy Caucasian population in Boston. Of the 12 haplotypes occurring in at least 1% of the

chromosomes from the Boston area, 11 were found in the Hungarian samples as well. Distribution pattern of these strongly correlated between the two populations (R=0.789, p=0.0023), supporting that sample taking was representative. In the Hungarian population, altogether 30 chromosomes lacking *C4B* genes were identified. This, the majority of haplotypes lacking *C4B* genes occurred in a single copy only.

Variations of the RCCX modular structure

As a next step, copy numbers of long and short *C4* gene variants together with that of *CYP21A2* active- and *CYP21A1P* pseudogenes were determined, as well as haplotypic arrangement of long and short *C4A* and *C4B* genes in the members of the above described families. Two samples were excluded from the studies due to inadequate DNA quality; the rest 184 chromosomes contained 15 different RCCX structural variants and altogether 355 *C4* genes. Through analysis of *C4* genes we confirmed the former assumptions that *C4A* genes are generally long, while *C4B* genes are long and short in quasi equal proportions. We confirmed that the *C4* gene located in the very first (telomeric) module of multimodular structures mostly belongs to the *C4A* variant, but, exceptionally, it can also belong to the *C4B*

variant. To our best knowledge, we are the first to describe structures with a short *C4* gene followed by a long one. However, no prominently frequent haplotype lacking *C4B* genes was found, as of the three such structures two had an almost identical rate of occurrence.

CYP21 regulation sequences located within C4 genes

In order to compare the order of nucleotide bases of the three transcription factor binding sites within *C4A* and *C4B* genes, we investigated 32 *C4A* genes of 12 homozygous C4B*Q0 carriers and 23 *C4B* genes of 10 homozygous C4A*Q0 carriers. No difference in the sequences of any *CYP21* regulatory elements was found between *C4* genes encoding either identical or different isotype proteins. Similarly, intron 35 sequences of *C4* genes were absolutely identical in six MHC Haplotype Project cell lines homozygous for the MHC region with that observed in our samples.

Relationship between C4 gene dosage and steroid hormones

Finally, in a retrospective study we determined *C4B* gene copy numbers in 76 individuals diagnosed with non-functional, benign adrenal incidentaloma and compared the

C4B gene dosages with hormonal data. Concerning the latter, information was available about basal and ACTH induced cortisol, aldosterone, corticosterone and 17-hydroxyprogesterone levels, as well as of basal ACTH and dexamethasone inhibited cortisol levels. Patients were divided into two groups according to their *C4B* gene dosages; 15 individuals with low (0 or 1) *C4B* gene copy numbers were certainly C4B*Q0 carriers, the rest 61 persons had at least 2 *C4B* gene copies. Basal corticosteroid levels were identical in the two patient groups, but basal ACTH concentration was significantly lower in C4B*Q0 carriers than non-carriers. Serum levels of cortisol following ACTH stimulation were significantly higher in C4B*Q0 carriers than in non-carriers; this difference became even more pronounced when quotients of stimulated and basal corticosteroid hormone levels were compared. When ratios of persons exhibiting high or low corticosteroid reactivity following ACTH stimulation were investigated in terms of C4B*Q0 carrier state (reactivity: quotients of induced/ basal corticosteroid levels), we found that within the C4B*Q0 carrier group, ratios of individuals with high cortisol, aldosterone and corticosterone reactivity were significantly higher compared to non C4B*Q0 carriers. In the case of cortisol and aldosterone these differences remained significant even after adjusting for confounding

variables. Administration of dexamethasone resulted in a less effective inhibition of cortisol secretion in C4B*Q0 carriers. (All these associations were even more prominent in the case of homozygous C4B*Q0 carriers, however, this subgroup could not be evaluated separately by statistical analysis due to the small number of cases.)

Conclusions

In summary, the most important novel findings are as follows:

- We ascertained that lack of *C4B* genes can be observed in several ancestral MHC haplotypes and RCCX structural variants. Thus, the possibility can be rejected that the relationship between C4B*Q0 carrier state and increased risk to cardiovascular diseases is due to joint effect of remote polymorphisms.

- We demonstrated that nearly 50% of all MHC haplotypes present in the Caucasian population occur repeatedly, and that remote, independent Caucasian populations exhibit strongly similar haplotype patterns.

- Through detailed characterizing of RCCX structural variants occurring in the Caucasian population, we gave evidence of the assumption that while *C4A* genes are generally long, virtually half of all *C4B* genes are long and half are short. In addition, we demonstrated that it is generally a long *C4A* gene that is located in the most telomeric module of multimodular structures.

- Analyzing the *CYP21* regulation sequences within *C4* genes, we could exclude the possibility that increased cardiovascular risk of C4B*Q0 carriers is caused by some difference in the sequences of the transcription factor binding sites within *C4A* and *C4B* genes.

- We examined the relationship between *C4B* gene dosage and some steroid hormones as well as ACTH in patients with non-functional, benign adrenal incidentaloma. Significant differences in basal ACTH levels, as well as in inducibility and inhibitability of some steroid hormones were observed between C4B*Q0 carriers and non-carriers. The results imply that lack of *C4B* genes is linked to some variation related to the *CYP21A2* gene, which leads to a higher inducible gene expression or a modified enzyme activity, thus a more efficient reactivity of the hypothalamus-pituitary-adrenal (HPA) axis. C4B*Q0 carriers presumably show an above average cortisol production during stress response, a known risk factor for cardiovascular diseases. In the case of smokers, effects of inherited and environmental effects might be cumulative, as nicotine also enhances reactivity of the HPA axis. Accordingly, the present observations may provide an

explanation to the increased cardiovascular morbidity and mortality seen in C4B*Q0 carriers.

List of own publications

Publications related to the thesis

Szilagyi A, Banlaki Z, Pozsonyi E, Yunis EJ, Awdeh ZL, Hosso A, Rajczy K, Larsen CE, Fici DA, Alper CA, Fust G. *Frequent occurrence of conserved extended haplotypes (CEHs) in two Caucasian populations.* Mol Immunol. 2010 Jun;47(10):1899-904; IF: 2,916

Banlaki Z, Raizer G, Acs B, Majnik J, Doleschall M, Szilagyi A, Racz K, Fust G, Patocs A. *ACTH-induced cortisol release is related to the copy number of the C4B gene encoding the fourth component of complement in patients with adrenal incidentaloma.* Clin Endocrinol (Oxf). 2012 Apr;76(4):478-84; IF:3,23

Banlaki Z, Doleschall M, Rajczy K, Fust G, Szilagyi A. *Fine-tuned characterization of RCCX copy number variants and their relation to extended MHC haplotypes.* (manuscript accepted in Genes & Immunity)

Other publications

Blasko B, Banlaki Z, Gyapay G, Pozsonyi E, Sasvari-Szekely M, Rajczy K, Fust G, Szilagyi A. *Linkage analysis of the C4A/C4B copy number variation and polymorphisms of the adjacent steroid 21-hydroxylase gene in a healthy population.* Mol Immunol. 2009 Aug;46(13):2623-9; IF: 3,202

Pozsonyi E, Gyorgy B, Berki T, Banlaki Z, Buzas E, Rajczy K, Hosso A, Prohaszka Z, Szilagyi A, Cervenak L, Fust G. *HLA-association of serum levels of natural antibodies.* Mol Immunol. 2009 Apr;46(7):1416-23; IF: 3,202