Association of ERAP1 Allelic Variants with Risk of Ankylosing Spondylitis

I. V. Zvyagin1#, V. Yu. Dorodykh1#, I. Z. Mamedov1, D. B. Staroverov1, A. G. Bochkova2, D. V. Rebrikov3,4, Y. B. Lebedev1
1Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences
2Institute of Rheumatology, Russian Academy of Medical Sciences
3Vavilov Institute of General Genetics, Russian Academy of Sciences
4DNA-Technology JSC
# Authors contributed equally to this work
*E-mail: izvyagin@gmail.com
Received 03.09.2010

Abstract: Ankylosing spondylitis (AS) belongs to a group of autoimmune diseases affecting the axial skeleton. Besides the hla-b*27 allele, several other human genes that control the variety processes of immune homeostasis are considered to be associated with AS manifestation in different human populations. Among strongly associated non-MHC genes erap1 encoding the endoplasmic reticulum aminopeptidase 1 isoform was recently identified by single nucleotide polymorphisms (SNPs) meta analysis. In our study we inspected the genetic association of five non-synonymous coding SNPs from erap1 with AS in Caucasians. We implemented the SSP-PCR system for precise genotyping of 87 hla-b*27 positive AS patients and 77 hla-b*27 healthy donors from the Russian population. Considerable differences in allele's frequencies within patients vs control cohort were shown for 3 of 5 SNPs under investigation. Using the EM-algorithm we reconstructed 3-marker haplotypes that distinguish with high probability two cohorts due to differences in the haplotypes frequencies. In such a way both the sensitive, CCT, haplotype and the protective, TTC, one were predicted. To verify the calculation we determined genuine frequencies of 5-marker haplotypes in AS cohort by haplotyping of individual cDNA samples using improved SSP-PCR primer set. We demonstrated that the frequencies of in silica reconstructed haplotypes and the frequencies of experimentally detected haplotypes are in a good agreement. Frequency of the risk haplotype CCT (rs17482078/10050860/2287987) detected within AS cohort reaches 88%, as well as the frequency calculated by EM-algorithm.

Keywords: aminopeptidase ERAP1, human single nucleotide polymorphism, allele-specific PCR, human haplotypes, ankylosing spondylitis.

Abbreviations: AS - ankylosing spondylitis, ER - endoplasmic reticulum, MHC - major histocompatibility complex, TNF - tumor necrosis factor, TNFR - tumor necrosis factor receptor, IL - interleukin.

Introduction: Ankylosing spondylitis (AS) is one of autoimmune diseases belonging to spondyloarthropathy. The disease is characterized by an inflammation of axial skeleton joints and a resulting formation of ossified regions in the joints. The first symptoms of the disease appear at the age of 20-25, and then it slowly progresses. Recently several genetic loci have been found to be associated with the disease. The products of these genes are involved in various stages of the immune response. According to results of twins studies the concordance in monozygote twins varies from 26 to 60%, whereas for dizygote twins the concordance varies from 4 to 20%. This evidence underlines the importance of the genetic background for AS development [1-3].

The strongest association was observed for the hla B*27 allelic variant of the MHC-I gene. Approximately 90% of patients of European descent suffering from AS bear this allele; however, only 5% of Europeans carrying the allele are affected by the disease [4]. Large-scale full-genome comparisons of diagnosed patients cohorts with healthy people show that several other SNPs in non-MHC-I genomic loci are also associated with the disease. These loci have a weaker association with the risk of developing AS. Among the non-MHC loci, the strongest association was observed for the erap1 aminopeptidase gene [5, 6].

ERAP1 aminopeptidase acts in a variety of immune response stages. The difference of ERAP1 functioning probably exists due to its different intracellular localizations. ERAP1 is mainly located in the endoplasmic reticulum (ER) in human and murine cells. However, the protein also might be expressed on the outer surface of the cytoplasmic membrane or be secreted into the extracellular environment [7]. One of the functions of ERAP1 is the N-terminal proteolysis of peptides, which are formed during the proteasome-dependent degradation of cellular proteins. The formed peptides are then exposed on the cell’s surface by MHC-I molecules. Also ERAP1 might be involved in the formation of the solu-
ble form of the type I tumor necrosis factor receptor (TNFR-1), along with the soluble forms of interleukin 1 and 6 (IL) receptors. ERAP1 thus plays an important role in the regulation of the immune response [8-10].

Studies of cohorts of different ethnicities and varying genetic homogeneities in terms of the hla-b allele content revealed in identification of several non-synonymous SNPs in the coding region of the erap1 gene which are associated with an increased risk of AS [11-15]. The sets of AS-associated polymorphisms detected in these studies are different. It is probably due to the genetic properties of the cohorts under investigation.

The presence of amino acid substitutions and/or their combinations in ERAP1 variants may result in changes in the aminopeptidase activity. These substitutions may also result in the deregulation of peptide processing and control over immune cells via TNF or IL-1 and -6. The putative ability of ERAP1 variants bearing substitutions to form inappropriate peptides on MHC-I is considered as the aminopeptidase impact on AS progression. The hypothesis assumes a misrecognition of the presented peptides as alien signals, and it leads, in turn, to an increase of empty MHC-I or its dimmers consisting exclusively of heavy chains on cell’s surface. Finally these processes could result in the formation of nonspecific immune responses against cells that expose the misfolded MHC-I [16]. Dysfunction of ERAP1 activity aimed at IL receptors could be another explanation of the mutant aminopeptidase involvement in AS progression.

In our research, we have determined the frequencies of erap1 allele variants and corresponding genotypes for 5 non-synonymous SNPs in cohorts of AS patients and healthy individuals from the Russian population. Both the AS patients cohort and healthy group consist of hla-B*27 positive individuals. Markers rs2287987 (Met349Val), rs30187 (Lys528Arg), rs17482078 (Arg725Gln), and rs27044 (Glu730Gln) were selected for the genotyping. Each of the markers have been previously associated with AS alleles [11-15]. The location of these non-synonymous substitutions within the coding region of the erap1 gene allows us to imply the existence of functional effects of the corresponding amino acid substitutions.

We used the genotyping results for haplotypes reconstruction and for association analysis of the erap1 allele variants with AS.

MATERIALS AND METHODS

Genomic DNA and RNA samples

The samples of genomic DNA were purified from the peripheral blood of 77 hla-B*27 positive healthy donors aged 21-63, and of 84 hla-B*27 positive patients with the axial form of AS. Genomic DNAs were extracted from the peripheral blood mononuclear cells using a Diatom Prep 100 kit (Izogen Laboratory, Moscow) according to the manufacturer’s protocol.

Samples of total RNA were extracted from peripheral blood mononuclear cells using TRIzol® Reagent (Invitrogen Lifetechnologies, USA), according to the manufacturer’s protocol.

cDNA synthesis

1-2 micrograms of total RNA were used for cDNA synthesis. The first strands of cDNAs were synthesized using the “Mint-Universal cDNA synthesis kit” (Eurogen, Russia) according to the manufacturer’s protocol.

Geno- and haplotyping

Genotyping was performed using a set of allele-specific PCR primers (Table 1). Each pair consisted of an allele-specific primer complementary to the allele variant and a primer specific to the nearest intron. The presence of the distinct allele in a sample was determined by the presence/absence of the specific PCR products. Allele-specific PCR on cDNA templates using subsets of four allele-specific primers corresponding to each pairs of SNPs was performed for haplotyping.

To enhance the allele discrimination ability a mismatch was introduced in the third position of the primer 3’-end of each of the allele-specific primers. Geno- and haplotyping results have been independently reproduced at least twice for each sample. Additionally concordance of the geno- and haplotyping results was inspected.

“HS Taq-DNA-polymerase kit” (Eurogen, Russia) was used for PCR amplification; the final volume of the reaction mixture was 15 microliters; the annealing tempera-
ture was set at 65°C; 30 and 33 cycles of PCR were implemented to amplify DNAs and cDNAs, respectively.

**Statistical analysis**

Haploview 4.1 (http://www.broadinstitute.org/haplovie/haploview) and Genetic Data Analysis (Lewis, P. O., and Zaykin, D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. http://lewis.eeb.uconn.edu/lewishome/software.html) software was used to analyze the results of geno- and haplotyping.

Hardy-Weinberg equilibriums for erap1 alleles were tested using Fischer exact test, as well as the exact test described by G. Abecasis and J. Wigginton [17].

The statistical significance of the differences in the allele frequencies were analyzed using the χ2 test with Yate’s correction. The value of the risk factor (OR) and the confidence interval (CI 95%) were calculated using an online calculator: www.openepi.com.

Expectation-maximization algorithm (EM-algorithm) of Haploview 4.1 software was used to detect linkage disequilibrium (LD) of tested SNPs and to calculate the frequencies of reconstructed haplotypes. Strong LD was defined as D’ ≥ 0.8 CI95% 0.7–0.99 [18].

Bayesian algorithm of PHASE 2.1 software was implemented to confirm the frequencies of reconstructed haplotypes.

**RESULTS AND DISCUSSION**

The frequencies of minor allelic variants of the erap1 gene are decreased in AS patients

By using a set of allele-specific primers (Table 1, Fig. 1), we identified the frequencies of 10 allelic variants of the erap1 gene and the corresponding genotypes in hla-B27 positive cohorts of 84 AS patients and 77 healthy donors. The allelic variants of the erap1 gene were distinguished by the set of non-synonymous SNPs: rs2287987 (C/T, Met349Val), rs30187 (C/T, Lys528Arg), rs10050860 (C/T, Asp575Asn), rs17482078 (C/T, Arg725Gln), and rs27044 (C/G, Glu730Gln). These polymorphysms code the amino acid substitutions in ERAP1. The protein variants bearing the substitutions might be functionally different and therefore play a positive or protective role in AS progression. Association of each of 5 markers with ankylosing spondylitis was published recently by several authors [11-15]. Two of the five SNPs are “located” near the putative active sites of the aminopeptidase; rs2287987 – is located near the Zn2+ ion-binding site [11], while rs30187 is located near the putative substrate-binding pocket.

Distribution of 5 allele pairs in both the AS patients and the control cohorts showed no significant deviations from the Hardy-Weinberg Equilibrium.

Analysis of the genotyping results showed that the frequencies of three minor alleles of rs2287987 (p< 0.002, OR = 0.35), rs10050860 (p < 0.004, OR = 0.39), and rs17482078 (p < 0.05, OR = 0.52) markers are significantly lower among AS patients compared to healthy donors in the Russian population (Table 2). An associative analysis test shows that three allelic variants...
of *erap1* are associated with an increased risk of AS (rs2287987 [T] (OR = 2.86), rs10050860 [C] (OR = 2.54), and rs17482078 [C] (OR = 1.91). However, our results also show a lack of a significant association of rs30187 and rs27044 polymorphisms with AS in the Russian population.

Decreased frequencies of minor alleles of rs2287987, rs10050860, and rs17482078 markers are accompanied by a decreased number of heterozygous individuals and an increased number of homozygous for the major allele individuals in the AS patient group as compared to the control group. Heterozygocity for the rs30187 and rs27044 markers, which are not associated with AS in the Russian population, increases in AS patients, although the minor allele frequencies do not differ significantly AS patients from healthy individuals.

In general our data are in a good agreement with those reported by other researchers. Several studies have shown that all five of the inspected markers are associated with the risk of AS [5, 6, 11], although the cohorts of AS patients and healthy individuals included a considerable number of hla-B27-negative donors. Two other studies [14, 15] were performed on cohorts consisting of hla-B27-positive donors exclusively or mostly (77-85%). No association with AS was found for the rs30187 and rs17482078 markers, while rs2287987, rs10050860, and rs27044 proved to be associated with the AS risk [14]. In another study [15] it was demonstrated that allelic variants for the markers rs10050860 and rs30187, but not rs27044, were associated with the AS risk.

Statistically significant association of rs27044 and rs30187 but not of rs2287987, rs10050860, and rs17482078 markers with AS was reported for the Korean population [12].

The differences between the results of the associative analyses are most likely due to considerable differences between the studied populations in *erap1* allele frequencies. Another reason for the incongruence between the cited results might be the different presentation of hla-B27 positive individuals in the cohorts under comparison. In our study, special attention was focused on the genetic homogeneity of the cohorts in terms of hla-b27 carriership and on the uniformity of clinical manifestations of the disease. Such considerations helped us avoid risk of false positive associations.

### Haplotype reconstruction based on obtained genotype data

The genotyping results for both cohorts were used to calculate haplotype frequencies. We searched for haplotypes whose frequencies increased statistically significantly in AS patients and could, thus, be associated with AS risk. We used Haplovie 4.1 software to identify the linkage disequilibrium between the pairs of the studied markers. Polymorphisms rs2287987, rs10050860, and rs17482078 proved to be in statistically

### Table 2. Frequencies of allelic variants and genotypes for the studied genetic markers in two cohorts

| Marker, polymorphic nucleotide | AS patients, n = 84 | Healthy donors, n = 77 |  |  |
|------------------------------|--------------------|------------------------|---|---|
|                              | MAF* [nucleotide in minor allele] | Number of genotypes (frequency) | MAF* [nucleotide in minor allele] | Number of genotypes (frequency) | P | OR (95%CI) |
| rs2287987 [C/T]              | 0.09 [C]          | CC 1 (0.01) CT 13 (0.15) TT 70 (0.84) | 0.22 [C] | CC 3 (0.04) CT 26 (0.36) TT 44 (0.60) | 0.002 | 0.35 (0.18–0.68) |
| rs30187 [C/T]               | 0.37 [T]          | CC 29 (0.35) CT 48 (0.57) TT 7 (0.08) | 0.33 [T] | CC 35 (0.45) CT 33 (0.43) TT 9 (0.12) | 0.55 | 1.18 (0.75–1.87) |
| rs10050860 [C/T]            | 0.11 [T]          | CC 66 (0.78) CT 18 (0.21) TT - | 0.23 [T] | CC 46 (0.60) CT 26 (0.34) TT 5 (0.06) | 0.004 | 0.39 (0.21–0.73) |
| rs17482078 [C/T]            | 0.13 [T]          | CC 65 (0.77) CT 17 (0.20) TT 2 (0.02) | 0.21 [T] | CC 47 (0.61) CT 27 (0.35) TT 3 (0.04) | 0.046 | 0.52 (0.29–0.95) |
| rs27044 [C/G]               | 0.33 [G]          | GG 7 (0.08) GC 42 (0.5) CC 35 (0.42) | 0.29 [G] | GG 9 (0.12) GC 26 (0.34) CC 42 (0.55) | 0.47 | 1.23 (0.76–1.97) |

*MAF – minor allele frequency*
Table 3. Calculated frequencies of reconstructed haplotypes including markers rs17482078/10050860/2287987, associated with risk of AS in Russian population.

| Haplotype | Frequency in AS patients, \( n = 84 \) | Frequency in healthy donors, \( n = 77 \) | \( p \) | OR (95%CI) |
|-----------|----------------------------------------|----------------------------------------|----------|------------|
| CCT       | 0.86                                   | 0.75                                   | 0.026    | 1.96 (1.12–3.46) |
| TTC       | 0.08                                   | 0.2                                    | 0.003    | 0.33 (0.17–0.67) |

significant pairwise linkage disequilibrium \( D' \geq 0.85, 95\%CI 0.76–0.99, \) according to [18]) and were grouped into a single block. The theoretical haplotype frequencies in both AS patients and healthy cohorts were calculated for markers belonging to this block. Haplotypes with theoretical frequencies below 5% were discounted from further analysis.

The CCT haplotype was found to be statistically significantly associated with AS risk \( (p < 0.03, \text{OR} = 1.96) \). The calculated frequency of the TTC haplotype was to be lower in AS patients as compared to the healthy individuals \( (p < 0.003, \text{OR} = 0.33) \) (Table 3).

**Haplotyping of AS patients in Russian population**

Using paired combinations of allele-specific primers and cDNA samples we identified erap1 gene haplotypes in the same cohort of AS patients. Table 4 lists the frequencies of the identified 5-marker haplotypes in the patients cohort. Haplotypes consisting of only 3 markers associated with AS in the Russian population were listed specifically. The table does not list haplotypes with frequencies of less than 5%, with the exception of TTC haplotype, which was determined as a putative protective haplotype based on theoretical calculations. The determined frequency of the putative risk haplotype CCT among AS patients was approximately 88%. The percentage is in agreement with the calculated frequency value. The frequency of the putative protective TTC haplotype is approximately 1%, while the calculated value is approximately 8% (Table 3). This difference could be attributed to the insufficient size of the AS patients cohort for a statistically representative identification of the actual frequency of the protective haplotype.

The identified risk haplotype includes mostly those SNP variants that were demonstrated to be risk-associated in other cohorts of Europeans (Table 5).

Notably recent research has yielded controversial data regarding the risk-associated haplotypes for the markers rs27044 and rs30187 in different populations. Maksymowych et al. [14] reported a risk association for the C-allele (marker rs27044), while a study by Pazar et al. [15] showed that both identified risk haplotypes include G-allele of the same marker. Moreover, the two risk haplotypes identified in the cited study [15] are in variant for the rs30187 marker. These results, as well as the controversial data on the association of erap1 gene allele variants for markers rs27044 and rs30187 with AS risk, may indicate that the amino acid substitutions encoded by these polymorphisms do not affect the activity of the aminopeptidase, nor its role in AS. The described associations may be due to the heterogeneity of the patient cohorts or/and genetic linkage of these polymorphisms with other non-synonymous substitutions which result in changes of aminopeptidase function that is essential for AS.

The published results of the few in vitro studies indicate that a Lys528Arg substitution, encoded by the SNP rs30187, has a strong effect on ERAP1 aminopeptidase activity. Goto et al. [19] demonstrated that the ERAP1 variant, which has Arg in position 528, displays low peptidase activity to synthetic substrates as well to the natural substrates (angiotensin II and kallidin). In the cited study [19] it was also shown that substitution Gln730Glu, encoded by the polymorphism rs27044, and substitution Asp575Asn, encoded by the polymorphism rs10050860, have almost no effect on peptidase activity to synthetic and natural substrates. Transient ex-

Table 4. Distribution of identified haplotypes in a AS patient cohort of the Russian population.

| Haplotype* | Number of haplotypes in AS patient (frequency), \( n = 69 \) |
|------------|----------------------------------------------------------|
| rs27044/17482078/10050860/30187/2287987 | |
| CCCCT | 70 (0.51) |
| GGCTT | 36 (0.26) |
| GGCCT | 8 (0.06) |
| CCCTT | 7 (0.05) |
| GTTGT | 7 (0.05) |
| CTTCC | 2 (0.01) |
| rs17482078/10050860/2287987 | |
| CCT | 121 (0.88) |
| TTT | 7 (0.05) |
| TTC | 2 (0.01) |

*Nucleotides which are a part of the predicted protective and risk –associated haplotypes are underlined and bold.*
pression of erap1 in a melanoma cell line demonstrated that the amino acid substitution Met349Val (polymorphism rs2287987) increased ERAP1 peptidase activity to a synthetic substrate [20]. Comparison of amino acid substitutions characterized in the previously cited articles with our associative analysis data, and other similar findings by other authors (Table 5), suggests that some of the AS-associated non-synonymous substitutions can alter ERAP1 peptidase activity. However, most of the disease-associated alleles are likely to have other functional manifestations, since the corresponding amino acid substitutions do not affect the enzyme’s activity on synthetic substrates.

Seemingly disagreement between the results of functional tests and the associative analysis can be overcomed by identification of haplotypes specific for healthy cohorts of patients. In our study, we have identified risk-associated CCT and protective TTC haplotypes for the markers rs17482078/10050860/2287987. The frequency of both haplotypes is statistically significantly increased in the corresponding cohort (Table 4). One of the substitutions encoded by the protective haplotype (Met349Val) increases ERAP1 activity on several substrates [20]. This fact suggests that increased ERAP1 activity might prevent AS. The protective effect of the TTC haplotype may be due to the activity of ERAP1 towards receptors of proinflammatory cytokines or due to the correct presentation of peptides on MHC-I molecules. In contrast, the non-synonymous substitutions included in the risk haplotype CCT could lead to increased levels of membrane-bound forms of TNFR-1 and receptors for IL-6 and -1.

Considering that AS is a multigene disease, further research for the ERAP1 role in AS should be focused on both identification of the function features of different erap1 allele variants and study of the putative interactions between the enzyme and the products of other genes associated with the disease. These interactions may indeed be affected by the non-synonymous substitutions in the risk haplotype of the erap1 gene which we have identified in the Russian population.

This study was supported by the Molecular and Cellular Biology program of Presidium of RAS and by grants from the Russian Foundation for Basic Research (№ 10-04-01771-a and 11-04-01585-a).

**REFERENCES**

1. van der Linden S.M., Valkenburg H.A., de Jongh B.M., Cats A. // Arthritis Rheum. 1984. V. 27. № 3. P. 241–249.
2. Pedersen O.B., Svedensen A.J., Ejstrup L., et al. // Scand. J. Rheumatol. 2008. V. 37. № 2. P. 120–126.
3. Brown M.A., Laval S.H., Brophy S., Calin A. // Ann. Rheum. Dis. 2000. V. 59. № 11. P. 883–886.
4. Reveille J.D., Maganti R.A. // Adv. Exp. Med. Biol. 2009. V. 649. P. 159–176.
5. The Australo-Anglo-American Spondyloarthritis Consortium (TASC) // Nature Genetics. 2010. V. 42. № 2. P. 123–127.
6. Wellcome Trust Case Control Consortium & The Australo-Anglo-American Spondylitis Consortium (TASC) // Nature Genetics. 2007. V. 39. № 11. P. 1329–1337.
7. Toshihiro T., Hattori A., Shinako M., et al. // J. Biol. Chem. 2003. V. 278. № 34. P. 32275–32283.
8. Cui X., Hawari F., Alsaaity S., et al. // J. Clin. Invest. 2002. V. 110. № 4. P. 515–526.
9. Cui X., Rouhani N.F., Hawari F., et al. // J. Immunol. 2003. V. 171. № 12. P. 6814–6819.
10. Cui X., Rouhani N.F., Hawari F., Levine J.S. // J. Biol. Chem. 2003. V. 278. P. 28677–28685.
11. Harvey D., Pointon J.J., Evans M.D., et al. // Hum. Mol. Genet. 2009. V. 18. № 21. P. 4204–4212.
12. Choi C.-B., Kim T.-H., Jun J.-B., et al. // Ann. Rheum. Dis. 2010. V. 69. № 3. P. 582–584.
13. Tsui F.W.L., Haroon N., Reveille D.J., et al. // Ann. Rheum. Dis. 2010. V. 69. № 4. P. 733–736.
14. Maksymowych W.P., Inman R.D., Gladman D.D., et al. // Arthr. & Rheum. 2009. V. 60. № 5. P. 1317–1323.
15. Pazar B., Safrani E., Gergely P., et al. // J. Rheum. 2010. V. 37. № 2. P. 379–384.
16. Haroon N., Inman D.R. // Nat. Rev. Rheum. 2010. Advanced online publication.
17. Wigginton J., Cutler D., Abecasis G.A. // Am. J. Hum. Genet. 2005. V. 76. № 5. P. 837–838.
18. Gabriel S.B., Schaffner S.F., Nguyen H., et al. // Science. 2002. V. 296. № 5576. P. 2225–2229.
19. Goto Y., Hattori A., Ishii Y., Tsujimoto M. // FEBS Lett. 2006. V. 580. P. 1833–1838.
20. Kamphausen E., Kellert C., Abbas T., et al. // Cancer Immun. Immunother. 2010. V. 59. № 8. P. 1273–1284.

---

**Table 5. Allelic variants of erap1, associated with increased risk of AS development in various populations**

| rs27044 | rs17482078 | rs10050860 | rs30187 | rs2287987 | reference |
|---------|------------|------------|---------|-----------|-----------|
| C       |            | C          | T       | -         | [14]      |
| G       | C          | C          | T       | T         | [15]      |
| G       | C          | C          | C       | T         | [12]      |
| G       | C          | C          | T       | -         | [13]      |
| G       | -          | -          | T       | -         | this study|

---

**Research Articles**

- Cui X., Rouhani N.F., Hawari F., et al. // J. Immunology. 2010. V. 171. № 12. Р. 6814–6819.
- Cui X., Rouhani N.F., Hawari F., Levine J.S. // J. Biol. Chem. 2003. V. 278. P. 28677–28685.