

HLA Typing by SNP Genotyping

A new Method for HLA Typing at High-throughput Level

Inaugural-Dissertation

zur

Erlangung des Doktorgrades am
Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
Ramón Kucharzak
aus Rostock-Warnemünde, Deutschland

Berlin
2006

1. Gutachter: Frau Professorin Petra Knaus
2. Gutachter: Herr Professor Hans Lehrach

Tag der Disputation: 27. April 2007

Die praktischen Arbeiten für diese Dissertation wurden am Centre National de Génotypage (Evry, Frankreich) in der Zeit von Januar 2001 bis Juli 2005 in der Arbeitsgruppe von Dr. Ivo Gut durchgeführt. Der Verfasser versichert, die Arbeit eigenständig durchgeführt und alle Hilfsmittel angegeben zu haben.

Berlin, den 18.08.2006

ABSTRACT	13
1 INTRODUCTION	15
1.1 General	15
1.2 Human Leukocyte Antigens (HLA)	17
1.2.1 Molecular structure and function of HLA molecules	17
1.2.2 HLA class II	18
1.2.3 HLA class I	19
1.2.4 MHC haplotypes	21
1.2.5 Nomenclature of the HLA antigens	22
1.3 History of HLA and HLA Typing	23
1.4 Current Methods for HLA Typing	24
1.4.1 Serological typing	25
1.4.2 Typing by hybridization to sequence-specific oligonucleotide probes	26
1.4.3 Reverse hybridization with immobilized sequence-specific oligonucleotide probe arrays – Reverse Line Strip Blot	29
1.4.4 Typing by differential primer extension and sequence-specific amplification	29
1.4.5 Typing based on the gel mobility of the PCR products	30
1.4.6 Sequence-based typing	32
1.5 New Methods for HLA Typing	32
1.5.1 HLA allele detection based on DNA array technology	32
1.5.2 Microsatellite genotyping and HLA alleles	33
1.5.3 HLA typing by TaqMan®	34
1.5.4 Pyrosequencing™ and HLA typing	37
1.6 The GOOD Assay a potential Method for HLA Screening	39
1.6.1 MALDI-TOF-MS technology and polymorphism analysis	41
1.7 Technologies in MADO	42
1.8 Objectives of this Thesis	43
1.8.1 Outlook on results	44
2 MATERIALS AND METHODS	45
2.1 Materials	45
2.2 Explanatory Remarks about DNA samples	47

2.3 Nomenclature of PCR- and Extension-Primers	47
2.4 Generic PCR Amplifications of Target Loci	48
2.5 Gel Electrophoresis	50
2.6 Exonuclease I (EXO I) and Shrimp Alkaline Phosphatase (SAP) Digest	51
2.7 Extension Reaction	51
2.8 Preparation of Extension Primers	54
2.8.1 Charge tagging of extension primers	54
2.8.2 Pooling of Extension Primers	56
2.9 PDE Digest	62
2.10 Alkylation	62
2.11 Target Preparation and MALDI-TOF-MS Analysis	63
2.12 BasePlate Liquid Handler	63
2.13 Sample organisation for high-throughput genotyping	63
2.14 Software developed and used for HLA typing	65
2.14.1 "Genalys" – Sequence analysis	65
2.14.2 "Trans_numerique" – transforming microhaplotypes into number codes	65
2.14.3 "select" – Marker selection software	68
2.14.4 "Helixir" – reading raw data	69
2.14.5 "HLAfamilies" – transforming microhaplotypes into HLA types	71
3 RESULTS	77
3.1 Explanatory Remarks and Definitions	77
3.2 Sequence Analysis and Marker Selection	80
3.2.1 Marker selection and description for screening of the alleles of the genes HLA-A and HLA-B	84
3.2.2 Marker selection and description for screening of the alleles of the gene HLA-DRB1	97
3.3 GOOD Assay for HLA Screening	104
3.3.1 Evaluation of molecular biology	105
3.3.1.1 Generic PCR amplifications	105
3.3.1.2 Extension reactions	106

3.4 Microhaplotype-based Pre-screening of CEPH Families	107
3.4.1 Pre-screening of HLA-DRB1	107
3.4.2 Pre-screening of HLA-A and HLA-B types	118
3.5 Microhaplotype-based Pre-screening at High throughput	124
3.5.1 HLA type estimation with missing data points	126
4 DISCUSSION	129
4.1 Comparison of microhaplotyping by the GOOD assay and by other SNP genotyping methods	134
4.2 Comparison of HLA typing by SBT and microhaplotyping	135
4.3 Software	137
4.4 Perspectives	139
4.5 A few comments to the cost of HLA typing by microhaplo-typing	140
5 APPENDIX	143
5.1 Zusammenfassung	143
5.2 Abbreviations	145
5.3 Output File of Marker Selection Software “select”	148
5.4 Allele Frequencies of HLA-A and HLA-B	148
5.5 High-throughput HLA Typing Results	150
5.6 Look-up Table or all known HLA-DRB Alleles and there Microhaplotype Combination (Status April 2006)	169
5.7 Look-up Table or all known HLA-A Alleles and there Microhaplotype Combination (Status April 2006)	185
5.8 Look-up Table or all known HLA-B Alleles and there Microhaplotype Combination (Status April 2006)	200
5.9 List of Publications	224
5.10 Acknowledgements	224

Abstract

Allogeneic stem cell transplantation has become an effective treatment for a number of haematological diseases, such as malignant diseases (acute and chronic leukaemia, myelodysplasia), aplastic anaemia, immune-deficiencies and inherited metabolic disorders. As a matter of fact, the best donor in terms of histocompatibility is a familial HLA identical donor, e.g. HLA-identical sibling. However, most of the patients do not have a suitable familial donor and grafts from unrelated donors must be used. In recent years stem cells collected from blood and especially umbilical cord blood cells have been used, but in most of the cases the source of grafted stem cells is bone marrow. A main problem is to find a donor compatible for each patient. This is difficult to achieve due to the large genetic heterogeneity of the Major Histocompatibility Complex (MHC) immunogenetic system and its variable distribution in populations. The human MHC are the human leukocyte antigens (HLA). Large numbers of potential volunteer bone marrow donors are needed. Today more than eight million donors are registered world wide, more than two-thirds of them in 24 European registries. Improving registry efficiency is a permanent aim in order to increase the likelihood of finding compatible donors, with acceptable economic conditions. Therefore in 2001 an EU-funded project was initiated to find a new strategy to increase the efficiency paired with a reduction of the cost.

This new strategy included to perform a screen for the present HLA types of the potential donors. Therefore the alleles of the transplantation relevant HLA genes, HLA-A, HLA-B and HLA-DRB1, are divided in two groups, rare and frequent alleles. Frequent alleles are those of which the 15 most frequent HLA haplotypes are made up of. Only donors with a rare HLA type will undergo a high resolution typing procedure. To perform the HLA type screening a robust, reliable and easy-to-use method for HLA type estimation at high throughput level was required.

Within this thesis a DNA analysis-based method was developed that allows a HLA type estimation of given DNA samples at high throughput. The HLA type identification is based on microhaplotype genotyping. Microhaplotypes in this term are short DNA sequences of 4 to 5 bases length. The informativity of a microhaplotype is significantly higher than the sum of informativity of the single base polymorphisms (SNP) the microhaplotype is made up of. A set of microhaplotype markers was selected which allows a medium to high resolution

identification of the given HLA types. As the technique to genotype these microhaplotypes the GOOD assay, a purification free genotyping method with mass spectrometric detection and with a high degree of automation, was chosen.

In the framework of this thesis new software tools for microhaplotype selection and data analysis were developed, which permit a highly flexible setup of this method for other highly polymorphic regions in the genome.

After proof of principle with DNA samples from the CEPH panel, a HLA type screen at high throughput level with randomly selected DNA samples from the registries was successfully performed.