Identifying Major Histocompatibility Complex Supertypes

Pingping Guan¹, Irini A. Doytchinova², and Darren R. Flower²

¹ John Innes Centre, Norwich, NR4 7UH, UK
² The Jenner Institute, University of Oxford, Compton, Berkshire, RG20 7NN,
  UKDarren.flower@jenny.ac.uk

Summary. Human leukocyte antigen (HLA) recognizes antigenic fragments and presents them to T cells. HLA is polymorphic. There are over 2000 different HLA alleles at present and the number is constantly increasing. However, antigen binding studies are limited to a small proportion of these alleles; the binding specificities of most alleles are unknown. Several research groups have attempted to partition different HLA alleles into groups. In this chapter previous classifications are reviewed and we present two chemometric approaches to classifying class I HLA alleles. The program GRID is used to calculate interaction energy between protein molecules and defined chemical probes. These interaction energy values are imported into another program GOLPE and used for principal component analysis (PCA) calculation, which groups HLA alleles into supertypes. Amino acids that are involved in the classification are displayed in the loading plots of the PCA model. Another method, hierarchical clustering based on comparative molecular similarity indices (CoMSIA) is also applied to classify HLA alleles and the results are compared with those of the PCA models.

10.1 Introduction

Major histocompatibility complex (MHC) molecules are polymorphic membrane glycoproteins [Zinkernagel 1986]. Human MHCs are also called human leukocyte antigen, often abbreviated to HLA [Clark & Forman 1984]. There are two classes of HLA, class I and class II. Class I HLA is present on most nucleated cells, including the surfaces of lymphocytes, which have 1000 to 10000 HLA molecules per cell [Goust 1993]. Class II HLA is mostly expressed on antigen presenting cells (APC) such as macrophages, B cells and dendritic cells. Partly as a result of their importance in mediating tissue rejection, sequencing has identified MHC proteins as amongst
the most polymorphic of all human gene products. According to the international ImMunoGeneTics information system (IMGT), there are over 2000 different HLA class I and II alleles and a significant number of new alleles are discovered every year [Robinson et al. 2003]. In Chapter 9, Borghans et al. explore the nature and origin of MHC diversity in more detail.

MHCs exhibit much polymorphic amino acid variation, and seemingly trivial alterations in the identity of binding site amino acid residues give rise to differences in peptide selectivity exhibited during peptide binding. Peptide binding assays are the most widely-used way of identifying T cell epitopes and measuring the affinities of peptides binding to MHC. Such assays include direct binding and the quantitative measurement of radio- or fluorescence-labeled peptides bound to the MHC molecules [Chen & Parham 1989, Schumacher & Heemels 1990, Cerottini & Luescher 1991, Christinck & Luscher 1991, Kast & Melief 1991, Mendez-Samperio & Jimenez-Zamudio 1991, Stuber & Dillner 1995, Wauben & van der Kraan 1997, Levitsky & Liu 2000]. Several databases have been set up to store peptide binding affinity data, such as MHCPEP [Brusic et al. 1998], MHCBN [Bhasin & Singh 2003], and AntiJen [Blythe et al. 2002, McSparron et al. 2003, Toseland et al. 2005].

Many HLA alleles have been demonstrated to bind peptides with similar anchor residues [Southwood et al. 1998]. This has led to the concept of MHC supertypes: the idea that MHCs with distinct sequences can be classified into separate groups, each of which displays equivalent, if not necessarily identical, specificities when binding peptides. The celerity of experimental research will be greatly accelerated if one could identify a procedure able to cluster HLA alleles with similar specificities. Several research groups have sought to classify HLA alleles in this way, using a wide variety of different methods. Examples of such disparate methodologies include sequence analysis [Lawlor & Warren 1991], structural analysis [Chelvanayagam 1997], use of geometrical similarity matrix methods [Cano & Fan 1998], and motif search [Sette & Sidney 1998, Lund et al. 2004].

We have recently developed and applied chemometric GRID/CPCA and hierarchical clustering methods to the identification of MHC supertypes [Doytchinova et al. 2004b]. Within vaccinology, HLA classification, using bioinformatics methods, can potentially reduce the overall experimental burden by rendering unnecessary the individual study of every allele. It can thus accelerate the discovery of both epitope-based vaccines, and other immunotherapies, that are targeted at multiple alleles. In the remainder of this chapter, we will explore attempts, both ours and those of others, to address the problem of finding and populating MHC supertypes.

10.1.1 Evolutionary Analysis

An early attempt to classify MHC molecules is from protein sequence studies [Lawlor & Warren 1991]. Lawlor compared the sequences of 14 gorilla class I MHC alleles with HLA-A, B and C alleles in human and MHC in chimpanzees. Sequences of human, gorilla and chimpanzee MHC alleles are similar but not identical, as most