A license to kill:
The evolution of NK cell receptors

Paola Carrillo Bustamante

Cover Illustration Metaphor of the co-evolutionary arms race between hosts and viruses. Inspired by Gustav Klimt's *Tree of Life* and Eric Cabebe's *Virus*. Concept: Paola Carrillo Bustamante. Design: Sanja Selaković and Jovan Hadži-Durić.
Print Gildeprint, Enschede
ISBN 978-90-393-6279-2
No part of this thesis may be reproduced in any form, by any print, microfilm, or any other means, without prior written permission of the author.

A license to kill:
The evolution of NK cell receptors

A license to kill:
De evolutie van NK cel receptoren
(met een samenvatting in het Nederlands)

Con licencia para matar:
La evolución de receptores de células NK
(con un resumen en Español)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 28 januari 2015 des middags te 4.15 uur

door

Nancy Paola Carrillo Bustamante

geboren op 16 mei 1984 te Quito, Ecuador

Promotor: Prof. dr. Rob J. De Boer
Copromotor: Dr. Can Keşmir

The studies described in this thesis were financially supported by the Computational Life Sciences program of the Netherlands Organisation for Scientific Research (NWO), grant number 635.100.025.

Contents

1 General Introduction **1**

1.1 Introduction 2

1.2 Evolution of NK cell receptors 3

1.3 Why polymorphic and polygenic NK cell receptors? 8

1.4 Studying the evolution of NK cell receptors 15

2 Virus encoded MHC-like decoys diversify the inhibitory KIR repertoire **17**

2.1 Introduction 19

2.2 Results 21

2.3 Discussion 31

2.4 Material and Methods 34

2.5 Acknowledgments 38

2.6 Supporting Information 40

3 Quantifying the protection of activating and inhibiting NK cell receptors during infection with a CMV-like virus **45**

3.1 Introduction 47

3.2 Results 49

3.3 Discussion 62

3.4 Materials and Methods 66

3.5 Acknowledgments 69

3.6 Supplementary Information 70

4 A co-evolutionary arms race between hosts and viruses drives polymorphism and polygenicity of NK cell receptors **73**

4.1 Introduction 75

4.2 Results 77

4.3 Discussion 86

4.4 Materials and Methods 88

4.5 Acknowledgments 91

4.6 Supplementary Information 92

5 Selective MHC downregulation can explain the specificity and genetic diversity of NK cell receptors 95

5.1 Introduction 97

5.2 Results 98

5.3 Discussion 108

5.4 Materials and Methods 109

5.5 Acknowledgments 112

6 Can peptide sensitivity explain the diversity of NK cell receptors? 113

6.1 Introduction 115

6.2 Results 116

6.3 Discussion 123

6.4 Acknowledgments 124

7 Summarizing Discussion 131

7.1 Evolution of specific inhibiting receptors 132

7.2 Evolution of specific activating NK cell receptors 133

7.3 Genetic diversity 135

7.4 Model limitations 136

7.5 Concluding remarks 136

Bibliography 139

Summary 153

Samenvatting 155

Resumen 157

Curriculum vitæ 161

List of Publications 163

Acknowledgments 165

Chapter 1

General Introduction

1.1. Introduction

Natural killer (NK) cells are large granular cells from the innate immune system that play a pivotal role in controlling viral infections and tumors (Vivier *et al.*, 2008). To be tolerant to healthy tissue, and yet attack infected cells, the activity of NK cells must be tightly regulated. Unlike B and T cells, NK cells no not undergo gene rearrangements to generate the repertoire of cell surface receptors. Instead, they use germline-encoded inhibiting and activating receptors.

Inhibiting NK cell receptors are characterized by the presence of immunoreceptor tyrosine-based inhibitory motifs (ITIM) in the cytoplasmic tail that may decrease the state of activation (Vivier *et al.*, 2004). Activating receptors lack ITIMs, but contain a positively charged amino acid (arginine or lysine) in their transmembrane region, and are associated with signaling adaptor molecules containing immunoreceptor tyrosine-based activating motifs (ITAM), such as DAP10, DAP12 or FcγR (Lanier, 2005). NK cells integrate signals derived from both types of receptors upon cellular contact, thereby determining whether or not their effector functions should be initiated.

Many inhibiting NK cell receptors interact with major histocompatibility complex (MHC) class I proteins, which are ubiquitously expressed on the surface of nucleated cells. Because of the abundant expression of MHC-I on many cells, NK cells remain non-responsive to healthy tissue. But when cells have a decreased expression of MHC-I, which can occur during certain viral infections or cancer, they can become target for NK cell killing. This process by which NK cells detect cells with aberrant MHC-I expression has been coined by Kärre et al. as “missing-self” detection (Ljunggren and Kärre, 1990).

For the development of functional NK cells interactions between inhibiting receptors and MHC-I are required ((Höglund and Brodin, 2010, Raulet and Vance, 2006, Raulet *et al.*, 1997)). This process is called NK cell education (or licensing) and determines the threshold for activation in mature NK cells. Several models have been proposed to explain NK cell education. The consensus is that every NK cell balances its activation threshold as a rheostat to adapt to the particular MHC phenotype of its host, depending on the strength of the inhibitory signaling it received during development (Bessoles *et al.*, 2014, Brodin *et al.*, 2009, Höglund and Brodin, 2010). Additionally, NK cell education is associated with phenotypic changes of NK cells, including changes in their surface receptors specific for MHC-I (Johansson *et al.*, 2005, Sternberg-Simon *et al.*, 2013, van Bergen *et al.*, 2013). The molecular processes underlying NK cell development, including the signaling pathways that are involved in receptor acquisition and education

remain poorly defined. Although several experimental and mathematical studies (Johansson *et al.*, 2005, 2009, Sternberg-Simon *et al.*, 2013) have contributed to the current understanding of this education process, further studies are necessary to completely assess the generation and functional differences of the 6000-30000 phenotypically distinct NK cell subsets in a single individual (Horowitz *et al.*, 2013).

1.2. Evolution of NK cell receptors

Genes encoding NK cell receptors are clustered in two main gene complexes: the natural killer complex (NKC) encoding C-type lectin-like molecules, and the leukocyte receptor complex (LRC), encoding the immunoglobulin-like receptors (Trowsdale, 2001). Although these gene clusters are present in several species, there is extensive evidence for species-specific expansion of different NK cell receptor genes (reviewed in (Kelley *et al.*, 2005), and see Figure 1.1), resulting in a fascinating complexity of interactions between MHC-I and NK cell receptors. The NK cell receptor expansions known so far are described in detailed below.

1.2.1. Expansion of NK cell receptors in primates

In humans, the main receptors for MHC-I are the killer immunoglobulin-like receptors (KIRs), which are located in the LRC on chromosome 19q13.4 (Wende *et al.*, 1999). The haplotypes encoding KIRs exhibit great differences in gene content, and allelic polymorphism, with up to 17 genes encoded over approximately 150 kilo bases (Parham, 2005). The marked differences in gene content are thought to be the result of non-reciprocal crossovers in the tandemly arranged genes, causing hybrid loci or contraction and expansion of the haplotype (Martin *et al.*, 2003, Vilches and Parham, 2002).

KIRs can have either 2 (KIR2D) or 3 (KIR3D) extracellular imunnoglobulin-like domains, and contain either long cytoplasmic tails with ITIM motifs, or short cytoplasmic tails comprising ITAMs (Vilches and Parham, 2002). An exception is KIR2DL4 which has a cytoplasmic long tail, and possesses a positively charged residue in the transmembrane region, allowing association with the activating protein FcγR (Kikuchi-Maki *et al.*, 2005).

KIRs expanded between 31-40 million years ago, showing a rapid and species-specific

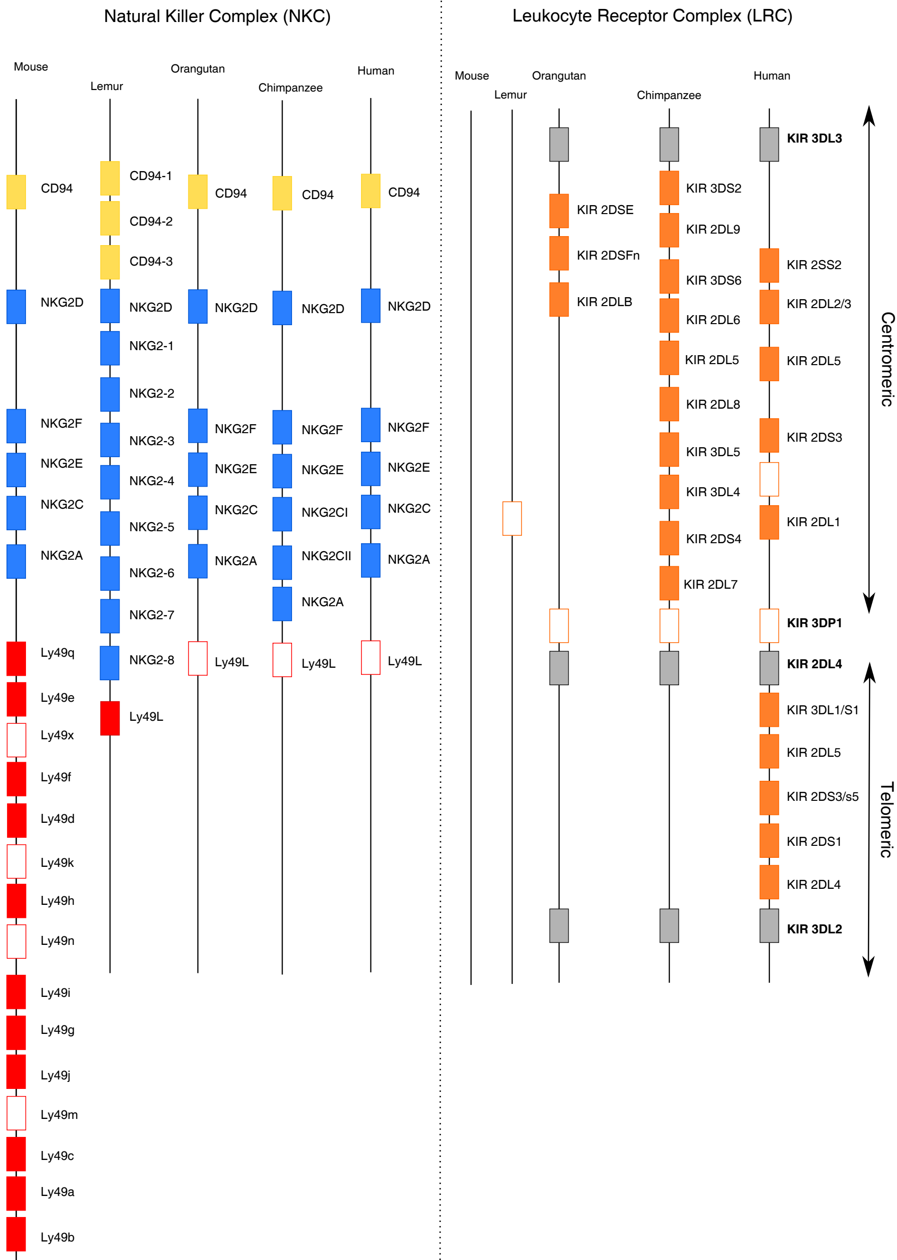

Figure 1.1. Cartoon of the NK cell receptor complexes. Legend continued on the following page.

Figure 1.1. Cartoon of the NK cell receptor complexes. This figure shows a schematic organization of some of the genes encoded in the NKC (left) and in the LRC (right) for different species. The NKC encodes genes from the CD94 (yellow boxes), NKG2 (blue), and Ly49 (red) families. While higher primates have one copy of a non-functional Ly49 gene (white boxes), lemurs have one functional Ly49, and mice encode 15 Ly49 genes, 11 of which are functional. Lemurs have expanded their CD94/NKG2 system, with three CD94 genes, and eight NKG2 genes. KIRs (orange boxes) are encoded in the LRC. All higher primates share a common organization within the gene complex. Between their four framework genes, i.e. KIR3DL3, KIR2DL4, KIR3DL2, (gray boxes), and the pseudogene KIR3DP1 (white boxes), the gene content varies across species. Lemurs have only one non-functional copy, and mice do not encode any gene in the LRC. The gene order was taken from the literature sources mentioned in the text.

diversification in primates (Martin *et al.*, 2000). Old world monkeys, apes, and humans have a common organization of the KIR gene complex, sharing four phylogenetic lineages (I,II,III, and V), which are characterized by their structure and specificity for MHC-I ((Guethlein *et al.*, 2007a, Rajalingam *et al.*, 2004)). In humans, a lineage I KIR binds epitopes of HLA-G (Rajagopalan, 2010), lineage II KIRs recognize epitopes A3/A11 on HLA-A, and Bw4 on HLA-A and B, and lineage III KIRs bind to HLA-C epitopes C1 and C2 (Trowsdale *et al.*, 2001) (a ligand for KIR lineage V has to still be identified).

In the LRC, there are some differences in gene content and specificity of the KIRs across primate species. Importantly, the marked differences go hand in hand with the evolution of MHC-I genes. In macaques, having duplicated sets of MHC-A and MHC-B genes (Daza-Vamenta *et al.*, 2004, Shiina *et al.*, 2006), the presence of the Bw4 motif (Adams and Parham, 2001) is related to a large collection of several lineage II KIRs (Bimber *et al.*, 2008, Blokhuis *et al.*, 2010, 2011, Kruse *et al.*, 2010). Consistent with the observation that macaques lack MHC-C molecules, they have only one lineage III KIR, which may not recognize any MHC-class I (Sambrook *et al.*, 2005). Orangutans, on the other hand, carry fewer MHC-A and MHC-B loci than macaques, and encode only one lineage II KIR accordingly. Orangutans were the first primates to evolve MHC-C (C1 epitope), corresponding to the expansion of their lineage III KIRs (Guethlein *et al.*, 2002, 2007a). Lineage III KIRs expanded further in chimpazees, correlating with the evolution of the C2 epitope in MHC-C. Chimpanzee have both inhibiting and activating KIRs, and eight of them recognize MHC-C specifically (Abi-Rached *et al.*, 2010).

Humans, in contrast, have only seven lineage III KIRs, and two lineage II KIRs. The specificity for MHC-C has also decreased in humans, where only three KIRs have specificity for HLA-C, including the inhibiting KIR2DL2/3, and KIR2DL1, and the activating KIR2DS2 (see Table 1.1). Additionally, humans are the only species that have

Table 1.1. Ligands of activating and inhibiting human KIRs

Activating KIR	Ligand
2DS1	HLA-C2
2DS2	HLA-C1, HLA-A*11:01
2DS3	unknown
2DS4	HLA-C*05:01, A*11:02, C*16:01
2DS5	unknown
3DS1	unknown
Inhibiting KIR	**Ligand**
2DL1	HLA-C2
2DL2 / 2DL3	HLA-C1, HLA-C2, HLA-B*46:01, and HLA-B*73:01 (C1 epitope)
2DL4	HLA-G (might be an intracellular interaction)
2DL5	unknown
3DL1	HLA-A with Bw4 motif, HLA-Bw4
3DL2	HLA-A3/A11
3DL3	unknown

undergone specific expansion in the telomeric part of the KIR complex (Parham and Moffett, 2013). While the centromeric part of human KIR haplotypes is more similar to chimpanzee KIR haplotypes (Abi-Rached *et al.*, 2010), the telomeric region in humans accumulated genes that show mainly activating potential and that have little or no binding affinity to HLA-I molecules, such as KIR2DS2, 2DS3, and 2DS5 (Moesta *et al.*, 2010, Pyo *et al.*, 2010). This clear distinction between centromeric and telomeric genes, allowed for the distinction of two haplotype groups, A and B. Both haplotypes are present in all human populations (Hollenbach *et al.*, 2010), differ in frequencies and are mantained by balancing selection (Yawata *et al.*, 2006).

The other main receptor cluster in primates is the NKC (Kelley *et al.*, 2005). The main members of these gene families are the Ly49 and the NKG2 genes. Primates have only one gene of the Ly49 gamily, which is a pseudogene, but their NKC encodes several NKG2 genes (Guethlein *et al.*, 2002, Khakoo *et al.*, 2000, LaBonte *et al.*, 2001, Renedo *et al.*, 1997). Members of the NKG2 family include the inhibiting NKG2A, the activating NKG2C, NKG2E, and NKG2D; and the NKG2F, for which no function has been yet determined (Lazetic *et al.*, 1996). NKG2 proteins dimerize with the invariant CD94 molecule on the cell surface, which contains a short cytoplasmic domain and transduces

the activating or inhibiting signal (Lazetic *et al.*, 1996). An exception is NKG2D, an activating receptor, which shares little sequence similarity with the other members, and associates with the activating molecule DAP10.

The ligands of NKG2A and NKG2C include the conserved and non classical HLA-E molecule in humans and Qa-1^b in mice (Borrego *et al.*, 1998, Braud *et al.*, 1998, Petrie *et al.*, 2008, Zeng *et al.*, 2012), which present peptides derived from the leader sequences of the classical HLA-A, -B, and -C molecules in humans, and from H2 molecules in mice. The engagement of NKG2A by HLA-E or Qa-1^b inhibits the activity of NK cells, preventing target cell lysis. In higher primates, both receptor and ligand are very well conserved (Shum *et al.*, 2002), presenting a system for detection of “gross” MHC-I expression, that unlike KIRs is highly conserved and stable.

Lemurs, on the other hand, exhibit only one single non functional KIR gene in their LRC, but they have diversified the genes encoding CD94 and NKG2 (Averdam *et al.*, 2009). Located in chromosome 7, the NKC comprises eight inhibiting and activating genes, and three CD94 genes. The equivalent molecule for HLA-E has not been yet identified in prosimians, but the ligands for the NKG2 receptors are expected to be MHC-I molecules (Averdam *et al.*, 2009). Additionally, Averdam et al. showed that all possible CD94/NKG2 combinations are able to form heterodimers at the cell surface, implying that an exchange of the CD94 or the NKG2 subunit can influence the binding specificity for MHC class I ligands (Averdam *et al.*, 2009). Thus, lower primates seem to have evolved an alternative system for variable NK cell receptors.

1.2.2. Expansion of NK cell receptors in rodents

The mouse LRC is located on chromosome 7 in mice and on chromosome 4 in rats (Iizuka *et al.*, 2003, Kirkham and Carlyle, 2014, Schenkel *et al.*, 2013), but it does not contain any of the KIRs that bind MHC-I in humans (Martin *et al.*, 2002a). Instead, rodents have expanded their Ly49 genes, resulting in a remarkable diversity across different inbred mouse strains (Iizuka *et al.*, 2003, Kirkham and Carlyle, 2014). While the mouse Ly49 complex comprises at least 20 genes and pseudogenes (Wilhelm *et al.*, 2002), the variation is even larger in rats, with 19 functional genes and 15 pseudogenes Table 1.2 shows the most important known receptors in mouse strains studied so far (Rahim *et al.*, 2014).

Ly49 receptors in mice are functionally similar to KIRs in humans, having both inhibiting and activating receptors, and genes encoding proteins that preferentially bind MHC-

I (Schenkel *et al.*, 2013). Although several ligands for activating Ly49 receptors remain unknown, some activating receptors bind viral encoded proteins (see below).

1.2.3. NK cell receptors in other species

At least five highly conserved polymorphic Ly49 genes have been found in some equids, including horses, asses and zebras (Futas and Horin, 2013, Takahashi *et al.*, 2004). By contrast, only one single Ly49 has been found in cattle (McQueen *et al.*, 1998), domesticated dogs and cats, and pigs (Gagnier *et al.*, 2003). Cattle have also functional KIRs, although their genes are divergent from the primate KIRs (Parham and Moffett, 2013). Opposite to primate KIRs, which diverged from these founder genes KIR3DL, cattle expanded the founder KIR3DLX. In each species, the gene that was not used became nonfunctional (Dobromylskyj and Ellis, 2007, Guethlein *et al.*, 2007b). No species studied so far is known to have two expanded systems (Parham and Moffett, 2013), but several species diversify neither, keeping both KIR and Ly49 as one single copy genes (Hammond *et al.*, 2009).

1.3. Why polymorphic and polygenic NK cell receptors?

The evolution of variable NK cell receptor genes is likely shaped by several factors determining fitness and survival, like pathogen resistance, detection of MHC-I ligands, and reproductive success (Parham and Moffett, 2013). The inter- and intra-species gene diversity indicates their rapid evolution. Importantly, the independent convergent evolution of variable NK cell receptors in several different species highlights their functional importance.

However, the evolutionary selection pressure whereby NK cell receptors became polymorphic and polygenic remains unresolved. The conserved inhibitory receptor NKG2A demonstrates that missing-self detection can be achieved without a polygenic and polymorphic NK cell receptor system. Why then have these polygenic and polymorphic receptors evolved?

1.3.1. Response to viral infections

Because of the evolutionary arms race between infectious pathogens and the immune system, one possible explanation for the diversification of NK cell receptors is the selection pressure on NK cells imposed by the successful immunoevasive mechanisms evolved by several pathogens (Lanier, 2008, Sun and Lanier, 2009). Indeed, there is extensive evidence of associations between particular NKRs and the viral control caused by viruses, including cytomegalovirus (CMV), human immunodeficiency virus (HIV-1), and hepatitis C virus (HCV).

KIR in human diseases

Several human studies have provided evidence that particular NK receptors may be directly involved in viral control. Associations between particular KIR alleles and disease outcome has been found in HIV-1, HCV, and Influenza (Jamil and Khakoo, 2011).

During HIV-1 infection, there is an expansion of $3DS1^{+}$NK cells, an expansion which is dependent on the presence of the Bw4-80I epitope, the cognate ligand for this receptor (Alter *et al.*, 2009, Pelak *et al.*, 2011). The expansion of selected NK cell subsets could be beneficial to the host due to an immediate and stronger NK cell response. Indeed, individuals carrying these KIR-HLA combinations showed lower viremia and a slower progression to AIDS (Flores-Villanueva *et al.*, 2001). Furthermore, an increased number of KIR3DS1 (caused by a higher copy number variants of KIR3DL1/S1) was correlated with a lower set viral point in the presence of HLA-Bw4-80I (Pelak *et al.*, 2011).

In response to HCV infected target cells, NK cells expressing 2DL2/L3 exhibit increased degranulation compared to $2DL2/L3^{-}$NK cells (Amadei *et al.*, 2010). Homozygous individuals for the 2DL2/L3-HLA-C1 pair control better HCV infection, and can even experience spontaneous clearance (Khakoo *et al.*, 2004, Romero *et al.*, 2008). Additionally, a protective effect of 3DS1 in combination with HLA-Bw4-I80 in hepatocarcinoma has been found in patients with chronic HCV infection (López-Vázquez *et al.*, 2005).

Studies of human influenza A virus (IAV) have also shed light on the protective effect of some KIR-HLA combinations. NK cells from individuals homozygous for the 2DL3-C1 pair had a stronger activation to IAV infected cells than those homozygous for 2DL1-C2 (Ahlenstiel *et al.*, 2008). Another study of patients infected with 2009 pandemic IAV

strain (H1N1/09), established that patients having all 3DL1, 2DL1, 2DL2/3 genes had a lower risk of death (La *et al.*, 2011).

These few examples mirror the large range of studies associating KIR-MHC combinations with the outcome of diseases. However, the lack of well-characterized ligands for several receptors and specific monoclonal antibodies for specific KIRs limit the understanding of the precise molecular mechanisms underlying these associations, and with it the precise functional role of these receptors upon infection.

Selective downregulation of MHC-I

Several viruses, including Epstein-Barr-Virus (EBV), CMV, and HIV, decrease the expression of MHC-I to escape from the T cell immune response. Interestingly, the downregulation does not always affect all MHC molecules in the same way, but some viruses have evolved proteins that target only particular alleles (reviewed in (Nash *et al.*, 2014)). Since the HLA molecules presenting peptides to T cells (A and B) tend to be downregulated, while those inhibiting NK cells (C and E) tend to remain expressed, selective MHC downregulation seems to be a viral strategy to avoid missing-self detection.

HCMV encodes several immunoevasin proteins that selectively downregulate the expression of MHC-I on the cell surface (Nash *et al.*, 2014), such as US2 and US11, targeting specific and non overlapping HLA-A, and HLA-B alleles, by promoting their export into the cytosol for proteosomal degradation (Gewurz *et al.*, 2001, Llano *et al.*, 2003, Schust *et al.*, 1998). In addition to selective MHC-downregulation, HCMV encodes proteins that induce MHC-I expression to inhibit NK cells. For example, UL40 has a high sequence similarity to peptides from HLA-C alleles (Tomasec *et al.*, 2000, Ulbrecht *et al.*, 2000). By binding to HLA-E, UL40 can promote its expression on the cell surface, providing a ligand for NKG2A.

HIV-1 also decrease the expression of particular HLA alleles. HIV Nef binds to the cytoplasmic tails of the HLA-A and HLA-B molecules in the ER, re-directing them to endo-lysosomal compartments for degradation (Schaefer *et al.*, 2008). Small differences in the tails of HLA-C and -E prevent Nef from hampering their transport to the cell surface, which in turn prevents HIV-infected cells to be lysed by NK cells (Cohen *et al.*, 1999, Collins *et al.*, 1998, Gall *et al.*, 1998).

Mouse CMV (MCMV) encodes glycoproteins that interfere with the expression of MHC-I molecules (Wagner *et al.*, 2002). For example, gp40 retains the MHC molecules in the

ER (Ziegler *et al.*, 1997), while gp48 re-routes mature MHC to endo-lysosomal compartments for degradation (Reusch *et al.*, 1999). Balancing this broad MHC-I downregulation is the protein gp34, which escorts some MHC alleles to the cell surface (Kleijnen *et al.*, 1997).

All these examples highlight the evolutionary importance of viruses partially downregulating the expression of MHC-I molecules. The non-overlapping targets for MHC downregulation (such as in HCMV or in HIV-1) illustrate the adaptation of the viruses to several MHC-I molecules. By selectively downregulating MHC-I molecules, some viruses escape from NK cell responses. Thus, it is possible that the evolution of these independent immunoevasins was driven by the selection pressure imposed by inhibiting NK cell receptors. The selective downregulation can in turn shape the evolution of the NK cell receptors, as inhibiting receptors with non overlapping specificities for different MHC molecules can help to detect missing-self.

MHC-I decoys in CMV

In addition to selectively downregulating the expression of MHC I molecules, some viruses use MHC-I like proteins, i.e. decoys, that can directly interact with the NK cell receptors to modulate the immune response. Examples of such decoys are the HCMV encoded UL18 binding to the inhibitory LIR-1 (Prod'homme *et al.*, 2007), and the MCMV encoded m144, mimicking key structural characteristics of H-2 molecules (Natarajan *et al.*, 2006, Prod'homme *et al.*, 2007). Interestingly, UL18 and m144 share more sequence similarity with MHC-I than they share with each other, showing that species-specific immune pressure led to independent acquisition of MHC-I mimics (Farrell *et al.*, 1997, Natarajan *et al.*, 2006).

The most extensively studied MHC decoy is the MCMV encoded m157 protein. m157 allows MCMV to avoid NK cell activation by engaging inhibitory receptors with high affinity, as shown in 129/J mice, which are highly susceptible to the infection (Smith *et al.*, 2002). Unlike 129/J mice, C57BL/6 mice exhibit spontaneous resistance against MCMV, a phenomenon that has been genetically mapped to one single gene encoding the activating receptor Ly49H, which also binds m157 with high affinity (Lee *et al.*, 2001, Smith *et al.*, 2002). The activating Ly49H evolved from its inhibitory counterpart Ly49I (Abi-Rached and Parham, 2005), indicating that the evolution of the activating receptors resulted from the novel selective pressure exerted by CMV after evolving MHC-I decoys (Arase and Lanier, 2002, Lanier, 2008, Sun and Lanier, 2009). The immuno-

Table 1.2. Ly49 haplotypes in four known mouse strains with their response to MCMV (modified from (Rahim *et al.*, 2014))

	Mouse strain			
	NOD	**129**	**B6**	**BALB**
	Response to MCMV			
	susceptible	susceptible	resistant	susceptible
Activating	Ly49D	Ly49P	Ly49D	Ly49L
	Ly49H	Ly49R	Ly49H	
	Ly49M	Ly4UP		
	$Ly49P_1$			
	$Ly49P_3$			
	Ly49U			
	Ly49W			
Inhibiting	Ly49A	$Ly49B^b$	Ly49A	Ly49A
	$Ly49B^b$	Ly49E	$Ly49B^b$	$Ly49B^b$
	Ly49C	$Ly49EC_2$	Ly49C	Ly49C
	Ly49E	$Ly49G^b$	Ly49E	Ly49E
	Ly49F	${Ly49I_1}^b$	Ly49F	Ly49G
	$Ly49G_2$	Ly49O	Ly49G	Ly49I
	Ly49I	$Ly49Q_1$	Ly49I	Ly49Q
	Ly49Q	Ly49S	Ly49J	
		Ly49T	Ly49Q	
		Ly49V		

evasive role of m157 is further supported by observations of MCMV infections in wild outbred mice, where several strong interactions between m157 and an array of inhibitory receptors were detected, while very few m157 variants engage the Ly49H encoded receptor (Corbett *et al.*, 2011). Accordingly, there were also variations in viral control.

C57BL/6 is not the only inbred strain resistant to MCMV. Inbred MA/My mice also have low viral titers after infection with MCMV, although they do not possess the Ly49H gene. Their resistance is mediated by the activating receptor Ly49P which specifically recognizes MCMV infected cells in a $\mathrm{H2} - \mathrm{D^k}$ dependent manner (Kielczewska *et al.*, 2009). Interestingly, this resistance requires the presence of the virally encoded protein m04, which escorts and binds newly assembled MHC-I molecules on the cell surface. Other activating receptors that recognize MCMV infection in a m04-H2 dependent manner include $\mathrm{LyL9L^{BALB}}$,$\mathrm{Ly49P1^{NOD}}$, and $\mathrm{Ly49W1^{NOD}}$ (Pyzik *et al.*, 2011). Like m157, the

original function of m04 might have been to counteract the effect of MHC downregulation and avoid "missing-self" detection by inhibiting receptors. These observations suggest that hosts evolved novel activating receptors to recognize the decoys evolved by viruses.

Peptide sensitivity

The general concept is that, unlike T cell receptors, inhibiting KIRs (iKIRs) are not highly specific for particular peptide-MHC (pMHC) complexes. However, several studies have shown that iKIRs can be sensitive to the specific peptides bound by the HLA molecules (Hansasuta *et al.*, 2004, Malnati *et al.*, 1995, Peruzzi *et al.*, 1996, Rajagopalan and Long, 1997, Thananchai *et al.*, 2007). Crystal structures of KIR2DL1 and KIR2DL2 in complex with their HLA-C ligands further supported this observation (Boyington *et al.*, 2000, Brooks *et al.*, 2000, Fan *et al.*, 2001, Li and Mariuzza, 2014), by revealing that specifically positions 7 (P7) and 8 (P8) of the bound peptide are in direct contact with residues of the iKIR. Other studies showed that peptides having weak or no binding to iKIRs can markedly reduce KIR-mediated inhibition (Cassidy *et al.*, 2014, Fadda *et al.*, 2010).

Because of the direct contact between iKIR and the MHC presented peptides, NK cell activation may vary in a peptide dependent manner, making iKIRs sensitive to changes in the peptide repertoire presented by MHC-I molecules. These observations calls for an extension of the current model of NK cell activation: "missing self" detection could be complemented by "altered self", where changes in the MHC-I peptide repertoire modulate the NK cell signaling.

In the context of viral infections, the importance of peptide sensitivity has been emphasized by recent HIV-1 studies, which demonstrate that sequence variations within HLA-C restricted HIV epitopes disrupt or promote the binding to inhibiting KIR2DL2, subsequently modifying NK cell activation (Fadda *et al.*, 2012, van Teijlingen *et al.*, 2014). Importantly, these studies show that a small number of naturally occurring variants of HIV-1 epitopes that are presented by HLA-C*03:04 can strongly engage KIR2DL2, inducing a strong inhibiting signal for the NK cells (van Teijlingen *et al.*, 2014). It is tempting to speculate, that viral variants are selected to avoid NK cell mediated immune responses in individuals expressing the adequeate KIR/HLA pair.

1.3.2. Reproductive success

Since the divergence from chimpanzees, hominids have evolved several changes in aspects of locomotion, anatomy, and reproduction (Parham and Moffett, 2013). Two key aspects of this human specific evolution affect the reproductive success: bipedalism and larger brain size. The evolution of bipedalism imposed drastic anatomic changes in the size and shape of the human female pelvis, affecting directly the size of the birth canal. While the size of the birth canal was decreasing, the evolution of larger brains was imposing additional challenges for a successful birth (Parham and Moffett, 2013).

The evolution of bigger brain sizes required more blood supply in the placenta (Leonard *et al.*, 2007), a process that has been achieved by an arterial remodelling of the uterine arteries (reviewed in (Wallace *et al.*, 2012)). Arterial remodelling occurs thanks to extravillous trophoblast (EVT) cells. EVT are fetal cells invading the uterus, transforming the spiral arteries into large vessels that are able to provide adequate blood supply to the growing fetus (Moffett-King, 2002). Indeed, pre-enclampsia and recurrent miscarriage have been associated with a compromised arterial remodelling (Wallace *et al.*, 2012).

A successful EVT invasion depends on the interactions of EVT with uterine NK cells. The activation of uterine NK cells is important for arterial remodelling, as it results in the release of cytokines, which in turn promote migration of the trophoblasts (Xiong *et al.*, 2013). Because EVT uniquely express HLA-C (lacking HLA-A, or -B), and uterine NK cells preferentially express HLA-C specific KIRs (Sharkey *et al.*, 2008), several correlations between reproductive success and particular KIR/HLA-C combinations have been found (Hiby *et al.*, 2004, 2010, 2014). The presence of HLA alleles binding more inhibiting than activating receptors results in compromised arterial remodelling and reduced fetal growth (Kieckbusch *et al.*, 2014). Accordingly, mothers being homozygous for KIR A haplotypes have a high risk of developing pre-enclampsia if the fetus carries one C2 allele, as AA homozygous individuals have two copies of the inhibiting KIR2DL1, which binds strongly to C2 (Hiby *et al.*, 2004, 2010, 2014).

The evolution of larger brain sizes started mainly in *Homo erectus* (Robson and Wood, 2008) and correlates with the emergence of B haplotypes. However, the evolution of more activating KIRs is not always beneficial for reproduction. NK cell-mediated placentation can lead to large babies that are not able to pass through the birth canal, causing obstructed labor. A successful placentation is hence dependent on a tight NK cell-mediated regulation. Therefore, the pressure for a successful reproduction can drive and maintain inhibiting and activating receptors specific for MHC-I.

1.4. Studying the evolution of NK cell receptors

Computational modeling is a useful tool to study long term evolution, and has been widely used to study host-pathogen co-evolution (Borghans *et al.*, 2004, Schmid *et al.*, 2010). In this thesis, we make use of computational and mathematical modeling to provide further insights into the evolution of NK cell receptors hoping to answer the question of why NK cell receptors have evolved to become specific, polygenic, and polymorphic.

We focus on the different hypothesis related to viral infections, i.e. viral evolution of decoys, peptide sensitivity, and selective MHC-downregulation. We first study the effect of MHC-I decoys on the evolution of the specificity of inhibiting receptors (Chapter 2). For this purpose, we developed an agent-based model (ABM) of human-like populations that are infected with herpes-like viruses causing chronic infections. We extended this model to include to activating receptors (Chapter 3). These studies allow us to quantify how specific activating and inhibiting KIRs must be to be protective to CMV-like viruses. In Chapter 4 we expand our model further and allow for host-pathogen evolution. Chapter 5 focuses on studying the selective MHC downregulation, and in Chapter 6 we use a mathematical model to study whether peptide sensitivity can be a possible explanation of polygenicity of NK cell receptors.

Chapter 2

Virus encoded MHC-like decoys diversify the inhibitory KIR repertoire

PAOLA CARRILLO-BUSTAMANTE, CAN KEŞMIR, ROB J. DE BOER (2013)

PLoS Computational Biology, 9(10): e1003264.

Abstract

Natural killer (NK) cells are circulating lymphocytes that play an important role in the control of viral infections and tumors. Their functions are regulated by several activating and inhibitory receptors. A subset of these receptors in human NK cells are the killer immunoglobulin-like receptors (KIRs), which interact with the highly polymorphic MHC class I molecules. One important function of NK cells is to detect cells that have down-regulated MHC expression (missing-self). Because MHC molecules have non polymorphic regions, their expression could have been monitored with a limited set of monomorphic receptors. Surprisingly, the KIR family has a remarkable genetic diversity, the function of which remains poorly understood. The mouse cytomegalovirus (MCMV) is able to evade NK cell responses by coding "decoy" molecules that mimic MHC class I. This interaction was suggested to have driven the evolution of novel NK cell receptors. Inspired by the MCMV system, we develop an agent-based model of a host population infected with viruses that are able to evolve MHC down-regulation and decoy molecules. Our simulations show that specific recognition of MHC class I molecules by inhibitory KIRs provides excellent protection against viruses evolving decoys, and that the diversity of inhibitory KIRs will subsequently evolve as a result of the required discrimination between host MHC molecules and decoy molecules.

2.1. Introduction

Natural killer (NK) cells constitute 5–25 % of the lymphocytes circulating in human peripheral blood (Caligiuri, 2008). Being part of the innate immune response, they play an important role in the defense against viral infections and in tumor surveillance (Lanier, 2008). In contrast to T and B cells, NK cells do not use somatic gene rearrangements to generate a diverse repertoire of cells expressing unique receptors. Instead, they sample a subset of receptors from a repertoire of activating and inhibitory receptors encoded by the germline.

Individual NK cells express several inhibitory and activating receptors that recognize, among others, major histocompatibility complex (MHC) class I and MHC class I related molecules as their ligands (Lanier, 2005). The interaction between these receptors and ligands generates signals that either allow the NK cell to attack target cells or prevent it from harming healthy tissue. Several viruses down-regulate the expression of host MHC class I molecules, and since these molecules are often inhibitory ligands of NK cell receptors, loss of their expression on the infected cell induces NK cell activation. This mechanism by which NK cells attack MHC-class I deficient cells was coined by Kärre et. al (Ljunggren and Kärre, 1990) as “missing-self” detection.

In humans there are two main receptor families contributing to missing-self detection. The inhibitory receptor CD94/NKG2A binds to complexes of the human leukocyte antigen (HLA)-E, presenting peptides derived from the leader sequences of HLA-A, -B, and -C (Braud *et al.*, 1998, O’Callaghan, 2000). In this inhibitory interaction both receptor and ligand are highly conserved, and the down-stream effects are remarkably similar in different individuals (Shum *et al.*, 2002). In contrast, killer immunoglobulin-like receptors (KIR), recognizing the highly polymorphic HLA-A, -B, and -C molecules, can be both inhibiting and activating, are very diverse, and rapidly evolving (Jamil and Khakoo, 2011). Engagement of either inhibitory KIR or NKG2A inhibits the activity of an NK cell, preventing target cell lysis. Phylogenetic studies have shown that the CD94/NKG2 system is relatively old, and that the KIR genes have evolved more recently (Parham *et al.*, 2010). Thus, there are two NK cell receptor systems, one conserved and one highly diverse, detecting abnormalities in MHC expression on cell surfaces.

KIRs are encoded by a large family of genes exhibiting a remarkable variability in gene content and allelic polymorphism. The KIR complex in humans contains up to 14 KIR genes and pseudogenes (Trowsdale *et al.*, 2001) that are arranged into two main groups of haplotypes, A and B, differing in size, gene content, function, and disease associ-

ation (Vilches and Parham, 2002). Since the MHC and KIR loci are on different chromosomes (in humans, on chromosome 6 and 19, respectively), a tremendous number of possible receptor-ligand combinations exists on the population level. Moreover, KIR-HLA interactions are rather specific, with four mutually exclusive epitopes on HLA molecules (A3/11, Bw4, C1 and C2) so far identified as inhibitory KIR ligands (Moretta *et al.*, 1996). KIR interactions with HLA-C are sensitive to polymorphisms at distal positions (Moesta *et al.*, 2008) and to bound peptides (Fadda *et al.*, 2010), affecting KIR binding, and with that the functionality of NK cells. It is widely accepted that the fine specificity and vast diversity of B and T cell receptors per individual render each host the capacity to recognize many different pathogens, and to distinguish them from healthy tissue. But how does the specificity and much smaller diversity of NK cell receptors per individual contribute to the host's survival? If missing-self detection were the main function of inhibitory KIRs, and since this can also be achieved by the conserved receptor NKG2A, why have these more recent NK cell receptors evolved to become specific, polymorphic, and diverse?

Specific KIR alleles have been associated with particular infections such as HIV, HCV, cerebral malaria, and with several pregnancy disorders (Alter *et al.*, 2011, Hiby *et al.*, 2004, 2010, Hirayasu *et al.*, 2012, Khakoo *et al.*, 2004, Knapp *et al.*, 2010, Moraru *et al.*, 2012). Indeed, population genetic studies have suggested that a high degree of KIR diversity is necessary for surviving epidemic infections and population bottlenecks (Gendzekhadze *et al.*, 2009), but no explicit evolutionary mechanism selecting for novel KIR alleles has been proposed so far. Why polymorphic KIRs would be required to just detect MHC down-regulation remains puzzling.

Cytomegaloviruses (CMV) and other viruses from the herpes family have large genomes that encode for a series of immuno-evasive mechanisms, targetting key molecular steps necessary for a successful immune response (Griffin *et al.*, 2010, McGeoch *et al.*, 2006, Revilleza *et al.*, 2011). Particularly important for the evasion of NK cell surveillance are MHC-I like molecules that can engage inhibitory NK cell receptors, like the mouse CMV (MCMV) encoded glycoprotein m157 binding to Ly49 receptors (Arase *et al.*, 2002, Smith *et al.*, 2002), and the human CMV (HCMV) UL18 engaging the inhibitory leukocyte immunoglobulin-like receptor LIR–1 (Wilkinson *et al.*, 2008). Not all of these evasion strategies have been elucidated yet, and it remains unclear whether m157 and UL18 are the only decoy molecules evolved by herpes viruses. Recent studies have revealed a strong imprint in the KIR repertoire of CMV seropositive individuals (Béziat *et al.*, 2013, Charoudeh *et al.*, 2013), suggesting that additional CMV evasion mechanisms interacting directly with KIRs (e.g. novel decoy molecules yet to be identified) exert

a strong selection pressure.

We investigated whether the presence of viral decoys like MCMV m157 or HCMV UL18 can drive the expansion of specific, inhibitory NK cell receptors, such as KIRs. We performed our study with an agent-based computer model of co-evolving hosts and viruses. Our results show that specific MHC recognition by inhibitory KIRs provides excellent protection against viruses evolving decoy molecules, and that diversity in the receptor system can be a consequence of this specific interaction between MHC and KIR molecules.

2.2. Results

To investigate the evolution of the KIR genes we developed an agent-based model consisting of a host population infected with a non-lethal herpes-like virus causing chronic infections. In this model, individuals were randomly selected during every time step to be confronted with one of the randomly chosen events: birth, viral infection, and death. We modeled a host population of simplified humans carrying one MHC locus and one KIR haplotype composed of five genes. Here, we only modeled inhibitory KIRs and work on activating receptors is in progress. All hosts were initialized with the same randomly generated KIR haplotype, but with different MHC genes. While we allowed for mutation of novel KIR genes during the birth event, mutation of MHC molecules was not considered, and the initial MHC polymorphism (14 alleles, mimicking the number of common HLA-C molecules) remained constant throughout the simulation. Although the host population considered here is inspired on humans, the model remains simple to allow it to be general and also apply to the evolution of KIR in 'higher' primates in which the KIR genes also expanded and diversified (Averdam *et al.*, 2009).

KIR, MHC, and viral MHC decoys were modeled with bit strings (i.e., randomly generated sequences of zero and ones), as a simplified representation of amino acid sequences (see Material and Methods). Receptor-ligand binding depended on the longest adjacent complementary match between their bit strings (Figure 2.1 A). If the length of the longest match reached a threshold L, the molecules could interact. Table 2.1 depicts the relation between the specificity, i.e., the likelihood of such an interaction, and L.

In every host, those KIRs that failed to recognize any of the two MHC molecules present in the individual were deleted from the host's repertoire, leaving each host with only a "licensed" KIR repertoire. Only those KIR molecules that were licensed participated in

Table 2.1. Relationship between specificity and crucial parameters for clearing the infection

Specificity (L)	**MHC recognition** (p)	**Licensed KIRs** (n)	**Protection** (P)	**Heterozygous advantage** (HA)
1	1.000	10.000	0.000	1.000
2	0.959	9.984	0.000	1.000
3	0.696	9.078	0.000	1.000
4	0.396	6.352	0.041	1.007
5	0.197	3.551	0.458	1.112
6	0.095	1.818	0.833	1.367
7	0.041	0.807	0.966	1.657
8	0.024	0.477	0.988	1.783
9	0.011	0.222	0.998	1.894
10	0.006	0.120	0.999	1.941

The specificity of a KIR molecule L is measured in bits. The analytical expectations of its probability of recognizing MHC, the expected repertoire of licensed KIRs, the expected protection against decoy viruses, and the expected heterozygous advantage of two KIR haplotypes were calculated as described in Material and Methods.

the immune response, mimicking the MHC-dependent education process during NK cell development (Anfossi *et al.*, 2006). The expected number of licensed KIRs per host, and consequently the probability of the host being protected, depends on the specificity of the KIR molecules, and can be calculated as described in Material and Methods (see Table 2.1).

Infection of a host started with a short acute phase, after which the individual could either recover, or become chronically infected. We considered 16 different viruses: one "wild type" virus, one "MHC down-regulating" virus, and 14 "MHC decoy" viruses (i.e., one for each of the 14 MHC molecules in our model). We did not allow for superinfection and hosts could be infected with one of the 16 viruses only. The evolution of decoy proteins was modeled by allowing the virus to adopt a randomly selected MHC molecule from its host. Therefore, each decoy protein was actually an MHC molecule. The host populations were first inoculated with the "wild-type" virus, which was typically cleared after the acute phase because of the assumed immune response of both, cytotoxic T and NK cells (Figure 2.1 B). The effect of the immune response was modeled by a parameter describing the probability of clearing the infection. For the wild-type virus, this parameter was set relatively high, i.e. 85 % of the wild type infections were cleared, resulting in approximately 30 % of the individuals becoming chronically infected (Figure 2.1 B, and

Figure 2.2 blue line). The clearance probability was lower for the down-regulating and decoy viruses, resembling their immune escape. Viruses establishing chronic infections spread over much longer periods of time than those that do not; and we therefore expect viruses capable of MHC down-regulation and carrying decoy molecules to outcompete wild-type viruses. We used this model to investigate the evolution of the KIR system, and compared the selection pressures imposed by different viral variants.

2.2.1. Specific inhibitory KIRs are disadvantageous in populations infected with a virus down-regulating MHC expression

If detection of MHC class I is the main function of inhibitory KIRs, we expect that KIRs do not need to be specific nor diverse because missing self detection can be achieved by a limited set of monomorphic receptors. To address this, we analyzed the effects of KIR specificity on populations infected with a virus that is capable of MHC down-regulation to escape T cell response. The immune escape of this mutant was modeled by decreasing the probability of clearing the infection (from 85% to 70%), resulting in a better spread of the virus and a larger fraction of individuals becoming chronically infected (Figure 2.1 B, and Figure 2.2 A red line).

We screened the average population size as an indicator of the hosts' protection after an infection. By comparing simulations with degenerate KIRs with those with specific KIRs, we observed significant differences in population sizes (from 4300 in $L = 1$ to 4100 in $L = 10$, $p = 0.0019$, Mann-Whitney U test). Although the effect of KIR specificity on protection against an MHC down-regulating virus was small, it clearly indicated that hosts with highly specific KIR–MHC interactions were more vulnerable than those having degenerate KIRs (Figure 2.2 A,B,E). Why is a high KIR specificity disadvantageous during an infection with an MHC down-regulating virus? Since degenerate KIRs (i.e. $L = 1$) are likely to recognize any MHC in the population, these receptors are perfectly capable of detecting the presence (and hence absence) of MHC molecules within one individual. But if the KIR-MHC interaction is specific enough (i.e. $L \geq 5$), the chance of a KIR to recognize any MHC within the same individual is small, impeding the host to detect MHC down-regulation, i.e. missing-self. Thus, the potential to recognize the absence of MHC molecules, and with it to clear the infection, decreases with a higher specificity of KIR-MHC interactions. Note that the inability of a specific inhibitory KIR to recognize missing-self is independent of the education process we implemented in the model. These results were consistent in all simulations we ran for each specificity setting ($L = 1 - 10$, Figure 2.2 E red line), confirming our reasoning that for

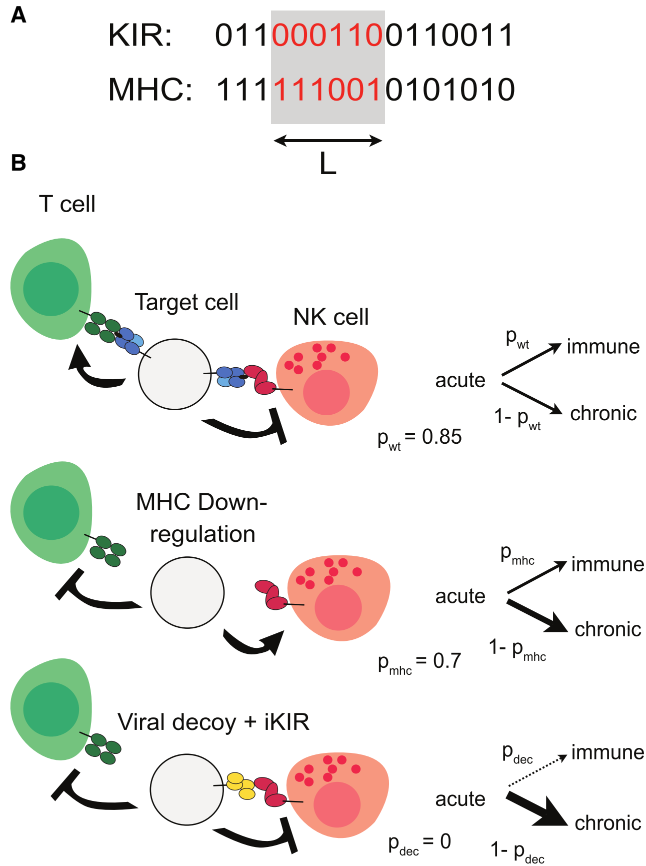

Figure 2.1. Schematic model description. (A) KIRs and MHC molecules are implemented as bitstrings. The receptor-ligand recognition is modeled via complementary matching. Here, the longest adjacent match is 6 bits long, and when the length of this match exceeds the threshold L, the MHC molecules can be recognized by the KIR. (B) Description of the viral evasion mechanisms. The wild type virus can evolve MHC-downregulation and a decoy ligand for KIRs. The immune escape of the virus after evolution of the immunoevasive mechanisms is modeled by decreasing the clearance probability p. The arrow-headed lines indicate the activation of the T cell or the NK cell, whereas the bar-headed lines indicate their inhibition.

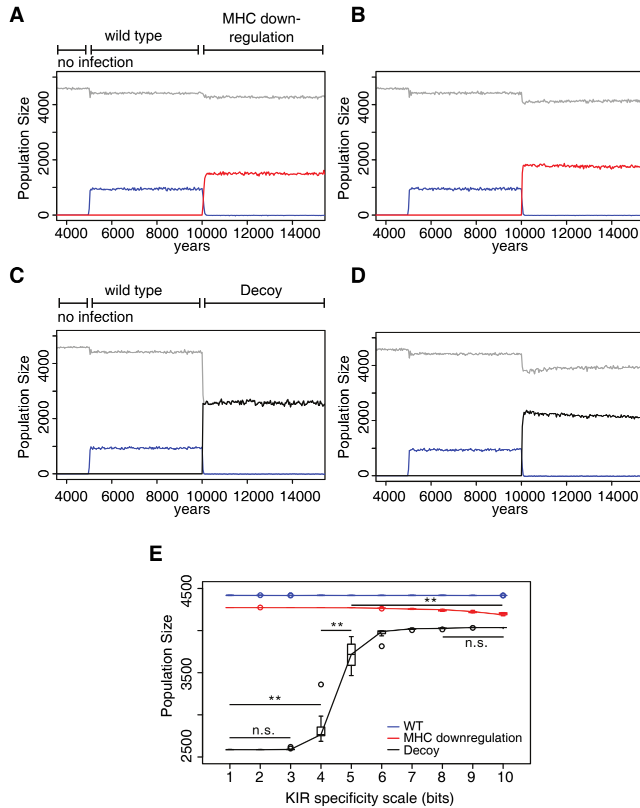

Figure 2.2. KIR specificity is protective against viruses evolving MHC-like decoys. Legend on the following page.

missing-self detection, inhibitory NK cell receptors do not need to be specific.

2.2.2. Specific inhibitory KIRs protect hosts against a virus evolving decoy molecules

To avoid elimination by the host immune response, viruses like CMV code decoy MHC molecules that can engage inhibitory NK cell receptors (Arase *et al.*, 2002, Smith *et al.*, 2002). As KIR specificity did not have a large effect on missing-self detection, we wondered whether high KIR specificity can be an adaptation to a CMV like virus. In our model, a virus down-regulating the MHC expression in one individual, can randomly select one of the MHC molecules of its host, incorporate it in its "genome", and express

Figure 2.2. KIR specificity is protective against viruses evolving MHC-like decoys. A host population was inoculated with a wild type virus after a "burn in" period of $t = 5000$ years; we allowed for mutation of the virus 5000 years after the initial epidemic. (A) The MHC down-regulating virus (red line) is fitter than the wild type virus (blue line), infecting more individuals and resulting in a larger decrease of the total population size (grey line). Here, the host population has degenerate KIRs (with $L = 1$ bit). (B) Population with a highly specific receptor-ligand recognition (with $L = 10$ bits). (C) Viruses evolving MHC down-regulation and decoy molecules (black line) take over the wild type virus (blue line), infecting more individuals and resulting in a larger decrease of the total population size (grey line). In a host population with degenerate KIRs, i.e., $L = 1$ bits, almost all individuals are infected with the virus. (D) The spread of the virus is reduced in host populations with specific KIRs, e.g., $L = 10$ bits. (A)-(D) single representative simulations of the first 15 000 years of the simulations. (E) Ten simulations were carried out per specificity value ranging from $L = 1$ bit to $L = 10$ bits. At the end of each simulation (i.e. 200 000 years), the mean of the total population size over the last 100 000 years was taken as a measure of the protection level of the host population. Protection against the wild type virus (blue) was not dependent on the specificity of the KIRs. Protection against an MHC down-regulating virus (red) was slightly, yet significantly worse in simulations with specific KIRs (i.e. $L \geq 7$), whereas protection against the virus evolving decoy molecules (black line) increased drastically in simulations with specific systems (i.e. $L \geq 5$). (** represent p values < 0.005 and were calculated using the Mann-Whitney U test. The boxes represent the interquartile range, the thick horizontal lines represent the median, and the circles are outliers.).

it as a decoy protein in the current and subsequent hosts. While in the current host this decoy is always successful, in the subsequent hosts its success will depend on the specificity of the KIRs. Viruses carrying successful MHC decoys can escape the immune response of both T and NK cells. The fitness cost of a host infected with one of these successful viruses was modeled by decreasing the probability of clearing the infection to zero (Figure 2.1 B). Thus, each individual with KIRs recognizing a foreign viral decoy like self MHC, became chronically infected in the model.

The better adaptation of a decoy virus compared to the MHC down-regulating virus was reflected in a higher fraction of chronically infected individuals and in a lower population size (Figure 2.2 C). But, opposite to what we observed with the virus down-regulating MHC, the effect of KIR specificity was drastic. The average population size increased from 2500 individuals in a degenerate system to 4100 in a very specific system ($p = 0.0019$, Mann-Whitney U test). Populations having specific KIR-MHC interactions were thus much better protected than those with degenerate or cross-reactive KIRs (Figure 2.2 C, D, E).

Why is a highly specific KIR-MHC interaction advantageous in this CMV like infection? To protect the host, KIRs face the challenge to detect MHC down-regulation but not re-

cognize the viral decoy masking MHC down-regulation. As seen in the previous section, a host with degenerate KIRs always has a large repertoire of licensed KIRs, and therefore always succeeds in detecting missing-self. But because of the same low specificity, the KIRs within that individual are expected to also recognize foreign decoy molecules as self MHC. On the other hand, a specific KIR system results in a smaller repertoire of licensed KIRs per individual, impeding the host's ability to detect missing-self (see Table 2.1 and previous section). However, because of their high specificity, it is also unlikely for any licensed KIRs to recognize foreign decoys. Therefore, a decoy virus typically fails to escape NK immune responses, allowing the infection to be cleared. Again, these results were consistent in all simulations we ran for each specificity setting ($L = 1 - 10$, Figure 2.2 E black line), showing that KIR specificity helps protecting individuals against viruses evolving MHC-like molecules.

2.2.3. KIR diversity evolves as a consequence of high specificity

We next studied the effect of specificity on the evolution of KIRs. To estimate the diversity of KIR molecules in the population, we calculated the Simpon's Reciprocal Index (SRI) (Simpson, 1949). The SRI is a diversity measure that is equal to the total number of KIR alleles if they are all equally distributed in the population, whereas the SRI is lower than that in a population where some alleles dominate (described in Material and Methods). This measurement of diversity has the advantage that it is not sensitive to fluctuations in the frequencies of rare KIR alleles in the population.

KIR polymorphism remained low in populations having degenerate and cross-reactive KIRs (i.e. $L < 5$), whereas it increased significantly in populations with specific KIR-MHC interactions (Figure 2.3 A-B). Why is there only selection for diversity in those populations having specific KIRs? Since every host needs to recognize at least one of its MHC-molecules to have a licensed KIR repertoire, and this is guaranteed with degenerate KIRs, there is hardly any selection pressure in these populations to evolve novel KIR molecules. But with specific KIRs, individuals do not always recognize their own MHC and hence they are more vulnerable during the infection with an MHC down-regulating virus. The chance of recognizing self MHC is higher in individuals carrying two different haplotypes of inhibitory KIRs. Therefore, heterozygous hosts have an advantage over homozygous hosts, an effect that becomes larger with increasing KIR specificity. We conclude that this "heterozygous advantage" is the main selection pressure driving the evolution of novel KIR haplotypes, and that the selection pressure is largest in populations with specific KIR-MHC interactions.

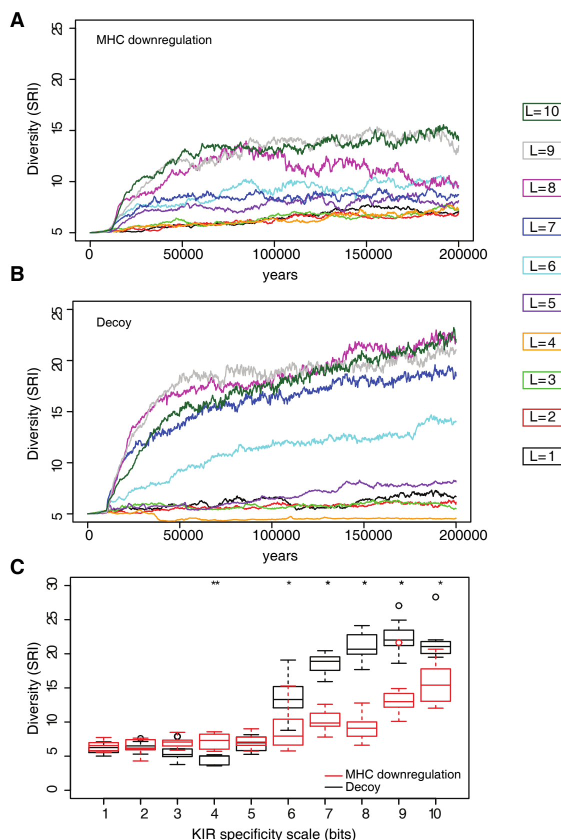

Figure 2.3. KIR polymorphism evolves only in populations having specific KIRs. The degree of KIR diversity (expressed as the Simspon's Reciprocal Index) increased only in populations having specific KIRs (i.e., for $L \geq 5$), whereas in populations with degenerate KIRs it remained close to its initial value, reflecting largely neutral drift. The time plot in (A) represents the KIR diversity in a population infected with an MHC-downregulating virus, and the time plot in (B) shows the KIR diversity in populations infected with viruses evolving decoy molecules. Each line represents the mean of the SRI score, i.e. averaged over 10 different simulations. The different colors represent the specificity values ranging from $L = 1$ to $L = 10$ bits. (C) Mean SRI score at the end of the simulation for each specificity value. There is a significant difference in the SRI score between the infection with an MHC down-regulating virus (red) and an infection with a virus evolving decoy molecules (black) (* represent p values < 0.05 and ** p values < 0.005, Mann-Whitney U test. The boxes represent the interquartile range, the thick horizontal lines represent the median, and the circles are outliers.).

2.2.4. Specific and orthogonal KIR haplotypes emerge after infections with CMV-like viruses

We argued that heterozygous advantage drives the selection of novel KIR molecules in populations with specific KIR-MHC interactions (i.e. $L \geq 5$). Yet, the KIR diversity

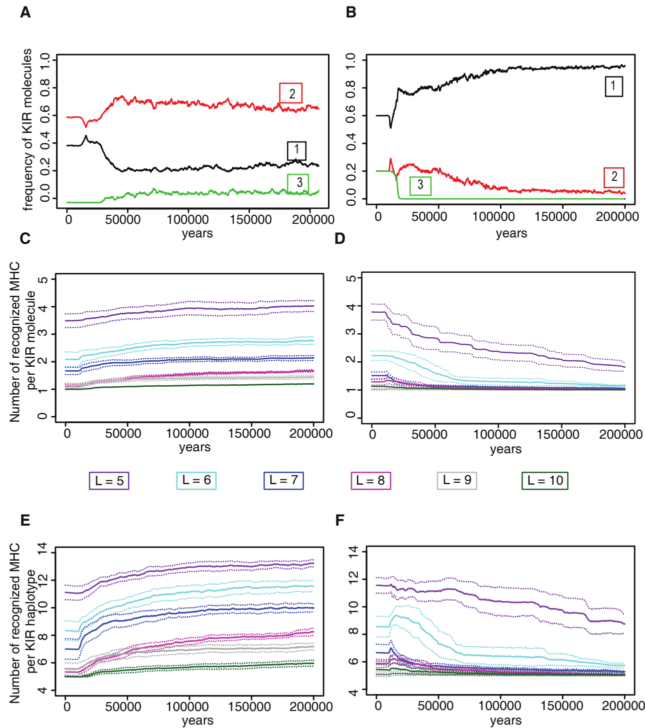

Figure 2.4. Viruses that down regulate MHC expression select for degenerate KIR molecules while viruses expressing decoy molecules drive the selection of specific receptors. Legend on next page.

differed significantly between simulations with MHC down-regulating and decoy viruses for exactly the same specificity ($p = 0.003$, and $p = 0.0002$, Mann-Whitney U test, Figure 2.3 C). This result was surprising, because the heterozygous advantage in the likelihood of recognizing self MHC should be equally strong in both types of infection. To address the possible mechanisms underlying this result, we studied the KIR molecules that were being selected after an infection with the CMV-like virus.

First, we analyzed the specificity of the KIR molecules. Although the specificity threshold L was fixed, some KIRs happened to recognize more MHC molecules than others. In populations with highly specific KIR-MHC interactions (e.g. $L = 7$), the initial haplotype was composed of 5 KIRs, each of them recognizing a different number of MHC in the population (Figure 2.4 A,B). This distribution remained constant until the mutant viruses emerged. When the CMV-like virus was introduced, there was a

Figure 2.4. Viruses that down regulate MHC expression select for degenerate KIR molecules while viruses expressing decoy molecules drive the selection of specific receptors. (A)-(B) Single representative simulations in a population infected with highly specific KIRs (i.e. $L = 7$ bits). (A) The number of MHC molecules recognized per receptor in a population infected with an MHC down-regulating virus. The initial KIR molecules happen to recognize different numbers of MHC molecules, as indicated by the numbers in the boxes. 40 % of the KIR molecules recognized one MHC molecule (black line), and 60 % recognized two MHC molecules (red line). Upon infection, new KIR receptors that happened to be more cross-reactive, i.e., recognizing three MHC molecules (green line) evolved, while the most specific KIRs decreased in frequency. (B) The specificity of the KIRs in a population infected with decoy viruses. 60 % of the inital KIR molecules recognized one MHC molecule (black line), 20 % recognized two MHC molecules (red line), and 20 % happened to recognize three different MHC molecules (green line). Over time, the frequency of those KIR molecules recognizing only one MHC molecule increased and they take over the population. (C)-(D) The average number of MHC molecules recognized per KIR molecule. In populations with highly specific KIRs ($L \geq 5$ bits), this number decreases if the individuals are infected with a virus down-regulating MHC expression (C), whereas it increases in populations infected with decoy viruses (D). (E)-(F): The average number of MHC molecules recognized per haplotype. We observed a similar behavior as in (C)-(D). The heavy lines represent the mean number of recognized MHC molecules over ten simulations. The different colors represent the specificity values ranging from $L = 5$ to $L = 10$ bits. The standard deviation is depicted by the dashed lines.

clear selection for those KIR molecules that were most specific, i.e. KIRs recognizing only one MHC molecule in the population (Figure 2.4 B). Similarly, populations infected with an MHC down-regulating virus evolved more cross-reactive KIR molecules (Figure 2.4 A). Thus, although the specificity threshold L was fixed, the system exploited the stochastic variation in cross-reactivity among KIR molecules, evolving towards the specificity that rendered most protection (Figure 2.4 C,D). During an infection with decoy viruses, this selection started already in populations having intermediate specific KIRs (i.e. $L = 4$). Surprisingly, a higher specificity was achieved by haplotypes implementing duplicate KIR genes, which effectively decreased the number of loci (Figure S2.2). The fact that the evolution of an even higher specificity affects the heterozygote advantage (Figure S2.1) explains the variation in KIR diversity between the infections with MHC down-regulating and decoy viruses (2.6.1-2.6.2).

If evolution selects for the most specific KIRs to protect against viruses evolving decoys, the challenge of recognizing self MHC is even larger. Is there any mechanism that allows for a higher chance of detecting self despite a high KIR specificity? To address this question, we studied the KIR haplotypes before and after the infection with a decoy virus, and again measured the number of recognized MHC per haplotype. In populations having specific KIRs (e.g. $L = 7$), a randomly generated initial haplotype recognized an average

of 8 MHC molecules, reflecting the expected cross-reactivity of its KIRs. Upon infection with decoy viruses, more specific haplotypes evolved due to the evolution of more specific KIR molecules. At the end of the simulation, approximately 80 % of the haplotypes recognized only five different MHC molecules in the population (Figure S2.3); a surprising result because this property is only expected for 45 % of the randomly created KIR haplotypes with $L = 7$. Hence, there was a clear selection for haplotypes that overlapped as little as possible in their MHC recognition, while keeping the highest specificity per KIR molecule. By evolving such “orthogonal“ haplotypes, the paradox of recognizing as many MHC in the population as possible without detecting foreign decoy molecules was solved. Together, our analysis suggests that, if the specificity of KIR increases, it becomes beneficial to have more loci to be able to detect missing-self, which provides an explanation for the observed polygeny in the KIR complex.

2.3. Discussion

The exact evolutionary advantage of the highly diverse KIRs has remained intriguing, especially because MHC class I detection, i.e. “missing-self” detection, can also be achieved by a limited set of monomorphic receptors. Our results show that for simple detection of MHC down-regulation, degenerate KIR molecules are advantageous, while a specific KIR-MHC interaction protects hosts against viruses evolving decoy molecules. In the presence of viruses expressing decoy molecules, the KIR became very specific, while at the same time the number of recognized MHC molecules per haplotype was maximized. The more specific the system becomes, the stronger the selection pressure on hosts to carry two different KIR haplotypes. As a result of this heterozygote advantage, KIR haplotypes evolve a high degree of diversity.

The results with viruses evolving decoy molecules depend strongly on the implemented MHC dependent NK education process, as we allow only for the “licensed” KIRs to participate in the immune response. Inhibitory receptors for MHC class I are very important for the education, repertoire development, and response of NK cells, and there is indeed good evidence that the failure to engage inhibitory receptors during development results in peripheral NK cells that are hyporesponsive (Anfossi *et al.*, 2006, Chalifour *et al.*, 2009, Fernandez *et al.*, 2005, Johansson *et al.*, 2005, Yokoyama and Kim, 2006). Yet, recent studies (Elliott *et al.*, 2010, Joncker *et al.*, 2009, Orr *et al.*, 2010) showed that NK cell populations that cannot ligate their inhibitory receptors –either because they are unlicensed cells, or because they have been transferred into a different MHC class I

environment– respond in a normal inflammatory manner after viral infection. This response of unlicensed NK cells appeared to be even more robust and protective than that of licensed NK cells. In all these studies, the NK cells were stimulated via their activating receptors. However, we only considered inhibitory receptors, modeling "functionality" as the capacity of the mature NK cells to respond to cells in which the expression of self MHC class I is decreased. MHC independent mechanisms for NK cell activation, such as activation via cytokines, is implicit in the model, and is taken into account in the overall probability of clearing the infection. Also note that we do not model the KIR molecules on individual NK cells, but define which KIRs are licensed in a host's whole repertoire.

The KIR system has evolved unusually rapidly, resulting in different levels of specificities across species. While KIRs in rhesus macaque have a broad specificity, orangutans, chimpanzees, and humans have evolved more specific KIR systems (Older Aguilar *et al.*, 2010, 2011). KIR recognition in humans is restricted to at least four epitopes (HLA-A11,-Bw4,-C1, and -C2), where HLA-C1 and -C2 have the highest avidity. The evolution of these particular MHC epitopes have left an imprint on the evolution of the KIR system. This is clearly shown in the differences of the KIR haplotypes starting from old world monkeys to humans (Guethlein *et al.*, 2007a). Humans, chimpanzees, bonobos, gorillas, orangutans, and rhesus macaques share four lineages of KIR genes, which expanded approximately 35–40 million years ago (Averdam *et al.*, 2009). Within these lineages, each species has independently evolved different numbers of KIRs with either an inhibiting or activating function, and their emergence is related to the evolution of their ligands. The expansion of lineage III KIRs in orangutans, chimpanzees, and humans is associated with the emergence of the C1-, and later with the C2 epitopes. In contrast, rhesus macaques have expanded lineage II KIR genes, corresponding to their complex MHC system, which is composed of several subsets of differentially expressed MHC-A, and -B genes in the absence of an MHC-C locus. Here, we do not model the four KIR epitopes present on several MHC alleles, but have randomly made MHC molecules. The composition of KIR haplotypes, as well as the evolution of novel MHC molecules, is not the focus of this manuscript. Our main question is why KIRs have evolved a degree of specificity, and our approach clearly reveals that specificity has a selective advantage because of its protective effect against CMV-like viruses evolving MHC decoys. Some of the high specificity values used in our model might seem contradictory to the small number of MHC epitopes identified as main KIR ligands. Yet, all results presented here are already obtained at specificity values $L = 5$, which corresponds to a recognition of 20 % of MHC molecules in the population (see Table 2.1), and is in agreement with the four MHC epitopes that have been identified so far in human KIRs.

Hosts are exposed to multiple challenges during their life span, and the immune system has evolved to respond to all of them rather than adapt to only one particular virus. For simplicity, we here consider only one type of infection at a time. Nevertheless, CMV seems to have an important role in NK cell mediated immunity. Recent studies revealed that there is a strong imprint in the NK cell repertoire of CMV seropositive individuals because a particular subset of NK cells with "self-specific" inhibitory KIRs is expanded (Béziat *et al.*, 2013, Charoudeh *et al.*, 2013). Furthermore, it has been shown that CMV plays an important role in viral driven evolution of NK cell receptors in mice (Arase and Lanier, 2002, Sun and Lanier, 2009). Mice possess the Ly49 receptor system, which is functionally similar to KIR but evolutionary and structurally different. The Ly49 receptors exhibit also high genetic diversity and also have mouse MHC-class I molecules as ligands. Mouse strains that are resistant to MCMV carry an activating receptor, Ly49H, binding to the "MHC-class I decoy" m157 with high affinity. Mice susceptible to MCMV lack the Ly49H gene but possess the inhibiting receptor Ly49I also binding strongly to the m157 glycoprotein. Because Ly49H evolved from its inhibitory homologue, Ly49I (Abi-Rached *et al.*, 2010), it seems that the m157 induced immune pressure led to the evolution of a new activating NK cell receptor, conferring resistance to the virus. Our results agree with this data, showing that a CMV encoded MHC-like decoy imposes a selection pressure to drive the evolution of novel NK cell receptors.

Fighting pathogens and successful reproduction are two crucial functions for survival. By their contribution to immune defense and reproduction, KIRs reveal various selection pressures imposed on NK cells, emphasizing the importance of diversity for surviving population bottlenecks and infections. For these reasons, it may seem intuitive that receptor diversity is beneficial for viral control. But we have seen that the mere detection of missing self is achieved best with degenerate KIRs.

Our agent-based model provides a solid explanation for one selection pressure driving the evolution of specific KIRs, namely viruses expressing MHC decoys. This does not need to be the only explanation, and our findings call for further studies into other possible mechanisms. The evolution of the specificity and number of loci per haplotype, as well as the evolution of activating receptors or other viral strategies, should now be integrated in our model to address additional questions.

2.4. Material and Methods

2.4.1. Agent-based model

We developed an agent-based model consisting of two types of actors (hosts and pathogens) and three types of events (birth, death, and infection). The basic time step of the model is one week, during which we run through all hosts in a random order and confront them to one of the randomly chosen events. Hosts age over time, and after each time step, their age, infection state, and infection type is updated. The cycle is repeated for many hosts generations to model the long-term evolution. All model parameters are fully described in Table 2.2. The following is a detailed description of the actors and the events:

Hosts and viruses

Hosts are modeled as simplified humans with a diploid genome encoding for one MHC class I locus and one KIR haplotype composed of five different inhibitory genes. For simplicity, we only model one MHC locus with 14 different possible alleles. KIR and their ligands are represented as bit strings, as a simplified representation of amino acid sequences (Farmer *et al.*, 1986). The simplest way to think about this representation is to see the proteins as being composed of two types of amino acids, e.g. hydrophobic, and hydrophylic. Interactions between two molecules are only allowed if the complementary matches between the strings exceed a predefined threshold. Since two molecules may interact in more than one way, the strings are allowed to match in any possible alignment. In our model we use a predefined threshold L, i.e. if the length of the longest adjacent complementary match between a ligand and a KIR is at least L bits long, we consider the ligand to be capable of binding to its receptor (Figure 2.1 A).

In our simulations, we consider in total 16 viruses: a wild type virus, one MHC down-regulating virus, and 14 decoy viruses. Each virus comes with a particular increase in the natural intrinsic death rate of its host VL_i, and a probability of clearing the infection p_{wt}, p_{mhc}, and p_{dec}. To escape an assumed T-cell response, the wild type virus evolves MHC down-regulation with a mutation rate μ_v. The decoy viruses evolve, in addition to MHC downregulation, a decoy protein mimicking an MHC molecule. We model the evolution of decoy molecules by allowing the virus to adopt a randomly selected MHC molecule from its host with the same rate μ_v. We do not allow for viruses to accumulate

these decoy proteins: when viruses lose the previous decoy, they can adopt a new MHC molecule. Because we fix the polymorphism of the MHC molecules to 14 alleles, the maximal number of decoy proteins that can evolve in the population is 14. We also do not allow for superinfection and hosts can be infected with one of the 16 viruses only. The immune escape of the mutant viruses is modeled by decreasing the initial probability of clearing the infection from p_{wt} to p_{mhc}, and p_{dec} (Figure 2.1 B, and Table 2.2).

The wild type virus is introduced once the population dynamics have reached a steady state. Similarly, the viral mutants are allowed to evolve 5000 years after introduction of the wild type virus.

Birth

We consider sexual reproduction, where each selected host passes one of its KIR haplotypes to its offspring. To maintain the initial polymorphism and the relative frequency of MHC alleles in the host population, we let the offspring inherit one MHC gene from one randomly selected parent while the other one is drawn out of a predefined pool of MHC alleles (see Model Initialization). The probability of a birth event taking place is described by $b(a, N)$ and decreases linearly with the population size (i.e. we consider logistic growth). Additionally, hosts with age between 20 and 45 are considered to be more fertile than children and the elderly, and at low population sizes fertile individuals are expected to give birth every two years. The resulting yearly birth rate (shown in Figure S2.4 A) is described by:

$$b(a, N) = 0.5 \left(1 - \frac{N}{N_{\max}}\right) \left(\frac{1}{1 + e^{(a-45)}} - \frac{1}{1 + e^{(a-20)}}\right), \tag{2.1}$$

where a is the age of the parent, N the actual population size, and $N_{\max}$ maximal population size.

KIR genes undergo mutation with a probability μ. To decrease computation time we model mutations by randomly drawing a new KIR gene out of a large pool of predefined KIR alleles. Other mutational operators, e.g. point mutations, recombination, deletion or duplication are not considered here. The NK cell education process takes place at birth, where only those KIRs in the newborn that recognize at least one of its own MHC molecules, are set to be licensed. Hosts without any licensed KIR can be born, but they will be unable to detect MHC down-regulation. Newborn children are given the age of one and are added to the host population.

Death

A death event takes place with an age and infection-dependent rate $\delta(a, VL_i)$. The yearly intrinsic death rate is a mathematical approximation of a human age-specific mortality curve (Carnes *et al.*, 2006). If the host is infected with a virus, the death rate is increased by a factor VL_i, representing the effect of viral load during the different infection stages i. The resulting death rate (Figure S2.4 B) is described by:

$$\delta(a, VL_i) = e^{(0.1a-10.5)} + e^{(-0.4a-8)} + VL_i, \tag{2.2}$$

where

$$VL_i = \begin{cases} 0, & \text{if } i = 0 \text{ (susceptible)} \\ 0.10, & \text{if } i = 1 \text{ (acute)} \\ 0.06, & \text{if } i = 2 \text{ (chronic).} \end{cases} \tag{2.3}$$

Infection

During each time step we challenge all susceptible hosts with possible infections by allowing every susceptible individual to contact one randomly selected partner. If this partner is infected, it will transmit its virus with a probability p_{ac} or p_{ch} depending on the infection state of the infecting host (see Table 2.2). Properties of the virus, i.e. its evasion strategies, are not changed upon transmission.

After transmission, every individual has an acute infection for a period of t_{inf} weeks, during which the intrinsic death rate is increased by VL_1 (see equation 2.2 and equation 2.3). After the initial acute phase the virus can be cleared with a probability p_{wt}, p_{mhc}, or p_{dec}, depending on the type of virus. If the individual fails to clear the infection, it will become chronically infected, and the death rate will decrease to VL_2. Recovered individuals are resistant to the virus, losing their immunity only after a long period, t_i.

Whether a virus can spread through the population depends on the number of newly infected cases it can generate during the course of its infectious period (R_0). In our model, the number of new cases depends on the transmission probabilities, and the infection state of the host. Because the acute phase is relatively short ($t_{\text{inf}} = 4$ weeks long) and a chronic infection lasts for several years (until the infected host dies), viruses inducing a chronic infection, spread much better through the population. Therefore, the probability of escaping the acute immune response (i.e. $1 - p_{wt}, 1 - p_{mhc}$, and $1 - p_{dec}$) plays a

major role in the R_0 of the virus.

2.4.2. Model Initialization

The model was initialized with a host population of 4500 hosts, with a random age between 1 and 70 years. Gene pools for MHC and KIR alleles were created at the start of each simulation. The pool of MHC consisted of 14 alleles according to the most frequent HLA-C alleles in the European population (dbMHC Anthropology (Meyer *et al.*, 2007)). For each MHC allele, ten different KIR were randomly generated, which could bind to the MHC with a specificity of at least L, resulting in a KIR pool of "functional" 140 alleles. To create the initial genome of each host, MHC and KIR genes were randomly drawn from the pools. The individuals were initialized with the same KIR haplotype, but with different MHC genes.

2.4.3. Genetic Diversity

The Simpson's Index is a measurement of diversity that can be interpreted as the probability that two randomly chosen molecules from two random hosts in the population are identical. The lower the Simpson's Index, the higher is the diversity of molecules in the population, and the reciprocal of the Simpson's Index (Simpson, 1949) defines a "weighted" diversity. This diversity measure has the advantage over the total number of unique KIR molecules because it is less sensitive to fluctuations in molecule numbers caused by random neutral drift. For instance, if all molecules are equally frequent in a population, the SRI score is equal to the number of alleles in the population. A population dominated by a single molecule will have an SRI score close to 1. The SRI was calculated as follows: $\mathrm{SRI} = \frac{1}{\sum_{i=1}^{K} {f_i}^2}$, where f_i is the fraction of the molecule i over all KIR molecules in the population, and K is the total number of unique KIR molecules.

2.4.4. Analytical Expectations

The probability that a host having a heterozygous diploid genome recognizes its own MHC molecules is defined by $p_{hap} = 1 - (1 - p)^{2N_{\mathrm{KIR}}}$, where p is the probability that a KIR recognizes a random MHC molecule in the population (which depends on L, see Table 2.1), and N_{KIR} is the number of KIR loci. The expected number of licensed KIR is determined by $n = 2N_{\mathrm{KIR}}(1 - (1 - p)^2)$. In our model each host has a genome

consisting of one MHC locus and five KIR loci, i.e. $N_{\text{KIR}} = 5$, hence that individual will recognize its own MHC molecules with a chance $p_{\text{hap}} = 1 - (1-p)^{10}$, and the expected number of licensed KIR for the same individual will be $n = 10(1 - (1-p)^2)$. The expected protection against a decoy virus, i.e. the probability of not recognizing the viral protein as self MHC molecule, depends on the size of the licensed KIR repertoire, n, and is described by $P = (1-p)^n$.

Heterozygous advantage is defined as: $HA = \frac{p_{\text{hap}}(het)}{p_{\text{hap}}(hom)}$, where $p_{\text{hap}}(het)$ and $p_{\text{hap}}(hom)$ represent the probability of recognizing self MHC molecules for a heterozygote and a homozygote individual, respectively. We obtained $p_{\text{hap}}(hom)$ by measuring the fraction of MHC molecules detected by a single KIR haplotype. To obtain $p_{\text{hap}}(het)$ we measured the fraction of recognized MHC by all pairwise combinations of KIR-haplotypes. The population has heterozygote advantage if $HA > 1$. Values of HA for different L, n, and p are given in Table 2.1.

2.4.5. Implementation

The model was implemented in the C++ programming language. The value of L was varied in a range from one to ten. For each value L, ten simulations were performed for 2000 centuries.

2.5. Acknowledgments

We thank Ronald Bontrop, and Hanneke van Deutekom for helpful discussions, technical support, and carefully reading the manuscript. This work was financially supported by the Netherlands Organization for Scientific Research (NWO), grant grant 635.100.025.

Table 2.2. Model Parameters

Parameter	**Value**
Time step	1 week
Simulation time	2000 centuries
Host parameters	
Maximal population size $N_{\max}$	4980 individuals
MHC diversity	1 locus with 14 alleles
Number of KIR loci	5
Bit string length	16 bits[a]
Specificity L (bit scale)	1–10 bits
Host mutation rate μ	0.00005 per allele per birth event
Infection[b]	
Infection state i	0 (susceptible), 1 (acute), 2 (chronic)
Probability of viral transmission during acute phase p_{ac}	0.85 per contact
Probability of viral transmission during chronic phase p_{ch}	0.15 per contact
Probability of clearing the wild-type virus p_{wt}	0.85
Probability of clearing the MHC downregulating virus p_{mhc}	0.75
Probability of clearing the virus evolving decoy molecules p_{dec}	0
Immunity time t_{i}	10 years
Acute infection time t_{inf}	4 weeks
Virus parameters	
VL_i	0 (for $i = 0$), 0.1 (for $i = 1$), 0.06 (for $i = 2$)
Virus mutation rate μ_v	0.0001 per week
Initial conditions	
Initial population size N_{init}	4500 individuals
KIR initial diversity (SRI)	5 (1 allele per locus)

a: The choice to use 16-bit strings represents a large enough theoretical repertoire of 65,536 sequences. *b*: The parameters used for the infection are chosen to maintain the epidemic. Changing the length of the acute phase or the probabilities of clearance do not affect our results on the evolution of the KIRs qualitatively (results not shown).

2.6. Supporting Information

2.6.1. Heterozygous advantage is higher in populations infected with viruses evolving decoy proteins

We measured significant differences in the SRI score between the viral strategies when KIRs are specific (i.e. $L \geq 5$), and for one intermediate specificity (i.e. $L = 4$). This observation strongly suggests that the different viral evasion routes impose different selection pressures. To study how this comes about, we first measured the fraction of MHC molecules recognized per haplotype (Figure S2.1 A), and calculated the advantage of heterozygotes over homozygotes in the detection of MHC molecules (Figure S2.1 B).

The observed haplotype specificity and heterozygous advantage confirmed our analytical expectations only in degenerate KIR systems, whereas, in specific systems, haplotypes evolved a different specificity than the threshold L (Figure S2.1 A). In response to the MHC-downregulating virus, KIR haplotypes recognizing more MHC molecules were selected. But during an infection with the decoy viruses, KIR haplotypes recognizing less MHC molecules dominated the population. This is in agreement with our previous observations that more cross-reactive KIRs evolve after an infection with an MHC down-regulating virus, and more specific KIRs evolve during an infection with a decoy virus.

Given that the heterozygous advantage becomes larger with increasing specificity (Figure S2.1 B), it is natural that the level of diversity is higher in populations infected with decoy viruses than in populations infected with an MHC down-regulating virus. Summarizing, these results show that the actual heterozygous advantage depends on the specificity of the KIRs, which can be higher or lower depending on the virus type the population is infected with.

The observed significant difference in the SRI in intermediate specificity $L = 4$ arises from yet another strategy. These systems evolved to become more specific by incorporating KIR genes as duplicates within the same haplotype (Figure S2.2). By decreasing the number of unique KIR genes per haplotype, the number of recognized MHC molecules per haplotype is lower, and the overall probability of recognizing the decoy decreases.

2.6.2. Haplotypes with maximal MHC coverage limit heterozygous advantage

The evolved orthogonal haplotypes have a strong effect on the heterozygous advantage (Figure S2.1 B, red line). In our model, a single orthogonal haplotype recognizes exactly five different MHC molecules because it is composed of highly specific KIR molecules recognizing only one of the 14 MHC in the population.

A single haplotype is expected to detect 5 out of 14 of the MHC molecules in the population (Figure S2.3). Consequently, the advantage that a heterozygous host carrying these complementary haplotypes has over a homozygote is smaller than the heterozygous advantage of host carrying non orthogonal haplotypes. Hence, the selection pressure reduces somewhat in orthogonal cases, and it reaches a plateau at $HA = 1 - \frac{(1-0.4)^2}{0.4} = 1.65$.

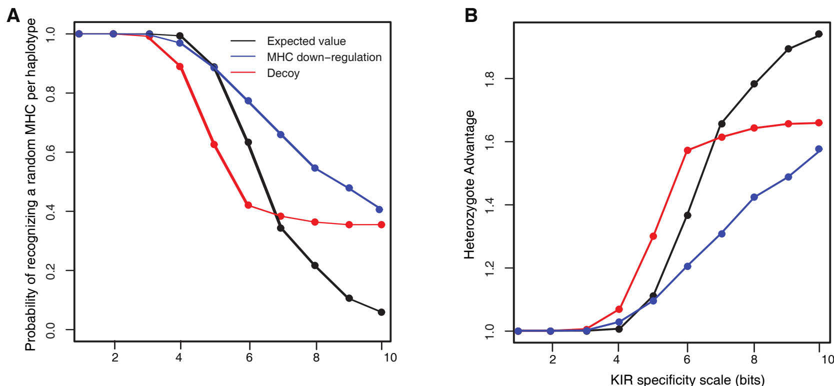

Figure S2.1. Heterozygous advantage is higher in populations infected with a virus evolving decoys. (A) Probability of recognizing a randomly selected MHC molecule per KIR haplotype, as a function of the KIR specificity L. The observed values after infection with an MHC down-regulating virus (blue line) or viruses evolving decoy molecules (red line) differ from the expected probability value (black line). (B) Heterozygous advantage (i.e. $\frac{p_{\text{hap}}(het)}{p_{\text{hap}}(hom)}$, see Material and Methods) as a function of the specificity L. KIR evolve to become more degenerate after infection with an MHC-downregulating virus, therefore the selection pressure (blue line) is lower than expected (black line). In contrast, KIR molecules evolve a higher specificity upon infection with a virus evolving decoys. Hence, heterozygous advantage is higher in populations infected with a virus evolving decoy molecules (red line). Data represent the mean out of ten simulations per specificity value.

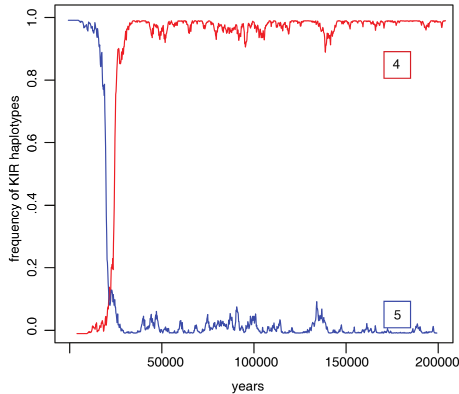

Figure S2.2. KIR haplotypes decrease their number of loci. Upon infection with a virus expressing decoy proteins (occurring at $t = 10000$), the initial haplotype composed of five unique KIR molecules (blue line) decreases in its frequency. The original haplotype is taken over by a new haplotype having only four unique KIRs, i.e., one KIR is a “duplicate“, which is indeed a strategy to become more specific. Single representative simulation of a population with intermediate specific KIRs ($L = 4$ bits).

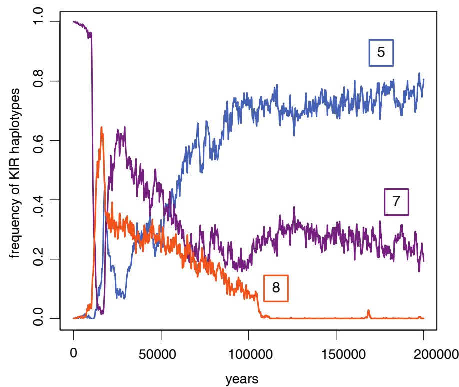

Figure S2.3. Viruses expressing decoy molecules drive the selection of specific “orthogonal” haplotypes. The time plot depicts a single representative simulation of a population with highly specific KIRs (i.e. $L = 7$ bits). The initial KIR haplotype happened to recognize exactly seven different MHC molecules. Upon infection with a virus expressing decoy proteins, new haplotypes recognizing different number of MHC molecules emerged. Haplotypes recognizing fewer MHC alleles increased over time and in this example the population acquired haplotypes covering five or seven MHC molecules.

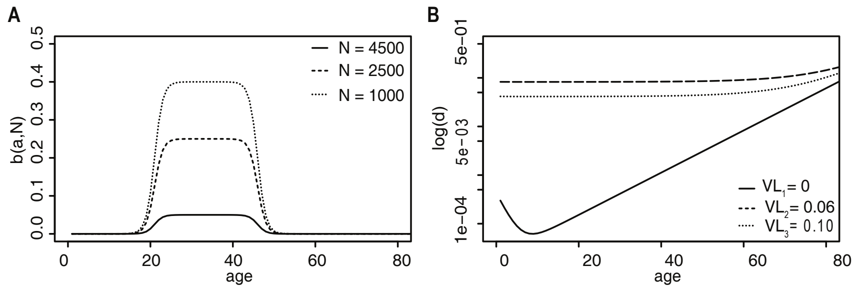

Figure S2.4. Birth and death rate of the model. (A) The age-dependent birth rate of a host population. The different lines show how the birth rate changes with different population size N. (B) The different lines depict the influence of the viral load VL_i on the intrinsic age-dependent death rate of a host population.

Chapter 3

Quantifying the protection of activating and inhibiting NK cell receptors during infection with a CMV-like virus

Paola Carrillo-Bustamante, Can Keşmir, Rob J. de Boer (2014)

Frontiers Immunology, 5:20

Abstract

The responsiveness of natural killer (NK) cells is controlled by balancing signals from activating and inhibitory receptors. The most important ligands of inhibitory NK cell receptors are the highly polymorphic major histocompatibility complex (MHC) class I molecules, which allow NK cells to screen the cellular health of target cells. Although these inhibitory receptor-ligand interactions have been well characterized, the ligands for most activating receptors are still unknown. The mouse cytomegalovirus (MCMV) represents a helpful model to study NK cell driven immune responses. Many studies have demonstrated that CMV infection can be controlled by NK cells via their activating receptors, but the exact contribution of the different signaling potential (i.e. activating vs. inhibiting) remains puzzling. In this study, we have developed a probabilistic model which predicts the optimal specificity of inhibitory and activating NK cell receptors needed to offer the best protection against a CMV-like virus. We confirm our analytical predictions with an agent-based model of an evolving host population. Our analysis quantifies the degree of protection of each receptor type, revealing that mixed haplotypes (i.e. haplotypes composed of activating and inhibiting receptors) are most protective against CMV-like viruses, and that the protective effect depends on the number of MHC loci per individual.

3.1. Introduction

Natural killer (NK) cells contribute to the host immune response by recognizing and killing viral infected and tumor cells (Lanier, 2008). Their activity is controlled by balancing signals from a vast repertoire of activating and inhibiting receptors enabling them to distinguish healthy from unhealthy cells (Lanier, 2005). The most important ligands for inhibitory NK cell receptors (iNKR) are MHC class I molecules on other cells. An infected cell may have lower MHC expression, altering the binding with inhibitory receptors, disrupting the balance of signals, and allowing for NK cell activation. The mechanism by which NK cells attack MHC-class I deficient cells was coined by Kärre et al. as "missing-self" detection (Ljunggren and Kärre, 1990).

There are several NKRs that contribute to missing-self detection. In humans for example, the inhibitory receptor CD96/NKG2A binds to complexes of the human leukocyte antigen E (HLA-E), which presents peptides derived from the leader sequences of HLA-A, -B, and -C molecules (Braud *et al.*, 1998, O'Callaghan, 2000). Both, the receptor and the ligand are highly conserved in these inhibitory interactions, and the downstream effects are remarkably similar across individuals (Shum *et al.*, 2002). The killer immunoglobulin-like receptors (KIRs) also contribute to monitor abnormalities in MHC class-I expression on cell surfaces. In contrast to the CD96/NKG2 superfamily, KIRs are highly polygenic and polymorphic, exhibit both inhibitory and activating potential, and bind to the highly polymorphic HLA-A, -B, and -C molecules (Jiang *et al.*, 2012, Moesta and Parham, 2012, Trowsdale *et al.*, 2001). Consequently, the interactions between KIRs and classical HLA-class I molecules are very diverse (Vilches and Parham, 2002). Thus, humans have two types of NKRs, one conserved and one highly diverse, performing seemingly the same function.

Humans are not the only species that has an expanded, and polymorphic NKR gene complex. During mammalian radiation, many different species have diversified alternative NKR gene families recognizing MHC class I. This example of convergent evolution includes three gene families from two structurally unrelated superfamilies: KIRs, the CD94/NKG2, and the Ly49 (Kelley *et al.*, 2005). Higher primates have expanded their KIR genes (Guethlein *et al.*, 2007a); a group of lower primates have expanded NKG2 (Anfossi *et al.*, 2006), whereas rodents and equids have expanded Ly49 (Wilhelm and Mager, 2004, Wilhelm *et al.*, 2002). These alternative genetic strategies illustrate the evolutionary complexity of these systems, and suggest that an expanded NKR gene complex is beneficial for survival. But, if conserved inhibitory receptor-ligand interactions

(such as NKG2A-HLA-E in humans) are capable to successfully detect missing-self, why have several NKR families evolved to become polygenic and polymorphic? Even more intriguing, why have they evolved receptors with activating potential?

In humans, some activating NKR (aNKR) are associated with the disease outcome of viral infections and malignancies (Körner and Altfeld, 2012). For example, in combination with HLA-Bw4, the activating KIR3DS1 has been associated with a delayed progression to AIDS in HIV-1 infected individuals (Martin *et al.*, 2002b, Pelak *et al.*, 2011). KIR3DS1 has also been linked to an increase rate of spontaneous recovery after hepatitis B infections (Zhi-ming *et al.*, 2007), a reduced risk of developing hepatocellular carcinoma in patients infected with HCV (López-Vázquez *et al.*, 2005), and a reduced risk of Hodgkin's lymphoma (Besson *et al.*, 2007). Moreover, maternal activating KIRs are related to protection against several pregnancy disorders (Hiby *et al.*, 2010). But because only a few ligands for activating KIRs have been identified so far, the exact mechanisms underlying the provided protection in humans remain puzzling.

Studies in mice have revealed important insights into the role of aNKR during viral infections (Arase *et al.*, 2002, Smith *et al.*, 2002). Viruses like the mouse cytomegalovirus (MCMV) down regulate the expression of MHC-class I molecules from the cell surface to escape T cell response, and may additionally code decoy MHC molecules (m157) that can inhibit NK cell activation (Smith *et al.*, 2002). Mouse strains that are resistant to MCMV carry the activating Ly49H gene, which binds with high affinity to the MHC-like viral protein m157. In contrast, mice susceptible to MCMV lack the activating gene but carry the inhibiting receptor Ly49I, which also binds strongly to the m157 protein. The activating Ly49H emerged from an inhibitory counterpart (Abi-Rached and Parham, 2005), suggesting that the evolution of an aNKR was due to the immune pressure induced by the "MHC-decoy" m157 during CMV infection (Arase and Lanier, 2002, Sun and Lanier, 2009).

Although these studies shed light into the importance of NKR in general, the specific contribution of activating and inhibitory receptors to the NK cell response is still unknown. In Chapter 2 studied the evolution of KIR diversity in a human population infected with CMV-like viruses by using a computational agent-based model. We showed that iNKRs require sufficient specificity to protect populations against viruses evolving MHC-like molecules, and that diversity in the NK cell genetic complex evolves as a result of the required discrimination between self MHC-molecules and viral decoy molecules. Here, we also consider aNKRs, and develop a probabilistic model to quantify the optimal specificity of inhibitory and activating NKRs needed to render maximal pro-

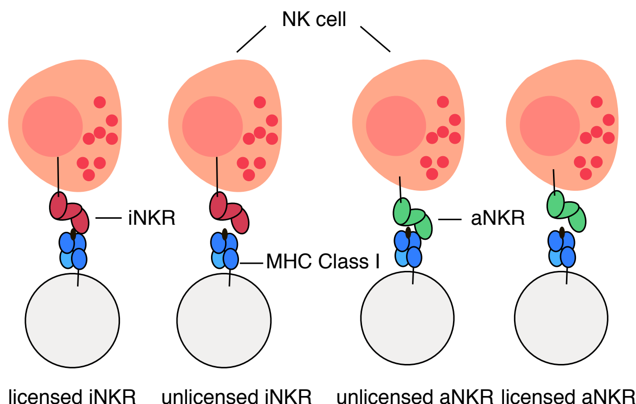

Figure 3.1. Cartoon of NK cell education. iNKRs (represented in red) recognizing at least one of the MHC molecules per individual will become licensed. In contrast, aNKRs (depicted in green) which do not recognize *any* of the host's MHC molecules will become licensed.

tection against CMV-like viruses. We also analyze the effect of mixed haplotypes (i.e. composed of aNKR and iNKR) on protection, and confirm the expectations of the probabilistic model with an agent-based computational model. Our studies reveal that mixed haplotypes composed of specific activating and inhibitory NKRs render high protection against CMV like viruses encoding for decoy molecules, and that the protective effect depends on the number of MHC loci per individual.

3.2. Results

We analyze the effect of the specificity of activating and inhibitory NKRs on the detection of a virus presenting MHC-like molecules with a simple probabilistic model. Our model estimates the chance of protection P, i.e. the probability of a host detecting an infection by NK cells, as a function of the haplotype size, specificity (i.e. the probability p of recognizing any random MHC molecule), and number of MHC loci.

The responsiveness of NK cells (i.e. their ability to discriminate cells with normal MHC expression from those lacking MHC) is regulated by a process called "education" or "licensing" taking place during NK cell development (Elliott and Yokoyama, 2011). During this process, the interactions of iNKRs with their MHC ligands render the NK cells with functional competence (Anfossi *et al.*, 2006, Chalifour *et al.*, 2009, Elliott and Yokoy-

ama, 2011). To prevent NK cell related autoimmunity, activating receptors also participate in the education process, where the chronic exposure of aNKR ligands during development results in hyporesponsive NK cells (Fauriat *et al.*, 2010, Sun and Lanier, 2008).

For simplicity, we do not model individual NK cells, each expressing a random set of tuned receptors. We rather consider for each individual a global repertoire of receptors, which have the potential to license NK cells. We will refer to these receptors henceforth as "licensed" receptors. We mimick the MHC-dependent education process during NK cell development by creating a repertoire of NKRs composed of iNKRs that recognize at least one of the MHC molecules of the host, and of aNKRs that recognize none of the MHC molecules of the host (Figure 3.1). By considering a global repertoire, we assume that there will be at least one subset of NK cells expressing at least one of the "licensed" NKRs. Upon infection, we consider only those NK cell subsets having licensed receptors. If these can successfully detect the virus, they will become activated, expand, and protect (for a full discussion see Supplementary Information). Therefore, only the licensed repertoire of NKRs is allowed to participate in the immune response.

Whether a decoy protein allows a virus to successfully escape the NK response, i.e. whether that individual is protected against the infection, depends on the receptor type and on the receptor specificity. iNKRs that bind the decoy molecule cannot detect missing-self and are "fooled" by the decoy. Conversely, aNKRs binding the foreign decoy protein can specifically recognize the infection and therefore protect the host. With this model we quantify the contribution of each receptor type and its specificity to the detection of CMV-like viruses.

3.2.1. Inhibitory and activating NK cell receptors differ in the protection level they provide

There are two crucial processes for a single iNKR to detect a virus evolving MHC-like molecules. First, iNKRs have to be licensed during the NK cell education to become fully functional during an immune response. Second, iNKRs should not bind decoy molecules upon infection. Because, in our model, iNKRs are only licensed if they recognize at least one of the MHC molecules in their host, and decoy molecules are similar to self MHC molecules, iNKRs face the challenge of distinguishing self MHC molecules from foreign decoy molecules. We previously demonstrated that this challenge can be solved by evolving sufficiently specific iNKRs (see Chapter 2). In that study, we defined spe-

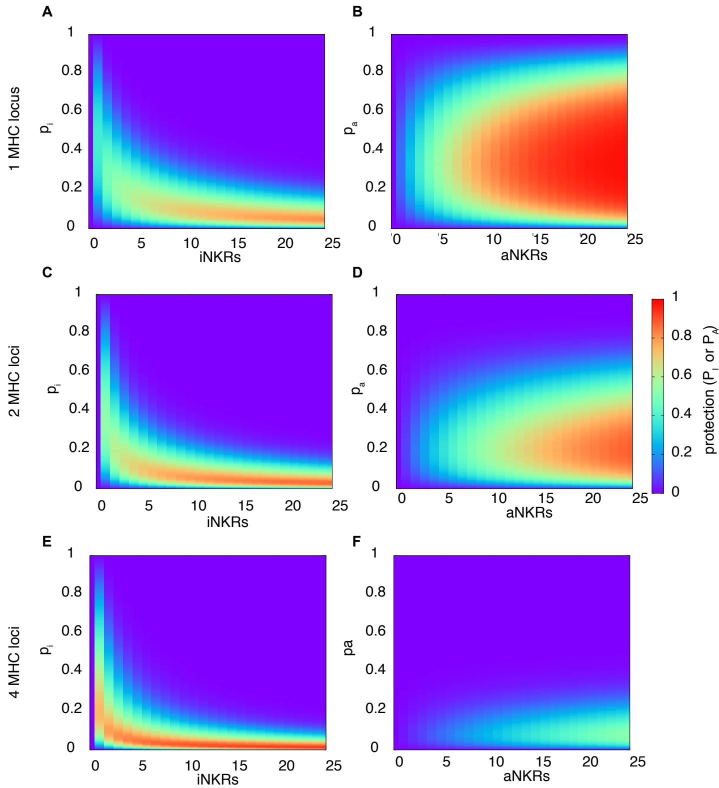

Figure 3.2. Range of protection differs between iNKRs and aNKRs. The heatmaps show the protection level as the probability of detecting the infection with a virus expressing a decoy molecule to mask MHC down regulation. In the left column, the protection for individuals carrying only iNKRs (calculated by Eq. 3.2) is shown, whereas in the right column, the protection for individuals carrying only aNKRs (calculated by Eq. 3.3) is depicted. The protection level is shown in the color bar from highest (red) to lowest (blue). p_i, and p_a correspond to the specificity of iNKR and aNKR, respectively. (A)-(B) The protective range for iNKRs is small and skewed towards a large haplotype size and high specific values. In contrast, aNKRs offer a broad range of protection for intermediate specificity values and a smaller haplotype size. Calculations were done with 1 MHC locus (A-B), 2 MHC loci (C,D), and 4 MHC loci (E,F).

cificity as the probability (p) of any NKR to recognize a random MHC molecule in the population. Herewith, degenerate receptors (i.e. with $p = 1$) are able to recognize all MHC molecules in the population, whereas specific receptors (i.e. with $p \ll 1$) recognize only a small fraction of them. Since the exact relation between ligand-receptor binding

affinity and signaling potential remains unknown, we do not consider different binding affinities here, and we model discrete MHC-NKR interactions.

To study whether there is an optimal specificity for which iNKRs are not inhibited by such "decoy viruses", we calculated the probability of licensed iNKRs detecting the infection. A single iNKR becomes licensed with a probability $q_{\mathrm{I}} = 1 - (1 - p_{\mathrm{I}})^{2N_{\mathrm{MHC}}}$, where p_{I} describes the specificity (i.e. the probability of any iNKR to recognize any MHC in the population), and N_{MHC} the number of MHC loci per individual. The probability of a haplotype composed of N_{iNKR} to have exactly ℓ licensed iNKRs is given by the binomial distribution as follows:

$$P(\mathrm{iNKR}_{\mathrm{licensed}} = \ell) = \binom{N_{\mathrm{iNKR}}}{\ell} (1 - q_{\mathrm{I}})^{N_{\mathrm{iNKR}} - \ell} q_{\mathrm{I}}^{\ell}. \qquad (3.1)$$

To successfully detect a decoy virus, none of the licensed iNKRs should bind the decoy molecule. Thus, the overall probability of detecting the infection is determined by the chance that none of the licensed iNKRs recognizes a decoy molecule, and can be calculated by:

$$P_{\mathrm{I}}(\mathrm{detection}) = \sum_{\ell=1}^{N_{\mathrm{iNKR}}} \binom{N_{\mathrm{iNKR}}}{\ell} (1 - p_{\mathrm{I}})^{\ell} (1 - q_{\mathrm{I}})^{N_{\mathrm{iNKR}} - \ell} q_{\mathrm{I}}^{\ell}. \qquad (3.2)$$

Our analysis confirms that for any haplotype size there is an optimal specificity. For $N_{\mathrm{iNKR}} \leq 25$, our model predicts a maximal level of protection (i.e. $P_{\mathrm{I}} = 0.85$), which can only be obtained with high specificity values ($p_{\mathrm{I}} \leq 0.2$) and a large number of genes per haplotype ($N_{\mathrm{iNKR}} \geq 20$) (Figure 3.2 A).

A host with degenerate iNKRs (e.g. $p_{\mathrm{I}} \geq 0.8$) has a large repertoire of licensed iNKRs. But because of the low specificity, the iNKRs within that individual are expected to also recognize any foreign decoy molecule as self, offering no protection. In contrast, when iNKRs are specific (e.g. $p_{\mathrm{I}} \leq 0.2$) the repertoire of licensed iNKRs per individual is lower, but if there are several genes per haplotype, the chance of having at least one licensed specific iNKR increases. Due to their high specificity, it is unlikely for a licensed iNKR to also recognize a foreign decoy molecule, impeding the virus to escape the NK immune response. Therefore, an infection with a decoy virus can be controlled with a probability of at least 70% in a haplotype composed of more than 10 iNKRs when

$p_{\rm I} \leq 0.25$. Thus, the probability of detecting the virus increases with both a higher specificity, and a larger number of genes per haplotype (i.e. $N_{\rm iNKR}$). This confirms our results of Chapter 2, suggesting that large haplotypes composed of non-overlapping specific iNKRs are most protective.

We next developed a model considering only aNKRs. Similar to the iNKRs, the two crucial processes for an aNKR to detect the virus depends on the probability of becoming licensed and recognizing the decoy molecule as a foreign antigen. However, the licensing process is almost opposite between aNKRs and iNKRs. An aNKR becomes licensed if it does not recognize any MHC molecule within an individual. The probability of a single aNKR to become licensed is therefore described by $q_{\rm A} = (1-p_A)^{2N_{\rm MHC}}$, where p_A is the specificity of an aNKR. Opposite to an iNKR, an aNKR detects a "decoy virus" if it binds the MHC decoy . Thus, the overall probability of protection in this case is determined by the chance of at least one licensed aNKR binding the decoy molecule, and is given by

$$P_{\rm A}(\text{detection}) \;=\; \sum_{\ell=1}^{N_{\rm aNKR}} \binom{N_{\rm aNKR}}{\ell} \left(1-(1-p_{\rm A})^{\ell}\right)(1-q_{\rm A})^{N_{\rm aNKR}-\ell} q_{\rm A}^{\ell}, \quad (3.3)$$

where $N_{\rm aNKR}$ is the number of aNKRs per haplotype.

This model reveals that there is again an optimal specificity, and the protection range for aNKR is much broader than that for iNKR, covering also less specific receptors (i.e. $0.1 \leq p_{\rm A} \leq 0.7$) (Figure 3.2 B). In these cases, the optimal protection (i.e. $P_A = 1$) is obtained with haplotypes composed of 12 genes, having intermediate specificity values ($0.2 \leq p_{\rm A} \leq 0.65$). To avoid self-reactivity, aNKRs become licensed only if they fail to recognize all self MHC molecules. Additionally, an aNKR must recognize foreign MHC-like molecules to detect the infection. Therefore, the challenge for an aNKR is opposite to that of an iNKR, since it must recognize foreign antigens but not self-MHC molecules. A degenerate aNKR will recognize every decoy in the population but it will never become licensed. As a result, the optimal protection is reached in large haplotypes composed of aNKRs with intermediate specificity.

Note that we consider individuals to be heterozygous for all MHC loci. Allowing individuals to be homozygous in some MHC loci does not qualitatively change our results on specificity and protection, since MHC homozygocity has only a mild effect on the number of licensed receptors ℓ (results not shown).

3.2.2. The protection level depends on the number of MHC loci

Above, we considered only one MHC locus per individual as a representation of HLA-C as the main identified ligand for inhibitory KIRs. However, HLA-A, and -B molecules have also been identified as KIR ligands, and HLA-E is the ligand for CD94/NKG2A. Therefore, we expanded our model to consider two MHC loci per individual. The distribution of protection levels is similar to the model with one MHC locus, showing a small protective area for individuals carrying only iNKRs (Figure 3.2 C), whereas individuals carrying aNKRs have a broader protective range (Figure 3.2 D). However, the area of maximal protection is skewed in both cases. Because iNKRs have to recognize at least one self MHC molecule to become licensed, the chance of having several licensed NKRs per haplotype increases by having 2 MHC loci (and thus 4 MHC molecules per heterozygous individual). Therefore, a high protection (e.g. $P_{\mathrm{I}} \geq 0.85$) can be reached already with a smaller haplotype, e.g. one composed of at least 11 iNKRs.

In contrast, the probability of an aNKR to become licensed decreases with 2 MHC loci because aNKRs should not recognize *any* of the MHC molecules within an individual. Consequently, the protection with aNKRs reaches high values (i.e. $P_{\mathrm{A}} \geq 0.85$) only with large haplotypes composed of at least 20 genes. and the optimal protection level ($P_{\mathrm{A}} = 1$) is never obtained. Thus, the protection of aNKRs is highly dependent on the number of MHC molecules per individual.

With even higher MHC complexity, i.e. with increasing the number of MHC loci per individual to 4, fewer iNKRs are sufficient to successfully clear the infection (Figure 3.2 E, F). Because of the education process in our model, hosts with 4 MHC loci have a much larger licensed iNKR repertoire compared to individuals having 1 MHC locus. These hosts reach the maximal protection already with a haplotype size of 4 receptors. Even for lower haplotype sizes, a good protection level (i.e. $0.3 \leq P_{\mathrm{I}} \leq 0.7$) can be reached at lower specificity values ($p_{\mathrm{I}} \leq 0.35$) (Figure 3.2 E). This effect was further increased when considering 8 MHC loci per individual, where the maximal protection was reached with only one specific iKIR (results not shown).

However, an expanded MHC haplotype is disadvantageous for individuals having only aNKRs. Because in our model the licensing process is more difficult with a higher number of MHC molecules, little protection can be provided. The infection can be controlled with a maximal probability of 50 % and 35 % in individuals with 4 (Figure 3.2 F) and 8 MHC loci (results not shown), respectively.

Taken together, these results show that aNKRs provide little protection against a virus evolving MHC decoy proteins in individuals having several MHC loci, and that a contracted haplotype of iNKRs is already protective when the MHC complexity increases.

3.2.3. Viral detection is maximal in mixed haplotypes

To predict the combined protection of activating and inhibitory NKRs, we expanded our model and considered mixed haplotypes, i.e. haplotypes composed of both iNKRs and aNKRs. We predict the combined probability of detecting the virus as follows:

$$P = 1 - (1 - P_{\mathrm{I}})(1 - P_{\mathrm{A}}). \tag{3.4}$$

We computed the protection in hosts carrying two MHC and 20 NKR loci, and varied the fraction of aNKRs in the NKR haplotype, while keeping the total number of loci constant. The best protection is reached in mixed haplotypes (Figure 3.3). As seen above, haplotypes with aNKRs only provide protection (i.e. $0.5 \leq P \leq 0.8$) for intermediate specificity values $0.15 \leq p_{\mathrm{A}} \leq 0.4$ (Figure 3.3 A). With increasing number of iNKRs per haplotype, the protection reaches higher values (approaching $P = 1$) (Figure 3.3 B-C), covering a larger range of specificity values and having a skewed distribution towards more specific inhibitory and activating receptors. A large number of iNKRs per haplotype reduces the contribution of aNKRs, and therefore the latter can have low specificity values without affecting the protection level (Figure 3.3 D-E). Note that the area of high protection shrinks when the fraction of aNKRs is decreased, where maximal protection can only be achieved for extremely high specificities (i.e. $p_{\mathrm{I}} \leq 0.1$) (Figure 3.3 F).

These results depend on a similar manner on the number of MHC loci per individual as those shown in Figure 3.2, with aNKRs having a lower protective effect with increasing MHC loci number (results not shown). Therefore, we conclude that the maximal protection against CMV-like viruses is easier to achieve in mixed haplotypes.

3.2.4. Agent-based model of activating and inhibitory NKRs

Our probabilistic model allows us to quantify the expected protection given a certain number of aNKRs and iNKRs. However, it is not clear whether a population with evolving NKRs would find the same basin of attraction for the specificity (i.e. p) when

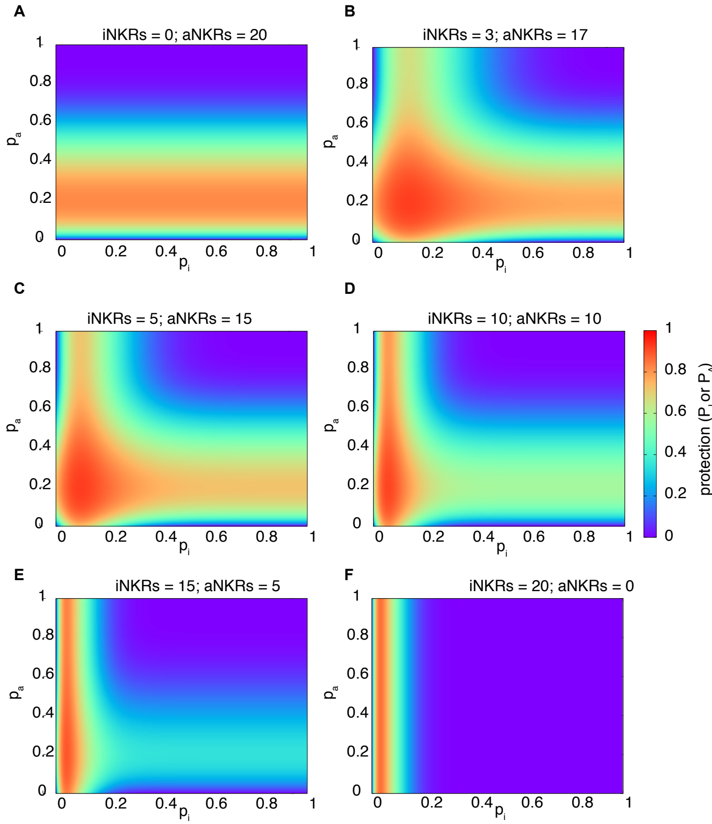

Figure 3.3. Mixed haplotypes render highest protection. The heatmaps show the protection level as the probability of detecting the infection for haplotypes composed of iNKRs and aNKRs (as calculated by Eq. 3.4). The protection level is shown in the color bar from highest (red) to lowest (blue). Panels (A)-(F) show the NKR haplotype composition for different fractions of iNKR and aNKR. Considering a haplotype size of $N_{NKR} = 20$, we first modeled haplotypes composed of aNKRs only (A), and reduced their number, while increasing the number of iNKRs (B-E), until we obtained a haplotype composed only of iNKRs (F). We here consider 2 MHC loci (i.e. $N_{mhc} = 2$).

infected with a CMV-like virus.

To study the evolution of NKR specificity in populations infected with a CMV-like virus, we developed an agent-based model similar to the one described in Chapter 2 (for a detailed description of the model see Material and Methods section). Briefly, our model considers a human population infected with a non-lethal herpes-like virus causing

chronic infections. The hosts carry a diploid genome with one or two MHC loci, and ten NKR loci. We consider 15 MHC alleles per locus (mimicking the common HLA-B, and C alleles in the European populations) and this polymorphism is kept constant throughout the entire simulation (i.e. we do not allow for mutation of the MHC genes). The initial NKR haplotype consists of ten different genes, and all individuals are homozygous for the same NKR haplotype.

Upon birth, novel receptors can be created, allowing for evolution within the NKR gene complex. Each new receptor comes with a randomly chosen receptor type (i.e. either inhibitory or activating), and a randomly chosen specificity value (corresponding to $0 \leq p \leq 1$, see Material and Methods). Receptors that are so specific that they are unable to recognize any MHC in the population will never be functional, and are considered to be pseudogenes. Thus, haplotypes expand by acquiring receptors with novel p values and signaling potential, but can also contract due to the accumulation of pseudogenes.

In this agent-based model, we also mimic the MHC dependent NK cell education process (Figure 3.1). We remove iNKRs which fail to recognize any MHC molecule within an individual from the licensed repertoire. Similarly, those aNKRs capable of recognizing self MHC molecules are deleted from the licensed repertoire. Only the licensed NKRs are able to participate during the immune response.

Infection of a host starts with a short acute phase, after which the individual either recovers or becomes chronically infected. We consider one wild-type virus and several decoy viruses (1 decoy per MHC molecule in the population). We do not allow for superinfection nor co-infection, thus hosts can be infected with only one of the viruses. A decoy virus down-regulates the expression of all MHC molecules within an individual, and expresses an MHC-like molecule. Thus, every virus expressing a decoy molecule has the potential to escape the immune response of both T and NK cells. The evolution of decoy proteins is modeled by allowing the virus to adopt a randomly selected MHC molecule from its host. Therefore, each decoy protein is actually an MHC molecule.

The population is first inoculated with the wild-type virus, which can be typically cleared after the acute phase because of the implicit response of both T and NK cells. We model the immune response with one parameter describing the probability of clearing the infection. For the wild-type virus this is set to $p_{wt} = 0.85$ (Table 2.2), resulting in approximately 85 % of the wild-type infections being cleared. Individuals clearing the infection become immune for a period t_{i} of ten years. At steady state, approximately 20 % of the population becomes chronically infected (Fig 3.4 A-B (green solid lines)), 65 % become immune (Fig 3.4 A-B (green dashed lines)), and 5 % are susceptible for infection. The

immune escape of the decoy viruses is modeled by decreasing the clearance probability to zero ($p_{dec,1} = 0$, Table 2.2), which occurrs if at least one of the licensed iNKRs or none of the aNKRs binds to the decoy molecule (Table 3.1). With this agent-based model, we can study the evolution of NKR specificity, and quantify the protection provided by activating and inhibitory receptors.

3.2.5. Inhibitory receptors evolve higher specificity than activating receptors after a CMV-like infection

We first study the protection provided by iNKRs against a CMV-like virus. After 5000 years of infection with the wild type virus, we allow for the emergence of decoy viruses. The initial specificity of the iNKRs is set to $p \approx 0.4$ (see Material and Methods). The decoy viruses spread easily among individuals carrying only iNKRs, resulting in a high fraction of chronically infected individuals (Figure 3.4 A red solid lines). Moreover, almost none of the hosts is able to control the infection (Figure 3.4 A red dashed lines), and the total population size decreases dramatically to 50 % of the carrying capacity (i.e. maximal population size), confirming the results from the probabilistic model. However, after centuries of infection, the fraction of recovered individuals increases, and with it the total population size, indicating a recovery of the population. This observation is consistent in all ten simulations we performed for iNKRs (Figure 3.4 A, C).

To study how these individuals evolve to control an infection with a virus having an MHC-like molecule, we analyze the average specificity of the NKRs over time. We determine how many MHC molecules in the population can be recognized by each receptor, and normalize it by the number of total MHC molecules in the population (Figure 3.4 E red line). We observe that the specificity increases after the emergence of the decoy viruses. At the end of the simulations, each iNKR recognize less than 20 % of the MHC molecules in the population, indicating that evolution selects for specific iNKRs.

We perform the same simulations and analysis for populations having only aNKRs. Compared to populations having iNKRs, the initial spread of the virus is somewhat impaired (Figure 3.4 B red solid lines). Already at the beginning of the infection, some individuals are able to control the virus (Figure 3.4 B red dashed lines). The population size decrases to 60 % of the carrying capacity, and is therefore fitter than in those simulations considering only iNKRs (Figure 3.4 A-C). Thus, aNKRs provide a better initial protection than iNKRs. Accordingly, the number of recovered individuals and with it of the total population increases rapidly, reflecting their fast recovery against viruses evolving decoy

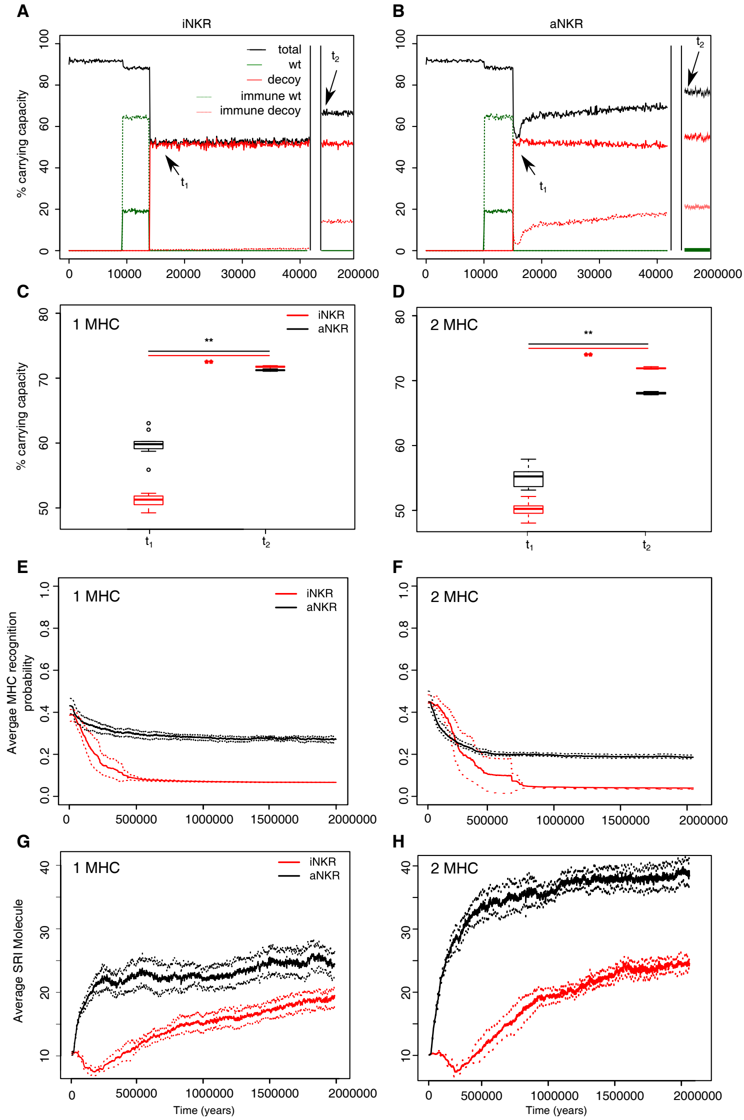

Figure 3.4. Agent based model confirms statistical model Continued on the following page.

proteins.

The higher protection of aNKRs compared to iNKRs can be explained by the initial

Figure 3.4. Agent based model confirms probabilistic model (A) A host population having only iNKRs was inoculated with a wild type virus (green lines) after a period of $t = 5000$ years (green solid lines show the chronically infected individuals, and the dashed lines the immune individuals). 10 000 years after the initial epidemic (i.e. t_1), we allowed for the evolution of decoy viruses (red lines). During the wild type infection, most individuals recover (dashed green line). In contrast, almost none of the individuals are initially capable of clearing a CMV-like infection (red dashed line), resulting in a large decrease of the total population size (black line). (B) A host population having only aNKRs is initially better protected against decoy viruses, resulting in a higher fraction of the population clearing the infection, and a lower decrease of the total population size. (A) and (B) show single representative simulations. (C) The average population size during the initial spread of decoy viruses (t_1) is lower than that at the end of the simulations (i.e. $t_2 = 3\,\text{mio.}$ years), indicating that over time, the populations learn to cope with the viral infection. Individuals in simulations considering only aNKRs (black) are initially better protected than those in simulations considering only iNKRs (red). In these simulations, all hosts carry only one MHC locus. (D) The initial advantage that aNKRs have over iNKRs receptors decreases in simulations considering two MHC loci per individual. (E) The probability of iNKRs recognizing any random MHC molecule in the population decreases over time (red line), indicating that more specific receptors are being selected for. In contrast, aNKRs (black line) do not evolve such high degree of specificity. (F) aNKRs evolve to become more specific in simulations where individuals have two MHC loci. (G) The degree of NKR polymorphism (expressed as the SRI score) increases in time, as a result of the evolved higher specificity. (H) SRI score in simulations considering two MHC loci. In (C)-(D), the boxes represent the interquartile range, and the thick horizontal lines the median out of ten simulations (** represent p values < 0.005, and were calculated using the Mann-Whitney U test). In (E)-(H), the solid lines represent the average out of ten simulations, and the dashed lines are the standard deviation.

specificity. Because we initialize all populations with intermediate specificity, individuals carrying only aNKR are initially better protected (Figure 3.2). Nevertheless, aNKRs also evolve to be more specific (Figure 3.4 E black lines). At the end of the simulation, aNKRs recognize on average approximately 35 % of all MHC molecules, and hence decoys in the population. Taken together, our agent-based model reveals that iNKRs need to be more specific than aNKRs to protect during an infection with a CMV-like virus, confirming the results from our probabilistic model.

Note that we do not explore all possible loci number in the agent-based model. To save computational time, we test the evolution of the specificity given a fixed loci number of NKRs. Populations carrying 10 NKR loci correspond to 20 NKRs in the probabilistic model, where the protection is maximal at very high specificity values for iNKR, and intermediate values for aNKR. These values corespond indeed to the specificity values that the populations evolve in our simulations. We carried out additional simulations for 5 and 15 NKR loci, the results of which confirmed the predictions of the mathematical model (results not shown).

3.2.6. Populations having only aNKRs evolve a larger NKR polymorphism than populations with only iNKRs

Our probabilistic model predicts that the protection by iNKRs and aNKRs increases with the number of receptors per individual (Figure 3.2), because a large receptor number increases the chance of a host carrying very specific NKRs to have licensed receptors. This observation suggests that heterozygous hosts should have an advantage over homozygotes. We therefore hypothesized that heterozygous advantage must be selecting novel NKRs in our agent-based model, driving polymorphism of NKRs in the population.

To measure the polymorphism at population level, we use the Simpson's reciprocal index (SRI, see Material and Methods). The SRI is a diversity measure that is equal to the total number of NKRs if they are equally distributed in the population, whereas the SRI is lower than that in a population where some alleles dominate (Simpson, 1949).

The initial polymorphism of aNKRs (i.e. SRI = 10) increases over time (Figure 3.4 G black line), reflecting that a high number of aNKRs provides indeed an advantage. Similarly, the SRI score of iNKRs increases over time. Because each evolved iNKRs recognizes on average one MHC molecule in the population, there is selection for haplotypes that do not overlap in the MHC molecules they recognize. Thus, the heterozygote advantage is large in these populations, driving the diversity of iNKRs.

3.2.7. Protection depends on the number of MHC loci

To confirm our results concerning the dependency on MHC loci number, we also perform simulations with individuals having two MHC loci. An increasing number of MHC loci has a large effect on the protection provided by aNKRs. Although these populations are initialized with intermediate specific NKRs, the initial protection is lower than in the population carrying only one MHC locus (Figure 3.4 D). For better protection, a higher specificity is required, and the selection for more specific aNKRs is stronger in these simulations (Figure 3.4 F). As a result of the higher specificity, a larger number of receptors per haplotype is necessary to become licensed and to recognize the foreign decoy molecules. Therefore, the advantage of heterozygotes over homozygotes is larger in these populations, resulting in a higher degree of polymorphism (Figure 3.4 H).

The protection and evolution of iNKRs is also sensitive to the number of MHC loci per individual. Like in the simulations considering one MHC locus, we observe a recovery

of the population as more specific receptors are evolving (Figure 3.4 D, F). Hereby, the specificity evolved to even higher values, as the evolved iNKRs recognize on average less than 5 % of all the MHC alleles in the population. Because of this high specificity and the larger number of MHC alleles in populations having two MHC loci, more iNKRs per haplotype are necessary to have at least one licensed receptor. Hence, the total SRI score is higher in these simulations, than in the case of single MHC locus (Figure 3.4 G-H, red line).

3.2.8. Basin of attraction: mixed haplotypes containing a majority of aNKRs

Finally, we performed simulations of populations having both iNKRs and aNKRs, in which we allow for the evolution of the specificity and also the receptor type. The initial specificity values for both receptor types was intermediate (i.e. $p \approx 0.4$) and we intialized the genotypes with a random number of activating and inhibitory receptors.

After the appearance of decoy viruses, the populations suffered similar effects to those having only iNKRs and aNKRs. The population size decreases dramatically at first, and with time it recovers. The final population size is higher than in the simulations considering only one type of receptor, approaching 70 % of the carrying capacity (Figure 3.5 A) because mixed haplotypes protect better than only one type of receptors. At the end of the simulations, we observe more aNKRs than iNKRs per haplotype (Figure 3.5 B), i.e. the final haplotypes are composed on average of 6 aNKRs, and 4 iNKRs. In agreement with the predictions of the analytical model, both receptor types evolve high specificity, (i.e. $p \leq 0.35$), and a high polymorphism (Figure 3.5 C,D). Summarizing, the agent-based model confirms the prediction of the probabilistic model.

3.3. Discussion

Our mathematical model predicts the optimal protection level provided by inhibiting and activating NKRs against viruses expressing MHC-like molecules. Haplotypes composed only of iNKRs detect the viral infection within a small range, requiring high specificity and large haplotype size. In contrast, the maximal protection is reached for intermediate specificity values and at a smaller haplotype size in individuals having only aNKRs. Mixed haplotypes, i.e. haplotypes carrying both iNKRs and aNKRs offer the highest

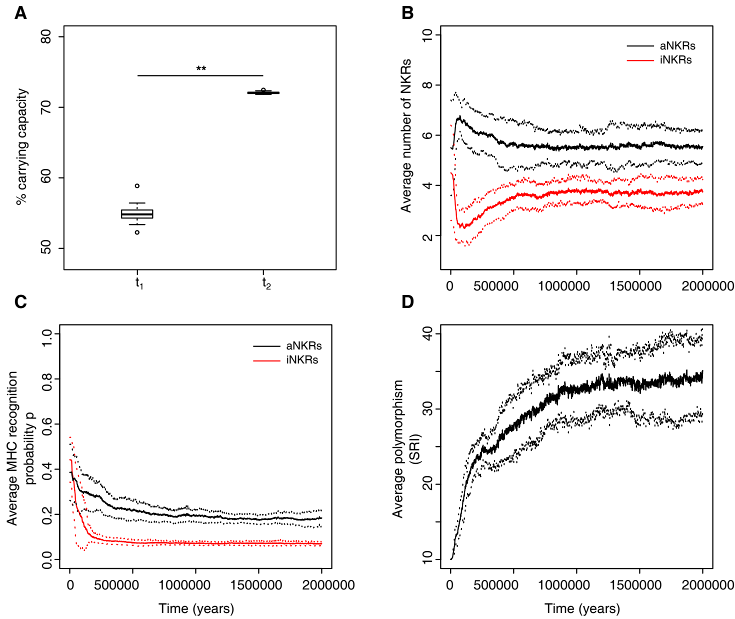

Figure 3.5. Mixed haplotypes offer the highest protection. A host population having iNKRs and aNKRs was inoculated with a wild type virus after a period of $t = 5000$ years; we allowed for the evolution of decoy viruses 10 000 years after the initial epidemic (i.e. t_1). (A) The population size during the initial spread of decoy viruses (t_1) is lower than that at the end of the simulations (i.e. $t_2 = 3$ mio. years), indicating that over time, the population recovers from the viral infection. (B) The initial haplotype is composed of five iNKRs and five aNKRs. The number of aNKRs and iNKRs per haplotype varies over time, resulting in a selection for haplotypes with a larger activating potential. (C) The probability of NKRs recognizing any random MHC molecule in the population, decreases over time, indicating that more specific receptors are being selected for. (D) The degree of NKR polymorphism (expressed as the SRI score) increases in time, as a result of the heterozygote advantage due to the evolved higher specificity. Averages taken out of ten different simulations. In (A), the boxes represent the interquartile range, and the thick horizontal lines the median (** represent p values < 0.005, and were calculated using the Mann-Whitney U test). In (B)-(D), the solid lines represent the average out of ten simulations, and the dashed lines are the standard deviation.

protection.

All these results are dependent on the number of MHC loci per individual. With increasing MHC loci, aNKRs lose their ability to become licensed and thus provide little or no protection. In contrast, haplotypes composed only of iNKR have a higher chance of having licensed receptors when the number of MHC loci is increased. In this case,

the protection level is maximal already with a contracted NKR haplotype. Thus, there seem to be several combinations of MHC-NKR genotypes that provide maximal protection. A high protection is reached with a simple MHC complex and a high number of NKR genes. With increasing complexity of the MHC, a contracted NK complex is sufficient to render protection. These last results are particularly interesting, as they provide a possible explanation of the differences in KIR and MHC gene content across primate species (Sambrook *et al.*, 2005), and the expansion of new KIR lineages corresponding to the contraction of the MHC gene complex, illustrating thus the co-evolution of MHC class I and KIRs (Parham and Moffett, 2013).

The model described here is inspired by viruses evolving decoys. However, its main outcome, namely the requirement for specificity, might be more general than the defense against such decoy viruses. Studies have shown that viral infections can change the repertoire of peptides presented by MHC class I molecules (Hickman *et al.*, 2003), and that these different peptides affect the NKR-MHC interactions, perturbing the binding of iNKRs and leading to NK cell activation (Fadda *et al.*, 2011). In such cases, specific recognition of the changes in peptide repertoire by NK cells seems advantageous for the host. Also, the specificity ranges obtained in our model for mixed haplotypes (Figure 3.3 E) are similar to those observed in reality, with iNKRs having a specificity of approximately 0.2. This corresponds to the four mutually exclusive epitopes that have been detected so far for inhibitory KIRs in humans: HLA-A11, -Bw4, -C1, and -C2.

The exact role of aNKRs remains intriguing. Since only a few aNKRs tend to recognize MHC class I molecules (Moesta *et al.*, 2010), we speculate that aNKRs could specifically recognize new ligands expressed upon viral infection (e.g. decoy molecules or stress ligands). Our model predicts that to face the challenge of not recognizing self but specifically recognize foreign antigens, aNKRs do not need to be so specific. Indeed, the haplotype providing the highest protection is a combined haplotype composed of more aNKRs than iNKRs, which disagrees with the most primate KIR haplotypes (Parham and Moffett, 2013). Most primate NKRs are inhibitory, and activating receptors have been linked to selection pressure induced by reproduction (Parham and Moffett, 2013). Our model predicts that aNKRs should evolve to an intermediate specificity upon CMV-like infections. However, not many activating ligands have been identified yet, and it remains puzzling what other roles aNKRs might play.

The expansion of NKR superfamilies, presumably in order to gain resistance against pathogens, illustrates the high evolutionary complexity of NKRs. We aimed to fully understand the effects of a single possible driving force of this evolutionary process, namely

that of a viral encoded MHC-like molecule. Therefore, we focused on modelling only the evolution of NKRs in a simple model, which requires simplifying assumptions. For instance, we fixed MHC polymorphism despite the evidence of the co-evolution between MHC class I and KIRs (Parham and Moffett, 2013, Sambrook *et al.*, 2005). Given their different evolutionary timescales, i.e., that MHC molecules are older than both Ly49 and KIRs, we chose to model the expansion and contraction of NKR systems within an already existing MHC diversity. Additionally, we assumed that decoy viruses downregulate the expression of all MHC molecules in the host. Even though we do not expect selective MHC downregulation to affect the evolution of aNKRs (since activating receptors cannot detect missing self), the evolution of iNKRs might be affected because more licensed iNKRs will be necessary to recognize a virus that downregulates only one of the host's MHC-I molecules. Note that if the licensed repertoire of iNKRs is larger, these receptors should be even more specific to avoid being "fooled" by the decoy molecule. The exact effect of selective MHC downregulation on the specificity of iNKRs is an open interesting question, which we are currently working on.

Other simplifying assumptions were also necessary, such as considering a global NKR repertoire and ignoring the synergy between NKRs or the direct interaction between immune cells. Additionally, we ignored mutational operators that conserve similarity between pre- and post-mutation receptors (e.g. point mutations), as we only model mutations that significantly change receptor functionality. Including point mutations, did not affect the results qualitatively (results not shown), however a longer evolutionary time was necessary to approach the same solution of specificity. Overall, since our main results are of a qualitative nature, it seems unlikely that relaxing any of these assumptions would affect our main results. Note also that our agent-based model is inspired on humans and KIRs, with the advantage of having realistic parameters for processes like birth and death. However, the model can be generalized to other species, and qualitatively it represents a model of the evolution of the expansion of the NKR complex.

All our analytical results were consistent with the agent-based model and our analysis allowed us to quantify the protection against an infection for both receptor types. It confirmed our results of Chapter 2 that iNKRs should become specific enough. Our new approach has shed light into the possible contribution that each receptor type confers upon infection, and allowed us to conclude that mixed haplotypes render the best protection.

3.4. Materials and Methods

3.4.1. Agent-based model

The agent-based model consists of two types of actors (hosts and viruses) and three events (birth, death, and infection). This model is virtually identical to the one in Chapter 2. Briefly, we screen all hosts in a random order during each time step of one week, and confront them to one of the randomly chosen events. Hosts age over time and at the end of each time step, their age, infection state, and type of infection is updated. This cycle is repeated for two million years to simulate long term evolution. All model parameters are given in Table 2.2.

We model simplified diploid individuals, carrying gene complexes for NKR, and MHC-class I. For simplicity, we consider 15 MHC alleles per locus, resembling the most common HLA alleles in the European population (Meyer *et al.*, 2007). NKRs and their ligands are modeled with randomly generated bit strings as a simplified representation of amino acids (Farmer *et al.*, 1986). If the longest adjacent complementary match between two strings exceeds a threshhold L, we allow for the receptor to interact with its ligand. The threshold L determines thus the specificity of each receptor: a receptor with a small L value will be very degenerate and the probability of a random NKR to recognize a random MHC molecule will be $p \approx 1$. In contrast, a receptor with a large L value will be specific, and accordingly, the probability of this receptor binding any MHC molecule in the population will be $p \ll 1$ (for a detailed description see Chapter 2).

Receptor types

In the present model, we allow for the evolution of aNKRs. When a novel NKR is generated, a random L value between 1–16 is assigned to it, and its type (i.e. whether it is activating or inhibitory) is also randomly chosen. Thus, each receptor has its particular specificity and funcionality. Recepors with L values larger than 13 will usually not recognize any MHC molecules in the population, and are typically not functional. Genes encoding such unfunctional NKRs are considered to be pseudogenes. Haplotypes containing pseudogenes are effectively shorter than haplotypes composed of fully functional NKRs. Thus, we can model the contraction and expansion of the NKR gene complex.

Viral infections

In our simulations, we consider one wild type virus and several "decoy viruses", i.e. viruses expressing MHC decoys. Each virus comes with a viral load, which is implemented as an increase of the host's death rate, VL_i depending on the infection state i (see Table 2.2), and a probability of clearing the infection p_{wt}, and $p_{dec,s}$ for the wild type and the decoy viruses, respectively. A decoy virus downregulates the expression of all MHC molecules in that host, and encodes one MHC-like molecule. The evolution of decoy molecules is modeled by allowing the virus to adopt a randomly selected MHC molecule from its host with a rate μ_v. The virus keeps this decoy for the rest of the simulation. Because we fix the MHC polymorphism to 15 alleles per locus, the maximal number of decoy proteins that can evolve in the population is 15 for the simulations considering 1 MHC locus, and 30 for those considering two MHC loci.

We consider different levels of protection against a decoy virus, depending on the success of the virus to escape the NK cell response, s. If at least one of the licensed iNKR binds to the decoy molecule, there will be an inhibitory signal, the host will not be able to detect "missing-self", and the decoy virus will be successful. Similarly, if none of the licensed aNKRs recognizes the decoy molecule, the decoy virus will evade the NK cell response. Thus, none of the iNKRs or at least one aNKRs should bind the decoy molecule to render protection (Table 3.1). We model the immune escape by setting the probability of clearing the infection to zero, letting the host become chronically infected. In the case that a decoy is not successful, the host will be able to detect "missing-self". Since this virus is nevertheless able to evade the response from T cells (due to the MHC downregulation), the probability of clearing the infection is lower than that of the wild type virus ($p_{wt} = 0.85$). The resulting probability of clearing the infection is described by:

$$p_{dec,s} = \begin{cases} 0, & \text{if } s = 0 \text{ (successful decoy)} \\ 0.5, & \text{if } s = 1 \text{ (unsuccessful decoy).} \end{cases} \tag{3.5}$$

The rest of the parameters defining the infection dynamics and immune escape of the decoy viruses (i.e. time of infection, immunity time, and transmission probabilities) were set like in Chapter 2 and are described in Table 2.2.

Mutation

During each birth event, NKRs undergo mutation with a probability μ. To decrease computation time, we model mutation by randomly creating a new receptor with its particular specificity and signaling type. We do not consider other mutational operators, e.g. point mutations, recombination, deletion, or duplication.

We first perform simulations where only the specificity can evolve (i.e. a random value L is assigned to each new receptor), while the receptor type remains fixed. Hereby, we are able to compare what the basin of attraction for the specificity will be, if a population has only aNKRs or only iNKRs. We also simulate populations with mixed haplotypes, by allowing the receptor type to mutate.

NK cell education

During the birth event, an NK cell education process takes place. Like in our probabilistic model explained above, iNKRs which recognize *at least one* of the MHC molecules within one individual, and aNKRs that fail to recognize *all* of the MHC molecules within the host, are set to be licensed. In our model, only the licensed repertoire of NKRs will participate in an NK cell response (Figure 3.1).

3.4.2. Model Initialization

The model is initializated with a host population of 4500 hosts, with random ages between 1 and 70 years corresponding to a uniform age distribution. After approximately 10 host generations, this age distribution corresponds to more modern age distributions with the majority of individuals having an age between 15-60.

At the start of every simulation, a gene pool for MHC alleles is generated, the size of which depends on the number of MHC loci per individual. It consists of 15 alleles in simulations considering one MHC locus per individual, and of 30 alleles in those simulations considering two MHC loci per individual. To create the initial genome of each individual, MHC genes were randomly drawn from the pool, while ten NKRs with intermediate specificity ($2 \leq L \leq 4$, i.e. $p \approx 0.4$) were generated. Thus, the initial haplotypes did not contain any pseudogenes. In the simulations considering mixed NKR haplotypes, the initial genes can be both activating and inhibitory. The type of each receptor was ran-

domly chosen as explained above, resulting in approximately 50 % of the receptors being activating. All individuals were initialized with the same NKR haplotype, but with different MHC genes.

3.4.3. Implementation

The model was implemented in the C++ programming language. We considered populations with haplotypes composed of only aNKRs, only iNKRs, or both. In every scenario, we compared the effects of one or two MHC loci per individual. For each of these settings, we performed ten simulations for 2 million years. The code is available upon request.

3.5. Acknowledgments

We thank Chris van Dorp, Leïla Perie and Hanneke van Deutekom for helpful discussions, and carefully reading the manuscript. We also thank Oussama Jarrousse and Johannes Textor for their technical support during the development of the code. This work was financially supported by the CLS program of the Netherlands Organization for Scientific Research (NWO), grant 635.100.025.

Table 3.1. Levels of protection against a decoy virus in the agent-based model

	$\#\ \mathbf{aNKR_{binding}} = \mathbf{0}$	$\#\ \mathbf{aNKR_{binding}} > \mathbf{0}$
$\#\ \mathbf{iNKR_{binding}} = \mathbf{0}$	$p_{dec,1}$	$p_{dec,1}$
$\#\ \mathbf{iNKR_{binding}} > \mathbf{0}$	$p_{dec,0}$	$p_{dec,1}$

$\mathrm{aNKR_{binding}}$ and $\mathrm{iNKR_{binding}}$ refer to the number of iNKRs and aNKRs binding the decoy molecule, respectively. The receptors here are considered to be licensed.

3.6. Supplementary Information

3.6.1. Global repertoire of licensed NKRs

Instead of considering individual NK cells expressing a tuned set of NKRs, we model a global repertoire of “licensed” NKRs in each host. Hereby, we assume that the expression of at least one “licensed” receptor is sufficient to educate NK cells, and that those functional NK cells can become activated upon viral infection. This assumption needs further explanation.

If an aNKR recognizes some self molecule (or self MHC), all developing NK cells expressing that receptor should additionally express at least one iNKR recognizing self to prevent self-reactivity. As a consequence, all NK cells expressing that aNKR should never become activated, even if a virus encodes a ligand which engages that aNKR. Therefore, we call such an aNKR“unlicensed”: it will never contribute to detect a viral infection. In contrast, if an aNKR that does not recognize any self molecule, all developing NK cells expressing that receptor will not be influenced by it, and these cells will express other iNKRs and aNKRs that balance their self-reactivity. NK cells expressing that aNKR will become activated when a virus expressing its ligand engages that aNKR. Therefore we call such an aNKR "licensed". Any NK cell expressing this aNKR should breach its activation threshold, expand, and protect, when the ligand is present.

Now consider an iNKR that recognizes some self MHC. All developing NK cells expressing that receptor will be tuned to balance its self-reactivity. As a consequence, all NK cells expressing that iNKR should become activated when a virus down-regulates this particular MHC. Therefore we call such an iNKR “licensed”: it detects MHC down-regulation. However, for an iNKR that does not recognize any self MHC, the developing

NK cells expressing that receptor will not be influenced by that iNKR. Consequently, these cells will express other iNKRs and other aNKRs in order to balance their self-reactivity. Therefore, NK cells expressing that iNKR will not become activated by this iNKR when a virus down regulates self MHC. Therefore we call such an iNKR "unlicensed": it would never contribute to the detection of virus infected cells.

What happens with the NK cells expressing such an unlicensed iNKR if it happens to recognize something else on virus infected cells, e.g. a decoy? Then the iNKR should deliver an extra inhibitory signal to those cells. Consequently, these NK cells cannot expand and protect, even if their other iNKR detect MHC down regulation, or if their other aNKR detect a new ligand. Hence, such an unlicensed iNKR will not protect, and can be considered to be non functional at the repertoire level.

Chapter 4

A co-evolutionary arms race between hosts and viruses drives polymorphism and polygenicity of NK cell receptors

PAOLA CARRILLO-BUSTAMANTE, CAN KEŞMIR, ROB J. DE BOER

Submitted

Abstract

Natural killer cell receptors (NKRs) monitor the expression of MHC class I and stress molecules to detect unhealthy tissue, such as infected or tumor cells. The NKR gene family shows a remarkable genetic diversity, containing several genes encoding receptors with activating and inhibiting signaling, and varying in gene content and allelic polymorphism. The expansion of the NKR genes is species-specific, with different species evolving alternative expanded NKR genes, which encode structurally different proteins, yet perform comparable functions. So far, the biological function of this expansion within the NKR cluster has remained poorly understood.

To study the evolution of NKRs, we have developed an agent-based model (ABM) implementing a co-evolutionary scenario between hosts and herpes-like viruses, that are able to evade the immune response by downregulating the expression of MHC-I on the cell surface. We show that hosts evolve specific inhibitory NKRs, specialized to particular MHC-I alleles in the population. Viruses in our simulations readily evolve proteins mimicking the MHC molecules of their host, even in the absence of MHC-I downregulation. As a result, the NKR locus becomes polygenic and polymorphic, encoding both specific inhibiting and activating receptors to optimally protect the hosts from co-evolving viruses.

4.1. Introduction

Natural killer (NK) cells are key players in the immune defense against viral infections and tumors (Vivier *et al.*, 2008). Classified as components of the non-specific innate immune system, NK cells recognize abnormal or viral infected cells with a sophisticated array of germline-encoded activating and inhibiting receptors. Inhibiting NK cell receptors (iNKRs) bind mostly MHC class I (MHC-I) molecules, providing a mechanism by which NK cells distinguish normal healthy cells from abnormal cells expressing reduced levels of MHC-I molecules on the cell surface. Activating NK cell receptors (aNKRs) target viral products, stress-induced ligands, or other tumor-specific molecules (Lanier, 2005). Thus, the engagement of iNKRs and aNKRs enables NK cells to become cytotoxic towards abnormal or viral infected cells, while remaining silent towards healthy cells with normal MHC-I expression.

Some of the NK cell receptors (NKRs) discriminating between healthy and unhealthy cells are highly conserved across species. Examples are the inhibitory CD94/NKG2 receptors recognizing the also monomorphic HLA-E in humans and Qa-1^b in mice (Borrego *et al.*, 1998, Braud *et al.*, 1998, Petrie *et al.*, 2008, Zeng *et al.*, 2012), and the activating NKG2D detecting MHC-I-like proteins that are induced upon stress, like MIC-A, MIC-B in humans and RAE-1, MULT1 and H60 proteins in mice (Bacon *et al.*, 2004, Bauer *et al.*, 1999, Cerwenka and Lanier, 2001, Cerwenka *et al.*, 2000). In addition to these conserved receptor-ligand interactions, there exists a variety of polymorphic multigene families performing equivalent functions. These include the killer immunoglobulin-like receptors (KIRs) in higher primates (Guethlein *et al.*, 2007a, Parham *et al.*, 2010), the CD94/NKG2 receptor family in lemurs (Anfossi *et al.*, 2006), and the Ly49 gene family in rodents and equids (Wilhelm and Mager, 2004, Wilhelm *et al.*, 2002).

The multigene NKR families exhibit a large variation on the population level, as haplotypes differ in gene content, allelic polymorphism, and the ratio between activating and inhibiting NKR genes. Because NKR genes segregate independently from MHC-I genes, the number of possible receptor-ligand combinations is enormous. But, if conserved receptor ligand interactions (such as NKG2A/HLA-E, NKG2A/Qa-1^b, and NKG2D/MIC-A) are able to successfully discriminate between self and viral molecules, what is the evolutionary advantage of having polymorphic, polygenic, and specific NKRs? The polymorphism of their ligands (i.e. classical MHC-I molecules) is not enough to explain the polymorphism of NKRs, as there are several non-polymorphic receptors binding all class I MHC molecules (e.g. CD8 on T cells). Thus, what is the selection pressure driving the

evolution of polygenicity and polymorphism within the NKR cluster?

The fact that different species show expansion of different structural NKR families suggests that the NKR cluster has been subject of rapid evolution and that a polygenic and polymorphic NKR gene complex provides an evolutionary advantage in many taxa. Moreover, the associations between specific NKRs and positive disease outcome in human viral infections, such as HIV-1 (Martin *et al.*, 2002b, Pelak *et al.*, 2011) and HCV (López-Vázquez *et al.*, 2005), further suggest that NKRs are beneficial against a large number of viral infections. The complex evolutionary interplay between NKRs and viruses has been elucidated by several studies of mice infected with the murine cytomegalovirus (MCMV) (Arase *et al.*, 2002, Smith *et al.*, 2002). MCMV downregulates the expression of MHC-I and additionally codes for MHC-like molecules that can engage iNKRs in 129/J mice, thereby impeding NK cell activation (Smith *et al.*, 2002). In contrast to 129/J mice, C57BL/6 mice are resistant to MCMV infection. The susceptibility of 129/J mice has been explained by the strong interaction of their iNKR Ly49I with the CMV encoded protein m157. In contrast, the resistance of C57BL/6 mice to MCMV has been genetically mapped to a gene coding for the aNKR Ly49H, which also binds m157 with high affinity (Lee *et al.*, 2001, Smith *et al.*, 2002). Because the activating Ly49H evolved from its inhibitory counterpart Ly49I (Abi-Rached and Parham, 2005), it has been proposed that the evolution of the aNKR resulted from the immune pressure exerted by CMV having escaped the NK cell responses by evolving "MHC-I decoys" (Arase and Lanier, 2002, Lanier, 2008, Sun and Lanier, 2009).

C57BL/6 is not the only strain resistant to MCMV. Inbred MA/My mice also have low viral titers after infection with CMV and do not possess the Ly49H gene. Their resistance is mediated by the aNKR Ly49P which specifically recognizes MCMV infected cells in a H2-D^k dependent manner (Kielczewska *et al.*, 2009). Interestingly, this resistance requires the presence of the virally encoded protein m04, which escorts and binds newly assembled MHC-I molecules on the cell surface. Other aNKRs that recognize CMV infection in a m04-H2 dependent manner include Ly49L^{BALB}, Ly49P1^{NOD}, and Ly49W1^{NOD} (Pyzik *et al.*, 2011). Like m157, the original function of m04 might have been to counteract the effect of MHC downregulation and avoid "missing-self" detection by iNKRs. Apparently, hosts evolved novel aNKRs allowing them to control the infection as a counter-strategy.

The hosts' adaptation exerts a strong selection pressure on the virus, as demonstrated by the high degree of polymorphism observed in the m157 and m04 genes from MCMV strains isolated from wild mice (Voigt *et al.*, 2003), as well as by the loss-of-function

mutations in m157 genes after serial passage of MCMV in C57BL/6 mice (Voigt *et al.*, 2003). Thus, the co-evoution between NKRs and viruses shows features of Red Queen dynamics, where viruses continue to evolve immuno-evasive mechanisms to escape from the NK cell response, and the hosts keep on adapting by evolving novel NKRs, exerting novel selection pressures on the virus.

We investigate this dynamic and complex co-evolutionary process with a computational agent-based model (ABM) of host populations infected with two different viral species, one of which is able to evade the adaptive immune response by downregulating MHC-I expression. We observe that viruses "spontaneously" evolve MHC-like molecules, "fooling" and hence escaping detection by both iNKRs and aNKRs. Consequently, hosts evolve haplotypes composed of specific iNKRs, which are specialized to recognize small groups of MHC-I molecules, and sufficiently specific aNKRs, recognizing viral molecules.

4.2. Results

4.2.1. Agent Based Model

To study the evolution of NKRs in a host population, we developed an ABM based on our previously described models (see Chapters 2 and 3). Briefly, the host population consists of simplified humans infected with non-lethal viruses causing chronic infections. The hosts are diploid, carrying two polymorphic MHC loci and an NKR cluster. NKRs and MHC molecules are modeled with random sequences of zeros and ones (i.e. bit strings) as a simplified representation of amino acids. Whenever the longest adjacent complementary match between two strings exceeds a binding threshold, the molecules can interact (Figure 2.1 A).

Initially, the NKR cluster is composed of one degenerate receptor pair, i.e. one iNKR and one aNKR, each recognizing every MHC molecule in the population. Upon birth, individuals inherit one NKR haplotype from each parent. During this process, NKRs can mutate their sequence, their binding threshold, and their signaling potential, allowing for the emergence of novel receptors. If a newly generated receptor is so specific that it fails to recognize any MHC molecule in the population, it will be a *pseudogene*. We focus on the evolution of NKRs, and therefore fix the MHC polymorphism throughout the simulations. The MHC alleles in the population are not completely random, but somewhat

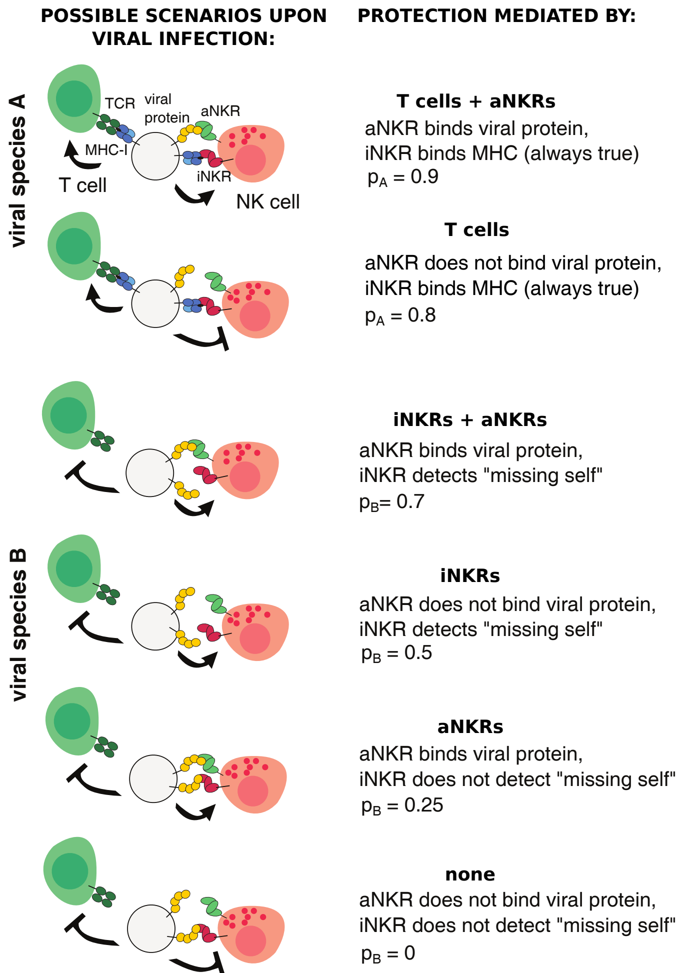

Figure 4.1. Model cartoon of the different levels of protection against viral species A and B. The probability of clearing the infection ($p_{A/B}$) depends on the interactions between the licensed receptors and the viral molecule. If at least one aNKR of one host binds the viral molecule, the NK cells of that individual will have an extra activating signal. However, the NK cells will not be able to detect the downregulation of MHC-I if the viral molecule of the B species engages at least one of the iNKRs. $p_{A/B}$ describes the probability of the hosts becoming immune. The probability of becoming chronically infected is hence $1 - p_{A/B}$.

similar to each other (see Material and Methods). During development, NK cells undergo an education process during which their reactivity is “tuned”. While the binding of iNKRs with their cognate MHC molecules render these NK cells functional capacity (Anfossi *et al.*, 2006, Chalifour *et al.*, 2009, Elliott and Yokoyama, 2011), the exposure of aNKRs to self ligands results in hyporesponsive cells (Fauriat *et al.*, 2010, Sun and Lanier, 2008). We implement this process at birth by evaluating the binding of NKRs with their MHC ligands in the host. For simplicity, we do not model individual NK cell subsets, but estimate the total repertoire of “licensed” receptors per host. In each host, the repertoire of licensed receptors consists of iNKRs binding at least one of the host’s MHC molecules, and aNKRs failing to recognize all of the host’s MHC molecules (Figure 3.1). Additionally, aNKRs can only become licensed if the host expresses at least one licensed iNKR. This last point was included to mirror the fact that the presence of licensed aNKRs only would result in NK cell-related autoimmunity; however this does not have a large impact on our results. In our model, only licensed NKRs participate during an immune response, assuming that at least one NK cell subset will express the set of licensed receptors, and will become activated and expand upon infection.

We model two viral species causing chronic infections (viral species A, and B). Both species carry molecules (also modeled with bit strings), representing products coded by viral genomes. Both viral species can mutate their viral molecules during the transmission to a new host, allowing for the emergence of new strains of each species. B viruses can escape the cytotoxicity of T cells by downregulating the expression of MHC class-I on infected cells. Upon transmission, the host will enter a phase of acute infection, after which it can either recover or become chronically infected. Individuals clearing an infection become immune against that particular viral species for a period of ten years. Hosts can be co-infected with one viral strain of each species, but we do not allow for superinfection with different strains of the same virus.

The probability of clearing the infection depends on the viral strategy of the virus to escape the immune response (illustrated in Figure 4.1). If the viral molecules can be detected by at least one aNKR of the host that gets infected, the NK cells of that individual will receive an extra activating signal. However, if the viral molecule of the B species engages at least one of the licensed iNKRs, the NK cells of that host will not be able to detect the altered MHC-I expression, and will be inhibited. Thus, we assume different levels of protection depending on the possible combinations of iNKRs and aNKRs per individual.

In a typical simulation, 100 different strains from each viral species are introduced after

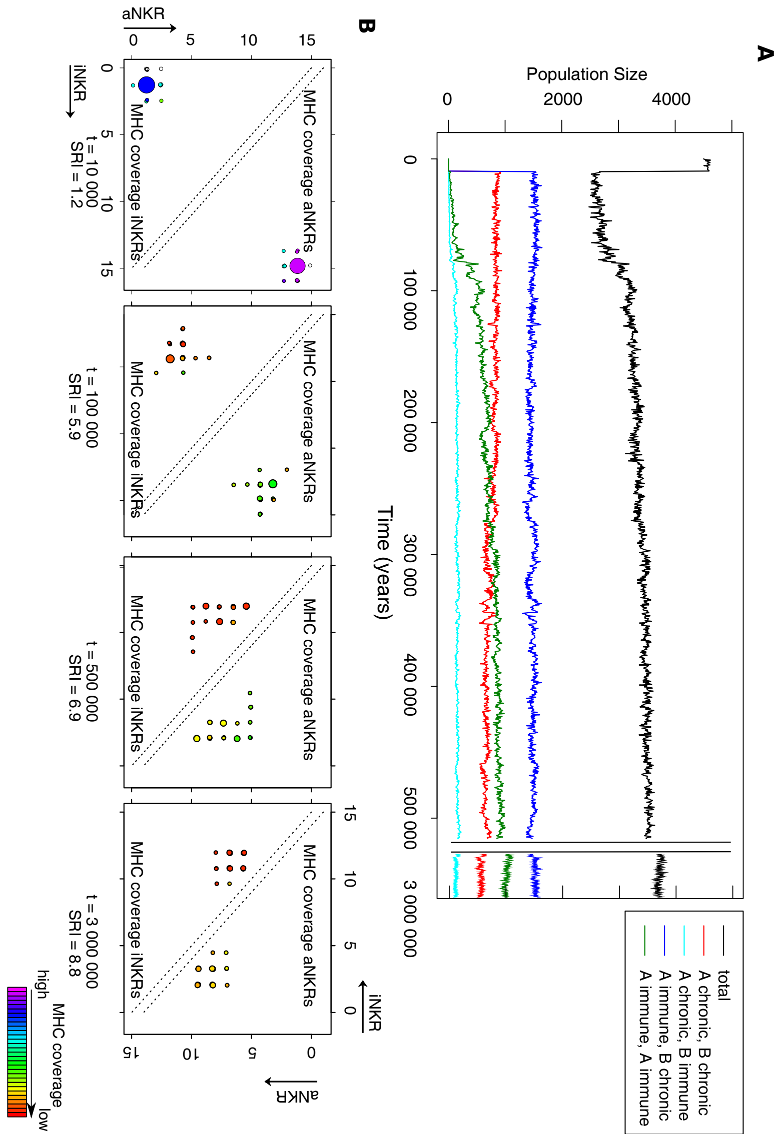

Figure 4.2. A host population recovers evolves immunity to two viral species due to an expansion of the NKR cluster. Legend on the next page.

Figure 4.2. A host population recovers evolves immunity to two viral species due to an expansion of the NKR cluster. (A) Single representative simulation of a population infected with both viral species. After their introduction at $t_1 = 10000$ years, the viruses from both species spread rapidly through the population, infecting most individuals and causing a marked decrease of the total population size (black line). A large fraction of hosts become immune against A viruses, and chronically infected with viruses of the B species (blue line). In contrast, only a few individuals are immune against the B species and chronically infected with A viruses (cyan line). Individuals chronically infected with both viral species (red line) decrease in time, while hosts able to clear both infections evolve (green line). (B) Haplotype composition in the NKR cluster at different time points (snapshots taken from the video provided in the Supporting Information). Each haplotype is represented by one dot, the size of which illustrates the frequency of the haplotype in the population. The position of each dot represents the number of iNKRs and aNKRs in the haplotype. In the lower triangle of each box, every haplotype is colored according to the average MHC coverage of its iNKRs.The MHC coverage is defined as the probability for an NKR to recognize any MHC molecule in the population. The degrees of MHC coverage are shown in the color bar from lowest (red) to highest (pink). The diagonal lines represent the maximal number of genes per haplotype we allow in our simulations. Mirrored on the diagonal, the haplotypes are colored according to the MHC coverage of the aNKRs. At $t_{end} = 3000000$ there are 9 common haplotypes as estimated by the SRI.

the population has stabilized ($t_1 = 10000$ years). Very rapidly, the viruses spread through the population, infecting almost every individual, and causing a drastic reduction of the population size (Figure 4.2 A, black line). In the infected subpopulation, most hosts are infected with strains of the B species because it is more challenging to become immune against B viruses compared to A viruses, as B viruses downregulate MHC-I molecules (Figure 4.1). After several host generations, a population of individuals that is immune against both viral strains evolves (Figure 4.2 A, green line). The number of immune individuals increases over time, causing a recovery of the total population size, suggesting that the initially susceptible host population evolves a NKR system providing immunity to both types of infections.

4.2.2. Immunity to infections is associated with the expansion of specific iNKRs and aNKRs

To understand how NKR evolution enables the immunity of the host population, we analyzed the receptors in the NKR cluster before and after a long period of co-evolution (Figure 4.2 B). The initial haplotype composed of one degenerate aNKR and one degenerate iNKR diversifies shortly after the introduction of the viruses, with a clear selection for large mixed haplotypes. Additionally, we observe selection for polymorphism, with

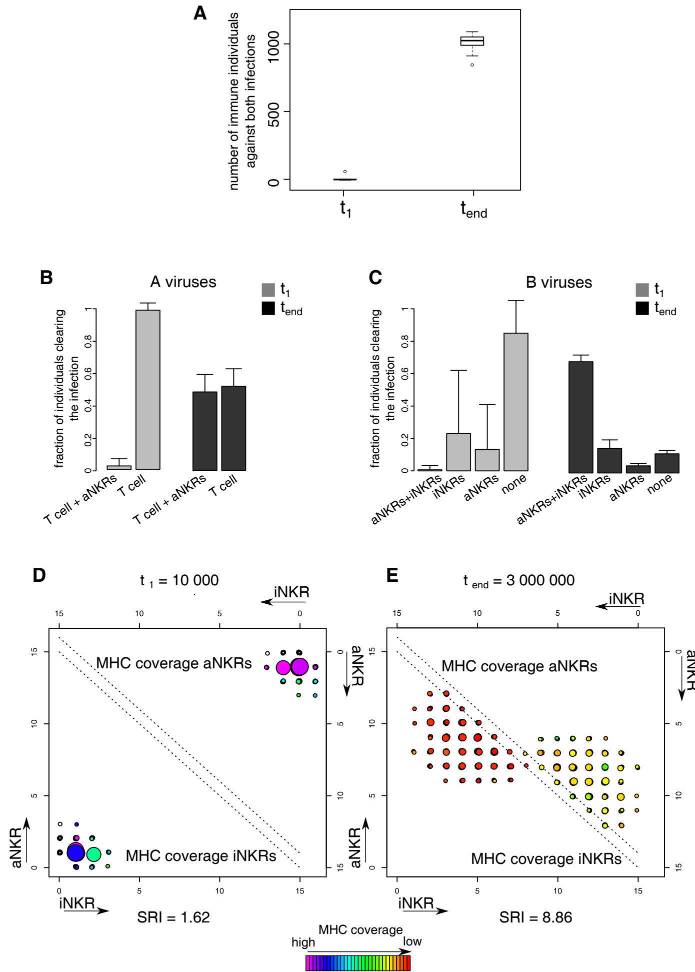

Figure 4.3. Expansion of the NKR haplotypes leads to better immunity. (A) Individuals immune to both viral species at $t_1 = 10000$ years and $t_{end} = 3000000$ years. (B)-(C) Observed fraction of individuals having different levels of protection mediated by T cells, aNKRs, and/or iNKRs as illustrated in Figure 4.1. At the beginning of the infection (t_1), A viruses tend to be cleared with $p_A = 0.8$, whereas most individuals cannot clear B viruses ($p_B = 0$). Consequently, the number of hosts immune against both viral species is close to zero. At the end of the simulation (t_{end}), the fraction of individuals with a high level of protection against B species increases to approximately 50 %, resulting in an increase in the number of double immune individuals. The boxplots and the barplots show the mean out of 15 simulations. NKR haplotypes from all 15 simulations at t_1 are shown in (D) and t_{end} in (E). The composition and specificity are as described in Figure 4.2 B. At t_{end} there are on average 9 common haplotypes per simulation as calculated by the SRI (several symbols are overlapping).

populations evolving more than 9 different haplotypes as estimated by the Simpson's Reciprocal Index (SRI) diversity measure (see Material and Methods).

As more receptors evolve, the number of immune individuals increases, indicating that an expanded haplotype is necessary for the hosts to clear the infections. At the end of the simulations, the fraction of hosts responding well to both types of viruses increases compared to hosts from the initially susceptible population (Figure 4.3 A,B,C), confirming that hosts evolve better immunity. The observation that the number of immune individuals increase as the haplotype composition changes was consistent for 15 simulations (Figure 4.3 A,B,C).

To better study the effects of each viral species, we also performed simulations of populations spreading strains of only one viral species (Figure S4.1). Populations infected with A strains only expand their aNKRs while keeping one degenerate iNKR (Figure S4.1 B). This is because iNKRs remain engaged by the host's MHC-I molecules as A viruses do not downregulate MHC. On the other hand, there is selection for aNKRs recognizing less MHC-I molecules, as they are advantageous only if they become licensed within an individual by not binding any self MHC-I molecule.

During infections with viruses downregulating MHC-I expression (i.e. B strains), the specificity of iNKRs plays a crucial role. If an iNKR binds the viral molecule, the virus successfully masks MHC downregulation, and escapes the NK cell response. Since iNKRs should learn to discriminate between self MHC-I and viral molecules, there is selection for iNKRs binding fewer MHC-I molecules in the presence of B strains (Figure S4.1 D). The disadvantage of having specific iNKRs is that they are less likely to bind an MHC-I in the same host, reducing the repertoire of licensed iNKRs in the host. The selection pressure to have licensed iNKRs that can detect MHC downregulation drives the evolution of large haplotypes, as a haplotype encoding several specific iNKRs is more likely to contribute licensed iNKRs .

Thus, our simulations show that in order to survive an infection with different viruses downregulating MHC-I, a single degenerate receptor pair is not sufficient. Immunity evolves by increasing the haplotype composition with specific activating and inhibiting NKRs.

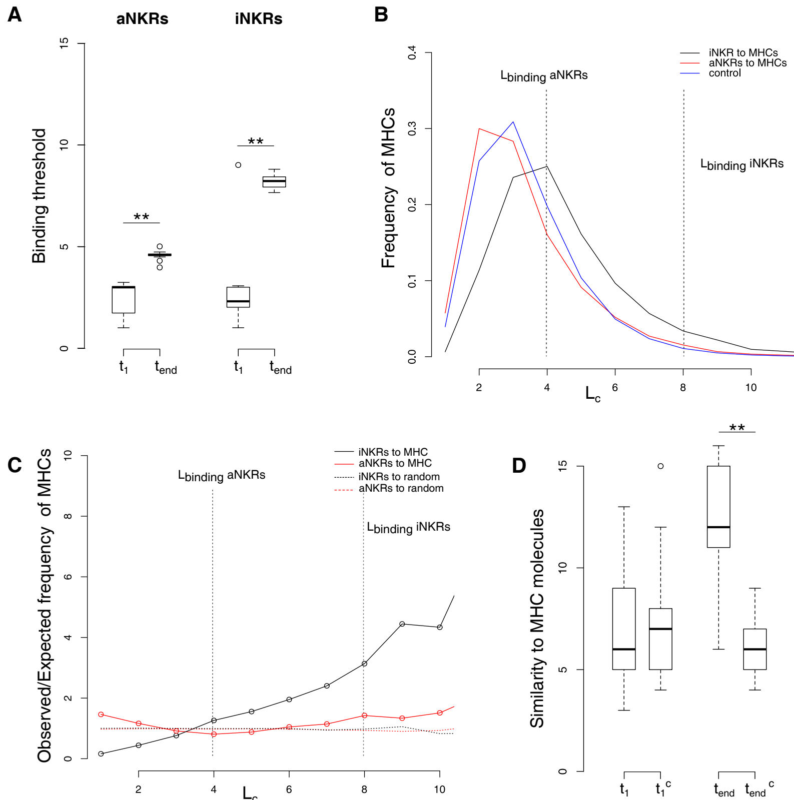

Figure 4.4. Evolution of NKR specificity binding when viruses evolve MHC-I mimicks. We study the evolution of NKR specificity by comparing the binding threshold of iNKRs and aNKRs (A) at the beginning of the infection ($t_1 = 10000$ years) and at the end of the simulation ($t_{end} = 3$ million years). (B) Fraction of MHC molecules that are recognized with a particular complementary adjacent length L_c. The blue curve depicts the expected frequency distribution determined by measuring the L_c between a 1000 random NKRs and a pool of 30 MHC alleles, and repeating this experiment 1000 times. The red and black curves represent the frequency distribution of the evolved aNKRs and iNKRs, respectively, at t_{end}. In both cases, the frequency distribution was determined by measuring the L_c of the evolved NKRs to the MHC alleles in their own population, and averaging over 15 simulations. (C) The fold change between the observed and expected frequency of MHC molecules recognized with a given L_c of the evolved iNKRs (black) and aNKRs (red). There is no difference to the expected values when analyzing the evolved NKRs to random MHC molecules (dashed lines). (D) Similarity between the viral and MHC molecules, determined by the maximal adjacent match between their bit strings. The boxplots represent the average over 15 simulations. t_x^c describes the control analysis we use at the different time points. As a control, we use randomly generated MHC molecules in (A) and randomly generated viral molecules in (D). In (A)-(D), the boxes represent the interquartile range, and the thick horizontal lines the median out of 15 simulations (** represent p values < 0.005, and were calculated using the Mann-Whitney U test)

4.2.3. iNKRs specialize to MHC molecules

The previous analysis shows that NKRs evolve to recognize a small fraction of MHC-I molecules in the population. Since the probability of a NKR interacting with its ligands, i.e. its specificity, depends on its binding threshold, we first measure the binding threshold of aNKRs and iNKRs before (i.e. $\mathrm{t}_1 = 10000$ years) and after the infections (i.e. $\mathrm{t}_{\mathrm{end}} = 3\mathrm{mio}$). As expected from the haplotype analysis, the binding threshold of both receptor types increases after the introduction of the viruses (Figure 4.4 A), confirming the evolution of specificity.

To further analyze the specificity, we next determine the complementarity of NKRs to the MHC-I molecules. To this end, we measure the maximal complementary adjacent match between each evolved NKR and the 30 MHC alleles in the same population (see Material and Methods). A random NKR is expected to bind maximally 30 % of random MHC molecules with a maximal complementary adjacent match of 3 bits, as predicted by the frequency distribution of MHC molecules recognized with a particular complementary adjacent length (Figure 4.4 B, blue curve). Similarly, a random NKR should not bind more than 5% of random MHC molecules with a maximal complementary match larger than six. In our simulations, the frequency distribution between the evolved NKRs and their own MHC molecules deviate from this expectation, with iNKRs recognizing more than two-fold more MHC molecules at their evolved binding threshold (Figure 4.4 B, black curve). Interestingly, the binding of the same evolved iNKRs to randomly generated MHC molecules does not deviate from the expected values (Figure 4.4 C), suggesting that iNKRs evolve to specialize to particular MHC molecules of the population. On the contrary, the complementarity of aNKRs to their cognate MHC molecules does not deviate from the expectation. These results suggest that there is selection pressure for iNKRs to specialize to particular MHC-I alleles. Indeed, carrying specialized iNKRs is advantageous for the hosts, as it enables iNKRs to detect a diverse array of MHC alleles, while minimizing the interactions with viral molecules.

4.2.4. Viruses evolve MHC-I mimicks

We next studied the evolution of the viruses. When analyzing the most dominant viruses, we found that they evolve to express molecules that are similar to the MHC molecules in the population (Figure 4.4 D), i.e. viruses evolve MHC-mimicks. The evolutionary advantage of MHC-like molecules is large, as MHC-mimicks hamper viral detection by

aNKRs, and impede iNKRs to detect missing self and provide protection against an MHC downregulating virus. Even after the evolution of more specific NKRs, hosts cannot fully adapt to the MHC-mimicks. At the end of the simulations, the iNKRs are "fooled" by the MHC-like molecule in approximately 30% of the infections (Figure 4.3 C (sum of protection by aNKRs and none)). Thus, self/non-self discrimination becomes more challenging when viruses evolve MHC-mimicks.

Viral species A and B consistently evolve MHC-I like molecules in all 15 simulations, indicating that MHC-I mimicks play a major role in their evasion of the NK cell immune responses, irrespective of the ability of the virus to downregulate MHC. We tested this by modeling host populations infected with viruses that are unable to mutate. These hosts are also able to recover from the infections, yet with a different evolutionary strategy (Figure S3).When viruses do not mutate, aNKRs can become specialized to particular viral molecules and hence evolve a high specificity. On the contrary, iNKRs evolve an intermediate specificity, which allows them to bind most of the MHC-I molecules in the population. Note that viruses downregulating MHC-I are less likely to "fool" iNKRs if their molecules cannot evolve to become MHC-like. Therefore, iNKRs do not need to be highly specific to distinguish self from non-self in this case. Because aNKRs become highly specific and iNKRs intermediate specific, it is much easier to have licensed receptors, resulting in the evolution of small haplotypes. As we obtain different results with non-mutating viruses, we conclude that viral evolution shapes the evolution of both specificity and polygenicity in the NKR cluster.

Taken together, we show that it is beneficial for viruses to evolve MHC-mimicks, and that populations adapt to such viruses by evolving expanded NKR clusters composed of several specific activating and inhibiting receptors.

4.3. Discussion

The exact evolutionary advantage of the polygenicity and polymorphism within the NKR cluster has remained puzzling. Since the detection of "missing-self" can be achieved by monomorphic and conserved receptor ligand-interactions, the evolution of polymorphic iNKR genes seems redundant. Even more intriguing is the unknown role of aNKRs, for which most of the ligands have not been identified yet. In this work, we study the evolution of NKRs in host populations infected with different viral species. We find that viruses have a large evolutionary advantage when they evolve MHC-like

molecules, especially if they are capable of downregulating the expression of MHC-I on infected hosts. In turn, hosts evolve expanded haplotypes composed of various specific aNKRs, and iNKRs that are specialized on particular MHC molecules. Thus, a polygenic and polymorphic NKR cluster encoding for specific receptors is beneficial to cope with co-evolving pathogens.

To model this highly complex evolutionary process, simplifying assumptions were necessary. As discussed in detail in chapters 2-3, our ABM is inspired on humans and KIRs, which has the advantage of having realistic parameters for processes like birth and death. However, by changing parameters, the ABM can be adapted to other species, as it qualitatively represents a model of the evolution of the expansion of the NKR complex. Indeed, our results hold true for different parameter settings and most of our assumptions.

Probably, the most critical assumption for our results is the asymmetry in the protection levels, as we assume that *at least one* aNKR, but *none* of the iNKRs should bind the viral molecule, to protect the host. We based this asymmetry on experimental data showing susceptibility and protection of mice to MCMV. C57BL/6 mice are resistant against MCMV because of their aNKR Ly49H binding the MCMV encoded protein m157, indicating that one aNKR directly recognizing decoy molecules is sufficient to provide protection. Similarly, the susceptibility of 129/J mice has been related to their iNKR Ly49I binding m157 with high affinity. This observation suggests that the contribution of one single inhibiting receptor dominates, making the mice susceptible. The latter is surprising because those NK cell subsets that do not express iNKRs recognizing the decoy should be able to detect missing self, proliferate, and protect its host. Whether Ly49I is the only iNKR binding m157 remains unknown, but the data clearly demonstrates that one iNKR-m157 interaction is sufficient for the virus to evade immunosurveillance.

The spontaneous evolution of MHC-like molecules by both types of viruses in our model suggests that this evasion mechanism provides an evolutionary advantage for many viruses. A virus expressing MHC-mimicks can avoid missing-self detection by engaging iNKRs, and will escape from aNKR-mediated NK cell responses as aNKRs should have low affinity for cognate MHC molecules. The few examples of CMV encoded MHC-I like molecules (including m157, and m04 in mice, and UL18 in humans) are in line with our findings. Given that MHC-like molecules evolve so easily in our model because of their evolutionary advantage, it remains intriguing why only few viruses MHC mimicks have been found in other viral species, but it is possible that these molecular mimicks are encoded by short sequences, and are thereby difficult to identify.

The characterization of specific ligands for aNKRs has remained difficult. Even though

some specific interactions (e.g. Ly49H/m157, or Ly19P/m04-H2^{k}) have been identified, the ligands for most activating KIRs remain unknown. Our model shows that it is indeed difficult to evolve specific interactions between aNKRs and particular viral proteins as viruses evolve rapidly, and become a moving target for aNKRs, impeding the adaptation to particular infectious pathogens. Note that the specific interactions found in the mouse CMV model are from particular inbred mice strains infected with one specific laboratory CMV strain. In outbred wild mice, Ly49H-mediated resistance is rather uncommon (Scalzo *et al.*, 2005), confirming that there is large heterogeneity in the interactions between aNKRs and viruses.

Although iNKRs and aNKRs share a large sequence similarity in their extracellular binding domain, aNKRs do not bind the ligands of iNKRs (Gillespie *et al.*, 2007, O'Connor and McVicar, 2013, Saulquin *et al.*, 2003). Studies have even shown that there is an evolutionary trend in humans and chimpanzees towards reducing the affinity of the aNKRs to the HLA epitopes C1 and C2 (Moesta *et al.*, 2010). On the contrary, inhibitory KIRs have rather specific interactions, binding preferentially four mutually exclusive epitopes on HLA molecules (A3/11, Bw4, C1 and C2) (Moretta *et al.*, 1996). We show that the specialization of iNKRs but not aNKRs on a small group of MHC-I molecules is advantageous to discriminate self/non-self more efficiently.

Summarizing, the evolution of both host receptors and viral immuno-modulatory molecules is an ongoing dynamic process. We have shown that the NKR genetic complex evolves polygenicity and specificity in response to rapidly co-evolving viruses.

4.4. Materials and Methods

4.4.1. Agent based model

The ABM consists of two types of agents, i.e. hosts and viruses, and three types of events: birth, death, and infection. During each time step of one week, we screen all hosts in a random order and confront them to one of the events. Hosts age over time and their ages, infection states, and infection types are updated at the end of every time step. This cycle is repeated for three million years to simulate long term evolution. The model used here is based on our previously described ABMs (see Chapters 2-3). Below is a detailed description of the extensions we implemented in the current model. The full description and model parameters are described in Chapters 2-3.

Bitstrings

All molecules, i.e. NKRs, MHC-I and viral proteins, are modeled with bit strings of length $L = 16$. Two molecules can interact whenever the longest adjacent complementary match between their strings exceeds a binding threshold $\mathrm{L}_{\mathrm{binding}}$ (Figure 2.1).

We create a gene pool of 30 MHC alleles which reflects the most common HLA B, and C alleles in the European population (Meyer *et al.*, 2007). The MHC molecules are not completely random, and they have a hamming distance (HD) of maximal 4 bits among each other. To create these alleles, we first create one random bit string and generate all possible strings that have a HD = 1 to the original sequence. For each of these "first generation mutants", we generate again all possible strings that have a HD = 1. This ensures that the HD is maximally 4. We randomly select 30 strings out of these "second-generation mutants" to fill the MHC pool.

Mutations

Only NKRs and viral molecules are allowed to mutate in our model. During a mutation event, NKRs evolve their bit string, binding threshold and signaling type. i.e. a new bit string will be randomly generated, to which a random $\mathrm{L}_{\mathrm{binding}}$ will be assigned. This new NKR can be inhibiting or activating with equal probability of 0.5. We do not model point mutations in order to decrease computational time. Similarly, viral molecules mutate into a new random bit string with a small probability ($p = 0.0005$) upon transmission to a new host.

Viral Infections

In these simulations, we consider two viral species. Both species, A and B, express a viral encoded molecule, representing viral proteins that are expressed on infected cells. The B species downregulates the MHC expression in the infected host, which allows us to model viruses like CMV, or HIV that use this mechanism to escape T cell responses (Cohen *et al.*, 1999, Collins *et al.*, 1998, Gall *et al.*, 1998, Gewurz *et al.*, 2001, Llano *et al.*, 2003, Schust *et al.*, 1998).

Viral molecules are randomly generated at the beginning of the simulation, and each virus can express only one. We model different levels of protection depending on the

viral strategy and its success to escape the immune response. Viral strains from the A species do not downregulate MHC. Hence, they do not escape the adaptive immune response, and can therefore be cleared with a high probability $p_A = 0.8$. However, if the host has at least one licensed aNKR binding a viral molecule of the A viruses, there is an extra NK cell activation and we increase $p_A = 0.95$. By downregulating the MHC expression, viruses from the B species escape the T cell response, and have a higher fitness than A viruses. Whether or not the host can clear the infection depends on its repertoire of licensed NKRs. We model four scenarios for the probability of clearing virus B:

$$p_{B,s} = \begin{cases} 0, & \text{if } s = 0 \text{ (inhibition, no activation)} \\ 0.25, & \text{if } s = 1 \text{ (inhibition, activation)} \\ 0.5, & \text{if } s = 2 \text{ (release from inhibition, no activation)} \\ 0.70, & \text{if } s = 3 \text{ (release from inhibition, activation).} \end{cases} \tag{4.1}$$

If at least one licensed iNKR but no aNKR binds the viral molecule, the NK cells of that host will not be activated, since the virus escapes the response of both T and NK cells and we set $p_B = 0$. If at least one licensed iNKR and one aNKR binds a viral molecule, we assume that the inhibiting signal is stronger than the activating. Therefore, we set $p_B = 0.25$. If none of the host's licensed NKRs binds the viral molecule, the NK cells of that host will not have any activating signal but they will be released from inhibition, and we set $p_B = 0.5$. If none of the host's licensed iNKR binds the viral molecule and at least one aNKR does, then the NK cells are not inhibited and they receive a strong activation signal, and we set $p_B = 0.7$. We set this value lower than p_A because the viruses from the B species always escape T cell responses and should therefore have an advantage over the A strains.

NK cell education

The education process of NK cells occurs during the birth event. iNKRs become licensed if they recognize at least one of the MHC molecules in its host. aNKRs become licensed if they do not bind any of the host's MHC molecules within an individual. An aNKR can only be licensed if that host has already one licensed iNKR (Figure 3.1).

Model initialization

The model is initialized with a host population of 4500 individuals, with random ages between 10 and 70 years. Hosts carry two MHC loci. Similarly, the initial NKR cluster is composed of two loci, encoding for one degenerate receptor pair, i.e. one iNKR and one aNKR, each being able to recognize all MHCs in the population. (Thus, initially none of the aNKRs will become licensed).

4.4.2. Analysis of recognized MHC molecules

To determine the expected distribution of the maximal complementary match L_c, we generate 1000 random NKRs and measure their longest complementary adjacent match to 30 randomly generated MHC alleles. The frequency of recognized MHCs with a particular L_c is averaged over 1000 of these “experiments”. We also perform the analysis using 30 “structured” MHC alleles, i.e. alleles generated with a maximal HD=4 as explained above, however the distribution of expected maximal complementary match remain the same (results not shown)

To analyze how this distribution changes after evolution, we measure the longest complementary adjacent matches between NKRs and the MHC molecules at $t_{end} = 3mio$. The final distribution is the average out of 15 simulations.

We also analyze the distribution of L_c between the evolved NKRs and random MHC molecules by measuring the L_c of a each NKR to 30 random MHC alleles, and repeating this experiment 1000 times. Here too, the frequency distribution does not deviate from the expected values (results not shown).

4.5. Acknowledgments

This work was financially supported by the CLS program of the Netherlands Organization for Scientific Research (NWO), grant 635.100.025.

4.6. Supplementary Information

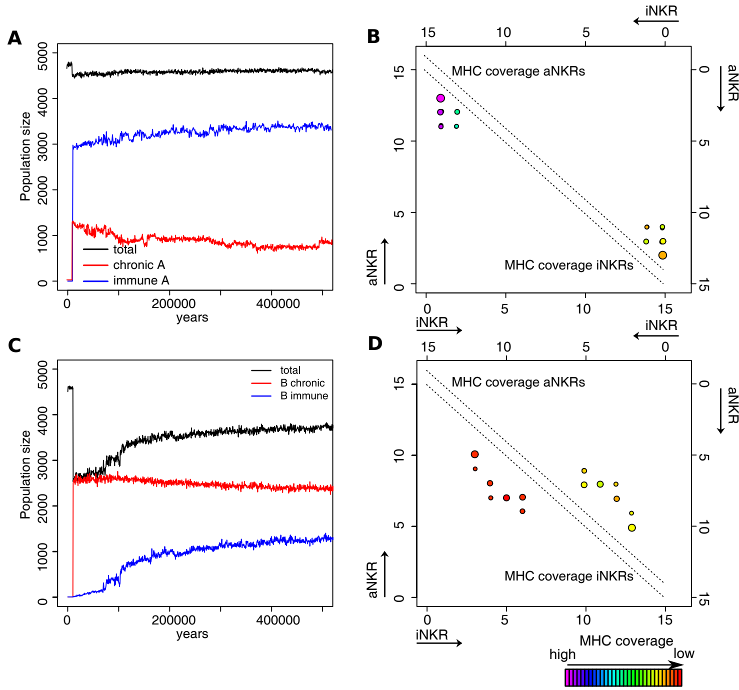

Figure S4.1. Populations infected with one viral strain. Single representative simulations of a host population infected with an A virus (A)-(B) and with a B virus (C)-(D). The infection dynamics (A)-(C) show that it is more difficult for a population to clear B strains than A strains, as shown with the total number of immune individuals (blue lines). The composition of the haplotypes at t_{end} (B)-(D) evolve depending on the viral species infecting the population. The haplotype composition and specificity are as described in Figure 4.2 B.

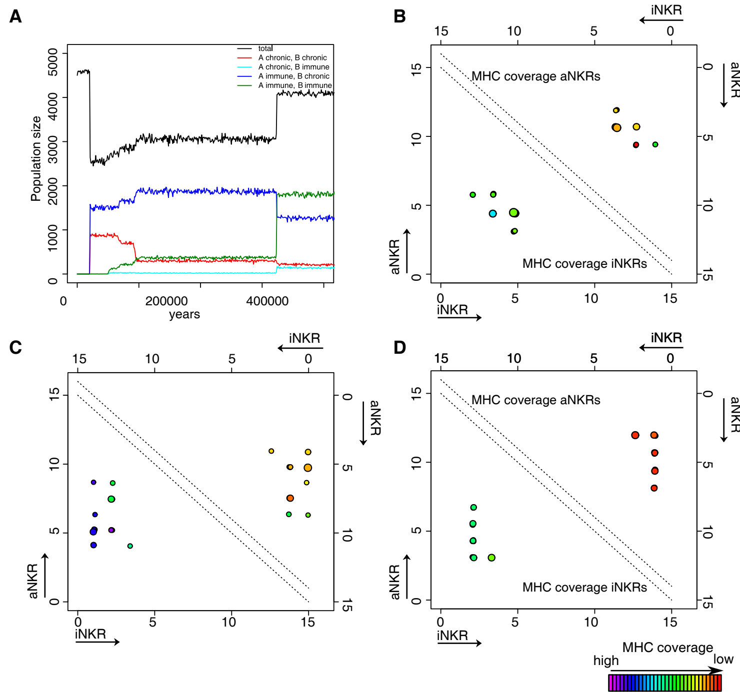

Figure S4.2. Viral evolution shapes the evolution of the NKR complex. The figure summarizes the results from simulations in which viruses are unable to mutate. (A) Single representative simulation of a population infected with both viral species. Although the initial viral spread causes a dramatic decrease of the total population size (black line), the hosts are able to recover fully, as shown by the increase in the immune individuals (green line). Haplotypes of a single host population infected with both viral species (B) or with only A strains (C), or only with B strains (D) at t_{end}. Because the virus is unable to mutate in this simulation, the evolved haplotypes show a different specificity than those shown in Figure 4.2 B, with aNKRs being highly specific and iNKRs more degenerate. The composition and specificity are as described in Figure 2 B.

Chapter 5

Selective MHC downregulation can explain the specificity and genetic diversity of NK cell receptors

Paola Carrillo-Bustamante, Can Keşmir, Rob J. de Boer

In preparation

Abstract

Natural killer cell receptors (NKRs) monitor the expression of MHC class I and stress molecules on the surface of cells to detect unhealthy tissue, such as infected or tumor cells. The NKR gene family shows a remarkable genetic diversity, containing several haplotypes varying in gene content and allelic polymorphism. The expansion of the NKR genes is species-specific, with different species evolving alternative expanded NKR genes, which encode structurally different proteins, yet perform comparable functions. So far, the biological function of this expansion within the NKR cluster has remained poorly understood.

Several viruses, including Epstein-Barr-Virus (EBV), CMV, HIV, decrease the expression of MHC-I to escape from the T cell immune response. This downregulation does not always affect all MHC molecules in the same way, but some viruses have evolved proteins that target only particular alleles. Since the HLA molecules presenting peptides to T cells (A and B) tend to be downregulated, while those inhibiting NK cells (C and E) tend to remain expressed, selective MHC downregulation seems to be a viral strategy to avoid missing-self detection.

To study the evolution of NKRs, we have developed an agent-based model (ABM) implementing an evolutionary scenario of hosts infected with herpes-like viruses that are able to evade the immune response by selectively downregulating the expression of MHC-I molecules on the cell surface. We show that hosts evolve specific inhibitory NKRs that are specialized to particular MHC-I molecules in the population.

5.1. Introduction

Natural killer (NK) cells are key players of the innate immune system, detecting and attacking viral infected and tumor cells. To detect aberrant cells but remain tolerant to healthy tissue, NK cells have several inhibiting and activating receptors, which tightly regulate their cytotoxicity. The recognition of major histocompatibility complex (MHC) class I molecules on the surface of target cells is central for NK cell mediated immune recognition. Several MHC molecules are ligands for inhibitory NK cell receptors (iNKRs) and signal NK cells to remain silent to self healthy cells. During some viral infections, the expression of MHC molecules decreases, resulting in a lack of inhibiting signals for the NK cell, and hence in NK cell activation (missing-self detection (Ljunggren and Kärre, 1990)).

Several viruses, including Epstein-Barr-Virus (EBV), cytomegalovirus (CMV), and the human immunodeficiency virus (HIV), decrease the expression of MHC-I. Interestingly, the downregulation does not always affect all MHC molecules in the same way, and some viruses have evolved proteins that target only particular alleles (reviewed in (Nash *et al.*, 2014)). HCMV encodes several immunoevasin proteins that selectively downregulate the expression of MHC-I on the cell surface (Nash *et al.*, 2014), such as US2 and US11, each targeting particular HLA-A and HLA-B alleles, by promoting their export into the cytosol for proteosomal degradation (Gewurz *et al.*, 2001, Llano *et al.*, 2003, Schust *et al.*, 1998). In addition to selective MHC-downregulation, HCMV encodes proteins that enhance MHC-I expression to inhibit NK cells, such as UL40 having a high sequence similarity to peptides from HLA-C alleles (Tomasec *et al.*, 2000, Ulbrecht *et al.*, 2000). HIV-1 also decreases the expression of particular HLA alleles. HIV Nef binds to the cytoplasmic tails of the HLA-A and HLA-B molecules in the ER, re-directing them to endo-lysosomal compartments for degradation (Schaefer *et al.*, 2008). Small differences in the cytoplasmic tails of HLA-C and -E prevent Nef from hampering their transport to the cell surface, which in turn prevents HIV-infected cells to be lysed by NK cells (Cohen *et al.*, 1999, Collins *et al.*, 1998, Gall *et al.*, 1998).

Since the HLA molecules presenting peptides to T cells (HLA-A and -B) tend to be downregulated, while those inhibiting NK cells (HLA-C and -E) tend to remain expressed, selective MHC downregulation seems to be a viral strategy to avoid missing-self detection. This "selective" downregulation can in turn shape the evolution of the NK cell receptors, as specific inhibiting receptors with non overlapping specificities for different MHC molecules can be advantageous to detect missing-self.

Here, we investigate whether selective MHC downregulation can exert enough selection pressure to drive the evolution of a polygenic and polymorphic NKR system. We develop an agent-based model (ABM) of host populations infected with herpes like viruses causing chronic infections. Our simulations show that NKRs readily evolve specificity for a single MHC loci. The evolution of these "MHC locus specific" detectors depends on the structure of the MHC molecules, i.e. how similar they are to each other.

5.2. Results

5.2.1. Agent Based Model

To study the evolution of NKRs in a host population, we developed an ABM based on our previously described models (see Chapters 2 and 3). Briefly, the host population consists of simplified humans infected with non-lethal viruses causing chronic infections. The hosts are diploid, carrying two polymorphic MHC loci and an NKR cluster. Of note, we only model inhibitory NKRs (iNKRs) in this work, as only iNKRs are involved in missing-self detection. iNKRs and MHC molecules are modeled with random sequences of zeros and ones (i.e. bit strings) as a simplified representation of amino acids. Whenever the longest adjacent complementary match between two strings exceeds a binding threshold, the molecules can interact. We model two different groups of MHC molecules, henceforth referred to as MHC-X and MHC-Y. The molecules within each MHC group are somewhat similar to each other, and we vary the hamming distance (HD) of the molecules within one group between 2, 4, 6, and 8 (see Material and Methods).

Initially, all hosts carry the same randomly generated NKR cluster, which is composed of one gene (i.e. we initialize the populations with homozygous individuals). This iNKRis degenerate, i.e. it can recognize every MHC molecule in the population. Upon birth, individuals inherit one NKR haplotype from each parent. During this process, NKRs can mutate their sequence, and their binding threshold (L), allowing for the emergence of novel receptors. If a newly generated receptor is so specific that it fails to recognize any MHC molecule in the population, it will be a *pseudogene*. In these simulations, the maximal number of NKRs that we allow per host is 5 genes per haplotype. We focus on the evolution of NKRs, and therefore fix the initial polymorphism of the MHC-X and -Y molecules throughout the simulations.

During development, NK cells undergo an education process during which their reactivity is “tuned”. The general consensus is that the binding of iNKRs with their cognate MHC molecules render these NK cells functional capacity (Anfossi *et al.*, 2006, Chalifour *et al.*, 2009, Elliott and Yokoyama, 2011). We implement this process at birth by evaluating the binding of iNKRs with their MHC ligands in the host. For simplicity, we do not model individual NK cell subsets, but estimate the total repertoire of “licensed” receptors per host. In each host, the repertoire of licensed receptors consists of iNKRs binding at least one of the host’s MHC molecules. In our model, only licensed NKRs participate during an immune response, assuming that at least one NK cell subset will express at least one licensed receptors, and will become activated and expand upon infection.

We model two viral species causing chronic infections (viral species A and B). Both viral species can escape the cytotoxicity of T cells by downregulating the expression of the MHC I molecules on the surface of the infected cells. Hereby, the downregulation is specific to each locus, i.e. virus species A downregulates the MHC-X molecules in the host, while virus B downregulates the host’s MHC-Y molecules.

Upon transmission, the host will enter a phase of acute infection, after which it can either recover or become chronically infected. Individuals clearing an infection become immune against that particular viral species for a period of ten years. Hosts can be co-infected with both viral species..

The probability of clearing the infection depends on the interactions between the iNKRs and the expressed MHC-I molecules in the host. For a virus to be cleared, the licensed repertoire of iNKRs must be able to detect MHC downregulation. If a virus downregulates the MHC X molecules, and all licensed iNKRs interact with the MHC Y molecules, the NK cells will have sufficient inhibition, and will not be able to detect the altered MHC expression. But if at least one licensed iNKR binds *none* of the expressed MHC Y molecules, the NK cell carrying that receptor will loose an inhibiting signal when MHC-X is downregulated, become activated, and provide protection (illustrated in Figure 5.1).

In a typical simulation, both viral species are introduced after the population has stabilized ($t_1 = 10000$ years). Very rapidly, the viruses spread through the population, infecting almost every individual, and causing a drastic reduction of the population size (Figure 5.2 A, black line). After approximately 200 host generations, individuals that are immune to each or both viral strains evolve (Figure 5.2 A, cyan, blue and green lines), causing a rapid recovery of the total population size. These results suggest that the initially susceptible host population evolves an NKR system providing immunity to both

types of infections.

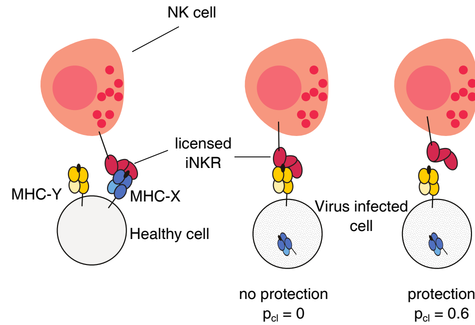

Figure 5.1. Model cartoon of the protection after infection with an MHC-downregulating virus. An iNKR becomes licensed if it binds at least one MHC molecule in the host. In this example, the iNKRs is licensed by binding the MHC-X molecule in that host. The virus downregulates the expression of all MHC-X molecules. The infection can be cleared with a probability $p_{cl} = 0.6$ if and only if the licensed iNKR fails to bind all MHC-Y molecules in that host.

5.2.2. Specific iNKRs protect hosts from (viral) infections

To study how the evolution of iNKRs allows the host population to recover, we next analyze the iNKRs before and after a evolutionary long period (Figure 5.2 B). The initial haplotype composed of one degenerate iNKR recognizes all MHC-I in the population, covering the entire space of both MHC-X, and -Y molecules. After the introduction of the viruses, there is selection for more specific iNKRs, as shown by the increase of their binding threshold L. As more specific iNKRs evolve, the number of immune individuals increases, indicating that the expansion of these specific iNKRs is necessary for the hosts to clear the infection.

Because of their higher specificity, the evolved iNKRs are expected to recognize fewer MHC molecules than the initial degenerate receptor. Interestingly, some iNKRs are still able to recognize most of the MHC-I molecules within one locus, despite their higher specificity (Figure 5.2 B). For example, at $t = 600000$, there is one iNKR recognizing all 15 MHC-X molecules, while binding none of the MHC-Y molecules. This specialization to the MHC-I loci allows the host to successfully detect the selective MHC-downregulation.

In our model, two processes are crucial in determining NK cell mediated protection.

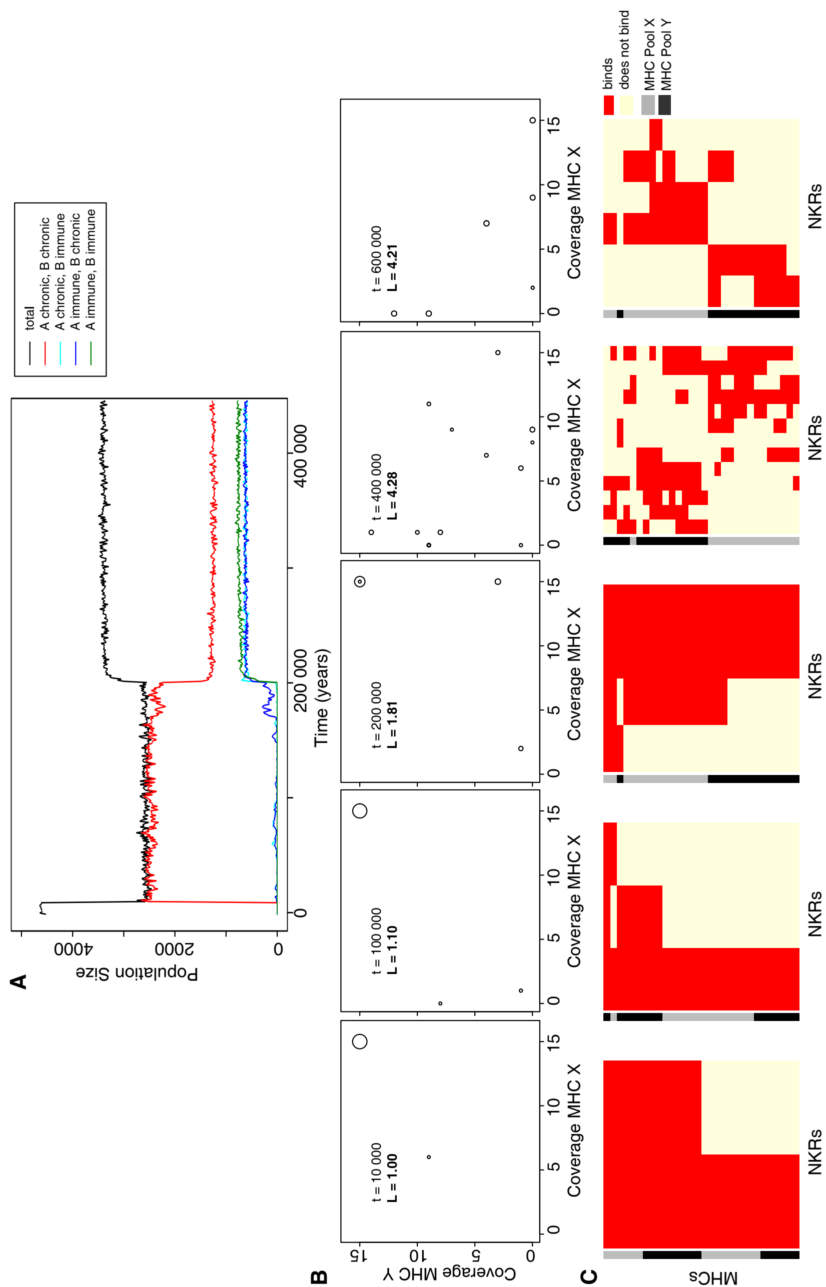

Figure 5.2. The evolution of MHC-X and MHC-Y detectors leads to an increase in immunity. Legend on next page.

Figure 5.2. The evolution of MHC-X and MHC-Y detectors leads to an increase in immunity. A. Representative simulation of a host population infected with both viral species. After their introduction at $t_1 = 10000$, the viruses spread rapidly through the population, infecting most individuals and causing a drastic decrease of the total population size (black line). Most hosts are chronically infected with both viruses simultaneously (red line). After several host generations, some hosts evolve immunity to viruses downregulating MHC-X molecules but remain chronically infected with viruses downregulating MHC-Y (blue line). Shortly afterwards, other individuals become immune to viruses downregulating MHC-Y molecules while being chronically infected with viruses downregulating MHC-X (cyan line). Finally, hosts that are able to clear both viruses evolve (green line), resulting in a recovery of the population. B. MHC coverage per iNKR and average binding threshold (L) at different time points (snapshots taken from the video provided in the Supporting Information). Each iNKR is depicted by one circle the size of which represents its frequency in the population. The position of each circle shows how many MHC-X and MHC-Y molecules in the population each iNKR can recognize. C. Clusters of iNKRs at different time points according to the iNKR-MHC-I binding matrix. We cluster both rows (MHC molecules) and columns (iNKRs), by using the Manhattan measure of distance and the Ward clustering algorithm. The heat map colors represent the iNKR-MHC-I binding; red illustrating binding and yellow no binding. The MHC molecules in this simulation were modeled with a hamming distance (HD) of two.

First, an iNKR must be licensed to participate in an immune response. Second, the licensed receptor should be able to detect missing-self. For an iNKR to become licensed, it must recognize as many MHC molecules as possible in the population. For instance, an iNKR binding all MHC-X molecules in the population will get licensed in every host. However, during an infection with a virus downregulating MHC-X, the licensed iNKR will only detect missing-self if it fails to recognize all MHC-Y molecules. Therefore, an "MHC-X detector" (i.e. an iNKR recognizing more MHC-X than Y molecules) is protective against viruses downregulating MHC-X molecules. Similarly, an MHC-Y detector is protective against MHC-Y downregulating viruses.

We further study the evolution of MHC-X and MHC-Y detectors by clustering the iNKRs according to their binding to all MHC-I molecules in the population (Figure 5.2 C). Our analysis reveals that after a long evolutionary period most iNKRs specialize to the MHC loci, having little overlap in the detection of MHC-X and MHC-Y molecules. Taken together, our results show that viruses downregulating selective MHC loci drive the evolution of iNKRs that are specific for different MHC-I loci.

5.2.3. Specialization to the MHC-I groups depends on the similarity of the MHC molecules

So far we analyzed results are from simulations with highly similar MHC molecules (i.e. MHC molecules having HD=2). To analyze how our results depend on the structure of the MHC molecules within each locus, we perform several simulations with a HD of 4, 6, and 8. Additionally, we carry out one set of simulations with totally random MHC molecules.

Regardless of the structural similarity of the MHC molecules, the initially susceptible host populations recover from the infection after a long evolutionary period. However, there are small differences in protection: populations having highly similar MHC molecules (i.e. HD = 2) recover better than those with random MHC molecules (Figure 5.3 A-B). Similarly, although all host populations evolve more specific iNKRs irrespective of the HD, the evolution of specific MHC detectors is more difficult as the similarity of the MHC molecules decreases (Figure 5.3C-F). Host populations carrying highly similar MHC-I molecules (i.e. HD=2) evolve four distinct clusters of iNKRs: a large one of MHC-Y detectors (cluster I in Figure 5.3 C), a small cluster of highly specific receptors (recognizing few MHC-X and -Y molecules, cluster II), a small cluster of degenerate receptors (cluster III), and a large cluster of MHC-X detectors (cluster IV). With decreasing similarity (i.e. increasing HD among MHC molecules), the distinction of the clusters becomes less clear (Figure 5.3 D-E). In simulations with random MHCs the classification of X or Y detectors disappears (Figure 5.3 F). In these cases, iNKR evolve still a high specificity, binding only a few molecules of each group. However, the "excellent" X or Y detectors no longer evolve. Because of the absence of excellent X and Y detectors, host populations carrying random MHCs are not able to recover as well as those having MHC molecules with HD = 2 (Figure 5.3 A,B).

These results were consistent for all 15 simulations we performed (n=15), showing that iNKRs become specific as a result of the selection pressure imposed by the virus. However, the specialization to the MHC loci occurs only if the latter have a structure that the iNKRs can adapt to (compare Figure 5.3 F with 5.3 C-E).

In all these simulations, the host populations evolve a similar binding threshold, L, regardless of the similarity among their MHC-I molecules. This is a surprising observation because it indicates that the binding threshold is not the only factor mediating specificity and specialization of iNKRs to particular MHC molecules. To analyze the NKR specificity further, we next determine the complementarity of iNKRs to the MHC-X and

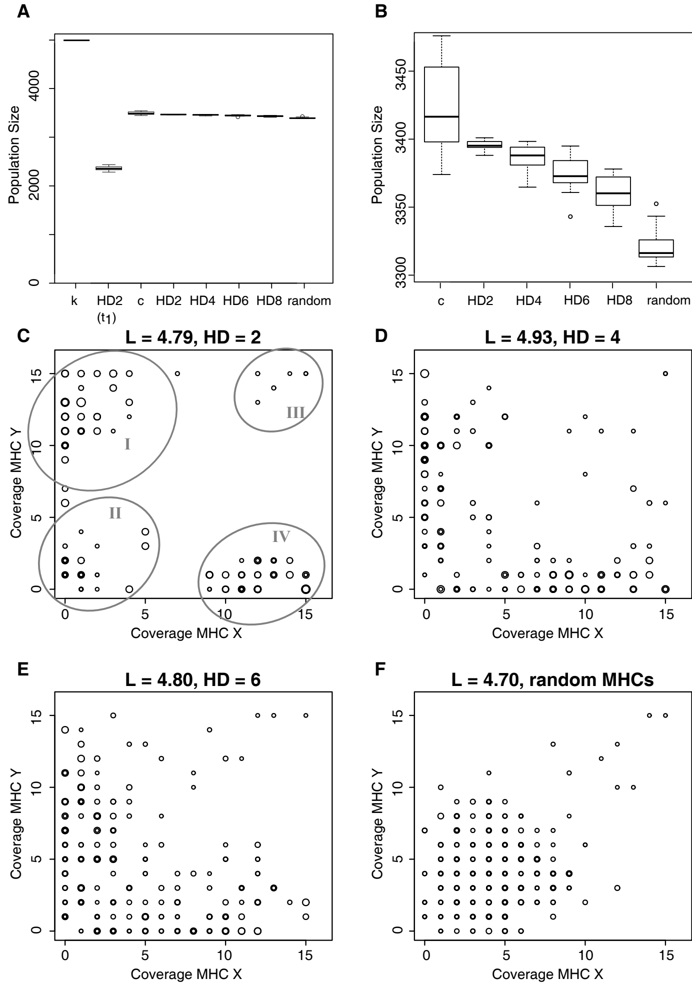

Figure 5.3. Evolution of MHC-X and MHC-Y detectors depends on the similarity of the MHC molecules. Legend on the next page.

Figure 5.3. Evolution of MHC-X and MHC-Y detectors depends on the similarity of the MHC molecules. Panel A depicts the average population size out of 15 simulations at the carrying capacity (k), immediately after the infection (i.e. $t_1 = 10000$) for HD 2, and at t_{end}, for simulations considering different similarity in the MHC molecules. As a control, we simulate populations infected with both viruses, where the probability of clearing the infection does not depend on the iNKR-MHC interactions (i.e. p_{cl} is always 0.6). After a long evolutionary period, all populations evolve immunity. However, the recovery was slightly higher in those simulations considering highly structured MHC molecules (more visible in B). Panels C-F show the MHC coverage per iNKR and the average binding threshold (L) at t_{end} out of all 15 simulations. Each host population has a different degree of MHC-I similarity, with HD = 2 (C), HD=4 (D), HD=6 (E), and completely random MHC molecules (F). Each iNKR is depicted by one circle, the size of which represents its frequency in the population. The position of each circle shows how many MHC-X and MHC-Y molecules in the population each iNKR can recognize. Note that the diagonal is only avoided in Panels C-E, i.e. if there is no pattern in the MHC-X and -Y molecules, there is no evolution of X and Y detectors.

MHC-Y molecules.

To have an expectation for the complementarity between iNKRs and MHC molecules, we generate a thousand random iNKRs and measure their longest complementary adjacent match (L_c) to randomly generated MHC molecules (see Material and Methods and Figure 5.4 A-C blue lines). This analysis reveals that a random NKR is expected to bind 30 % of random MHC molecules with a maximal complementary adjacent match of 3 bits, and no more than 5 % of random MHC molecules with a maximal adjacent complementary match of 6 bits. We next determine the distributions of the *evolved* iNKRs in our simulations by measuring the longest complementary adjacent match between each evolved X and Y detectors and the MHC-X and -Y molecules in the same population. The frequency distributions ofL_c between the evolved iNKRs and the MHC-X and -Y molecules deviate from the expected distribution: the distribution of X detectors to MHC-X molecules is shifted to the right (Figure 5.4 A, black curve), indicating that X detectors evolve a high complementarity to MHC-X molecules. At the same time, the distribution of X detectors to MHC-Y molecules is shifted to the left (Figure 5.4 A, red curve), suggesting that X detectors loose complementarity to MHC-Y molecules. Indeed, at their evolved binding threshold, X detectors recognize more than two-fold more MHC-X molecules and less than half MHC-Y molecules than expected. Similarly, MHC-Y detectors evolve high complementarity to MHC-Y molecules (Figure 5.4 B, red curve), while losing affinity to MHC-X molecules (Figure 5.4 B, black curve).

The deviation from the expected complementarity depends on the similarity of the MHC molecules. In simulations with less similar MHC molecules (e.g. HD=6, Figure 5.4 C-

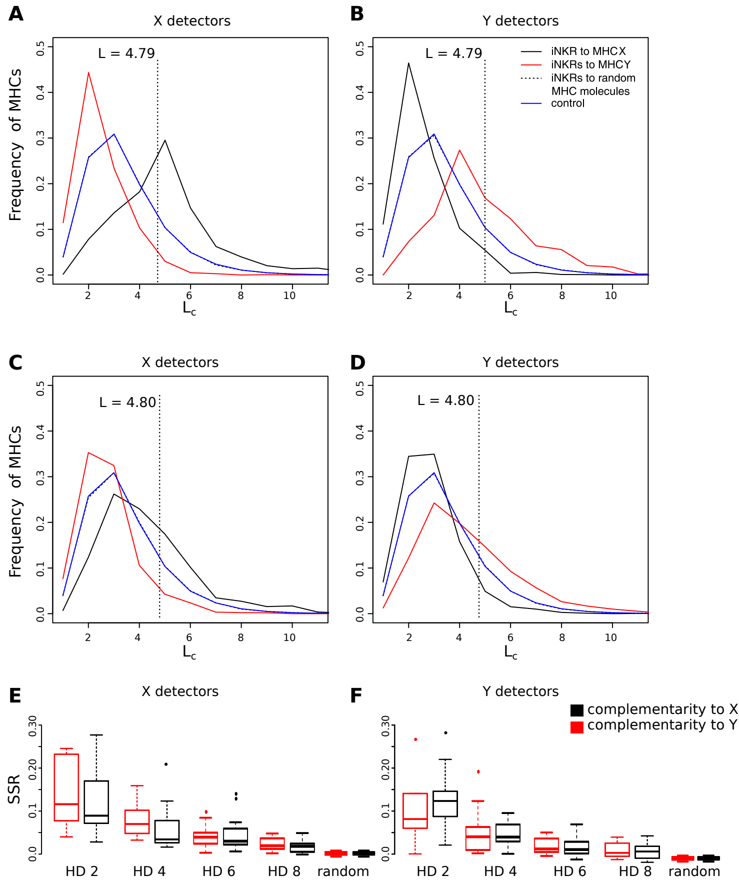

Figure 5.4. Evolution of NKR binding affinity depends on the similarity of the MHC molecules. We study the evolution of the binding affinity by comparing the complementarity of iNKRs to their MHC-I molecules. Panels A-D depict the fraction of MHC molecules that are recognized with a particular complementary adjacent length L_c. The blue curve describes the expected frequency distribution determined by measuring the L_c between 1000 random NKRs and a pool of random MHC molecules, and repeating this experiment 1000 times. The black and red curves represent the frequency distribution of the evolved iNKRs to MHC-X and MHC-Y molecules, respectively, at t_{end}. The frequency distribution was determined by measuring the L_c of the evolved MHC-X detectors (A,C) and MHC-Y detectors (B,D) to the MHC-X and MHC-Y molecules in their own populations, and averaging over 15 simulations. Panels A-B show simulations considering very similar MHC molecules (i.e. HD=2), and panels C-D show simulations with less similar MHC molecules (i.e. HD=6). The sum of squared residuals (SSR) from the observed complementarity distributions to the expected L_c distribution is summarized for X detectors (E) and Y detectors (F) for all 15 simulations performed per setting.

D), the evolved receptors still specialize to their MHC molecules, albeit with lower complementarity. To quantify the difference to the expected distribution, we calculate the sum square residuals (SSR) in each simulation (Figure 5.4 F-G). The higher the similarity of the MHC molecules, the more challenging it is for an iNKR to specialize to one MHC locus. Therefore, the complementarity distribution hardly deviates from the expected distribution with an increasing HD (e.g. HD = 8). The SSR is zero in those simulations considering random MHC molecules.

Taken together, our results show that iNKRs evolve specificity to detect viruses downregulating particular MHC-molecules. However, the iNKRs specialization to the different MHC loci depends on the structure of the MHC molecules. If the MHC molecules within one locus share a structural motif, iNKRs become complementary to one MHC-I locus, while losing capacity to bind molecules of the other locus.

5.2.4. Similarity of the MHC molecules shapes the evolution of NKR diversity

We next analyze the genetic diversity that evolves in our simulations. As a measure of genetic diversity we use the Simpson's reciprocal index (SRI). The SRI is a measure of diversity that can be interpreted as the probability that two random genes from two random hosts in the population are identical (Simpson, 1949). We measure the diversity of the different NKR haplotypes (Figure 5.5 A) and of the genes (Figure 5.5 B) that evolve after one million years of evolution. In all simulations, more genes evolve as a result of the selection pressure exerted by the viruses. Indeed, at least two iNKRs (one binding all X molecules, and another one binding all Y molecules) are necessary for an optimal missing-self detection. In populations with highly structured MHC molecules (i.e. HD = 2) these optimal X and Y detectors readily evolve, resulting in little selection for polygenicity and polymorphism.

In contrast, the selection for diversity increases in simulations when MHC molecules have a vague structural motif. In this case, iNKRs evolve a high specificity to detect the particular MHC-molecules that have been selectively downregulated. However, a very high specificity hampers their licensing process. Individuals carrying two different haplotypes (e.g. 10 different genes per individual) increase their chances of having at least one licensed iNKR, as they have a higher probability of recognizing one of the four MHC molecules in that same host. This heterozygote advantage in populations with less structured MHC molecules drives the evolution of a larger genetic diversity.

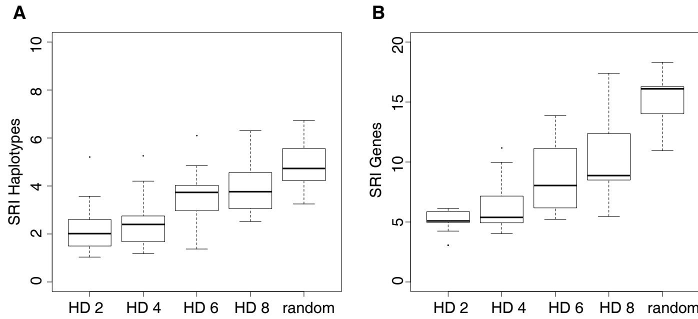

Figure 5.5. Evolution of NKR diversity depends on the similarity of the MHC molecules. We estimate the genetic diversity by computing the SRI of all haplotypes (A) and iNKRs (B) present in the population at $t_{end} = 3mio$. The SRI score is higher in populations with a decreased structure within the MHC molecules (i.e. an increased HD). The panels depict the summary out of 15 simulations for each setting.

5.3. Discussion

The functional relevance of the remarkable genetic diversity of NKRs remains intriguing. In this work, we investigate the evolution of NKRs in host populations that are infected with viruses that downregulate non-overlapping MHC-I molecules. We find that this selective MHC downregulation drives the evolution of specific iNKRs, which can specialize to the different MHC loci if the MHC molecules in one locus share structural motifs.

In order to develop an understandable model of a complex evolutionary process, simplifying assumptions were necessary. Therefore, we ignore the synergy between NKRs or the direct interaction between immune cells, and focus here only on the NK mediated immune response. Furthermore, we consider a global NKR repertoire composed of iNKRs only. Although the role of activating receptors and their genetic diversity is equally fascinating, we did not include them here, as they are not involved in the recognition of missing-self.

Several hypothesis have been proposed to describe the factors affecting the evolution of NKRs, including reproduction success (Parham and Moffett, 2013) and viral evasion strategies (Lanier, 2008, Sun and Lanier, 2009). Probably, the most intuitive explanation is that NKRs must be able to detect decreases in the expression of MHC molecules on the cell surface, particularly when viruses target specific non-overlapping MHC-I mo-

lecules. We show that if MHC molecules share structural similarity, iNKRs can readily evolve the ability to recognize them. These findings are in line with the at least four structural motifs known to be main ligands for inhibiting KIRs (A3/A11, Bw4, C1, and C2) (Trowsdale *et al.*, 2001), indicating that optimal detectors of these common MHC structural motifs have indeed evolved. According to our model, the specialization to these different MHC loci can exert selection pressure to evolve novel NKRs, providing thus a plausible explanation for the polygenicity and polymorphism of NKRs.

However, the degree of genetic diversity we observe in this study is limited. Indeed, in Chapters 2 and 3 we have previously shown that the selection pressure driven by viruses encoding MHC-like molecules (i.e. MHC "decoys") has a larger effect on the evolution of NKRs. Host populations require a lager NKR specificity to clear decoy-encoding viruses, which in turn exerts a stronger selection pressure on the NKRs. This suggests that the evolution of NKRs is subject to many more processes than only one viral immunoevasive mechanism.

Summarizing, we have shown that selective MHC downregulation drives the evolution of specific NKRs. However, the evolution of optimal MHC-loci detectors does not require an extensive degree of genetic diversity. Therefore, selective MHC downregulation is unlikely to be only explanation underlying the extensive genetic diversity observed in the NKR families.

5.4. Materials and Methods

5.4.1. Agent based model

The ABM consists of two types of agents, i.e. hosts and viruses, and three types of events: birth, death, and infection. During each time step of one week, we screen all hosts in a random order and confront them to one of the events. Hosts age over time and their ages, infection states, and infection types are updated at the end of every time step. This cycle is repeated for three million years to simulate long term evolution. The model used here is based on our previously developed ABMs (see Chapters 2, and 3). Below is a detailed description of the extensions we implemented in the current model. The full description and model parameters are described in Chapters 2, and 3.

MHC molecules

We create two gene pools of 15 MHC molecules each which reflects the most common HLA B, and C alleles in the European population (Meyer *et al.*, 2007). We perform simulations differing in the similarity of their MHC molecules. In the structured simulations, MHC molecules can have a hamming distance (HD) of maximal 2, 4, 6, or 8 bits to each other. To create the MHC molecules with HD = 2, we first create one random bit string and generate all possible (16) strings that have a HD = 1 to the original sequence. We take all of these strings and fill the pool of MHC X molecules. The pool for the Y molecules is filled similarly, yet with a different randomly generated bit string. For the populations with MHC molecules with HD = 4, we repeat this procedure: for each of the “first generation mutants” (i.e. those with HD = 1 to the original, HD = 2 to each other), we generate again all possible strings that have a HD = 1. This ensures that the HD between the strings in the second generation is maximally 4. We randomly select 15 strings out of these “second-generation mutants” to fill one MHC pool. Accordingly, we chose 15 strings from the “third-generation mutants”, and from the “fourth-generation mutants” to fill the MHC pools with HD=6, and HD=8, respectively. For the simulations with random MHC molecules, each gene pool is filled with 15 randomly generated bit strings.

Viral Infections

In these simulations we consider two viral species. Both species, A and B, downregulate the MHC expression in the infected host, but species A downregulates the molecules in the MHC-X locus, whereas species B downregulates those in the Y locus. Viruses do not evolve in this model. The infection will be cleared with a probability p_{cl} depending on the interactions of the licensed iNKRs with the expressed MHC-I molecules. If a virus downregulates all MHC-A molecules, and the iNKRs binds none of the MHC-B molecules, the NK cell will perceive a reduced inhibiting signal, resulting in a successful missing-self detection. In this case, the infection is cleared with $p_{cl} = 0.6$ (see Figure 5.1).

Model initialization

The model is initialized with a host population of 4500 individuals, with random ages between 10 and 70 years. Hosts carry two MHC loci, each of them encoding two genes from the MHC-X gene pool, and two genes from MHC-Y pool, respectively. The initial

host population is homozygous for NKRs, carrying a NKR haplotype composed of two copies of one degenerate iNKR, i.e. an iNKR being able to recognize all MHCs in the population.

5.4.2. Clustering of NKRs

We use a $(n \times m)$ matrix A to describe the binding of all n iNKRs and m MHC-I molecules in the population. If the i-th iNKR binds the j-th MHC-I molecule, the entry of the matrix A_{ij} will be one, otherwise A_{ij} will be zero. We cluster the iNKRs and the MHC-I molecules according to this binding matrix by clustering both rows and columns. We use the Manhattan measure of distance and the Ward clustering algorithm.

5.4.3. Analysis of recognized MHC molecules

To determine the expected distribution of the maximal complementary match L_c, we generate 1000 random NKRs and measure their longest complementary adjacent match to randomly generated MHC molecules. The frequency of recognized MHCs with a particular L_c is averaged over 1000 of these “experiments”.

To analyze how this distribution changes after evolution, we measure the longest complementary adjacent matches between iNKRs and the MHC-A, and MHC-B molecules after one million years of evolution. An iNKR is classified as an MHC-X detector if it binds to more MHC-X molecules than MHC-Y molecules at its evolved L. Similarly, MHC-Y detectors are those iNKRs that recognize a larger number of Y molecules than X molecules. The rare set of iNKRs recognizing the same number of X and Y MHC molecules are classified as neutral receptors and are excluded from this analysis. The final distribution is the average out of 15 simulations.

We also analyze the distribution of L_c between the evolved iNKRs and random MHC molecules by measuring the L_c of each iNKR to random MHC molecules, and repeating this experiment 1000 times. This frequency distribution does not deviate from the expected values (results not shown).

5.5. Acknowledgments

This work was financially supported by the CLS program of the Netherlands Organization for Scientific Research (NWO), grant 635.100.025. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Chapter 6

Can peptide sensitivity explain the diversity of NK cell receptors?

PAOLA CARRILLO-BUSTAMANTE, ROB J. DE BOER, CAN KEŞMIR

In preparation

Abstract

The activity of natural killer (NK) cells is tightly regulated by inhibiting and activating receptors. Particularly, inhibitory killer immunoglobulin-like receptors (iNKRs) survey the peptide-HLA complexes expressed on the surface of target cells. The binding of iNKRs has been shown to be sensitive to the peptide presented by HLA-I, implying that iNKRs have the ability to recognize changes in the peptide-HLA repertoire, which may occur during viral infections. To study how the peptide repertoire changes upon infection, and whether an iNKR is able to detect these changes, we combine a mathematical model with data analysis of cells prior and after infection with measles virus. Our study reveals that most changes in the repertoire of iNKR ligands are caused by the altered expression of self-peptides. Additionally, we show that an iNKR can detect changes in the presented peptides only if it is sufficiently specific. This required specificity subsequently exerts a selection pressure for the evolution of diverse iNKRs.

6.1. Introduction

Natural killer (NK) cells are key players of the innate immune response. Their activity is tightly regulated by several germline encoded inhibiting and activating receptors, including the highly polymorphic killer immunoglobulin-like receptors (KIRs) which interact with the classical HLA class I molecules expressed on their target cells. Due to a dominance of inhibiting over activating signals, healthy cells do not initiate NK cell cytotoxicity. However, virus-infected or malignant cells may either up-regulate ligands for activating receptors ("induced self detection") or down-regulate ligands for inhibiting receptors ("missing self detection"), thereby inducing NK cell activation and target cell lysis (Ljunggren and Kärre, 1990, Raulet and Vance, 2006).

Unlike T cell receptors, inhibiting KIRs (iNKRs) are not highly specific for particular peptide-MHC (pMHC) complexes. Instead, they recognize subsets of HLA molecules sharing structural motifs, including the four mutually exclusive epitopes on HLA alleles: A3/11, Bw4, C1, and C2 (Moretta *et al.*, 1996). However, several studies have shown that iNKRs are also sensitive to the peptides bound by HLA molecules (Fadda *et al.*, 2010, Hansasuta *et al.*, 2004, Malnati *et al.*, 1995, Peruzzi *et al.*, 1996, Rajagopalan and Long, 1997, Thananchai *et al.*, 2007). Crystal structures of KIR2DL1 and KIR2DL2 in complex with their HLA-C ligands further supported this observation by revealing that specifically positions 7 (P7) and 8 (P8) of the MHC bound peptide are in direct contact with residues of the inhibitory KIR (Boyington *et al.*, 2000, Brooks *et al.*, 2000, Fan *et al.*, 2001, Li and Mariuzza, 2014). This suggests that NK cell activation may be regulated in a peptide dependent manner. Indeed, MHC-peptide complexes that have weak or no binding to iKIRs can efficiently reduce KIR-mediated inhibition (Cassidy *et al.*, 2014, Fadda *et al.*, 2010). Recent studies also show that sequence variations within HLA-C restricted HIV epitopes have a large impact on the binding of inhibiting KIR2DL2, with some peptides enhancing and others disrupting the binding, and with it the inhibition of NK cells (Fadda *et al.*, 2012, van Teijlingen *et al.*, 2014).

The previous mentioned studies clearly underline potential functional consequences of the peptide in KIR-MHC binding, and thereby in NK cell regulation. This calls for an extension of the current model of NK cell activation: "induced" or "missing self" detection should be complemented by "altered self", where changes in the MHC-I peptide repertoire modulates the NK cell signaling. MHC bound peptides can promote KIR mediated activation either by binding stronger to activating KIRs, or by abrogating inhibiting signals. Alternatively, particular peptides could enhance the binding of iKIRs to the MHC,

resulting in a stronger inhibiting signal. If viruses had the ability to encode peptides that increase the iNKR-MHC binding, they would escape from NK cell activation (Alter *et al.*, 2011, Fadda *et al.*, 2012, van Teijlingen *et al.*, 2014).

In this study we will address two questions. First, is it likely that iKIRs can detect changes in the peptide repertoire bound to MHCs? And secondly, can this "peptide sensitivity" provide an explanation for the high polymorphism observed in the KIR gene family? To investigate the possibility of NK cell regulation via altered peptide repertoires, we analyze how the iNKR-ligands change after infection with measles virus (MV) using peptide elution data from cells prior and after infection with this virus. Our analysis shows that iNKRs need to be specific to detect changes in the peptide repertoire, and that this required specificity might be sufficient to select for iNKR diversity.

6.2. Results

6.2.1. How do iKIR ligands change after a viral infection?

It is natural to think that the MHC presented peptide repertoire changes during a viral infection. Although several studies have tried to address this issue (Wahl *et al.*, 2010), there are no quantitative estimates of the changes of iNKR ligands upon infection. Therefore, we make use of recent data that has been obtained in a collaboration with the group of Schellens et al. (Schellens *et al.*, 2014). In this study, peptide-HLA complexes were purified from four B lymphoblastoid cell lines (BLCLs), after which acid elution was applied to separate the peptides from their HLA molecules. All BLCLs were left uninfected, or were infected with MV. To isolate the HLA-peptide complexes, the authors used an antibody specific for HLA molecules, having a comparable affinity for all three major HLA class I molecules (HLA-A, -B and -C). As a result, the HLA assignment of the eluted peptides was unknown, and peptide-HLA binding affinity prediction programs (i.e. NetMHC) were used to determine which of the HLA class I molecules expressed by BLCLs most likely presented the identified peptides. By using this approach, the number and abundance of self and viral peptides associated to particular HLA-alleles could be determined. The results are summarized in Table 6.1 and Figure 6.1.

There is a large variation in the change of the peptide pools after the infection with MV. While there is an increase in the number of eluted peptides after the infection in BLCLs 1053 and 1112, in BLCLs 1090 and 1077 the total peptide abundance decreases after

Table 6.1. Number and abundance of self and viral peptides in uninfected BCLC and MV infected BCLC. Summarized from (Schellens *et al.*, 2014).

	BLCL identifier							
	1053		**1112**		**1090**		**1077**	
HLA alleles[a]	A*0201 A*0301 B*0701 B*0701 Cw*0702 Cw*0702		A*0201 A*0201 B*1501 B*4402 Cw*0304 Cw*0501		A*0201 A*1101 B*3501 B*4402 Cw*0401 Cw*0501		A*0101 A*2402 B*0801 B*4001 Cw*0304 Cw*0701	
	unique peptides (n)	**copy number (T)**	**unique peptides (n)**	**copy number (T)**	**unique peptides (n)**	**copy number (T)**	**unique peptides (n)**	**copy number (T)**
Self before MV	903	99656	1926	832801	2270	734854	1329	405715
Self after MV	532	82188	1851	1126080	1463	283201	643	181285
MV	26	32084	25	48124	22	10178	12	2679
Overlap[b]	0.55		0.42		0.41		0.52	

a: The HLA-alleles sharing the structural motifs A3/A11 and Bw4 are: A*0301, A*1101, A*2402, and B*4402. All HLA-C alleles belong to either the C1 or C2 epitope.
b: Overlap of self presented peptides before and after MV infection.

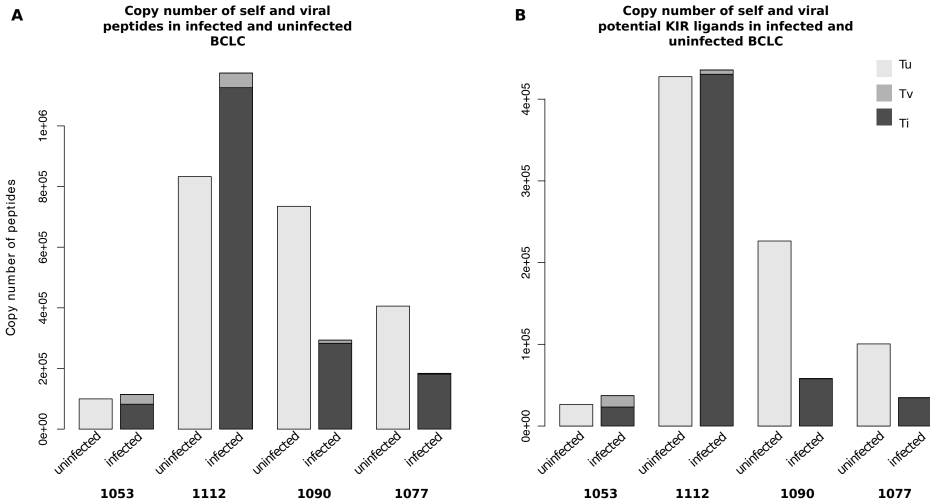

Figure 6.1. Abundance of self and viral peptides. In panel A the bars depict the total copy number of eluted self peptides in uninfected (T_u) and MV infected BLCLs (T_v), as well as the abundance of viral peptides (T_i). Panel B depicts the copy number of potential ligands for iKIR, i.e. those peptides that are restricted to HLA alleles expressing any of the four motifs A3/A11, Bw4, C1, and C2. In half of the four cell lines, the total number of peptides (i.e. T_i plus T_v) increases after MV infection. Data adopted from (Schellens *et al.*, 2014).

MV infection (Figure 6.1 A). Of note, the majority of peptides were derived from self proteins, whereas only very few peptides were derived from viral proteins. Since the main ligands for iNKRs include at least four motifs, i.e. A3/A11, Bw4, C1, and C2 (Trowsdale *et al.*, 2001), we performed a second analysis of only those peptides that are presented by the HLA molecules carrying any of these four structural motifs (Figure 6.1 B). The distribution of these restricted peptides is very similar to that of all peptides, suggesting that the ligand concentration of iNKRs would increase in two out of the four cell lines studied here, and decrease in the other two.

6.2.2. iKIRs can detect decreases in ligand concentration if they have sufficient specificity

The observation of an increased peptide/ligand concentration after MV infection suggests that some iNKRs could receive an even stronger inhibiting signal, hampering NK cell activation. Note however that different iNKRs will recognize different subsets from all these peptides. If there was one iNKR detecting a decrease in its set of ligands, the NK

cells carrying this iNKR would receive a reduced inhibiting signal and could become activated. Thus, the probability of a single iNKR detecting a decrease of ligands should determine whether or not there could be an NK cell response.

To study "altered-self" recognition, we develop a simple mathematical model that considers how the iNKR ligand pool changes upon infection, and estimates the chance of an iNKR detecting these changes. First, we estimate the probability of a peptide being a iNKR ligand. Given that iNKRs are in contact with P7 and P8 of the presented peptide (Boyington *et al.*, 2000, Brooks *et al.*, 2000, Fan *et al.*, 2001, Li and Mariuzza, 2014), and there can be 20 amino acids per position, an iNKR distinguishing between any pair of amino acids in those positions will recognize a particular peptide with a chance $p = 1/400$. Therefore, the probability that i out of a pool of n unique peptides form a set of iNKR ligands can be described with the following binomial distribution:

$$P(i) \quad = \quad \binom{n}{i} p^i (1-p)^{(n-i)}. \tag{6.1}$$

With this model, we can analytically determine the probability distribution of an individual iNKR binding i self peptides before MV infection ($P_u(i)$), binding i self peptides after MV infection ($P_i(i)$), and binding i viral peptides ($P_v(i)$) for each cell line (Figure 6.2). Moreover, for every value of i (i.e. from $i = 0$ to $i = n$) we can obtain the corresponding ligand density that an individual iNKR can bind in the set. Using the total abundance of peptides T as described in Table 6.1, this density can be expressed as $i \times T/n$. Hereby, we assume that all unique peptides are presented at equal frequencies.

To compute these probability distributions and the corresponding ligand densities, we use the values of unique peptides, n, and total abundance, T, as shown in Table 6.1, thereby including all peptides, and not only those restricted to HLA-A3/A11,- Bw4, -C1 or -C2 alleles. We include all peptides for two reasons. First, the changes in the number of peptides upon infection between all peptides and those restricted to HLA-A3/A11,- Bw4, -C1 or -C2 are similar (Figure 6.1). Second, in BLCL 1077, no viral peptides are presented by HLA molecules expressing these four HLA epitopes. Thus, despite having an overestimation of the number of peptides, the incorporation of all peptides allows us to model a more general scenario, comprising changes in the ligand repertoire induced by both self and viral peptides. The probability distribution before MV infection predicts that, at a specificity $p = 1/400$, an iNKR will most likely recognize between one and nine unique self-peptides as ligands (Figure 6.2 A). After the infection, an iNKR is likely to detect between one and six self-peptides (Figure 6.2 B). However, because the number of unique viral peptides is very low compared to that of self-peptides, the distribution of

iNKR detecting viral ligands is skewed, indicating that an iNKR will at most recognize one unique viral peptide with a probability of 0.1 (Figure 6.2 C).

To estimate how the abundance of iNKR ligands changes after MV infection, we next sample from each of these distributions, and compare in each sample whether the number of ligands sampled from $P_u(i)$, i.e. in uninfected cells, is larger, or smaller, than the abundance of ligands obtained after sampling from $P_i(i)$ and $P_v(i)$, i.e. after the infection. We repeat this procedure 10000 times for all four BLCLs to robustly estimate the chance of one random iNKR detecting a decrease of ligands after the infection (Figure 6.3). As expected, the abundance of iNKR ligands decreases in BLCLs having a decreased peptide presentation after infection, i.e. BLCL 1090 and 1077 (Figures 6.3 C,D). In these cell lines, at least 75 % of the sampled iNKRs detect a lower number of ligands (pie charts in Figures 6.3 C,D). Interestingly, subsets of iNKRs can also detect decreases in their ligands in BLCLs 1053 and 1112, even though these cell lines have an enhanced peptide presentation after MV infection (Figures 6.3 A,B). The most striking case is BLCL 1112, whereby one third of the iNKRs are expected to detect a decrease in their ligands. Altogether, these results suggest that highly specific iNKRs have the ability to detect decreases in their ligands, even if the amount of presented peptides after infection increases.

Until now, we have considered the maximal specificity of iNKRs, i.e. that they can discriminate between any pairs of amino acids in P7 and P8 (i.e. $p = 1/400$). iNKRs are probably less specific and could, for instance, recognize groups of amino acids (e.g. non polar, polar, basic and acidic) in P7 and P8. If there were indeed four such amino acid groups, the chance of any peptide becoming an iNKR ligand would increase ($p = 1/16$). Our sampling model shows no major changes when decreasing the specificity to $p = 1/16$, and subsets of iNKRs are still able to detect changes in the ligand repertoire after the infection in all BLCLs (results not shown). To have an estimate of the actual iNKRs specificity, we adopt data published by Fadda et al (Fadda *et al.*, 2010), where the binding of KIR2DL2, KIR2DL3, and KIR2DS2 to HLA-C-peptide complexes was studied. Out of 59 peptides where amino acids in P7 and P8 form a unique pair, 13 were described as KIR binders (Supplementary Table 1 in (Fadda *et al.*, 2010)), suggesting that the probability of an iNKR recognizing a peptide as a ligand is approximately $p = 0.2$. For this specificity, iNKRs can no longer detect "altered-self" in cells presenting a larger number of peptides during viral infections (Figure 6.4). In BLCL 1053, one third of the sampled iNKRs detect a decrease in the ligand density, whereas none of the iNKRs detects a decrease in ligands in BLCL 1112. Thus, in cells with an enhanced peptide presentation after a viral infection, subsets of iNKRs can only be sensitive to decreases

in the ligand repertoire, if they are more specific than estimated by Fadda et al. (compare Figure 6.3 with Figure 6.4).

6.2.3. Sequence based analysis of iNKR ligands

In our model the identity of peptides before and after the infection is assumed to be independent. However, the presentation of some self-peptides might not change at all after the infection. Indeed, there is an overlap of at least 40 % among the self peptides presented before and after MV infection (see Table 6.1). To study the actual change of iNKR ligands in the experimental data set, we performed a sequence based analysis of the eluted peptides before and after the infection. We first analyze the 9-mers among the peptides presented by HLAs carrying either the A3/A11, Bw4, C1 or C2 motif. We then group the peptides according to the unique amino acid combinations in P7 and P8, and observe in how many cases there is a decrease, or an increase, in the abundance of the peptides carrying these particular amino acid combinations after the infection with MV. Since we analyze the actual peptides, we can here use the real copy numbers of each peptide (or amino acid combination) as described by Schellens et al. (Schellens *et al.*, 2014). Assuming that iNKRs can distinguish between any pair of amino acids in those residues (i.e. $p = 1/400$), we thus again consider how likely individual iNKRs can detect changes in their ligand pool. The results are summarized in Figure 6.5.

In all BLCLs, the number of “KIR motifs” (i.e. unique pairs of amino acids in P7 and P8) is smaller than 400, indicating that not all possible amino acid pairs are present in P7 and P8 of the peptides presented by these cell lines. Of these motifs, the majority are derived from self-peptides, confirming that indeed most changes in the ligand pool are not caused by viral proteins. Additionally, the decrease of KIR motifs after MV infection is large in all BLCLs (Figure 6.5 A, BLCL 1112), suggesting that at least 50 % of specific individual iNKRs would have the ability to detect a loss in their ligands. However, iNKRs are not merely sensitive to the presence of a peptide but also to the total amount of pMHC complexes it interacts with (reviewed in (Cassidy *et al.*, 2014)). Therefore, it is important to quantify by how much the abundance of peptides changes after the viral infection, to predict the functional consequences of “altered-self” recognition. By computing the percentual decrease of KIR motifs after the viral infection, we observe that, indeed, a large fraction of the KIR motifs disappear in all cell lines, even in BLCLs with an enhanced peptide presentation (e.g., in BLCL 1112, at least 25 % of KIR motifs disappear, and in one third of the motifs the density decreases by half, Figure 6.5 A)

By restricting the peptides to those presented by the four epitopes A3/A11, Bw4, C1, and C2, the size of the data set was reduced by approximately half. Therefore, we extended our analysis and include all 9-mers in the data set. The results are remarkably similar to those from the potential ligands, indicating that highly specific iNKRs can easily detect substantial changes in the peptides presented by all BLCLs studied here (Figure 6.5 B). Since NK cell degranulation increases linearly with decreasing ligand concentration (Cassidy *et al.*, 2014), this strong decrease in ligand density is likely to induce a strong NK cell activation.

6.2.4. Specific recognition of peptides requires several iNKRs

Having an estimation of the probability that an iNKR recognizes changes in the ligand pool (Figure 6.5), we can speculate about the number of receptors required for a successful iNKR mediated NK cell response. Consider the "worse-case" scenario, where approximately one third of iNKRs detects a decrease in their ligands after MV infection (i.e. BCLC 1112 in Figure 6.5 B, dark gray area). After interacting with this cell line, approximately 10 % of the specific iNKRs would result in a strong NK cell activation, as 10 % of KIR motifs disappear upon the viral infection (BCLC 1112 in Figure 6.5 B, large white bullets). Because an individual basically needs at least one iNKR detecting altered self, as NK cells expressing that iNKR should become activated and respond, it is tempting to speculate that ten different iNKRs per individual would be sufficient to successfully detect "altered-self" when 10 % of KIR motifs are completely downregulated. Note that NK cell degranulation increases linearly when the concentration of peptide-MHC complexes on the cell surface decreases (Cassidy *et al.*, 2014). If a reduction of iNKR ligands by half would also be sufficient to mount a successful NK cell response, we obtain for BLCL 1112 that approximately 25 % of iNKRs detect a decrease in their ligands(BCLC 1112 in Figure 6.5 B, small white bullets), implying that about four iNKRs would be required for NK cell activation.

To have a lower estimate for the required KIR diversity, consider now the "best-case scenario" (i.e. BLCL 1077 in Figure 6.4 B). Hereby, one third of iNKRs would completely lose their ligands (BLCL 1077 in Figure 6.4 B, large white bullets), and two thirds of iNKRs would detect ligands decreasing in their abundance by half (BLCL 1077 in Figure 6.4 B, small white bullets). Thus, two or three iNKRs would be sufficient to detect a decrease in ligands, and result in activation of some NK cells. This analysis of the best and worse case scenarios indicate that one single iNKR is not sufficient for "altered-self" detection.

6.3. Discussion

The potential functional consequences of the presented peptide in KIR-MHC interactions have remained unresolved. Here we use a simple modeling approach and data analysis to study 1) how the peptide repertoire changes after a viral infection, and 2) the probability of iNKRs to recognize these changes. We show that most changes in the peptide pool are derived from self-proteins, and that iNKRs can detect decreases in the peptide pool if they are sufficiently specific.

The importance of peptide sensitivity for iKIRs has been mostly related to viral infections, especially in HIV-1 infections (Fadda *et al.*, 2012, van Teijlingen *et al.*, 2014). Recent studies demonstrate that sequence variations within HLA-C restricted HIV epitopes strongly engage inhibiting KIR2DL2, inducing a strong inhibiting signal for NK cells (van Teijlingen *et al.*, 2014). Although these studies certainly have important functional implications, especially regarding viral escape mechanisms, our study shows that changes in the presented self-peptides can be sufficient to detect a viral infection. If the HIV-1 peptides are not highly abundant on productively infected cells, their contribution to the total changes in the peptide repertoire might be small. Indeed, all changes in the peptide repertoire (i.e. derived from self and viral proteins) as well as the abundance of the peptides are necessary to estimate the implications of peptides on iNKR mediated NK cell regulation.

The BLCL data set allowed us to investigate how the peptide pool that is relevant for iNKR-binding (i.e. the peptides restricted to HLA-A3/A11, -Bw4, C1, and C2) changes after a viral infection. However, it is important to notice that the assignment of the eluted peptides to the HLA alleles was predicted *in-silico*, and possibly some eluted peptides were not assigned to any MHC molecule. Furthermore, the HLA-C restricted peptide pool might have been underestimated, as there is a preference in assignment to HLA-A and -B alleles, because the prediction performance for the better-studied HLA-A and HLA-B alleles is higher than that for HLA-C alleles. Even if the contribution of these "possible ligands" are underestimated, the main outcome of our study does not change much, as the analysis including all 9-mers (considering also peptides that have not been assigned to any HLA molecule) show results very similar to those considering only the group of peptides presented by HLA-A3/A11, -Bw4, C1 and C2.

In the sequence based analysis, we considered that an iNKR can discriminate any pair of amino acids in P7 and P8 (i.e. $p = 1/400$). However, neither in the data set of Schellens at al., nor in that from Fadda et al., all 400 pairs of amino acids are present among the

peptides studied. It is indeed likely that not all pairs of amino acids are presented by the HLA molecules studied in those studies, or that some combinations are not naturally occurring. What an iNKR detects in those two specific positions is important to estimate the required specificity for "altered-self" recognition. Our sampling model shows that, even if the iNKRs recognized four groups of amino acids instead of unique pairs (i.e. $p = 1/16$), the detection of a lower ligand concentration would still be possible. This observation calls for a further analysis of the amino acid groups in P7 and P8 of the peptides eluted from BLCLs.

The actual iNKR specificity is necessary to make a proper conclusion about diversity. In the sequence based analysis (where we again consider the case that an iNKR can detect all pairs of amino acids) we show that the required genetic KIR diversity varies widely among BLCLs, with ten specific iNKRs being necessary to detect changes in the peptide repertoire in BLCL 1112, and two iNKRs being sufficient in BLCL 1077. If the real specificity were lower, like suggested by Fadda's study (where 13/59 peptides were KIR-binders (Fadda *et al.*, 2010)), the detection of the altered peptide repertoire would become more challenging, as shown by our sampling model (Figure 6.4). In this case, several iNKRs per individual would be necessary to detect the viral infection.

Our study is unique in that it comprises a detailed analysis of the variation in the iNKR-ligands after a viral infection. By combining a mathematical model with experimental data, we were able to quantify the absolute numbers and direction of change in the peptide repertoire, shedding light on the complex mechanisms of peptide sensitivity and its possible functional consequences. However, further studies are indispensable to get a better estimate about the genetic diversity and optimal specificity of iNKRs. In Chapters 2-5 we studied the evolution of KIRs by using an evolutionary computational model. By expanding those models with the understanding of the peptide repertoire prior and after infection we have gained in this current study, we can more accurately study the selection pressure on iNKRs caused by peptide sensitivity.

6.4. Acknowledgments

We thank Ingrid Schellens, Ilka Hoof, Hugo Meiring, Sanna Spijkers, Mertien Poelen, Kees van der Poel, Ana Costa, Cecile van Els, and Debbie van Baarle for providing us the unpublished experimental data set. We are grateful to Enrico Sandro Colizzi for helpful discussions, Rutger Hermsen for his highly valuable ideas in a previous version of the

model, as well as Nikolas Satravelas for starting the project. We also thank Saikrishna Gadhamsetty for carefully reading the manuscript. This work was financially supported by the CLS program of the Netherlands Organization for Scientific Research (NWO), grant 635.100.025.

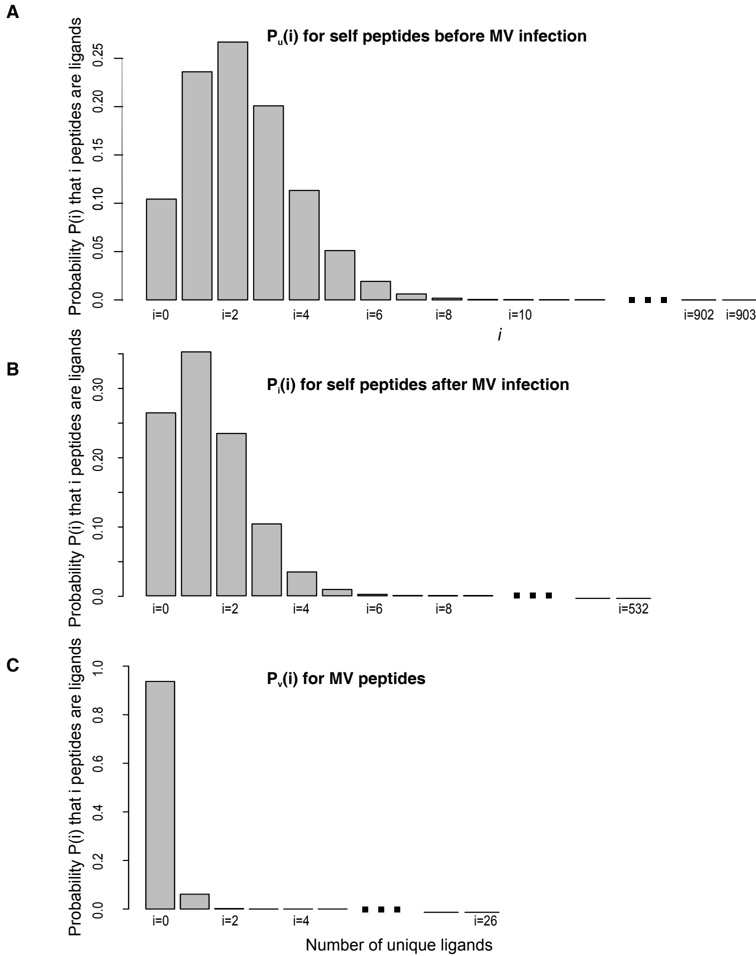

Figure 6.2. Probability distribution of an iNKR recognizing *i* ligands in BLCL 1053. Panel A depicts the probability distribution $P_u(i)$ of a iNKR detecting i self peptides before the MV infection. For instance, an iNKR will most likely recognize 2 out of the 903 unique self peptides as ligands. Similarly, panel B and C depict the probability distribution $P_i(i)$ and $P_v(i)$ of a iNKR detecting i self peptides after MV infection and detecting i MV peptides, respectively. $P_u(i)$, $P_i(i)$ and $P_v(i)$ are all calculated by equation 6.1. Note that an iNKR detecting 2 unique peptides out of the 903 self-peptides, with a total copy number of $T = 99556$, has an expected ligand density of $(2/903) \times 99556$ peptides. After MV infection, an iNKR detecting 2 unique peptides out of the 532 self-peptides, with a total copy number of $T = 85188$ is expected to see $(2/532) \times 85188$. Similarly, the expected ligand density of one iNKR binding 1 unique MV peptide out of the pool of 26 viral peptides is $(1/26) \times 32084$.

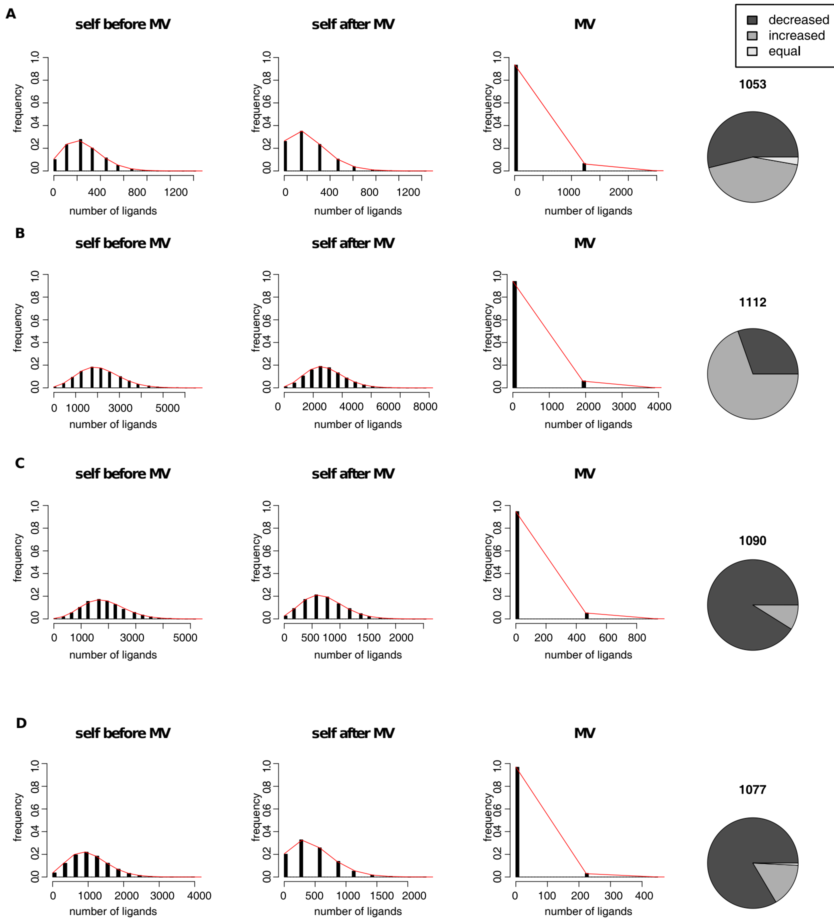

Figure 6.3. Probability distribution of iNKR ligands in four BLCLs. The red lines depict the probability distributions an iNKR detecting x ligands on the surface of cells before and after MV infection ($P_u(i)$, $P_i(i)$, and $P_v(i)$ calculated by equation 6.1) for BLCL 1053 (A), 1112 (B), 1090 (C), and 1077 (D). The bars depict the same distributions, yet obtained by sampling 10000 from each binomial distribution, whereby we obtain for every sample the total number of iNKR ligands on one cell. For every sample, we determine whether there is an increase, or a decrease, in the number of iNKR ligands before and after the infection, i.e. we compare the number of ligands after every sample from $P_u(i)$ with the sum of the ligand numbers obtained after sampling from $P_i(i)$ and $P_v(i)$. The pie charts depict the occurrence of having an increased (light gray), decreased (dark gray), or equal (gray) ligand concentration after MV infection. This sampling model assumes that iNKRs have a specificity of $p = 1/400$ (i.e. iNKR detect two amino acid combinations at P7 and P8).

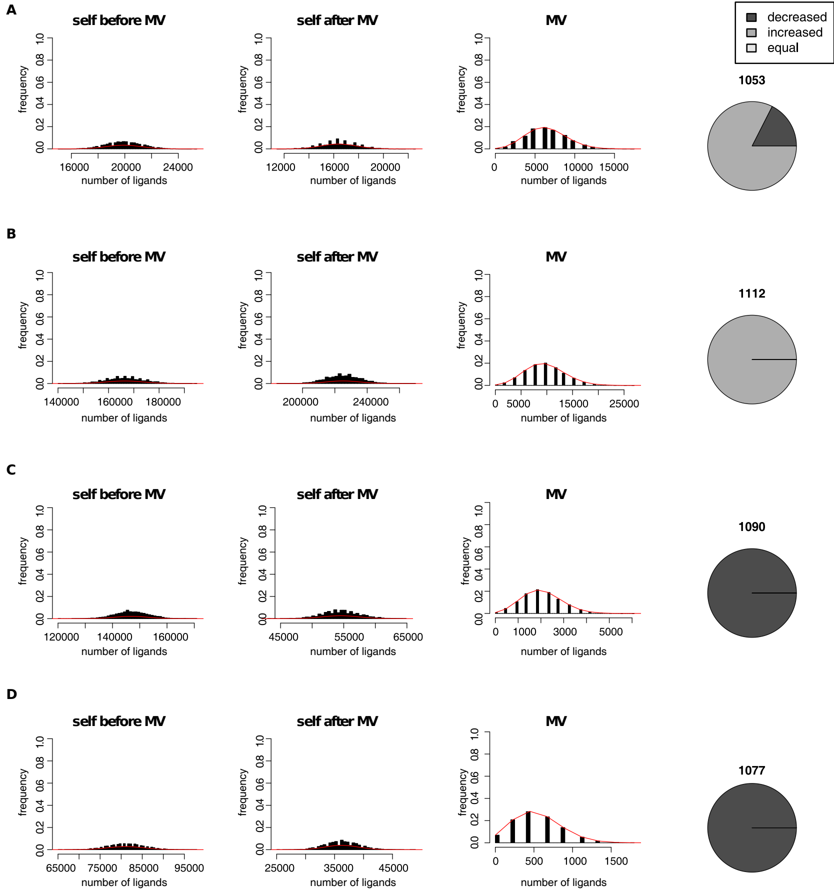

Figure 6.4. Distribution of iNKR ligands in BLCLs using a sampling model. Legend as in Figure 6.3 but using $p = 0.2$ (specificity estimated by Fadda et al. (Fadda *et al.*, 2012)).

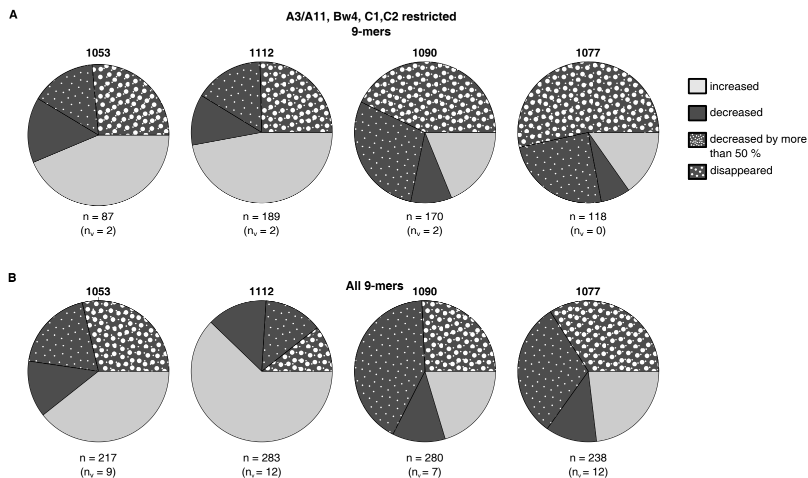

Figure 6.5. Changes in the iNKR ligand abundance after the sequence based analysis. In panel A, only those peptides restricted to HLA molecules having the A3/A11, Bw4, C1 or C2 motifs were analyzed, and in panel B or all peptides were considered. In all cases, the peptides are grouped according to their unique amino acid combinations in P7 and P8, and we monitor in how many cases there is an increase (light gray), or a decrease (dark gray), in these peptide motifs after the infection with MV. By considering the percentual change, we then quantify by how much the abundance of KIR motifs decreases after MV infection. Depicted are the KIR motifs that decrease their abundance by half (small white bullets), and those peptides that disappear (large white bullets).

Chapter 7

Summarizing Discussion: Why polymorphic and polygenic NK cell receptors?

The evolution of NK cell receptors (NKRs) in different species led to a fascinating complexity. Most receptors interact with highly polymorphic MHC molecules, resulting in receptor-ligand pairs that individualize the immune system, a process that is thought to improve the survival of individuals and populations. Yet, the exact evolutionary advantage of these expanded genotypes has remained unresolved. In this thesis, we have extensively studied how viruses exert selection pressures on NKRs, shedding light and providing understanding for the evolution of these polygenic and polymorphic inhibiting and activating NKRs.

7.1. Evolution of specific inhibiting receptors

Inhibiting NKRs (iNKRs) are typically involved in the recognition of missing-self, i.e. in the detection of viral infected or transformed cells that have a reduced expression of MHC-I molecules to evade cytotoxic T cells (Ljunggren and Kärre, 1990). Some viruses have evolved several mechanisms counteracting missing-self detection by NK cells, e.g. by evolving MHC-like decoy molecules that engage iNKRs, or by downregulating particular MHC loci. In this thesis, we studied the evolution of NKRs *in-silico* and showed that hosts infected with such viruses naturally evolve iNKRs that are specific for subsets of MHC molecules. Apparently, iNKR specificity allows for better NK cell responses.

The selection pressure for specificity varies depending on the strategy the viruses use to escape missing-self detection by NK cells. In Chapters 2 – 4, we show that viruses evolving MHC-decoys drive the evolution of highly specific iNKRs, i.e., receptors that specialize to single MHC molecules in the population. Because iNKRs face the challenge to avoid being “fooled” by the viral molecules mimicking MHC-I, a high degree of specificity is optimal during infections with viruses evolving decoys.

Viruses that selectively downregulate particular MHC-I loci, on the other hand, drive the evolution of less specific iNKRs. Hosts infected with these type of viruses evolve iNKRs that specialize to groups of MHC-I molecules rather than to single alleles (shown in Chapter 5), an effect that depends on the structural similarity of the MHC molecules. If the MHC molecules within each locus are similar to each other, iNKRs can evolve to specialize to all MHC alleles in one locus, while losing affinity to the MHC alleles in the other locus. In contrast, if MHC molecules share little or no similarity within one locus, iNKRs can no longer distinguish between the different groups of MHCs, which results in the evolution of a higher degree of iNKRs specificity (Table 6.1).

iNKR signaling may also depend on the peptide presented by the MHC molecules, since they are in direct contact with residues P7 and P8 (Boyington *et al.*, 2000, Brooks *et al.*, 2000, Fadda *et al.*, 2010, Fan *et al.*, 2001, Li and Mariuzza, 2014). In Chapter 6, we show that iNKRs need to be sufficiently specific to detect the changes in the peptide pool induced after a viral infection: if they are able to discriminate between any unique amino acid pairs or even amino acid groups in P7 and P8, iNKRs can easily detect "altered self", even when the total abundance of peptides increases upon infection. Whether the actual specificity lays in this range remains unresolved, and experimental elucidation of the KIR motifs in the MHC presented peptides is necessary to validate our predictions.

Altogether, our results show that the specificity of iNKRs is not only beneficial to counteract viral evasion mechanisms such as MHC-decoys or selective MHC downregulation, but specific iNKRs can also successfully detect changes in the altered peptide repertoire after viral infection.

7.2. Evolution of specific activating NK cell receptors

Activating NKRs (aNKRs) are not involved in missing self, and their role has remained intriguing because of the current lack of characterized ligands. In humans, aNKRs are associated with reproductive success, as they ensure the activation of NK cells required for arterial remodeling and with it a sufficient blood supply in the placenta (see Introduction and (Moffett-King, 2002, Wallace *et al.*, 2012)). The arterial remodeling required for a deep placentation is mostly important in humans, as a sufficient blood supply is necessary for the development of large brain sizes. Thus, reproductive success is not likely to exert such a strong selection pressure on the expansion and maintenance of aNKRs in other species.

A natural alternative hypothesis is that aNKRs bind viral encoded proteins, peptides presented from these proteins, or particular host proteins signaling stress, and thereby activate NK cells. Lanier and colleagues proposed a model of NKR evolution, in which aNKRs evolved from their inhibiting counterparts based upon the observations of the selection pressure exerted by CMV after evolving MHC-I decoys (Lanier, 2008, Sun and Lanier, 2009). In Chapter 4 we tested and confirmed this hypothesis, showing that aNKRs are selected because of the protection they provide when they bind MHC-I decoys, thereby activating NK cells. Given their involvement in pathogen defense and host survival, it is natural that aNKRs provide an evolutionary advantage, explaining their fast

expansion in the NKR cluster in our simulations.

If aNKRs are indeed advantageous by targeting viral products, why is it so challenging to find and characterize ligands for aNKRs? It is possible that these viral ligands are encoded by short undetermined sequences in viruses having large genomes, such as CMV, EBV and other viruses from the Herpes family. Another possible explanation for the lack of identified viral ligands is that viruses evolve rapidly, becoming a moving target for aNKRs, to impede the adaptation of aNKRs.

Despite the high sequence similarity in the extracellular binding domain of inhibiting and activating KIRs, activating KIRs do not bind most HLA molecules (Gillespie *et al.*, 2007, O'Connor and McVicar, 2013, Saulquin *et al.*, 2003). There is even an evolutionary trend in humans and chimpanzees towards reducing the affinity of activating KIRs to the HLA epitopes C1 and C2 (Moesta *et al.*, 2010). Our model of host-pathogen co-evolution provides an explanation for this phenomena, showing that aNKRs evolve to reduce their binding affinity to MHC-I molecules. The selection for lower binding affinity to MHC molecules is driven by NK cell education, as aNKRs would not be licensed if they were self reactive (i.e. they recognized any MHC-I molecule in a host, see Chapters 3 – 4).

aNKRs might additionally play a role in the detection of presented peptides, i.e. "altered-self" detection, as the changes in the MHC presented peptides after a viral infection might also enhance the binding to aNKRs, allowing for NK cell activation. In mice, the interactions between the activating CD94/NKG2E and the peptides loaded onto the non-classical MHC-I Qa-1^b have indeed been associated with enhanced viral control against mousepox infections (Fang *et al.*, 2011). Qa-1^b is normally loaded with small peptides derived from other self MHC-I molecules and is the natural ligand for the inhibiting CD94/NKG2A receptor. After mouse pox infection there is a shift in the recognition of peptide-MHC complexes from inhibiting to activating receptors. The molecular mechanisms underlying this aNKR "altered-self" detection remain unknown, as it is not clear whether it is a viral or a newly expressed self peptide that is engaging the aNKR. However, this observation clearly highlights the importance of aNKRs in recognizing peptide-MHC complexes during viral infections, calling for an extension of our analysis performed in Chapter 6.

Table 6.1. Comparison of specificity, haplotype composition, and genetic diversity among Chapters 3-4, and the Yucpa (Gendzekhadze *et al.*, 2009).

	# aNKRs / haplotype	# iNKRs / haplotype	maximal # loci	SRI[a]
Yucpa	2.5	5.5	14	3.4
Chapter 3	6	4	10	6.6
Chapter 4	8.6	4.15	15	8.9

a: The SRI index was calculated for the haplotypes, i.e. we compute how many different haplotypes there are in a population.
We exclude Chapters 2, and 5 from this comparison, as in those models aNKRs are not considered.

7.3. Genetic diversity

In all the models of NKR evolution studied in this thesis, the gene complex encoding NKRs becomes polygenic and polymorphic. But how realistic is the level of genetic diversity we obtain? Several studies have described the NKR diversity in numerous human populations, including Japanese (Yawata *et al.*, 2006), Amerindian (Gendzekhadze *et al.*, 2009), African (Norman *et al.*, 2013), and Polynesian (Nemat-Gorgani *et al.*, 2014) populations. Particularly informative are the studies of the Amerindian population Yucpa, as these individuals are estimated to have been isolated for almost 3000 years, during which they have suffered several epidemics and population bottlenecks (Gendzekhadze *et al.*, 2009). The Yucpa population retained several KIR haplotypes, indicating that NKR diversity is essential for long-term survival (Table 6.1) (Gendzekhadze *et al.*, 2009).

The seven haplotypes found in this population have mainly expanded the genes encoding iNKRs, obtaining approximately 5-7 loci encoding iNKRs per haplotype. While the number of iNKRs per haplotype evolved in our simulated populations is in agreement with the number of iNKRs per haplotype found in the Yucpa population, the number of aNKRs per haplotype and the haplotype diversity (SRI) in our simulations are larger than those observed in the Yucpa population (see Table 6.1). In our simulations, we do not allow for any fitness cost for the expansion of the genome. Hence, genotypes can expand until the maximum number of genes is reached, explaining the high NKR diversity obtained in our simulations. Additionally, the maximal number of loci per

individual affects both polymorphism and polygenicity in our models: in simulations where individuals have fewer loci per haplotype, the heterozygote advantage to have a licensed NKR repertoire is larger, thereby exerting a larger pressure for polymorphism. Since it was only for reasons of simplification that we do not model a fitness cost for additional loci, it is difficult to compare the obtained genetic diversity to that found in the Amerindian population. Nevertheless, all our studies clearly underline the importance of the expansion of iNKR genes for survival.

7.4. Model limitations

In our agent-based models, we made one important assumption: we consider a global repertoire of licensed receptors, instead of individual NK cells, which forced us to define the protection of a host in terms of the number of protective receptors, rather than the presence of protective NK cells. This modeling approach ignores distinct NK cell subsets, their development, and activation during and immune response. Additionally, this assumption affects the evolution of specificity and hence diversity, as it directly determine the host's protection and its survival. Since there is still a lack of detailed knowledge about the role of MHC-I molecules in NK cell maturation, the exact mechanisms of NKR repertoire acquisition, and the expansion of different NK cell subsets during viral infections remain poorly understood. Therefore, we deemed it to be unfeasible to model the NK cell dynamics in detail at this stage of our NK cell understanding. Determining these complex molecular and cellular processes is crucial for future extensions of these models, which in turn can allow the study of the effects of NK cell education, and the diversity of NK cell subsets, in fighting pathogens and survival.

7.5. Concluding remarks

With evolutionary and mathematical modeling, we have revealed several processes that are involved in the evolution of NKRs. We have shown that specificity plays a pivotal role for NKRs to be functional and to distinguish self from non-self. Additionally, we have demonstrated how hosts naturally evolve diverse NKR clusters as a consequence of the required specificity. These results mirror the complexity of the biological process, as we cannot not provide one simple explanation for the complex biological question why NKRs are specific, polygenic and polymorphic. In fact, we have identified several

plausible explanations, which contribute, albeit in different degrees, to the evolution of polygenicity and polymorphism in the NKR cluster. All processes revised in this thesis are most likely intertwined, simultaneously exerting pressures on hosts to evolve the functional NKR-MHC pairs, which provide only protective NK cells a license to kill.

Bibliography

Abi-Rached, L. and Parham, P. 2005. Natural selection drives recurrent formation of activating killer cell immunoglobulin-like receptor and Ly49 from inhibitory homologues. *J Exp Med*, 201(8): 1319–1332. (Cited on pages 11, 48, and 76.)

Abi-Rached, L., Moesta, A. K., Rajalingam, R., Guethlein, L. A., and Parham, P. 2010. Human-specific evolution and adaptation led to major qualitative differences in the variable receptors of human and chimpanzee natural killer cells. *PLoS Genet*, 6(11): e1001192. (Cited on pages 5, 6, and 33.)

Adams, E. J. and Parham, P. 2001. Species-specific evolution of MHC class I genes in the higher primates. *Immunol Rev*, 183: 41–64. (Cited on page 5.)

Ahlenstiel, G., Martin, M. P., Gao, X., Carrington, M., and Rehermann, B. 2008. Distinct KIR/HLA compound genotypes affect the kinetics of human antiviral natural killer cell responses. *J Clin Invest*, 118(3): 1017–1026. (Cited on page 9.)

Alter, G., Rihn, S., Walter, K., Nolting, A., Martin, M., Rosenberg, E. S., Miller, J. S., Carrington, M., and Altfeld, M. 2009. HLA class I subtype-dependent expansion of KIR3DS1+ and KIR3DL1+ NK cells during acute human immunodeficiency virus type 1 infection. *J Virol*, 83(13): 6798–6805. (Cited on page 9.)

Alter, G., Heckerman, D., Schneidewind, A., Fadda, L., Kadie, C. M., Carlson, J. M., Oniangue-Ndza, C., Martin, M., Li, B., Khakoo, S. I., Carrington, M., Allen, T. M., and Altfeld, M. 2011. HIV-1 adaptation to NK-cell-mediated immune pressure. *Nature*, 476(7358): 96–100. (Cited on pages 20 and 116.)

Amadei, B., Urbani, S., Cazaly, A., Fisicaro, P., Zerbini, A., Ahmed, P., Missale, G., Ferrari, C., and Khakoo, S. I. 2010. Activation of natural killer cells during acute infection with hepatitis C virus. *Gastroenterology*, 138(4): 1536–1545. (Cited on page 9.)

Anfossi, N., André, P., Guia, S., Falk, C. S., Roetynck, S., Stewart, C. A., Breso, V., Frassati, C., Reviron, D., Middleton, D., Romagné, F., Ugolini, S., and Vivier, E. 2006. Human NK cell education by inhibitory receptors for MHC class I. *Immunity*, 25(2): 331–342. (Cited on pages 22, 31, 47, 49, 75, 79, and 99.)

Arase, H. and Lanier, L. L. 2002. Virus-driven evolution of natural killer cell receptors. *Microbes Infect*, 4(15): 1505–1512. (Cited on pages 11, 33, 48, and 76.)

Arase, H., Mocarski, E. S., Campbell, A. E., Hill, A. B., and Lanier, L. L. 2002. Direct recognition of cytomegalovirus by activating and inhibitory NK cell receptors. *Science*, 296(5571): 1323–1326. (Cited on pages 20, 25, 48, and 76.)

Averdam, A., Petersen, B., Rosner, C., Neff, J., Roos, C., Eberle, M., Aujard, F., Münch, C., Schempp, W., Carrington, M., Shiina, T., Inoko, H., Knaust, F., Coggill, P., Sehra,

H., Beck, S., Abi-Rached, L., Reinhardt, R., and Walter, L. 2009. A novel system of polymorphic and diverse NK cell receptors in primates. *PLoS Genet*, 5(10): e1000688. (Cited on pages 7, 21, and 32.)

Bacon, L., Eagle, R. A., Meyer, M., Easom, N., Young, N. T., and Trowsdale, J. 2004. Two human ULBP/RAET1 molecules with transmembrane regions are ligands for NKG2D. *J Immunol*, 173(2): 1078–1084. (Cited on page 75.)

Bauer, S., Groh, V., Wu, J., Steinle, A., Phillips, J. H., Lanier, L. L., and Spies, T. 1999. Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science*, 285(5428): 727–729. (Cited on page 75.)

Bessoles, S., Grandclément, C., Alari-Pahissa, E., Gehrig, J., Jeevan-Raj, B., and Held, W. 2014. Adaptations of Natural Killer Cells to Self-MHC Class I. *Front Immunol*, 5: 349. (Cited on page 2.)

Besson, C., Roetynck, S., Williams, F., Orsi, L., Amiel, C., Lependeven, C., Antoni, G., Hermine, O., Brice, P., Ferme, C., Carde, P., Canioni, D., Brière, J., Raphael, M., Nicolas, J.-C., Clavel, J., Middleton, D., Vivier, E., and Abel, L. 2007. Association of killer cell immunoglobulin-like receptor genes with Hodgkin's lymphoma in a familial study. *PLoS One*, 2(5): e406. (Cited on page 48.)

Béziat, V., Liu, L., Malmberg, J.-A., Ivarsson, M. A., Sohlberg, E., Björklund, A. T., Retière, C., Sverremark-Ekström, E., Traherne, J., Ljungman, P., Schaffer, M., Price, D. A., Trowsdale, J., Michaëlsson, J., Ljunggren, H.-G., and Malmberg, K.-J. 2013. NK cell responses to cytomegalovirus infection lead to stable imprints in the human KIR repertoire and involve activating KIRs. *Blood*. (Cited on pages 20 and 33.)

Bimber, B. N., Moreland, A. J., Wiseman, R. W., Hughes, A. L., and O'Connor, D. H. 2008. Complete characterization of killer Ig-like receptor (KIR) haplotypes in Mauritian cynomolgus macaques: novel insights into nonhuman primate KIR gene content and organization. *J Immunol*, 181(9): 6301–6308. (Cited on page 5.)

Blokhuis, J. H., van der Wiel, M. K., Doxiadis, G. G. M., and Bontrop, R. E. 2010. The mosaic of KIR haplotypes in rhesus macaques. *Immunogenetics*, 62(5): 295–306. (Cited on page 5.)

Blokhuis, J. H., van der Wiel, M. K., Doxiadis, G. G. M., and Bontrop, R. E. 2011. The extreme plasticity of killer cell Ig-like receptor (KIR) haplotypes differentiates rhesus macaques from humans. *Eur J Immunol*, 41(9): 2719–2728. (Cited on page 5.)

Borghans, J. A. M., Beltman, J. B., and Boer, R. J. D. 2004. MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, 55(11): 732–739. (Cited on page 15.)

Borrego, F., Ulbrecht, M., Weiss, E. H., Coligan, J. E., and Brooks, A. G. 1998. Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med*, 187(5): 813–818. (Cited on pages 7 and 75.)

Boyington, J. C., Motyka, S. A., Schuck, P., Brooks, A. G., and Sun, P. D. 2000. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature*, 405(6786): 537–543. (Cited on pages 13, 115, 119, and 133.)

Braud, V. M., Allan, D. Söderström, K., D'Andrea, A., Ogg, G. S., Lazetic, S., Young, N. T., Bell, J. I., Phillips, J. H., Lanier, L. L., and McMichael, A. J. 1998. HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature*, 391(6669): 795–799. (Cited on pages 7, 19, 47, and 75.)

Brodin, P., Kärre, K., and Höglund, P. 2009. NK cell education: not an on-off switch but a tunable rheostat. *Trends Immunol*, 30(4): 143–149. (Cited on page 2.)

Brooks, A. G., Boyington, J. C., and Sun, P. D. 2000. Natural killer cell recognition of

HLA class I molecules. *Rev Immunogenet*, 2(3): 433–448. (Cited on pages 13, 115, 119, and 133.)

Caligiuri, M. A. 2008. Human natural killer cells. *Blood*, 112(3): 461–469. (Cited on page 19.)

Carnes, B. A., Holden, L. R., Olshansky, S. J., Witten, M. T., and Siegel, J. S. 2006. Mortality partitions and their relevance to research on senescence. *Biogerontology*, 7(4): 183–198. (Cited on page 36.)

Cassidy, S. A., Cheent, K. S., and Khakoo, S. I. 2014. Effects of Peptide on NK cell-mediated MHC I recognition. *Front Immunol*, 5: 133. (Cited on pages 13, 115, 121, and 122.)

Cerwenka, A. and Lanier, L. L. 2001. Ligands for natural killer cell receptors: redundancy or specificity. *Immunol Rev*, 181: 158–169. (Cited on page 75.)

Cerwenka, A., Bakker, A. B., McClanahan, T., Wagner, J., Wu, J., Phillips, J. H., and Lanier, L. L. 2000. Retinoic acid early inducible genes define a ligand family for the activating NKG2D receptor in mice. *Immunity*, 12(6): 721–727. (Cited on page 75.)

Chalifour, A., Scarpellino, L., Back, J., Brodin, P., Devèvre, E., Gros, F., Lévy, F., Leclercq, G., Höglund, P., Beermann, F., and Held, W. 2009. A Role for cis Interaction between the Inhibitory Ly49A receptor and MHC class I for natural killer cell education. *Immunity*, 30(3): 337–347. (Cited on pages 31, 49, 79, and 99.)

Charoudeh, H. N., Terszowski, G., Czaja, K., Gonzalez, A., Schmitter, K., and Stern, M. 2013. Modulation of the natural killer cell KIR repertoire by cytomegalovirus infection. *Eur J Immunol*, 43(2): 480–487. (Cited on pages 20 and 33.)

Cohen, G. B., Gandhi, R. T., Davis, D. M., Mandelboim, O., Chen, B. K., Strominger, J. L., and Baltimore, D. 1999. The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. *Immunity*, 10(6): 661–671. (Cited on pages 10, 89, and 97.)

Collins, K. L., Chen, B. K., Kalams, S. A., Walker, B. D., and Baltimore, D. 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature*, 391(6665): 397–401. (Cited on pages 10, 89, and 97.)

Corbett, A. J., Coudert, J. D., Forbes, C. A., and Scalzo, A. A. 2011. Functional consequences of natural sequence variation of murine cytomegalovirus m157 for Ly49 receptor specificity and NK cell activation. *J Immunol*, 186(3): 1713–1722. (Cited on page 12.)

Daza-Vamenta, R., Glusman, G., Rowen, L., Guthrie, B., and Geraghty, D. E. 2004. Genetic divergence of the rhesus macaque major histocompatibility complex. *Genome Res*, 14(8): 1501–1515. (Cited on page 5.)

Dobromylskyj, M. and Ellis, S. 2007. Complexity in cattle KIR genes: transcription and genome analysis. *Immunogenetics*, 59(6): 463–472. (Cited on page 8.)

Elliott, J. M. and Yokoyama, W. M. 2011. Unifying concepts of MHC-dependent natural killer cell education. *Trends Immunol*, 32(8): 364–372. (Cited on pages 49, 79, and 99.)

Elliott, J. M., Wahle, J. A., and Yokoyama, W. M. 2010. MHC class I-deficient natural killer cells acquire a licensed phenotype after transfer into an MHC class I-sufficient environment. *J Exp Med*, 207(10): 2073–2079. (Cited on page 31.)

Fadda, L., Borhis, G., Ahmed, P., Cheent, K., Pageon, S. V., Cazaly, A., Stathopoulos, S., Middleton, D., Mulder, A., Claas, F. H. J., Elliott, T., Davis, D. M., Purbhoo, M. A., and Khakoo, S. I. 2010. Peptide antagonism as a mechanism for NK cell activation.

Proc Natl Acad Sci U S A, 107(22): 10160–10165. (Cited on pages 13, 20, 115, 120, 124, and 133.)

Fadda, L., O'Connor, G. M., Kumar, S., Piechocka-Trocha, A., Gardiner, C. M., Carrington, M., McVicar, D. W., and Altfeld, M. 2011. Common HIV-1 peptide variants mediate differential binding of KIR3DL1 to HLA-Bw4 molecules. *J Virol*, 85(12): 5970–5974. (Cited on page 64.)

Fadda, L., Körner, C., Kumar, S., van Teijlingen, N. H., Piechocka-Trocha, A., Carrington, M., and Altfeld, M. 2012. HLA-Cw*0102-restricted HIV-1 p24 epitope variants can modulate the binding of the inhibitory KIR2DL2 receptor and primary NK cell function. *PLoS Pathog*, 8(7): e1002805. (Cited on pages 13, 115, 116, 123, and 128.)

Fan, Q. R., Long, E. O., and Wiley, D. C. 2001. Crystal structure of the human natural killer cell inhibitory receptor KIR2DL1-HLA-Cw4 complex. *Nat Immunol*, 2(5): 452–460. (Cited on pages 13, 115, 119, and 133.)

Fang, M., Orr, M. T., Spee, P., Egebjerg, T., Lanier, L. L., and Sigal, L. J. 2011. CD94 is essential for NK cell-mediated resistance to a lethal viral disease. *Immunity*, 34(4): 579–589. (Cited on page 134.)

Farmer, J. D., Packard, N. H., and Perelson, A. S. 1986. The immune system, adaptation, and machine learning. *Physica D: Nonlinear Phenomena*, 22: 187–204. (Cited on pages 34 and 66.)

Farrell, H. E., Vally, H., Lynch, D. M., Fleming, P., Shellam, G. R., Scalzo, A. A., and Davis-Poynter, N. J. 1997. Inhibition of natural killer cells by a cytomegalovirus MHC class I homologue in vivo. *Nature*, 386(6624): 510–514. (Cited on page 11.)

Fauriat, C., Ivarsson, M. A., Ljunggren, H.-G., Malmberg, K.-J., and Michaëlsson, J. 2010. Education of human natural killer cells by activating killer cell immunoglobulin-like receptors. *Blood*, 115(6): 1166–1174. (Cited on pages 50 and 79.)

Fernandez, N. C., Treiner, E., Vance, R. E., Jamieson, A. M., Lemieux, S., and Raulet, D. H. 2005. A subset of natural killer cells achieves self-tolerance without expressing inhibitory receptors specific for self-MHC molecules. *Blood*, 105(11): 4416–4423. (Cited on page 31.)

Flores-Villanueva, P. O., Yunis, E. J., Delgado, J. C., Vittinghoff, E., Buchbinder, S., Leung, J. Y., Uglialoro, A. M., Clavijo, O. P., Rosenberg, E. S., Kalams, S. A., Braun, J. D., Boswell, S. L., Walker, B. D., and Goldfeld, A. E. 2001. Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proc Natl Acad Sci U S A*, 98(9): 5140–5145. (Cited on page 9.)

Futas, J. and Horin, P. 2013. Natural killer cell receptor genes in the family Equidae: not only Ly49. *PLoS One*, 8(5): e64736. (Cited on page 8.)

Gagnier, L., Wilhelm, B. T., and Mager, D. L. 2003. Ly49 genes in non-rodent mammals. *Immunogenetics*, 55(2): 109–115. (Cited on page 8.)

Gall, S. L., Erdtmann, L., Benichou, S., Berlioz-Torrent, C., Liu, L., Benarous, R., Heard, J. M., and Schwartz, O. 1998. Nef interacts with the mu subunit of clathrin adaptor complexes and reveals a cryptic sorting signal in MHC I molecules. *Immunity*, 8(4): 483–495. (Cited on pages 10, 89, and 97.)

Gendzekhadze, K., Norman, P. J., Abi-Rached, L., Graef, T., Moesta, A. K., Layrisse, Z., and Parham, P. 2009. Co-evolution of KIR2DL3 with HLA-C in a human population retaining minimal essential diversity of KIR and HLA class I ligands. *Proc Natl Acad Sci U S A*, 106(44): 18692–18697. (Cited on pages 20 and 135.)

Gewurz, B. E., Wang, E. W., Tortorella, D., Schust, D. J., and Ploegh, H. L. 2001. Human cytomegalovirus US2 endoplasmic reticulum-lumenal domain dictates asso-

ciation with major histocompatibility complex class I in a locus-specific manner. *J Virol*, 75(11): 5197–5204. (Cited on pages 10, 89, and 97.)

Gillespie, G. M. A., Bashirova, A., Dong, T., McVicar, D. W., Rowland-Jones, S. L., and Carrington, M. 2007. Lack of KIR3DS1 binding to MHC class I Bw4 tetramers in complex with CD8+ T cell epitopes. *AIDS Res Hum Retroviruses*, 23(3): 451–455. (Cited on pages 88 and 134.)

Griffin, B. D., Verweij, M. C., and Wiertz, E. J. H. J. 2010. Herpesviruses and immunity: the art of evasion. *Vet Microbiol*, 143(1): 89–100. (Cited on page 20.)

Guethlein, L. A., Flodin, L. R., Adams, E. J., and Parham, P. 2002. NK cell receptors of the orangutan (Pongo pygmaeus): a pivotal species for tracking the coevolution of killer cell Ig-like receptors with MHC-C. *J Immunol*, 169(1): 220–229. (Cited on pages 5 and 6.)

Guethlein, L. A., Older Aguilar, A. M., Abi-Rached, L., and Parham, P. 2007a. Evolution of killer cell Ig-like receptor (KIR) genes: definition of an orangutan KIR haplotype reveals expansion of lineage III KIR associated with the emergence of MHC-C. *J Immunol*, 179(1): 491–504. (Cited on pages 5, 32, 47, and 75.)

Guethlein, L. A., Abi-Rached, L., Hammond, J. A., and Parham, P. 2007b. The expanded cattle KIR genes are orthologous to the conserved single-copy KIR3DX1 gene of primates. *Immunogenetics*, 59(6): 517–522. (Cited on page 8.)

Hammond, J. A., Guethlein, L. A., Abi-Rached, L., Moesta, A. K., and Parham, P. 2009. Evolution and survival of marine carnivores did not require a diversity of killer cell Ig-like receptors or Ly49 NK cell receptors. *J Immunol*, 182(6): 3618–3627. (Cited on page 8.)

Hansasuta, P., Dong, T., Thananchai, H., Weekes, M., Willberg, C., Aldemir, H., Rowland-Jones, S., and Braud, V. M. 2004. Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific. *Eur J Immunol*, 34(6): 1673–1679. (Cited on pages 13 and 115.)

Hiby, S. E., Walker, J. J., O'shaughnessy, K. M., Redman, C. W. G., Carrington, M., Trowsdale, J., and Moffett, A. 2004. Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J Exp Med*, 200(8): 957–965. (Cited on pages 14 and 20.)

Hiby, S. E., Apps, R., Sharkey, A. M., Farrell, L. E., Gardner, L., Mulder, A., Claas, F. H., Walker, J. J., Redman, C. W., Redman, C. C., Morgan, L., Tower, C., Regan, L., Moore, G. E., Carrington, M., and Moffett, A. 2010. Maternal activating KIRs protect against human reproductive failure mediated by fetal HLA-C2. *J Clin Invest*, 120(11): 4102–4110. (Cited on pages 14, 20, and 48.)

Hiby, S. E., Apps, R., Chazara, O., Farrell, L. E., Magnus, P., Trogstad, L., Gjessing, H. K., Carrington, M., and Moffett, A. 2014. Maternal KIR in combination with paternal HLA-C2 regulate human birth weight. *J Immunol*, 192(11): 5069–5073. (Cited on page 14.)

Hickman, H. D., Luis, A. D., Bardet, W., Buchli, R., Battson, C. L., Shearer, M. H., Jackson, K. W., Kennedy, R. C., and Hildebrand, W. H. 2003. Cutting edge: class I presentation of host peptides following HIV infection. *J Immunol*, 171(1): 22–26. (Cited on page 64.)

Hirayasu, K., Ohashi, J., Kashiwase, K., Hananantachai, H., Naka, I., Ogawa, A., Takanashi, M., Satake, M., Nakajima, K., Parham, P., Arase, H., Tokunaga, K., Patarapotikul, J., and Yabe, T. 2012. Significant association of KIR2DL3-HLA-C1 combination with cerebral malaria and implications for co-evolution of KIR and HLA. *PLoS Pathog*, 8(3): e1002565. (Cited on page 20.)

Höglund, P. and Brodin, P. 2010. Current perspectives of natural killer cell education by MHC class I molecules. *Nat Rev Immunol*, 10(10): 724–734. (Cited on page 2.)

Hollenbach, J. A., Meenagh, A., Sleator, C., Alaez, C., Bengoche, M., Canossi, A., Contreras, G., Creary, L., Evseeva, I., Gorodezky, C., Hardie, R.-A., Karlsen, T. H., Lie, B., Luo, M., Martinetti, M., Navarette, C., de Oliveira, D. C. M., Ozzella, G., Pasi, A., Pavlova, E., Pinto, S., Porto, L. C., Santos, P., Slavcev, A., Srinak, D., Tavoularis, S., Tonks, S., Trachtenberg, E., Vejbaesya, S., and Middleton, D. 2010. Report from the killer immunoglobulin-like receptor (KIR) anthropology component of the 15th International Histocompatibility Workshop: worldwide variation in the KIR loci and further evidence for the co-evolution of KIR and HLA. *Tissue Antigens*, 76(1): 9–17. (Cited on page 6.)

Horowitz, A., Strauss-Albee, D. M., Leipold, M., Kubo, J., Nemat-Gorgani, N., Dogan, O. C., Dekker, C. L., Mackey, S., Maecker, H., Swan, G. E., Davis, M. M., Norman, P. J., Guethlein, L. A., Desai, M., Parham, P., and Blish, C. A. 2013. Genetic and environmental determinants of human NK cell diversity revealed by mass cytometry. *Sci Transl Med*, 5(208): 208ra145. (Cited on page 3.)

Iizuka, K., Naidenko, O. V., Plougastel, B. F. M., Fremont, D. H., and Yokoyama, W. M. 2003. Genetically linked C-type lectin-related ligands for the NKRP1 family of natural killer cell receptors. *Nat Immunol*, 4(8): 801–807. (Cited on page 7.)

Jamil, K. M. and Khakoo, S. I. 2011. KIR/HLA interactions and pathogen immunity. *J Biomed Biotechnol*, 2011: 298348. (Cited on pages 9 and 19.)

Jiang, W., Johnson, C., Jayaraman, J., Simecek, N., Noble, J., Moffatt, M. F., Cookson, W. O., Trowsdale, J., and Traherne, J. A. 2012. Copy number variation leads to considerable diversity for B but not A haplotypes of the human KIR genes encoding NK cell receptors. *Genome Res*, 22(10): 1845–1854. (Cited on page 47.)

Johansson, S., Johansson, M., Rosmaraki, E., Vahlne, G., Mehr, R., Salmon-Divon, M., Lemonnier, F., Kärre, K., and Höglund, P. 2005. Natural killer cell education in mice with single or multiple major histocompatibility complex class I molecules. *J Exp Med*, 201(7): 1145–1155. (Cited on pages 2, 3, and 31.)

Johansson, S., Salmon-Divon, M., Johansson, M. H., Pickman, Y., Brodin, P., Kärre, Mehr, R., and Höglund, P. 2009. Probing natural killer cell education by Ly49 receptor expression analysis and computational modelling in single MHC class I mice. *PLoS One*, 4(6): e6046. (Cited on page 3.)

Joncker, N. T., Fernandez, N. C., Treiner, E., Vivier, E., and Raulet, D. H. 2009. NK cell responsiveness is tuned commensurate with the number of inhibitory receptors for self-MHC class I: the rheostat model. *J Immunol*, 182(8): 4572–4580. (Cited on page 31.)

Kelley, J., Walter, L., and Trowsdale, J. 2005. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet*, 1(2): 129–139. (Cited on pages 3, 6, and 47.)

Khakoo, S. I., Rajalingam, R., Shum, B. P., Weidenbach, K., Flodin, L., Muir, D. G., Canavez, F., Cooper, S. L., Valiante, N. M., Lanier, L. L., and Parham, P. 2000. Rapid evolution of NK cell receptor systems demonstrated by comparison of chimpanzees and humans. *Immunity*, 12(6): 687–698. (Cited on page 6.)

Khakoo, S. I., Thio, C. L., Martin, M. P., Brooks, C. R., Gao, X., Astemborski, J., Cheng, J., Goedert, J. J., Vlahov, D., Hilgartner, M., Cox, S., Little, A.-M., Alexander, G. J., Cramp, M. E., O'Brien, S. J., Rosenberg, W. M. C., Thomas, D. L., and Carrington, M. 2004. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science*, 305(5685): 872–874. (Cited on pages 9 and 20.)

Kieckbusch, J., Gaynor, L. M., Moffett, A., and Colucci, F. 2014. MHC-dependent

inhibition of uterine NK cells impedes fetal growth and decidual vascular remodelling. *Nat Commun*, 5: 3359. (Cited on page 14.)

Kielczewska, A., Pyzik, M., Sun, T., Krmpotic, A., Lodoen, M. B., Munks, M. W., Babic, M., Hill, A. B., Koszinowski, U. H., Jonjic, S., Lanier, L. L., and Vidal, S. M. 2009. Ly49P recognition of cytomegalovirus-infected cells expressing H2-Dk and CMV-encoded m04 correlates with the NK cell antiviral response. *J Exp Med*, 206(3): 515–523. (Cited on pages 12 and 76.)

Kikuchi-Maki, A., Catina, T. L., and Campbell, K. S. 2005. Cutting edge: KIR2DL4 transduces signals into human NK cells through association with the Fc receptor gamma protein. *J Immunol*, 174(7): 3859–3863. (Cited on page 3.)

Kirkham, C. L. and Carlyle, J. R. 2014. Complexity and Diversity of the NKR-P1:Clr (Klrb1:Clec2) Recognition Systems. *Front Immunol*, 5: 214. (Cited on page 7.)

Kleijnen, M. F., Huppa, J. B., Lucin, P., Mukherjee, S., Farrell, H., Campbell, A. E., Koszinowski, U. H., Hill, A. B., and Ploegh, H. L. 1997. A mouse cytomegalovirus glycoprotein, gp34, forms a complex with folded class I MHC molecules in the ER which is not retained but is transported to the cell surface. *EMBO J*, 16(4): 685–694. (Cited on page 11.)

Knapp, S., Warshow, U., Hegazy, D., Brackenbury, L., Guha, I. N., Fowell, A., Little, A.-M., Alexander, G. J., Rosenberg, W. M. C., Cramp, M. E., and Khakoo, S. I. 2010. Consistent beneficial effects of killer cell immunoglobulin-like receptor 2DL3 and group 1 human leukocyte antigen-C following exposure to hepatitis C virus. *Hepatology*, 51(4): 1168–1175. (Cited on page 20.)

Körner, C. and Altfeld, M. 2012. Role of KIR3DS1 in human diseases. *Front Immunol*, 3: 326. (Cited on page 48.)

Kruse, P. H., Rosner, C., and Walter, L. 2010. Characterization of rhesus macaque KIR genotypes and haplotypes. *Immunogenetics*, 62(5): 281–293. (Cited on page 5.)

La, D., Czarnecki, C., El-Gabalawy, H., Kumar, A., Meyers, A. F. A., Bastien, N., Simonsen, J. N., Plummer, F. A., and Luo, M. 2011. Enrichment of variations in KIR3DL1/S1 and KIR2DL2/L3 among H1N1/09 ICU patients: an exploratory study. *PLoS One*, 6(12): e29200. (Cited on page 10.)

LaBonte, M. L., Hershberger, K. L., Korber, B., and Letvin, N. L. 2001. The KIR and CD94/NKG2 families of molecules in the rhesus monkey. *Immunol Rev*, 183: 25–40. (Cited on page 6.)

Lanier, L. L. 2005. NK cell recognition. *Annu Rev Immunol*, 23: 225–274. (Cited on pages 2, 19, 47, and 75.)

Lanier, L. L. 2008. Evolutionary struggles between NK cells and viruses. *Nat Rev Immunol*, 8(4): 259–268. (Cited on pages 9, 11, 19, 47, 76, 108, and 133.)

Lazetic, S., Chang, C., Houchins, J. P., Lanier, L. L., and Phillips, J. H. 1996. Human natural killer cell receptors involved in MHC class I recognition are disulfide-linked heterodimers of CD94 and NKG2 subunits. *J Immunol*, 157(11): 4741–4745. (Cited on pages 6 and 7.)

Lee, S. H., Girard, S., Macina, D., Busà, M., Zafer, A., Belouchi, A., Gros, P., and Vidal, S. M. 2001. Susceptibility to mouse cytomegalovirus is associated with deletion of an activating natural killer cell receptor of the C-type lectin superfamily. *Nat Genet*, 28(1): 42–45. (Cited on pages 11 and 76.)

Leonard, W. R., Snodgrass, J. J., and Robertson, M. L. 2007. Effects of brain evolution on human nutrition and metabolism. *Annu Rev Nutr*, 27: 311–327. (Cited on page 14.)

Li, Y. and Mariuzza, R. A. 2014. Structural basis for recognition of cellular and viral ligands by NK cell receptors. *Front Immunol*, 5: 123. (Cited on pages 13, 115, 119, and 133.)

Ljunggren, H. G. and Kärre, K. 1990. In search of the 'missing self': MHC molecules and NK cell recognition. *Immunol Today*, 11(7): 237–244. (Cited on pages 2, 19, 47, 97, 115, and 132.)

Llano, M., Gumá, M., Ortega, M., Angulo, A., and López-Botet, M. 2003. Differential effects of US2, US6 and US11 human cytomegalovirus proteins on HLA class Ia and HLA-E expression: impact on target susceptibility to NK cell subsets. *Eur J Immunol*, 33(10): 2744–2754. (Cited on pages 10, 89, and 97.)

López-Vázquez, A., Rodrigo, L., Martínez-Borra, J., Pérez, R., Rodríguez, M., Fdez-Morera, J. L., Fuentes, D., Rodríguez-Rodero, S., Gonzáez, S., and López-Larrea, C. 2005. Protective effect of the HLA-Bw4I80 epitope and the killer cell immunoglobulin-like receptor 3DS1 gene against the development of hepatocellular carcinoma in patients with hepatitis C virus infection. *J Infect Dis*, 192(1): 162–165. (Cited on pages 9, 48, and 76.)

Malnati, M. S., Peruzzi, M., Parker, K. C., Biddison, W. E., Ciccone, E., Moretta, A., and Long, E. O. 1995. Peptide specificity in the recognition of MHC class I by natural killer cell clones. *Science*, 267(5200): 1016–1018. (Cited on pages 13 and 115.)

Martin, A. M., Freitas, E. M., Witt, C. S., and Christiansen, F. T. 2000. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics*, 51(4-5): 268–280. (Cited on page 5.)

Martin, A. M., Pontarotti, J. K., and Christiansen, F. T. 2002a. Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol*, 23(2): 81–88. (Cited on page 7.)

Martin, M. P., Gao, X., Lee, J.-H., Nelson, G. W., Detels, R., Goedert, J. J., Buchbinder, S., Hoots, K., Vlahov, D., Trowsdale, J., Wilson, M., O'Brien, S. J., and Carrington, M. 2002b. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet*, 31(4): 429–434. (Cited on pages 48 and 76.)

Martin, M. P., Bashirova, A., Traherne, J., Trowsdale, J., and Carrington, M. 2003. Cutting edge: expansion of the KIR locus by unequal crossing over. *J Immunol*, 171(5): 2192–2195. (Cited on page 3.)

McGeoch, D. J., Rixon, F. J., and Davison, A. J. 2006. Topics in herpesvirus genomics and evolution. *Virus Res*, 117(1): 90–104. (Cited on page 20.)

McQueen, K. L., Freeman, J. D., Takei, F., and Mager, D. L. 1998. Localization of five new Ly49 genes, including three closely related to Ly49c. *Immunogenetics*, 48(3): 174–183. (Cited on page 8.)

Meyer, D., Singe, R., Mack, S., Lancaster, A., Nelson, M., Erlich, H., Fernandez-Vina, M., and Thomson, G. 2007. Single Locus Polymorphism of Classical HLA Genes. In *J.A. Hansen, ed. Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference,*, volume 1, pages 653–704. Seattle, WA: IHWG Press. (Cited on pages 37, 66, 89, and 110.)

Moesta, A. K. and Parham, P. 2012. Diverse functionality among human NK cell receptors for the C1 epitope of HLA-C: KIR2DS2, KIR2DL2, and KIR2DL3. *Front Immunol*, 3: 336. (Cited on page 47.)

Moesta, A. K., Norman, P. J., Yawata, M., Yawata, N., Gleimer, M., and Parham, P. 2008. Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for HLA-C than KIR2DL3. *J Immunol*, 180(6): 3969–3979. (Cited on page 20.)

Moesta, A. K., Graef, T., Abi-Rached, L., Older Aguilar, A. M., Guethlein, L. A., and Parham, P. 2010. Humans differ from other hominids in lacking an activating NK cell receptor that recognizes the C1 epitope of MHC class I. *J Immunol*, 185(7): 4233–4237. (Cited on pages 6, 64, 88, and 134.)

Moffett-King, A. 2002. Natural killer cells and pregnancy. *Nat Rev Immunol*, 2(9): 656–663. (Cited on pages 14 and 133.)

Moraru, M., Cisneros, E., Gómez-Lozano, N., de Pablo, R., Portero, F., Cañizares, M., Vaquero, M., Roustán, G., Millán, I., López-Botet, M., and Vilches, C. 2012. Host genetic factors in susceptibility to herpes simplex type 1 virus infection: contribution of polymorphic genes at the interface of innate and adaptive immunity. *J Immunol*, 188(9): 4412–4420. (Cited on page 20.)

Moretta, A., Bottino, C., Vitale, M., Pende, D., Biassoni, R., Mingari, M. C., and Moretta, L. 1996. Receptors for HLA class-I molecules in human natural killer cells. *Annu Rev Immunol*, 14: 619–648. (Cited on pages 20, 88, and 115.)

Nash, W. T., Teoh, J., Wei, H., Gamache, A., and Brown, M. G. 2014. Know Thyself: NK-Cell Inhibitory Receptors Prompt Self-Tolerance, Education, and Viral Control. *Front Immunol*, 5: 175. (Cited on pages 10 and 97.)

Natarajan, K., Hicks, A., Mans, J., Robinson, H., Guan, R., Mariuzza, R. A., and Margulies, D. H. 2006. Crystal structure of the murine cytomegalovirus MHC-I homolog m144. *J Mol Biol*, 358(1): 157–171. (Cited on page 11.)

Nemat-Gorgani, N., Edinur, H. A., Hollenbach, J. A., Traherne, J. A., Dunn, P. P. J., Chambers, G. K., Parham, P., and Norman, P. J. 2014. KIR diversity in MÄĄori and Polynesians: populations in which HLA-B is not a significant KIR ligand. *Immunogenetics*, 66(11): 597–611. (Cited on page 135.)

Norman, P. J., Hollenbach, J. A., Nemat-Gorgani, N., Guethlein, L. A., Hilton, H. G., Pando, M. J., Koram, K. A., Riley, E. M., Abi-Rached, L., and Parham, P. 2013. Co-evolution of human leukocyte antigen (HLA) class I ligands with killer-cell immunoglobulin-like receptors (KIR) in a genetically diverse population of sub-Saharan Africans. *PLoS Genet*, 9(10): e1003938. (Cited on page 135.)

O'Callaghan, C. A. 2000. Molecular basis of human natural killer cell recognition of HLA-E (human leucocyte antigen-E) and its relevance to clearance of pathogen-infected and tumour cells. *Clin Sci (Lond)*, 99(1): 9–17. (Cited on pages 19 and 47.)

O'Connor, G. M. and McVicar, D. 2013. The yin-yang of KIR3DL1/S1: molecular mechanisms and cellular function. *Crit Rev Immunol*, 33(3): 203–218. (Cited on pages 88 and 134.)

Older Aguilar, A. M., Guethlein, L. A., Adams, E. J., Abi-Rached, L., Moesta, A. K., and Parham, P. 2010. Coevolution of killer cell Ig-like receptors with HLA-C to become the major variable regulators of human NK cells. *J Immunol*, 185(7): 4238–4251. (Cited on page 32.)

Older Aguilar, A. M., Guethlein, L. A., Hermes, M., Walter, L., and Parham, P. 2011. Rhesus macaque KIR bind human MHC class I with broad specificity and recognize HLA-C more effectively than HLA-A and HLA-B. *Immunogenetics*, 63(9): 577–585. (Cited on page 32.)

Orr, M. T., Murphy, W. J., and Lanier, L. L. 2010. 'Unlicensed' natural killer cells dominate the response to cytomegalovirus infection. *Nat Immunol*, 11(4): 321–327. (Cited on page 31.)

Parham, P. 2005. Influence of KIR diversity on human immunity. *Adv Exp Med Biol*, 560: 47–50. (Cited on page 3.)

Parham, P. and Moffett, A. 2013. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. *Nat Rev Immunol*, 13(2): 133–144. (Cited on pages 6, 8, 14, 64, 65, and 108.)

Parham, P., Abi-Rached, L., Matevosyan, L., Moesta, A. K., Norman, P. J., Older Aguilar, A. M., and Guethlein, L. A. 2010. Primate-specific regulation of natural killer cells. *J Med Primatol*, 39(4): 194–212. (Cited on pages 19 and 75.)

Pelak, K., Need, A. C., Fellay, J., Shianna, K. V., Feng, S., Urban, T. J., Ge, D., De Luca, A., Martinez-Picado, J., Wolinsky, S. M., Martinson, J. J., Jamieson, B. D., Bream, J. H., Martin, M. P., Borrow, P., Letvin, N. L., McMichael, A. J., Haynes, B. F., Telenti, A., Carrington, M., Goldstein, D. B., Alter, G., and , N. I. A. I. D. C. f. H. I. V. I. D. S. V. I. 2011. Copy number variation of KIR genes influences HIV-1 control. *PLoS Biol*, 9(11): e1001208. (Cited on pages 9, 48, and 76.)

Peruzzi, M., Wagtmann, N., and Long, E. O. 1996. A p70 killer cell inhibitory receptor specific for several HLA-B allotypes discriminates among peptides bound to HLA-B*2705. *J Exp Med*, 184(4): 1585–1590. (Cited on pages 13 and 115.)

Petrie, E. J., Clements, C. S., Lin, J., Sullivan, L. C., Johnson, D., Huyton, T., Heroux, A., Hoare, H. L., Beddoe, T., Reid, H. H., Wilce, M. C. J., Brooks, A. G., and Rossjohn, J. 2008. CD94-NKG2A recognition of human leukocyte antigen (HLA)-E bound to an HLA class I leader sequence. *J Exp Med*, 205(3): 725–735. (Cited on pages 7 and 75.)

Prod'homme, V., Griffin, C., Aicheler, R. J., Wang, E. C. Y., McSharry, B. P., Rickards, C. R., Stanton, R. J., López-Botet, Wilkinson, G. W. G., and Tomasec, P. 2007. The human cytomegalovirus MHC class I homolog UL18 inhibits LIR-1+ but activates LIR-1- NK cells. *J Immunol*, 178(7): 4473–4481. (Cited on page 11.)

Pyo, C.-W., Guethlein, L. A., Vu, Q., Wang, R., Abi-Rached, L., Norman, P. J., Marsh, S. G. E., Miller, J. S., Parham, P., and Geraghty, D. E. 2010. Different patterns of evolution in the centromeric and telomeric regions of group A and B haplotypes of the human killer cell Ig-like receptor locus. *PLoS One*, 5(12): e15115. (Cited on page 6.)

Pyzik, M., Charbonneau, B., Gendron-Pontbriand, E.-M., Babić, M., Krmpotić, A., Jonjić, S., and Vidal, S. M. 2011. Distinct MHC class I-dependent NK cell-activating receptors control cytomegalovirus infection in different mouse strains. *J Exp Med*, 208(5): 1105–1117. (Cited on pages 12 and 76.)

Rahim, M. M. A., Tu, M. M., Mahmoud, A. B., Wight, A., Abou-Samra, E., Lima, P. D. A., and Makrigiannis, A. P. 2014. Ly49 receptors: innate and adaptive immune paradigms. *Front Immunol*, 5: 145. (Cited on pages 7 and 12.)

Rajagopalan, S. 2010. Endosomal signaling and a novel pathway defined by the natural killer receptor KIR2DL4 (CD158d). *Traffic*, 11(11): 1381–1390. (Cited on page 5.)

Rajagopalan, S. and Long, E. O. 1997. The direct binding of a p58 killer cell inhibitory receptor to human histocompatibility leukocyte antigen (HLA)-Cw4 exhibits peptide selectivity. *J Exp Med*, 185(8): 1523–1528. (Cited on pages 13 and 115.)

Rajalingam, R., Parham, P., and Abi-Rached, L. 2004. Domain shuffling has been the main mechanism forming new hominoid killer cell Ig-like receptors. *J Immunol*, 172(1): 356–369. (Cited on page 5.)

Raulet, D. H. and Vance, R. E. 2006. Self-tolerance of natural killer cells. *Nat Rev Immunol*, 6(7): 520–531. (Cited on pages 2 and 115.)

Raulet, D. H., Held, W., Correa, I., Dorfman, J. R., Wu, M. F., and Corral, L. 1997. Specificity, tolerance and developmental regulation of natural killer cells defined by expression of class I-specific Ly49 receptors. *Immunol Rev*, 155: 41–52. (Cited on page 2.)

Renedo, M., Arce, I., Rodríguez, A., Carretero, M., López-Botet, L. L., and Fernández-Ruiz, E. 1997. The human natural killer gene complex is located on chromosome 12p12-p13. *Immunogenetics*, 46(4): 307–311. (Cited on page 6.)

Reusch, U., Muranyi, W., Lucin, P., Burgert, H. G., Hengel, H., and Koszinowski, U. H. 1999. A cytomegalovirus glycoprotein re-routes MHC class I complexes to lysosomes for degradation. *EMBO J*, 18(4): 1081–1091. (Cited on page 11.)

Revilleza, M. J., Wang, R., Mans, J., Hong, M., Natarajan, K., and Margulies, D. H. 2011. How the virus outsmarts the host: function and structure of cytomegalovirus MHC-I-like molecules in the evasion of natural killer cell surveillance. *J Biomed Biotechnol*, 2011: 724607. (Cited on page 20.)

Robson, S. L. and Wood, B. 2008. Hominin life history: reconstruction and evolution. *J Anat*, 212(4): 394–425. (Cited on page 14.)

Romero, V., Azocar, J., niga, J. Z., Clavijo, O. P., Terreros, D., Gu, X., Husain, Z., Chung, R. T., Amos, C., and Yunis, E. J. 2008. Interaction of NK inhibitory receptor genes with HLA-C and MHC class II alleles in Hepatitis C virus infection outcome. *Mol Immunol*, 45(9): 2429–2436. (Cited on page 9.)

Sambrook, J. G., Bashirova, A., Palmer, S., Sims, S., Trowsdale, J., Abi-Rached, L., Parham, P., Carrington, M., and Beck, S. 2005. Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome Res*, 15(1): 25–35. (Cited on pages 5, 64, and 65.)

Saulquin, X., Gastinel, L. N., and Vivier, E. 2003. Crystal structure of the human natural killer cell activating receptor KIR2DS2 (CD158j). *J Exp Med*, 197(7): 933–938. (Cited on pages 88 and 134.)

Scalzo, A. A., Manzur, M., Forbes, C. A., Brown, M. G., and Shellam, G. R. 2005. NK gene complex haplotype variability and host resistance alleles to murine cytomegalovirus in wild mouse populations. *Immunol Cell Biol*, 83(2): 144–149. (Cited on page 88.)

Schaefer, M. R., Wonderlich, E. R., Roeth, J. F., Leonard, J. A., and Collins, K. L. 2008. HIV-1 Nef targets MHC-I and CD4 for degradation via a final common beta-COP-dependent pathway in T cells. *PLoS Pathog*, 4(8): e1000131. (Cited on pages 10 and 97.)

Schellens, I. M., Hoof, I., Meiring, H. D., Spijkers, S. N., Poelen, M. C., van der Poel, K., Costa, A. I., van Els, C. A., van Baarle, D., and Keşmir, C. 2014. The peptide repertoire presented by HLA-B molecules is more diverse compared to HLA-A and HLA-C. *Submitted*. (Cited on pages 116, 117, 118, and 121.)

Schenkel, A. R., Kingry, L. C., and Slayden, R. A. 2013. The ly49 gene family. A brief guide to the nomenclature, genetics, and role in intracellular infection. *Front Immunol*, 4: 90. (Cited on pages 7 and 8.)

Schmid, B. V., Keşmir, C., and de Boer, R. J. 2010. Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway. *Epidemics*, 2(3): 99–108. (Cited on page 15.)

Schust, D. J., Tortorella, D., Seebach, J., Phan, C., and Ploegh, H. L. 1998. Trophoblast class I major histocompatibility complex (MHC) products are resistant to rapid degradation imposed by the human cytomegalovirus (HCMV) gene products US2 and US11. *J Exp Med*, 188(3): 497–503. (Cited on pages 10, 89, and 97.)

Sharkey, A. M., Gardner, L., Hiby, S., Farrell, L., Apps, R., Masters, L., Goodridge, J., Lathbury, L., Stewart, C. A., Verma, S., and Moffett, A. 2008. Killer Ig-like receptor expression in uterine NK cells is biased toward recognition of HLA-C and alters with gestational age. *J Immunol*, 181(1): 39–46. (Cited on page 14.)

Shiina, T., Ota, M., Shimizu, S., Katsuyama, Y., Hashimoto, N., Takasu, M., Anzai, T., Kulski, J. K., Kikkawa, E., Naruse, T., Kimura, N., Yanagiya, K., Watanabe, A., Hosomichi, K., Kohara, S., Iwamoto, C., Umehara, Y., Meyer, A., Wanner, V., Sano, K., Macquin, C., Ikeo, K., Tokunaga, K., Gojobori, T., Inoko, H., and Bahram, S. 2006. Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. *Genetics*, 173(3): 1555–1570. (Cited on page 5.)

Shum, B. P., Flodin, L. R., Muir, D. G., Rajalingam, R., Khakoo, S. I., Cleland, S., Guethlein, L. A., Uhrberg, M., and Parham, P. 2002. Conservation and variation in human and common chimpanzee CD94 and NKG2 genes. *J Immunol*, 168(1): 240–252. (Cited on pages 7, 19, and 47.)

Simpson, E. 1949. Measurement of diversity. *Nature*, 163: 688. (Cited on pages 27, 37, 61, and 107.)

Smith, H. R. C., Heusel, J. W., Mehta, I. K., Körner, B. G., Naidenko, O. V., Iizuka, K., Furukawa, H., Beckman, D. L., Pingel, J. T., Scalzo, A. A., Fremont, D. H., and Yokoyama, W. M. 2002. Recognition of a virus-encoded ligand by a natural killer cell activation receptor. *Proc Natl Acad Sci U S A*, 99(13): 8826–8831. (Cited on pages 11, 20, 25, 48, and 76.)

Sternberg-Simon, M., Brodin, P., Pickman, Y., Onfelt, B., Kärre, Malmberg, K.-J., Höglund, P., and Mehr, R. 2013. Natural killer cell inhibitory receptor expression in humans and mice: a closer look. *Front Immunol*, 4: 65. (Cited on pages 2 and 3.)

Sun, J. C. and Lanier, L. L. 2008. Tolerance of NK cells encountering their viral ligand during development. *J Exp Med*, 205(8): 1819–1828. (Cited on pages 50 and 79.)

Sun, J. C. and Lanier, L. L. 2009. The Natural Selection of Herpesviruses and Virus-Specific NK Cell Receptors. *Viruses*, 1(3): 362. (Cited on pages 9, 11, 33, 48, 76, 108, and 133.)

Takahashi, T., Yawata, M., Raudsepp, T., Lear, T. L., Chowdhary, B. P., Antczak, D. F., and Kasahara, M. 2004. Natural killer cell receptors in the horse: evidence for the existence of multiple transcribed LY49 genes. *Eur J Immunol*, 34(3): 773–784. (Cited on page 8.)

Thananchai, H., Gillespie, G., Martin, M. P., Bashirova, A., Yawata, N., Yawata, M., Easterbrook, P., McVicar, D. W., Maenaka, K., Parham, P., Carrington, M., Dong, T., and Rowland-Jones, S. 2007. Cutting Edge: Allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B. *J Immunol*, 178(1): 33–37. (Cited on pages 13 and 115.)

Tomasec, P., Braud, V. M., Rickards, C., Powell, M. B., McSharry, B. P., Gadola, S., Cerundolo, V., Borysiewicz, L. K., McMichael, A. J., and Wilkinson, G. W. 2000. Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science*, 287(5455): 1031. (Cited on pages 10 and 97.)

Trowsdale, J. 2001. Genetic and functional relationships between MHC and NK receptor genes. *Immunity*, 15(3): 363–374. (Cited on page 3.)

Trowsdale, J., Barten, R., Haude, A., Stewart, C. A., Beck, S., and Wilson, M. J. 2001. The genomic context of natural killer receptor extended gene families. *Immunol Rev*, 181: 20–38. (Cited on pages 5, 19, 47, 109, and 118.)

Ulbrecht, M., Martinozzi, S., Grzeschik, M., Hengel, H., Ellwart, J. W., Pla, M., and Weiss, E. H. 2000. Cutting edge: the human cytomegalovirus UL40 gene product contains a ligand for HLA-E and prevents NK cell-mediated lysis. *J Immunol*, 164(10): 5019–5022. (Cited on pages 10 and 97.)

van Bergen, J., Thompson, A., van Pel, M., Retière, C., Salvatori, D., Raulet, D. H., Trowsdale, J., and Koning, F. 2013. HLA reduces killer cell Ig-like receptor expression level and frequency in a humanized mouse model. *J Immunol*, 190(6): 2880–2885. (Cited on page 2.)

van Teijlingen, N. H., Hölzemer, A., Körner, C., García-Beltrán, W. F., Schafer, J. L., Fadda, L., Suscovich, T. J., Brander, C., Carrington, M., Evans, D. T., van Baarle, D., and Altfeld, M. 2014. Sequence variations in HIV-1 p24 Gag-derived epitopes can alter binding of KIR2DL2 to HLA-C*03: 04 and modulate primary natural killer cell function. *AIDS*, 28(10): 1399–1408. (Cited on pages 13, 115, 116, and 123.)

Vilches, C. and Parham, P. 2002. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu Rev Immunol*, 20: 217–251. (Cited on pages 3, 20, and 47.)

Vivier, E., Nunès, J. A., and Vély, F. 2004. Natural killer cell signaling pathways. *Science*, 306(5701): 1517–1519. (Cited on page 2.)

Vivier, E., Tomasello, E., Baratin, M., Walzer, T., and Ugolini, S. 2008. Functions of natural killer cells. *Nat Immunol*, 9(5): 503–510. (Cited on pages 2 and 75.)

Voigt, V., Forbes, C. A., Tonkin, J. N., Degli-Esposti, M. A., Smith, H. R. C., Yokoyama, W. M., and Scalzo, A. A. 2003. Murine cytomegalovirus m157 mutation and variation leads to immune evasion of natural killer cells. *Proc Natl Acad Sci U S A*, 100(23): 13483–13488. (Cited on pages 76 and 77.)

Wagner, M., Gutermann, A., Podlech, J., Reddehase, M. J., and Koszinowski, U. H. 2002. Major histocompatibility complex class I allele-specific cooperative and competitive interactions between immune evasion proteins of cytomegalovirus. *J Exp Med*, 196(6): 805–816. (Cited on page 10.)

Wahl, A., Schafer, F., Bardet, W., and Hildebrand, W. H. 2010. HLA class I molecules reflect an altered host proteome after influenza virus infection. *Hum Immunol*, 71(1): 14–22. (Cited on page 116.)

Wallace, A. E., Fraser, R., and Cartwright, J. E. 2012. Extravillous trophoblast and decidual natural killer cells: a remodelling partnership. *Hum Reprod Update*, 18(4): 458–471. (Cited on pages 14 and 133.)

Wende, H., Colonna, M., Ziegler, A., and Volz, A. 1999. Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. *Mamm Genome*, 10(2): 154–160. (Cited on page 3.)

Wilhelm, B. T. and Mager, D. L. 2004. Rapid expansion of the Ly49 gene cluster in rat. *Genomics*, 84(1): 218–221. (Cited on pages 47 and 75.)

Wilhelm, B. T., Gagnier, L., and Mager, D. L. 2002. Sequence analysis of the ly49 cluster in C57BL/6 mice: a rapidly evolving multigene family in the immune system. *Genomics*, 80(6): 646–661. (Cited on pages 7, 47, and 75.)

Wilkinson, G. W. G., Tomasec, P., Stanton, R. J., Armstrong, M., Prod'homme, V., Aicheler, R., McSharry, B. P., Rickards, C. R., Cochrane, D., Llewellyn-Lacey, S., Wang, E. C. Y., Griffin, C. A., and Davison, A. J. 2008. Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol*, 41(3): 206–212. (Cited on page 20.)

Xiong, S., Sharkey, A. M., Kennedy, P. R., Gardner, L., Farrell, L. E., Chazara, O., Bauer, J., Hiby, S. E., Colucci, F., and Moffett, A. 2013. Maternal uterine NK cell-activating receptor KIR2DS1 enhances placentation. *J Clin Invest*, 123(10): 4264–4272. (Cited on page 14.)

Yawata, M., Yawata, N., Draghi, M., Little, A.-M., Partheniou, F., and Parham, P. 2006. Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function. *J Exp Med*, 203(3): 633–645. (Cited on pages 6

and 135.)

Yokoyama, W. M. and Kim, S. 2006. Licensing of natural killer cells by self-major histocompatibility complex class I. *Immunol Rev*, 214: 143–154. (Cited on page 31.)

Zeng, L., Sullivan, L. C., Vivian, J. P., Walpole, N. G., Harpur, C. M., Rossjohn, J., Clements, C. S., and Brooks, A. G. 2012. A structural basis for antigen presentation by the MHC class Ib molecule, Qa-1b. *J Immunol*, 188(1): 302–310. (Cited on pages 7 and 75.)

Zhi-ming, L., Yu-lian, J., Zhao-lei, F., Chun-xiao, W., Zhen-fang, D., Bing-chang, Z., and Yue-ran, Z. 2007. Polymorphisms of killer cell immunoglobulin-like receptor gene: possible association with susceptibility to or clearance of hepatitis B virus infection in Chinese Han population. *Croat Med J*, 48(6): 800–806. (Cited on page 48.)

Ziegler, H., Thale, R., Lucin, P., Muranyi, W., Flohr, T., Hengel, H., Farrell, H., Rawlinson, W., and Koszinowski, U. H. 1997. A mouse cytomegalovirus glycoprotein retains MHC class I complexes in the ERGIC/cis-Golgi compartments. *Immunity*, 6(1): 57–66. (Cited on page 11.)

Summary

Natural killer (NK) cells are cells from the innate immune system that patrol the body, identifying infected and cancer cells. To distinguish healthy from pathological cells, the activity of NK cells is tightly regulated with a sophisticated detection system which is composed of activating and inhibitory cell surface receptors. The best characterized mechanism of NK cell activation is “missing self” detection, i.e. the recognition of virally infected or transformed cells that reduce the expression of self molecules (MHC-I molecules) to evade other cytotoxic immune cells. To monitor the expression of MHC-I on target cells, NK cells have NK cell receptors (NKRs), which are encoded by gene families showing a remarkable genetic diversity. The NKR haplotypes contain several genes encoding for either activating or inhibiting receptors, varying in gene content and allelic polymorphism. However, for the recognition of missing MHC molecules this diversity seems redundant as one set of receptors should be sufficient. Why then have these polygenic and polymorphic receptors evolved?

Because of the evolutionary arms race between infectious pathogens and the immune system, NKRs have possibly diversified due to the selection pressure imposed by the successful immunoevasive mechanisms evolved by several pathogens. In this thesis, we use computational evolutionary modeling, data analysis, and mathematical modeling to answer the question of why NKRs have evolved to become specific, polygenic, and polymorphic. We focus on the different hypothesis related to viral mechanisms to escape the NK cell immune response. We first studied the effect of “MHC-I decoys” on the evolution of the specificity of inhibiting receptors (Chapter 2). For this purpose, we developed an agent-based model of human-like populations that are infected with herpes-like viruses causing chronic infections. Next to downregulating MHC-I molecules, these viruses are able to escape the NK cell response by expressing MHC-decoys engaging the inhibitory NKRs. We show that specific NKR-MHC interactions are needed to protect against

viruses expressing such decoys. Because of the provided protection, specific inhibitory NKRs have an evolutionary advantage, giving rise to a high level of diversity.

In Chapter 3, we extended this model and included activating receptors, allowing us to quantify how specific activating and inhibiting NK cell receptors must be to be protective to these Herpes-like viruses. We confirm our findings from Chapter 2 and observe that activating receptors also evolve specificity in order to be protective. In Chapter 4 we expand our model further and allow for host-pathogen co-evolution. Here, we show that viruses readily evolve MHC-decoys and that inhibiting and activating receptors adapt by evolving specificity, acquiring the ability to discriminate between self MHC molecules and decoy molecules encoded in the viral genome. As a result, the genes encoding NK cell receptors becomes polygenic, encoding both inhibiting and activating receptors.

In Chapter 5 we shift our focus to other viral strategies of immune evasion and study the effect of selective MHC downregulation on the evolution of inhibiting receptors. Finally, we use a mathematical model to study whether peptide sensitivity can be a possible explanation of polygenicity of NK cell receptors in Chapter 6. In both these studies, the specificity of inhibiting receptors is again crucial to successfully discriminate healthy cells from the infected ones.

Taken together, we demonstrate that NKRs can diversify due to the selection pressure imposed by the successful immunoevasive mechanisms evolved by several pathogens. Our studies show that NK cell receptors evolve to become specific to distinguish self from non-self and that the required specificity selects for the evolution of polygenicity and polymorphism. The results in this thesis shed light on the evolutionary processes of NK cell receptors, providing plausible explanations for this fascinating complex system.

Samenvatting

Natural Killer (NK) cellen behoren tot het aspecifieke immuunsysteem en patrouilleren het lichaam op zoek naar geïnfecteerde cellen en kankercellen. Om onderscheid te kunnen maken tussen gezonde en afwijkende cellen wordt de activiteit van NK cellen strikt gereguleerd met behulp van een geraffineerd detectie mechanisme bestaande uit activerende en inhiberende receptoren op het celoppervlak. Het best beschreven mechanisme dat NK cellen activeert, is het detecteren van *"missing self"*: het detecteren van geïnfecteerde of getransformeerde cellen die een verminderde expressie van *self* moleculen (MHC-I moleculen) vertonen, waarmee de cellen andere cytotoxische immuunreacties ontwijken. Om de expressie van MHC-I op een verdachte cel waar te kunnen nemen, bezitten NK cellen NK-cel receptoren (NKRs), die gecodeerd worden door genfamilies met een opmerkelijke diversiteit. NKR haplotypes bevatten een aantal genen die coderen voor receptoren met activerende of inhiberende signalen, en de haplotypes zijn polymorf omdat ze verschillende aantallen genen en verschillende varianten (allelen) van die genen bevatten. Deze diversiteit lijkt echter overbodig om het ontbreken van MHC moleculen te detecteren, aangezien één stel receptoren voldoende zou moeten zijn om MHC moleculen te detecteren. Waarom heeft evolutie dan toch geleid tot polygene en polymorfe receptoren?

Omdat infectieuze pathogenen en het immuunsysteem verwikkeld zijn in een evolutionaire wapenwedloop, lijkt het waarschijnlijk dat diversiteit van NKRs veroorzaakt is door de evasieve mechanismen die pathogenen evolueren. In dit proefschrift gebruiken we evolutionaire modellering van evolutie, data analyse, en wiskundige modellen om de vraag te beantwoorden waarom evolutie geleid heeft tot specifieke, polygene en polymorfe NKRs. De focus ligt op virale infecties en de mechanismen die virussen evolueren om te ontsnappen aan NK-cellen. We hebben eerst het effect van "MHC-I imitaties" op de evolutie van de specificiteit van inhiberende receptoren onderzocht (Hoofdstuk

2). We hebben een agent-gebaseerd model ontwikkeld waarin gesimuleerde menselijke populaties geïnfecteerd worden met chronische Herpes-achtige virussen. Naast het onderdrukken van de expressie van MHC-I moleculen, zijn deze virussen ook nog in staat om aan de NK-cel reactie te ontsnappen door imitaties van MHC moleculen tot expressie te brengen, die de inhiberende NKRs stimuleren. We laten zien dat specifieke MHC-NKR interacties bescherming biedt tegen virussen die dergelijke imitaties tot expressie kunnen brengen. Dit evolutionaire voordeel leidt tot specifieke NKRs. Daarna vraagt een goede bescherming van de gastheer om een hoge diversiteit van de NKRs.

In Hoofdstuk 3 hebben we dit model met activerende receptoren uitgebreid. Dit heeft ons in staat gesteld om te quantificeren welke specificiteit van activerende en inhiberende NK-cel receptoren noodzakelijk is om bescherming te bieden tegen soortgelijke Herpes-achtige virussen. Hiermee bevestigen we onze resultaten uit Hoofdstuk 2, en vinden we dat ook bij activerende receptoren er evolutie van specificiteit plaatsvindt. In Hoofstuk 4 breiden we ons model nog verder uit en staan we co-evolutie tussen gastheer en pathogeen toe. We laten zien dat virussen gemakkelijk MHC-imitaties evolueren en dat activerende en inhiberende receptoren zich op hun beurt aanpassen door specificiteit te evolueren, waardoor ze onderscheid kunnen maken tussen authentieke MHC moleculen en virale imitaties. Deze specificiteit heeft tot gevolg dat de NK-cel receptorgenen polygeen worden en dat haplotypes zowel activerende als inhiberende receptoren evolueren.

In Hoofdstuk 5 hebben we onze aandacht gericht op andere virale strategieën om het immuunsysteem te ontwijken. We bestuderen het effect van selectieve onderdrukking van MHC-1 expressie op de evolutie van inhiberende receptoren. Ten slotte gebruiken we in Hoofdstuk 6 een wiskundig model om te bestuderen of sensitiviteit voor peptiden een mogelijke verklaring biedt voor het polygene van NKR haplotypes. Uit beide studies blijkt dat de specificiteit van inhiberende receptoren wederom cruciaal is om succesvol onderscheid te kunnen maken tussen *self* en *non-self*.

We stellen voor dat de diversiteit van NKRs mogelijk veroorzaakt is door de selectiedruk van succesvolle immuno-evasieve mechanismen die sommige pathogenen hebben geëvolueerd. Ons onderzoek laat zien dat de noodzaak om onderscheid te kunnen maken tussen *self* en *non-self* de evolutie van specifieke NK-cel receptoren tot gevolg heeft, en dat deze vereiste specificiteit op haar beurt leidt tot evolutie van polygeniteit en polymorfismen. De resultaten in dit proefschrift geven inzicht in de evolutionaire processen waar aan NK-cel receptoren onderhevig zijn, en geven plausibele verklaringen voor dit fascinerende en complexe systeem.

Resumen

Las células asesinas naturales (NK por sus siglas en inglés, *natural killer*) participan activamente en la defensa del organismo contra infecciones virales y contra el crecimiento de células tumorales. Para que las células NK desarrollen su función citotóxica, éstas tienen que identificar a la célula blanco. Para ello, las células NK poseen diferentes tipos de receptores en su superficie, que pueden ser del tipo activador o inhibidor, y que reconocen moléculas en las células dianas. El ligando de muchos receptores inhibidores son las moléculas de complejos mayores de histocompatibilidad (CMH-I) que normalmente están presentes en la superficie de toda célula nucleada. Debido a que las moléculas CMH-I indican a otras células inmunes acerca de la salud de la célula (es decir, si está infectada o no), varios virus y algunos cánceres impiden la expresión de CMH-I. Pero en ausencia de moléculas CMH-I, las células NK se activan y reconocen a la célula como aberrante o infectada. Éste es el modelo mejor caracterizado de activación de las células NK y es llamado la teoría de "missing-self".

Para reconocer la presencia de moléculas CMH-I, las células NK utilizan receptores que son codificados por genes que presentan una enorme variabilidad genética. De hecho, los haplotipos que codifican dichos receptores se diferencian en contenido genético, polimorfismo, y en el tipo de receptores, es decir si son activadores o inhibidores. Sin embargo, esta gran diversidad genética parece redundante, ya que para detectar la presencia de moléculas CMH-I un sólo receptor debería ser suficiente. ¿Por qué entonces las células NK han evolucionado receptores con tanta complejidad genética?

Debido a la carrera armamentista que ocurre entre varios agentes patógenos y el sistema inmunitario, los receptores de células NK pudieron haberse diversificado debido a las fuerzas evolutivas impuestas por los mecanismos de evasión desarrollados por varios patógenos. En esta tesis, hemos utilizado modelos evolutivos computacionales, análisis de datos y modelos matemáticos para estudiar el por qué los receptores de células NK han

evolucionado especificidad y una gran variabilidad genética. Nuestro enfoque ha sido el de las hipótesis relacionadas con infecciones virales y los mecanismos de evasión de la activación de las células NK. Primero hemos estudiado cuál es la fuerza de selección impuesta por virus que codifican moléculas "imitadoras" de CMH-I para la evolución de especifidad en los receptores (capítulo 2). Para esto, hemos desarrollado un modelo basado en agentes que simula poblaciones humanas que han sido infectadas con virus con características del virus herpes. En nuestro modelo, esta clase de virus impide la expresión de CMH-I y además codifica "imitadores" de CMH-I para evadir ser detectado por las células NK. Este estudio demuestra que los receptores deben tener interacciones específicas con las moléculas CMH-I para poder eliminar el virus. Como consecuencia natural de la ventaja evolutiva de los receptores específicos, podemos observar la evolución natural de una gran variedad genética.

A continuación hemos extendido este modelo computacional e incluimos receptores activadores en el capítulo 3. Utilizando un modelo matemático, hemos podido cuantificar la especificidad óptima que los receptores inhibidores y activadores deben tener para poder proteger poblaciones de estos virus característicos de la familia herpes. Aquí confirmamos los hallazgos encontrados en el segundo capítulo, demostrando que la especificidad de las dos clases de receptores es necesaria para sobrevivir la epidemia. En el capítulo 4 continuamos con la extensión de nuestro modelo basado en agentes, permitiendo la co-evolución de virus y huéspedes. En este estudio también pudimos demostrar como los virus fácilmente evolucionan "imitadores" de CMH-I, ya que éstos proveen una ventaja al escapar de la vigilancia inmunitaria. Adaptándose a los virus, las poblaciones evolucionan receptores activadores e inhibidores que son específicos a las moléculas CMH-I, para poder discernir entre moléculas CMH-I o sus imitadoras. Como resultado, los huéspedes evolucionan haplotipos de receptores que contienen una gran cantidad de genes codificadores de receptores de células NK.

En el quinto capítulo nos enfocamos en otros mecanismos de escape viral y estudiamos el efecto de los virus que interfieren con la expresión de solamente aquellas moléculas CMH-I que interact uan con los receptores inhibidores. Finalmente, en el sexto capítulo, analizamos el efecto de los péptidos en la activación de las células NK. En estos dos últimos estudios podemos ilustrar, una vez más, la importancia de la especificidad que los receptores deben tener para poder detectar agentes foráneos de los propios.

Basándonos en los resultados de esta tesis, proponemos que los receptores de las células NK se han diversificado como consecuencia de la fuerza de selección impuesta por ciertos virus y sus mecanismos de escape. Nuestros estudios demuestran claramente que,

para discernir entre agentes foráneos y patógenos de los propios y saludables, los receptores de las células NK deben ser específicos. La especificidad require que cada individuo tenga múltiples receptores, explicando así los haplotipos variables y polimórficos encodificando dichos receptores. Nuestra investigación sin duda ha clarificado los procesos tan complejos detrás de la evolución de los receptores de las células NK.

Curriculum vitæ

Paola Carrillo-Bustamante was born in Quito, Ecuador on May 16th 1984. She attended the German School in Quito (Deutsche Schule Quito), where she obtained the International Baccalaureate in 2003. Sponsored by a full scholarship of the German Academic Exchange Service (DAAD), she started her studies in Electrical Engineering at the Karlsruhe Institute of Technology in Karlsruhe, Germany. She specialized in Biomedical Engineering, where she started using computational modeling to study biological processes. At the Institute of Biomedical Engineering, she conducted several research projects focused on the electrical propagation in the heart tissue under the supervision of Prof.dr. Olaf Dössel. In 2009, she obtained her diploma in Electrical Engineering and Information technology. In May 2010, she started her PhD research at the Theoretical Biology & Bioinformatics group at Utrecht University, under the supervision of Prof.dr. Rob J. De Boer and Dr. Can Keşmir. The results of her PhD research are described in this thesis.

List of Publications

Paola Carrillo-Bustamante, Rob J. De Boer, Can Keşmir. Can peptide sensitivity explain the diversity of NK cell receptors? 2014. *In preparation*

Paola Carrillo-Bustamante, Can Keşmir, Rob J. De Boer. Selective MHC downregulation can explain the specificity and genetic diversity of NK cell receptors. 2014. *In preparation*

Paola Carrillo-Bustamante, Can Keşmir, Rob J. De Boer. A co-evolutionary arms race between hosts and viruses drives polymorphism and polygenicity of NK cell receptors. 2014. *Submitted*

Paola Carrillo-Bustamante, Can Keşmir, Rob J. De Boer. Quantifying the protection of activating and inhibiting NK cell receptors during infection with a CMV-like virus. 2014. *Frontiers Immunology 5:20.*

Paola Carrillo-Bustamante, Can Keşmir, Rob J. De Boer. Virus encoded MHC-like decoys diversify the inhibitory KIR repertoire. 2013. *PLoS Computational Biology; 9(10)*

Eberhard P. Scholz, **Paola Carrillo-Bustamante**, Fathima Fischer, Mathias Wilhelms, Edgar Zitron, Olaf Dössel, Hugo A. Katus, Gunnar Seemann. Rotor termination is critically dependent on kinetic properties of I kur inhibitors in an in silico model of chronic atrial fibrillation. 2013. *PLoS One; 8(12)*

Gunnar Seemann, **Paola Carrillo-Bustamante**, Stefan Ponto, Mathias Wilhelms, Eberhard P. Scholz, Olaf Dössel. Atrial fibrillation-based electrical remodelling in a computer model of the human atrium. 2010. *Computing in Cardiology, 417-420*.

Paola Carrillo, Gunnar Seemann, Eberhard P. Scholz, Daniel L. Weiss, Olaf Dössel. Impact of the hERG channel mutation N588K on the electrical properties of the human atrium. 2009. *4th eruopean Conference of the International Federation for Medical and Biological Engineering, 2583-2586. .*

Acknowledgments

When I first started my PhD, I naively thought that modeling heart cells could not be *that* different than modeling the evolution of NK cells. After all, in the mind of an Electrical Engineer, heart cells are pretty much the same thing as immune cells. Little I knew how mistaken I was, and the way to figure that out (and solve it!) was indeed steep. In those steep and courageous steps, many people were involved. And now, time has come to say thanks.

First and foremost, my deep thanks goes to Rob and Can, my supervisors. You have witnessed and guided my development from a square minded engineer to a scientist thinking in the colorful terms of biology. You taught me to ask the right questions and be critical about science. I was amazed by the freedom you granted me, and the trust you showed in me, despite the hurdles during this developmental process. Thank you for keeping the encouragement and motivation when I had lost mine. But most importantly, Rob and Can, thank you for all your immense support, not only scientifically but also personally. I am lucky to have been mentored by the great team the two of you form.

Rob and Can always supported me to attend several conferences and seminars. I am very grateful because in these meetings I had the opportunity to expose my theoretical models to a biological (somewhat skeptical) audience. Paul Norman, Becca Asquith, Jim Kaufmann, Ramit Mehr, and Mary Carrington, thank you for answering to all my questions, discuss my work, and for all the input I received from you. Thank you also to the Dutch experts, Linde Meyaard and Ronald Bontrop. Especially Ronald, who always had a kind word and new insights to share. Paul, Becca, Jim, Linde, and Ronald, thank you also for taking time to read my thesis, your critical input was invaluable.

The PhD years were challenging, but all difficulties were balanced with the great deal of fun I had at the TBB lab. It takes a lot of colorful people to make this lab as special

as it is. After all, not every lab has a kitchen bitch, a coffee czar, a choir performing at least once per year with self composed songs, a handicraft workshop when required, lab uitjes in the mud, and dinners made with so much dedication cooked by the bosses themselves. It has been fun, entertaining, and stimulating to talk to you about science and other random stuff. Thank you TBB for these unforgettable memories. A few (ex) members I need to specifically mention here. Jan Kees, you are the one that makes all of our work possible. Thank you for being the wheel behind all TBB research. Also many thanks to the translating team, Chris and Alejandra, for taking the time to help me out with translating my summary into Dutch (Chris) and pimping my Spanish version (Ale). Ale, I am also so grateful because you were the first one to patiently explain to me the basics of biology I needed to know. Hanneke, you have always been incredibly supportive. Not only because you taught me R and the hash of a hash of a hash, but because you were there, always. You would know with just one look if I needed you, and you would know immediately how to react to it. Thank you so much for that.

Lidija and Sandro, my dear paranymphs, our time together started during those dark winter months, working long evenings during Paulien's course. In time, our conversations developed into more substantial discussions about science, life and curly brackets, while enjoying dinners and one or two drinks. You have born with my frustrations and celebrated with me the successes; endless are the memories I have with each of you, thank you for each and every single one of them.

Naturally, many persons outside academia have also been crucial in the success of the last years: my family. And I am lucky to have many of those :-) Francesca, Sanja, and Abel, my Utrecht family, you have brought tons of warmth and sunshine into the cold rainy Netherlands, thanks for being a home for me. Isabel, although we have spent so many years aside, it is remarkable how small an ocean can feel. Thanks for the almost daily messages, thanks for being an anchor.

Despite sometimes not understanding my choices, Paul, Christian, Washington and Nancy have been there, pushing me, lifting me, caring even though the distance makes it so frustrating difficult sometimes. Spread around 3 continents, it has been wonderful to see how close we keep growing. You are that safe place I can always come back to.

And you Oussama, my current and future family, thank you for all these adventurous years. For being a source of encouragement, even when all my frustration about object oriented programming and debugging was coming out of me with rage. For all your enlightening ideas. For being the ideal partner. Ouss, thank you for being my accomplice, my love, and everything... and because together we are many more than two.