

## ***Searching for Molecular Solutions* – Additional Material**

### **CHAPTER 3**

These Files contain additional material relevant to **Chapter 3** of *Searching for Molecular Solutions*. The page numbers of the book pertaining to each section are shown in the Table below, the corresponding page number for this file, and the title of each relevant section.

#### **Contents:**

<b>Most Relevant Book Page Numbers</b>	<b>Page Number in this File</b>	<b>Section</b>	
		<b>No.</b>	<b>Title</b>
72	2	<a href="#">A1</a>	RAG Genes and Proteins
81	7	<a href="#">A2</a>	Repertoire Holes
84	21	<a href="#">A3</a>	Autoimmunity
91	30	<a href="#">A4</a>	Immune / Nervous System Comparisons

## Section A1: ***RAG Genes and Proteins***

Cited on p. 38 of *Searching for Molecular Solutions*; most relevant to immune somatic recombination processes depicted on p. 72

### *The riches of RAGs, and other movers of adaptive immunity*

Genetic recombination exemplifies principles of biological economy and parsimony<sup>▼</sup>, where different combinations of enzymes from a broad recombinational 'toolbox' can be used for a variety of applications. But there are limits to how much can be accomplished with one set of tools. The special instance of immune system recombination which assembles B and T cell antigen recognition receptors from separate germline segments would, from first principles, appear to require a special recombinase to provide the necessary level of recombinational specificity. And this is indeed the case. The products of two genes, termed recombination-activating gene (RAG)-1 and 2, were identified as the key players in this somatic rearrangement process<sup>1,2</sup>. The RAG-1 and RAG-2 gene products are indeed essential for immune receptor recombination, and the absence of either blocks the differentiation of T and B lymphocytes<sup>3,4</sup>. Although the RAG proteins are necessary for recombination, they are not sufficient, and a number of other ubiquitous DNA processing enzymes (of the non-homologous end-joining pathway) are required in the recombinational complex<sup>5</sup>. RAG proteins recognize specific sequence motifs (recombination signal sequences, or RSS) and catalyze the V(D)J rearrangements in both T and B cells, which involves the formation and re-ligation of double-stranded DNA breaks<sup>6-8</sup>. Sequence alterations at the rejoining junctions can also occur, either through the agency of the RAG-mediated recombinational mechanisms themselves (P-nucleotide addition) or through the activity of terminal deoxynucleotidyl transferase (N-region addition).

---

<sup>▼</sup> See SMS—Extras—Ch9/Section A15.

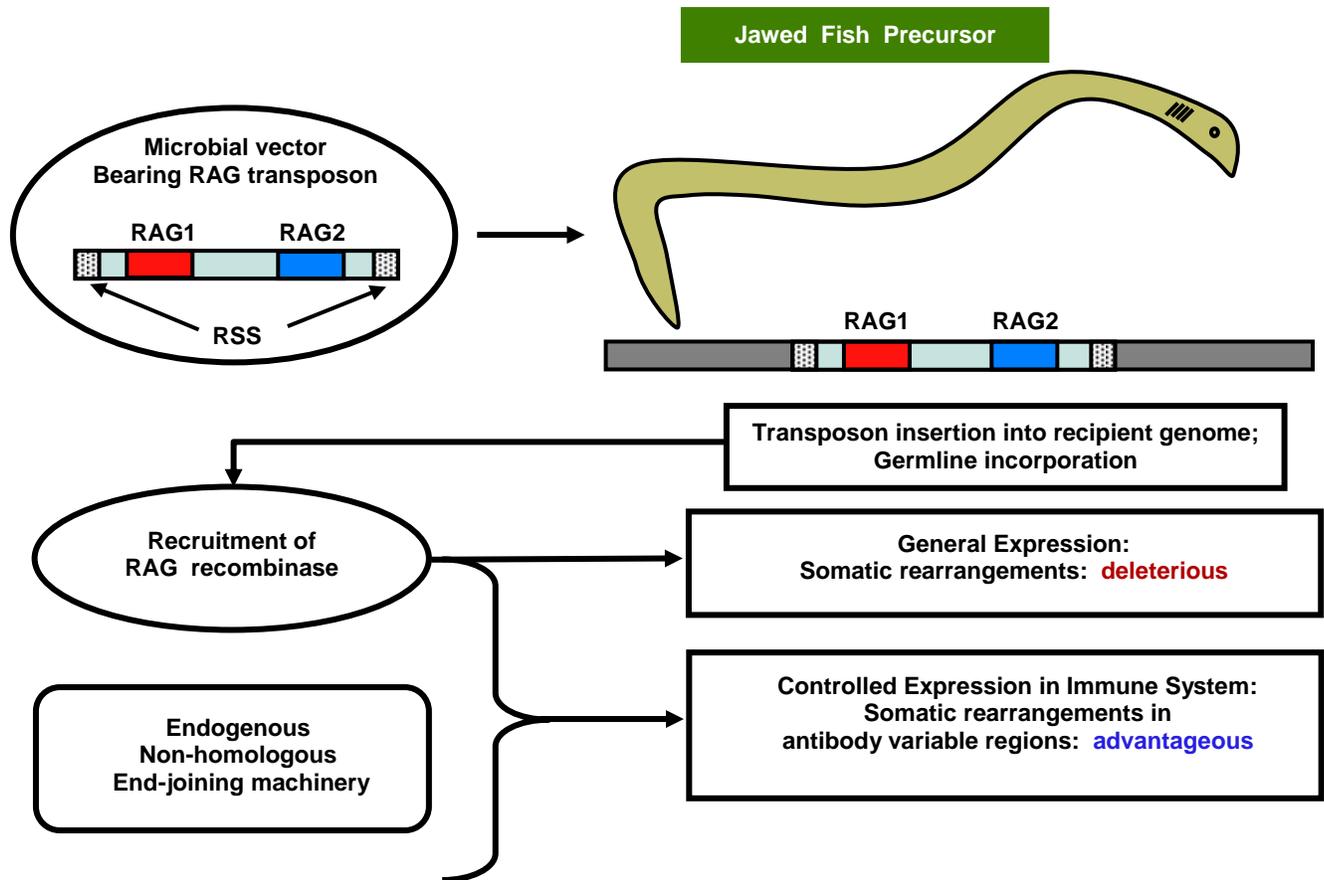


Fig. 3A1.1

Proposed origin of vertebrate RAG1 and RAG2 (recombination-activating genes) from a prokaryotic transposable element<sup>9</sup> (transposon). RSS = recombination signal sequences. 'Jawed fish precursor' denotes the point of acquisition of the hypothetical prokaryotic transposon into the vertebrate germline soon after the divergence of jawed fish (the gnathostome lineage) from the agnathans, or jawless fish. (The latter lineage are represented by modern lampreys and hagfish, which have a distinct type of immune recognition system<sup>10,11</sup>).

RAG genes are detectable in all vertebrates with the exception of the primitive jawless fishes (Agnathans). Considerable evidence favors the interpretation that

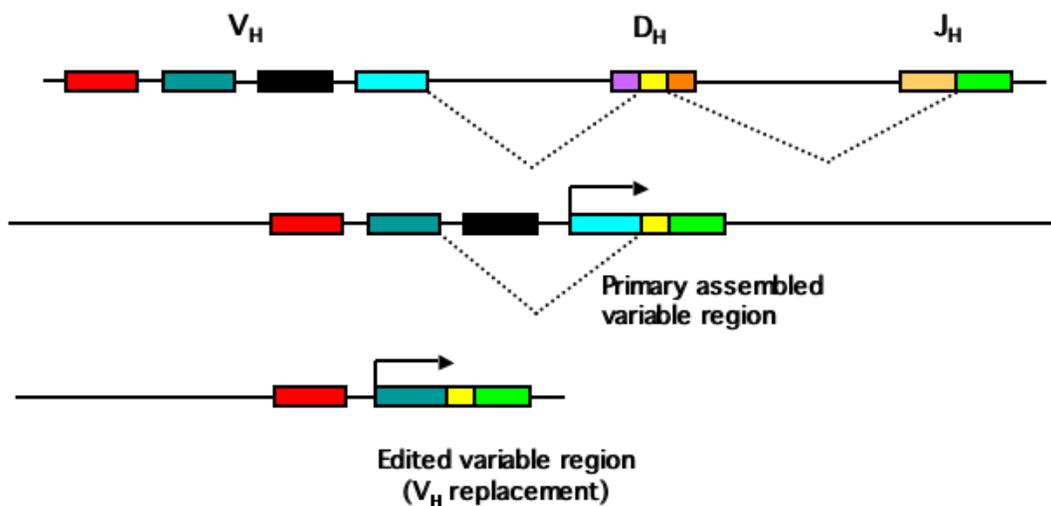
RAG genes were originally acquired by an ancestral vertebrate by means of a lateral (horizontal) transfer event mediated through a prokaryotic transposon. This surmise was based in part on the apparent sudden appearance of the RAG recombinational system (at least in evolutionary terms) into the vertebrate lineage and its lack of clear evolutionary antecedents. (For this reason, the advent of the adaptive immune system has been termed the 'Big Bang' of immunology). Also, certain similarities exist between RAGs and known bacterial transposases and integrases in their structures, genetic organization, and modes of action<sup>9</sup>. Furthermore, RAG proteins themselves can be shown catalyze transposition events which have transposon-associated features<sup>12,13</sup>. It is thus proposed that a prototypical RAG transposon laterally entered the vertebrate germline shortly after the divergence of the jawed fishes from the agnathans (the lineage which includes modern jawless fish such as lampreys), by interspersing itself into an immune receptor gene. The interrupted gene could nevertheless be reassembled through the action of the RAG proteins themselves. When adapted into the vertebrate environment with ubiquitous end-joining proteins<sup>5</sup> and other regulatory components<sup>14</sup>, the RAG usurpers became 'domesticated'<sup>15</sup> into the service of adaptive immunity<sup>9</sup> (depicted in Fig. 3A1.1). Although this model has not been universally accepted<sup>16,17</sup>, the weight of evidence in its favor has resulted in its being widely considered to be the most probable scenario<sup>18,19</sup>.

The RAG proteins thus mediate the primary recombinational events which assemble immunoglobulin (and T cell) variable regions, but their roles in the adaptive immune response do not end there. It is known that functional pre-assembled immunoglobulin variable regions can be replaced with another upstream V region; a processing event known as V-gene replacement or receptor editing<sup>▼</sup><sup>22</sup> (depicted in Fig. 3A1.2 below). Transgenic mouse models have clearly revealed the relevance of immunoglobulin receptor editing to the immune repertoire<sup>23,24</sup>. For example, mice have been derived with functional

---

▼ This editing process appears to be a normal pathway for altering the specificity of otherwise-dangerous self-reactive B cells<sup>20,21</sup>.

germline ‘knock-in’ assembled variable regions (encoding a known antigen-binding specificity to a small chemical hapten) into the immunoglobulin heavy chain locus, such that their immunoglobulin heavy chain expression would be expected to be monoclonal if no other factors were operative <sup>23</sup>. Extensive replacements of the  $V_H$  region were in fact noted <sup>23</sup>, and this diversification (combined with somatic hypermutation) was sufficient for the mice to produce protective antibodies against a virus <sup>25</sup>. A compelling case for the role of RAG recombination here was the observation that if such transgenic mice were also deficient in RAG proteins, no  $V_H$  replacement occurred <sup>26</sup>.



**Fig. 3A1.2**

Depiction of receptor editing process at the immunoglobulin heavy chain locus. (Numbers of  $V_H$ ,  $D_H$ , and  $J_H$  genes are for schematic purposes only and do not reflect actual gene numbers for humans or other mammals).

---

But despite such findings indicating the primacy of RAG recombination, another mechanism of receptor editing exists. This process is the type of homologous

recombination called gene conversion, the primary pathway leading to variable region diversity in some species (such as chickens). In some situations, gene conversion may also have a role in V-region replacement<sup>27,28</sup>. Although gene conversion *per se* can operate in varied biological circumstances, in the context of immune receptor diversification (whether primary, or via secondary replacement), the involvement of another pivotal driver of the adaptive immune system has been demonstrated. This is a nucleic acid editing enzyme termed AID (for ‘activation-induced deaminase’), which elicits the deamination of cytidine residues (forming uridines). The deamination activity is preferentially manifested on single-stranded DNA<sup>29,30</sup>, although RNA can also be an AID substrate<sup>31</sup>. In fact, AID has been shown to be essential for the three known somatic modifiers of the primary rearranged immunoglobulin repertoire: somatic hypermutation<sup>▼</sup>, gene conversion, and heavy chain class-switch recombination<sup>33</sup>.

Class switch recombination allows an assembled immunoglobulin heavy chain variable region to switch expression to a different downstream constant region, thus changing the class of the expressed antibody (eg. from IgM to IgA1<sup>\*</sup>). Switching to alternative constant region genes therefore confers different functional properties onto the same antibody combining specificity, a biologically useful activity. Switch recombination is a complex and distinct process from V(D)J rearrangements<sup>35</sup>, but does not involve diversification and selection, so it will not be considered further here.

---

▼ Other enzymes, principally DNA polymerase- $\eta$ , appear to be involved in somatic hypermutation, such that other types of mutations (other than the AID-induced pattern) can occur<sup>32</sup>.

\* Switching follows the order of the constant region genes in the immunoglobulin heavy chain locus. The C $\mu$  constant region gene (from which IgM is produced) is at the end closest to the assembled VDJ variable region and thus expressed first; switching of the *same* specific VDJ segment to downstream constant regions can subsequently occur. The only exception is for the constant region C $\delta$  (the most proximal C-gene to C $\mu$ ; from which IgD is produced); which is expressed by an alternate splicing mechanism<sup>34</sup> rather than through switch recombination.

## Section A2: *Repertoire Holes*

This section contains further thoughts on limitations of antibody repertoires. It is most relevant to the section 'Antibodies and Molecular Recognition in General' on p. 81 of *Searching for Molecular Solutions*, and also for the comparisons of antibodies and DNA-binding proteins, made on pp. 83-84, 128, 168, 219, 252, and 270

### *Antibody Repertoire Restrictions*

While antibodies are wonderfully capable in their ability to bind diverse structures, they have limitations with respect to certain molecular recognition demands. In other words, a single type of molecular framework is unlikely to be a universal solution to molecular binding design problems. This issue is depicted schematically in Fig. 3A2.1 below. Here it is postulated that some structures exist towards which no effective antibody can be obtained, but within a universal molecular space some other binding structure exists which is up to the job. This postulate is based on known properties of antibodies (as discussed in Chapter 3 of *Searching for Molecular Solutions*), and it seems probable that it could be generalized for any *single* molecular framework alternative to antibodies (albeit a difficult proposition to prove).

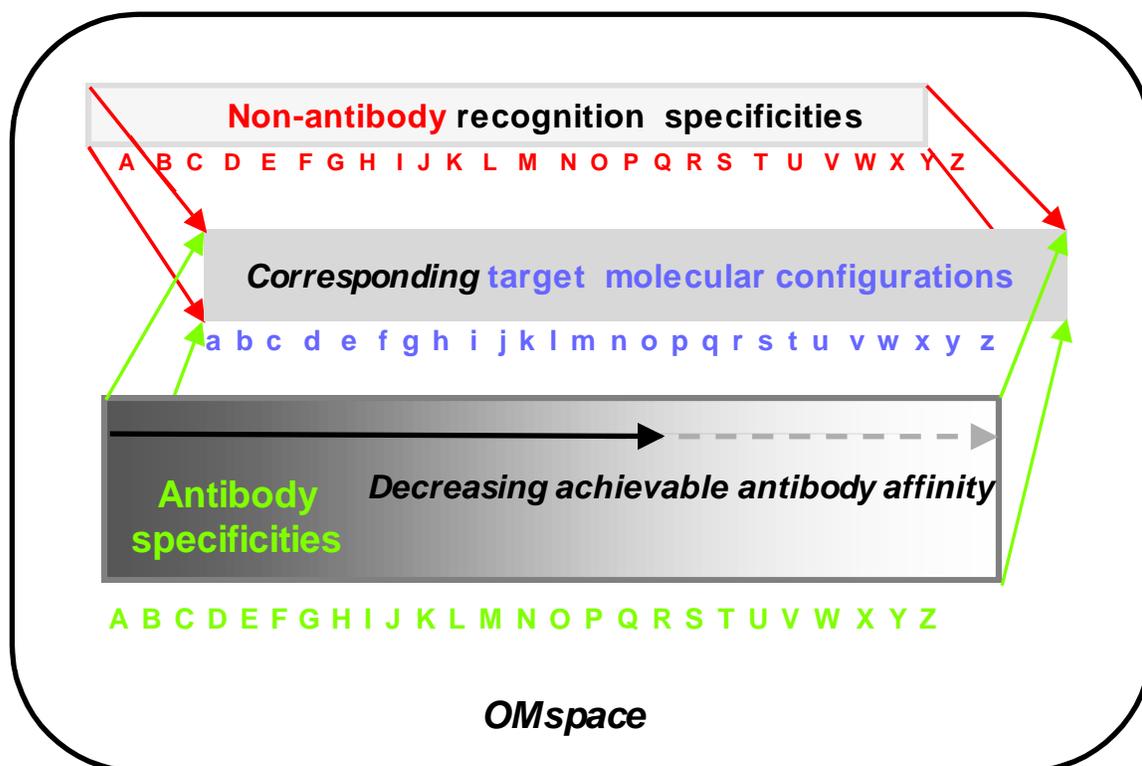


Fig. 3A2.1

Models of the range of the potential antibody repertoire. Molecular receptors and target molecules are depicted within all-encompassing molecular space (OMspace). The range of antibody specificities against all potential molecular target moieties is depicted as an array (A-Z; green), mapped onto the corresponding shapes recognized (a-z; blue). In this model molecular shapes exist against which no high-affinity antibody specificity can ever be achieved (for example, shapes y and z), but other types of non-antibody binding molecules may provide superior recognition against the same shape (A-Z, red). Note that an antibody specificity Z binding weakly to z might cross-react with a higher affinity against a different shape, but no better antibody against z exists in this model.

### *Repertoire Restrictions and Sources of Holes: Thinking Hole-istically*

Now, in the context of antibody repertoire limitations, let's take a short side-step and think about what physical factors might be involved. These considerations are applicable to antibodies, but are generalizable to any protein-based receptor system. 'Physical limitation' refers to any aspect of a receptor molecule's structure itself, or a restriction during its biosynthesis, which causes certain receptor sequences to be absent or under-represented. This in turn could have ramifications as to the potential range of ligands to which any possible variants of the receptor could feasibly bind. The theme of limitations in biological receptor expression serves as a springboard for a wide-ranging consideration of restrictions on protein expression in general. Some of the potential limitations raised may seem trivial, especially given the vast size of protein sequence space, and the fact that only a very small minority of protein sequences can assume useful folds in any case (as noted in *Searching for Molecular Solutions*). Still, it is worthwhile to list as many potential sources of sequence under-representation in biological systems as possible, even if only to conclude that some are not significant limitations in practice.

Since biological proteins are of course encoded by nucleic acids, an initial question could be directed to possible biases in coding sequences at the nucleotide level. Could a specific protein sequence require a string of genomic DNA (or transcribed RNA) codons whose sequence is unstable or otherwise problematic? If that was the case, some protein sequences (even if a tiny fraction of the total possible) might be excluded from the potential biologically-accessible range. Certainly some DNA sequences (such as highly repetitive or palindromic tracts) can cause problems for DNA replication or other DNA transactions<sup>36</sup>, but in at least some cases secondary structures that potentially arise as a consequence of such tracts may be suppressed by *in vivo* mechanisms<sup>37</sup>.

Sequence-based interference with the stability of a coding sequence is in any case inherently improbable due to the degeneracy of the genetic code. Thus while a palindrome might exist within a coding sequence <sup>▼</sup>, it could be readily eliminated by the use of alternate codons, and the same selection process can eliminate nuclease or other unwanted DNA-binding protein target sites. There is evidence for restrictions on the usage of codons in highly conserved genes which is presumed to be due to sequence-based requirements at the DNA or RNA levels <sup>38</sup>. So the nucleic acid coding level is unlikely to place any significant constraint on the natural repertoire of expressed protein sequences. Even if such a limitation was imposed at the level of biological protein production, it would in principle be possible to circumvent it by complete *in vitro* chemical synthesis of a protein (albeit a difficult prospect in some cases).

But what about 'holes' in general protein repertoires which might be incurred at the protein level itself? There are several potential areas where such a restriction might occur. Initially it is worth considering restrictions on protein sequences in general, and then see if any of these factors are relevant to receptor repertoires. Some peptide sequences (or even single residues of specific amino acids) are incompatible with specific types of protein secondary structures, or place special constraints upon them. An example of this is the effects of proline or glycine residues on  $\alpha$ -helices <sup>\*</sup>. Context-dependent sequence incompatibility can be contrasted with sequences which are not tolerated by any known natural polypeptide, or 'intrinsically restricted' (Table 3A2.1 below). A simple example for

---

<sup>▼</sup> For example, if it was advantageous for a protein to have a long tract of aspartic acid followed by valine residues ( $D_nV_n$  respectively; where  $n$  = the number of residues of each), then this could be encoded by  $(GAC)_n(GTC)_n$ , a palindromic tract. But if synonymous codons (GAT for aspartic acid and GTA, GTG or GTT for valine) are substituted appropriately, a palindromic sequence is avoided.

<sup>\*</sup> Proline (prolyl) residues have the unique structure such that the side chain is cyclized onto the amide in the protein backbone. This results in destabilization, kinking or bending of  $\alpha$ -helices <sup>39,40</sup>. Glycine residues (where the side chain is simply a hydrogen atom) are also  $\alpha$ -helix-disrupting through loss of hydrogen bonding or hydrophobic packing stabilization <sup>41-43</sup>.

the latter are long homopolymer tracts, especially of bulky hydrophobic or charged residues <sup>44</sup>. Thus hydrophobic tryptophan residues of >4 in length have not been observed in known natural proteins <sup>44</sup> (and the current Swiss-Prot protein database).

A sequence change which causes serious structural disruptions to the binding site of an immune receptor will have little to contribute to the repertoire. Even more obviously, sequences incompatible with general protein folding (such as the above-mentioned hydrophobic tracts) are irrelevant in this context. (No protein, no binding site). Thus to have any possible bearing on a variable receptor's repertoire, a sequence must be compatible with the basic receptor framework. (But a repertoire 'hole' at the physical level might in principle result from a deficiency in the molecular design of a particular framework structure itself towards certain targets, as we have seen with antibodies towards specific nucleic acid sequences).

---

Table 3A2.1	
<b>Potential Sources of Sequence-Related Holes in Biological Protein Recognition Repertoires</b>	
<b>Sequences Incompatible within Specific Protein Contexts</b>	
<b>A.</b>	Short sequences incompatible with a functional specific protein structural framework
<b>B.</b>	Sequences incompatible with specific oligomeric protein assembly
<b>Sequences Conditionally Restricted or Unstable in Original Form</b>	
<b>Dependent on Cellular Environment</b>	
<b>C.</b>	Sequences incompatible with processing or trafficking (transport or export) of a specific protein, including sequences generating an inappropriate processing signal
<b>D.</b>	Sequences corresponding to inappropriate enzymatic target sites (for proteases, kinases, glycosyl transferases and others)
<b>E.</b>	Sequences significantly interfering with any aspect of host organism function
<b>Dependent on Biosynthesis at Protein Level</b>	
<b>F.</b>	Sequences incompatible with ribosomal processing of nascent proteins
<b>Directly Sequence-dependent</b>	
<b>G.</b>	Self-splicing, self-cleaving, or other self-modifying protein sequences
<b>Sequences Intrinsically Restricted</b>	
<b>H.</b>	Polypeptides unfoldable in aqueous environment irrespective of molecular chaperones

**Footnotes to Table 3A2.1.** ‘Restricted’ peptide sequences are defined as sequences which do not permit specific protein functioning (whether through direct structural perturbation or indirectly) or sequences which prevent the stable maintenance of a defined original polypeptide itself. These are divided into categories as shown, from top:

**Sequences Incompatible with Specific Protein Contexts:** A sequence incompatible with a specific protein fold or secondary structure (Example for **A**: Proline-containing sequences and regular  $\alpha$ -helices) or oligomeric protein assembly (Example for **B**: Sequences at protein-protein assembly interfaces). ‘**Conditional**’ cases are protein or cell-context dependent (**C-F**), or are inherently sequence-dependent but controllable by inhibitors or co-factors (**G**). Conditional sub-groups: **Dependent on Cellular Environment:** An indirect sequence restriction arising from some interaction with another component of the cell in which the protein is synthesized. (Example for **C**: Sequences interfering with intracellular processing or transport, or generating spurious processing signals; Example for **D**: Sequences generating inappropriate enzyme target sites. [Note: this is conditional not only on the presence of an enzyme recognizing specific peptide motifs, but the efficiency of protein compartmentalization. A spurious enzyme site may not be significant if a protein bearing it is rapidly transported to an intracellular environment where it is partitioned from the enzyme itself. Also, a spurious enzymatic modification may not necessarily interfere with a protein’s function.]; Example for **E**: “Toxic” sequences which deleteriously affect host cell operations and indirectly prevent protein synthesis). **Dependent on biosynthesis at protein level:** sequences which interfere with normal peptide chain extension (Example for **F**: ‘2A’ sequences which cause ‘cleavage’ of a nascent polypeptide chain [see text below]). **Directly sequence-dependent:** a sequence which promotes protein self-processing (especially splicing or cleavage). Example for **G**: inteins). An ‘**Intrinsic**’ restriction (**H**) is a sequence incompatible with any normal protein folding or assembly (as with a very long hydrophobic homopolymer tract).

---

But in some rare cases, the synthesis of proteins *in vivo* places conditional restrictions on the use of certain peptide sequences. Comprehensive computational searches for general ‘forbidden’ pentapeptide sequences (those not found in any known proteins) have revealed many candidates<sup>45</sup>. Although

identification of a potentially excluded sequence in such a manner is not proof of any true restriction without further data, it is likely that a subset of them are biologically meaningful. 'Restricted' sequences which are potentially significant as losses to a repertoire are those which would have preserved the essential fold characteristics of a receptor but are absent (or under-represented) through some *in vivo* restriction. In other words, they are sequences which, if present, may have been potentially useful in extending the repertoire's range.

A general overview of types of peptide sequence incompatibilities is shown in Table 3A2.1. Let's consider these effects as applied towards a variable receptor type as found with the vertebrate adaptive immune system. Category A and H of this Table (sequences incompatible with specific structural features and 'intrinsically restricted' sequences, respectively) are those which are (as we have seen) not compatible with formation of a functional receptor in the first place. This too applies to Category B (sequences adversely affecting oligomeric protein assembly), although in a slightly more subtle fashion. In this case, a specific peptide sequence within a variable receptor is not necessarily directly disruptive of the binding site for the molecular target, but prevents successful higher-order association between receptor subunits (required for full binding affinity or specificity). Failure of association between specific immunoglobulin heavy and light chains, for example, has been noted as one factor potentially reducing the actual number of antibody variants below the maximal theoretical figure <sup>46</sup>.

But circumstances where specific peptide sequences might be excluded or severely under-represented can be cited, listed as 'conditional' Categories C-G of Table 3A2.1. In the subset of Categories C-E, a 'problem sequence' can potentially arise from some interaction with another component of the cellular system in which the receptor protein bearing such a sequence was synthesized. Such intermolecular events could result in abnormal processing or modification of the protein such that its normal expression is significantly impeded. As noted in Table 3A2.1, mere possession of a potentially interfering enzymatic site (for

example) may not necessarily doom such a sequence to repertoire exclusion, since rapid protein compartmentalization or export may prevent access to intracellular enzymes. Precedents for sequence-related conditional biases in repertoires do exist, however. A well-studied example is the generation of surface peptide display libraries in filamentous bacteriophage <sup>▼</sup>, where an essential step is recognition and cleavage of an N-terminal signal sequence in the phage protein used for display. This allows transport of the phage proteins into the bacterial host periplasmic space prior to secretion, but at the same time dictates a functional constraint on the expressible peptide repertoire in such systems <sup>47</sup>. Phage-displayed peptides under such circumstances thus have conditional restrictions upon their expression, in that the limitation is imposed by phage biology and not from any inherent problem from display of 'restricted' peptides *per se*. Indeed, the use of different phage which have a lytic life-cycle (and do not require secretion) can circumvent some of these constraints <sup>48</sup>.

The next Category (F) of Table 3A2.1 is wholly dependent on a specific set of protein sequences, but is conditional in the sense that the restriction is imposed at the level of eukaryotic ribosomal protein synthesis. This refers to a phenomenon only recognized relatively recently, where certain nascent protein sequences cause the termination of synthesis for the elongating amino acid residue chain at a specific point, but allow synthesis to resume from the C-terminally proximal amino acid through the rest of the C-terminal portion. First noted in the '2A' region of small RNA viruses, protein sequences in general which confer such unusual properties have become known as '2A-like' <sup>49</sup>. The end result of the 2A-mediated process is a 'cleavage' event within the polypeptide chain, although it is not proteolytic cleavage in the usual sense. It is probable that a hydrolysis event acting in *cis* (within the same polypeptide chain) occurs such that the translated N-terminal peptide segment within the ribosomal

---

<sup>▼</sup> This topic is also raised in the file SMS-CitedNotes/Section 7A; from the same ftp site.

tunnel is released from a specific tRNA<sup>gly</sup> without further peptide bond formation<sup>▼</sup><sup>50</sup>. Peptide synthesis effectively resumes at a required proline residue immediately C-terminal to this; in effect a single glycine-proline peptide bond is 'skipped' during the protein synthesis, resulting in the formation of two separate polypeptides. Efficiency of 2A-based 'cleavage' is dependent on the specific sequence involved (generally 13-18 amino acid residues), but the effect can go essentially to completion<sup>50,51</sup>. The conditional nature of the 2A-based effect is a consequence of its obligate linkage with the translational mechanism. This has become quite clear from studies with *in vitro* translation of proteins bearing inserted 2A or derivative sequences<sup>49,51,52</sup>. A protein with a 2A-like sequence which 'escapes' the translational skipping effect and emerges intact from the ribosome (if the 2A-sequence effect is <100% efficient) is stable. Moreover, it appears to depend on the specific ribosomal environment, in that expression of 2A-bearing sequences in prokaryotic systems does not result in the 'cleavage' effect<sup>52</sup>.

Much interest in the 2A-peptide effect has stemmed from its exploitation as a tool in biotechnology, for allowing multiple proteins to be potentially made from a single polypeptide sequence<sup>50</sup>. Yet with knowledge of the properties of 2A-like sequences, a logical conclusion is that the presence of a 2A-like sequence in an encoded eukaryotic protein will not be stably synthesized on eukaryotic ribosomes, but rather will result in two separate daughter polypeptides. Or to put into the context of adaptive immunity, if a 2A-like sequence embedded within a

---

▼ The ribosome contains a structural feature known as the 'tunnel' where nascent peptide chains exit during protein synthesis; residues within the 2A sequence are believed to interact with the tunnel in an unusual fashion. tRNA<sup>gly</sup> refers to a specific tRNA molecule charged with its cognate amino acid (glycine in this case; a specific *aminoacyl-tRNA*). Within the ribosome, normally a growing peptide chain is covalently linked to a peptidyl-tRNA molecule (at each successive specific mRNA codon) before being transferred (by peptide bond formation) to the appropriate aminoacyl-tRNA at the next codon site. If the peptidyl-tRNA bond is cut (hydrolysed) before the latter peptide bonding, N-terminal peptide chain growth will be stopped, as with the 2A effect.

contiguous eukaryotic immune receptor conferred a useful specificity, this would not be available to the receptor's repertoire due to the 'cleavage' effect.

The remaining restricted-peptide sequence Category (G) of Table 3A2.1 is also sequence-dependent, but not linked to translation directly. This is the special case of proteins which contain segments which are capable of self-splicing or self-cleaving. Internal sequences within proteins which can excise themselves are termed inteins, which have certain mechanistic similarities to C-terminal regions of proteins which self-cleave (C-terminal autocatalytic domains)<sup>53-55</sup>. Inteins were initially described in a yeast protein, but have now been identified in diverse forms of life<sup>54,56,57</sup>. These self-splicing protein sequences appear to have evolved to assist mobile DNA elements, as many excised inteins contain an endonuclease activity directed at sites within their encoding genes flanking the intein insertion<sup>▼</sup>. Through such specific DNA cleavage events, intein endonucleases can thereby direct transposition of the intein-encoding DNA sequence into homologous (previously intein-less) gene copies by gene conversion processes<sup>58,59</sup>. However, the protein splicing activity<sup>\*</sup> is entirely separable from intein-mediated endonuclease, either naturally<sup>62</sup> or through efforts to define a minimal intein sequence<sup>63</sup>.

It was necessary to include inteins within Table 3A2.1 in order to be comprehensive for protein sequences which are conditionally restricted from stable inclusion as part of a longer polypeptide sequence. (Rapid removal of an internal protein segment by self-splicing [or C-terminal self-cleavage] is not compatible with stable maintenance of a sequence which bears such elements). Inteins are included in the 'conditional' category in the sense that the self-splicing

---

▼ The role of inteins in homing endonucleases is considered further in the file SMS-CitedNotes-Ch3/Section 4, from the same ftp site.

\* It can be noted that protein *trans*-splicing events have been recorded, and are another potential genomic diversifier (see SMS-CitedNotes-Ch9/Section 30). This phenomenon has been implicated as in the generation of novel T cell epitopes<sup>60,61</sup>.

activity is known to be regulatable by inhibitors such as zinc ions<sup>64</sup>. As with the 2A-like sequences noted above, in principle an intein sequence would be excluded from a receptor repertoire under normal conditions. In practice, since a minimal intein sequence is >100 amino acids<sup>62,63</sup> the chance inclusion of an intein sequence in a variable receptor region is of negligible likelihood. (Here we might recall in the preceding chapter the calculation of the huge number of possible amino acid sequence combinations in a even a small protein of 100 residues).

This probability-related issue prompts a general consideration of the practical significance of 'restricted' sequences in receptor repertoires. Clearly, only relatively short potential problem sequences need be considered in this regard. It was noted above that many pentapeptide sequences are not represented in current protein databases<sup>45</sup>, and a random peptide of five residues is one of  $20^5$  ( $3.2 \times 10^6$ ) possibilities. Or we might therefore state that an arbitrary pentapeptide would be expected to occur in a random peptide string every 3.2 million residues or so. A large protein database such as [SWISS-PROT](#) contains (at present time, but constantly expanding) >495,000 sequence entries and >174 million amino acid residues. If probability was the only operating factor, any random pentapeptide would therefore have an excellent chance of being found in the known biological 'sequence space'. The absence of specific pentapeptides from such databases is then an indication that negative biases against certain sequence combinations exist, which might arise from biological restrictions as considered above. Of course, as the length of a sequence string rises, the expectations of finding it by chance go down commensurately. But even an arbitrary hexameric peptide string should be found every  $6.4 \times 10^7$  residues ( $20^6$ ), suggesting that most of the biologically permissible (non-restricted) set of hexamers are also 'out there' in known proteins, based on database sizes. Consider then a peptide string of 6 residues of chosen sequence 'INTEIN' (in the single-letter amino acid code<sup>♥</sup>), and search the protein databases. At least two

---

<sup>♥</sup> Isoleucine-asparagine-threonine-glutamic acid-isoleucine-asparagine

known proteins with this sequence turn up <sup>▼</sup>. Randomly scramble the letters (for example; 'TIIENN') and other proteins with such a sequence are found. Since the 'core' 2A-like sequence has six defined amino acid residues (DXEXNPGP; albeit operating at less than full efficiency <sup>50</sup>), it is not inconceivable that some potential vertebrate immune repertoires have certain sequences excluded on this basis. Nevertheless, once a random peptide 'window' is increased to 7 residues, the likelihood of finding such a specific sequence in known proteins declines substantially <sup>\*</sup>.

These kinds of estimations show that it is important not to overstate the case for the potential significance of 'restricted' peptide sequences in immune repertoires. The overall impact of this kind of physical limitation is likely to be minor and overshadowed in any case by other constraints such as individual-specific genetic repertoire deficiencies. Also, 'allowable' conservative amino acid replacements may (at least in some circumstances) remove a 'problem' sequence but still produce a receptor with reasonable target affinity. And yet the case of the 2A-like sequences in particular raises an interesting point. This type of effect with relatively short peptide sequences was not predicted in any way prior to its discovery, and it would have remained obscure were it not for its natural exploitation by small RNA viruses. How then, could one confidently assert that no other (as yet uncharacterized) peptide sequence motifs exist with unexpected consequences for protein processing or stability? These 'unknown problem sequences', if they exist, would also be potential repertoire limitations. Finding 'holes' in any very large set of biological sequences is a hard nut to crack

---

<sup>▼</sup> These are: Vacuolar protein sorting-associated protein 72 (yeast) and Intraflagellar transport 81 homolog (human). Thus these proteins contain an 'INTEIN' sequence but lack intein activity. (We must avoid confusion between different semantic levels, *n'est-ce pas?*).

<sup>\*</sup> Consider the heptapeptide 'PEPTIDE' (proline-glutamic acid-proline-threonine-isoleucine-aspartic acid-glutamic acid), which is not found in current databases (although 6/7 match strings are noted). Thus no known natural polypeptide sequence contains a 'PEPTIDE' sequence (watch those semantics again).

by directed experimentation (remember the old maxim, “Absence of evidence is not evidence of absence”), and the empirical approach of checking the range of natural sequences remains the best bet <sup>45,65</sup>.

Any such unrecognized constraints are likely to be conditional in one way or another, and some (as with 2A-like sequences) may depend on the biological expression system used. A logical practical extension of ‘biological conditionality’ for empirical molecular selection is that the more the process can be performed *in vitro*, or with alternative host cells, the more the chances of unexpected sequence constraints are minimized. A case in point in this regard is the generation of superior antibody affinities by phage display compared with conventional monoclonal technologies <sup>66</sup> (as noted in *Searching for Molecular Solutions*).

This discussion has proceeded from the (unproven) proposition that certain antibody (and by extension other protein recognition frameworks) might suffer functional constraints in a small minority of cases through restriction on their usage of certain amino acid sequences. But we should also recall that the natural protein alphabet itself could act as a restricting element for molecular recognition, a theme considered in some detail in *Searching for Molecular Solutions*. The modern ability to use expanded genetic codes provides an experimental platform for providing solid answers to this question. Some results already suggest modified amino acids allow the selection of better antibodies than those obtained with the natural alphabet <sup>67</sup>.

### Section A3: ***Autoimmunity***

As an extension of the self / nonself theme, this section is relevant to the 'Selfishness revisited' section of *Searching for Molecular Solutions* on p. 84, but autoimmunity as such was noted on pp. 67, 74, 81, and 83 within Chapter 3.

#### *Collateral Damage – Autoimmunity and Self-Recognition*

If the immunological distinction of self vs. non-self is not always easy to frame, neither is it a simple matter for adaptive immune systems to establish and to enforce. Yet it may seem difficult to accept that a system evolved largely to protect an organism from foreign assaults and preserve self could confuse the two. This is encapsulated in the well-known phrase of the microbiological and immunological pioneer Paul Ehrlich ♡, "horror autotoxicus", which refers not to the nasty results of self-inflicted alcoholic excess, but rather the conceptual conundrum of an organism attacking itself through its own immune system. Unfortunately, such self-attack (or autoimmunity) is a very real phenomenon, despite prolonged resistance towards its acceptance (until the 1960s) within the immunological community<sup>68</sup>. Yet at the present time, given knowledge of the exquisite complexities of adaptive immunity, if anything, it may seem more pertinent to ask why autoimmunity does not occur more often than it does.

What then is the source of Paul Ehrlich's logical error in assuming an organism could not immunologically attack itself? Firstly, one must distinguish between recognition of non-self *per se* and the ensuing action taken. Even if the recognition process is exclusively directed towards non-self molecules, some

---

♡ Paul Ehrlich (1854-1915) worked also on anti-microbial chemotherapy, as noted in the Introduction of *Searching for Molecular Solutions*. He was awarded a Nobel Prize in 1908.

immune effector responses may have non-specific host-damaging consequences. This 'self-harm' effect has been documented with invertebrate innate immunity in the face of pathogenic attack <sup>69</sup>, and over-reaction of even human immune systems struggling with viral challenges can result in the excessive production of immune mediators (a 'cytokine storm' <sup>70</sup>) and extensive self-damage. Yet this kind of 'autoreactivity' is not what is meant by true autoimmunity, since an acknowledged autoimmune response indeed involves a misfiring of the fundamental control over self / non-self recognition.

Even so, the existence of both autoreactivity and autoimmunity logically represent the evolutionary outcome of immunological cost-benefit equations, a point seemingly overlooked by Ehrlich. Insect self-harm from active innate immune responses thus appears to be an evolutionarily acceptable compromise <sup>69</sup>. In humans, severe autoreactivity problems and autoimmunity are not common, so from first principles it could be proposed that both pathologies result from low levels of random errors in the complex immune biosystem. In other words, the trade-off for a fully effective adaptive immune response might be a baseline level of autoreactive or autoimmune pathologies in some members of the species on a random basis, an admissible price for selective forces given the overall positive population fitness benefit. In this view, any adaptive immune system would have a low but finite probability of sliding into dysregulation by sheer chance, somewhat analogously to stochastic models of the generation of tumors <sup>71</sup>.

Yet in reality autoimmunity at least is heavily influenced by genetic factors. Generally speaking, since autoimmune phenomena are the outcome of the expression of many genes, a low level of pathology in some genetic backgrounds will not affect positive selection for the beneficial aspects of immunity in the population as a whole. For example, the autoimmune skeletal disease ankylosing spondylitis has a very high correlation with a specific human MHC gene allele, HLA-B27, <sup>72</sup> although details of its mechanistic basis remain unresolved. On the

other hand, observations that the incidence of autoimmune disease is never completely concordant in even genetically identical animals (including human monozygotic twins) strongly implicate additional factors in pathogenesis<sup>73</sup>. For at least some autoimmune conditions, epigenetic changes may also contribute to disease incidence<sup>74</sup>. As a case study of autoimmunity where environmental factors are believed to be significant, multiple sclerosis has an epidemiology which is consistent with viral co-factors<sup>75,76</sup>. Of course, these proposals are not mutually exclusive, and the consensus of opinion is that the pathogenesis of most (if not all) autoimmune disease is likely to result from a complex interplay of genes and environment<sup>77,78</sup>.

Although the simplistic notion of clinical autoimmunity arising solely from a chance perturbation of immune networks is thus untenable, the potential role of stochastic effects in immune system regulation in general has been noted<sup>79</sup>. Such chance-based effects, originating even at the single-cell level, could be amplified by clonal proliferation and contribute to autoimmunity in a favorable genetic and environmental background. Multiple sclerosis again can be cited as a complex example of the interplay between genes, environment and chance in the origin of Ehrlich's nightmare<sup>76,80</sup>. The 'trade-off' referred to above may still apply as an evolutionarily acceptable price to be paid for adaptive immunity, but the specific terms of the evolutionary 'bargain' are likely to be as complex as the pathogenesis of autoimmunity itself.

The role of the adaptive immune system in autoimmune phenomena has been accepted through clear-cut demonstrations of pathogenic mediators such as defined autoantibodies. In contrast, it was initially believed that innate immune systems are not linked to destructive self-reactivity<sup>81</sup>. Since innate immune mediators are genomically fixed and subject to the winnowing effects of natural selection, any such factors which were self-reactive (and thereby fitness-reducing) should be rapidly weeded out. Unfortunately for this eminently logical point of view, the real picture is not so simple, which is perhaps becoming a

familiar refrain as knowledge of the immune system extends further and further. The recognition of foreign nucleic acids by the innate immune system is a good case in point. At first it might seem a difficult problem to produce a recognition system that can distinguish nucleic acids in *general* between pathogens and their vertebrate hosts, since specific sequence motifs would tend to be limited to a narrow set of foreign organisms. Moreover, through mutation such pathogens could readily escape recognition of a specific sequence within their constituent nucleic acids. Yet there are avenues for broadly separating at least some major pathogen subsets through nucleic acid-based discrimination.

Mammalian DNA sequences with CpG tracts are typically methylated, and unmethylated CpG sequences (as found in bacterial DNAs) can trigger the innate immune system<sup>82</sup>. But at the molecular level this effect appears to operate through intracellular transport and compartmentalization, rather than direct receptor discrimination. Some of the Toll-like receptors (TLR) molecules involved with innate immunity (Chapter 3 of *Searching for Molecular Solutions*) act intracellularly. A specific Toll-like receptor molecule, TLR9, is located within intracellular endosomes (membrane-bound compartments derived from the outer plasma membrane<sup>83,84</sup>, and binds unmethylated CpG motifs<sup>82,85</sup>). An interesting observation has been that if short DNA segments are provided artificial passage into the endosomal compartment through chemical means, TLR9 nucleic acid selectivity is no longer observed<sup>86</sup>. In other words, it appears that the discrimination by this innate receptor molecule between bacterial and host DNAs is not an inherent feature of direct molecular recognition, but rather lies in differential cellular processing and transport of host vs. pathogen DNA fragments<sup>86,87</sup>. The TLR9 innate receptor thus is capable of responding to self-nucleic acids but normally does not, simply because it is denied access to them.

---

▼ Host nucleic acids are normally kept sequestered from endosomes (as with nuclear DNA within chromatin) or are rapidly degraded<sup>86,87</sup>.

Under normal circumstances, certain receptors of the innate immune system may thus be physically prevented from interacting with self nucleic acids (which they would otherwise bind), but still allowed access to foreign DNA or RNA.

Accordingly, such receptors can provide innate pathogen sensing without fitness loss in normal individuals, but recognition of self-nucleic acids or nucleic acid-protein complexes remains a potential pathway towards self-reactivity in susceptible members of the species. Some systemic autoimmune diseases (most notably systemic lupus erythematosus) are in part characterized by autoantibodies against ribonucleoprotein complexes<sup>88,89</sup> or chromatin components<sup>90,91</sup>. Co-signaling of (otherwise quiescent) self-reactive B cells through their antigen receptors (surface immunoglobulin) and TLR9<sup>92</sup> or TLR7<sup>93</sup> leads to direct B cell activation and potential autoimmunity<sup>87</sup>. The evolutionary retention of genomically-encoded innate receptors with potential autoreactivity will be a balance between the population fitness benefit from nucleic-acid based pathogen recognition and fitness loss due to the development of autoimmunity in a minority of individuals. It is clear, given the ancient ties between innate and adaptive immunity, that a low background incidence of pathologic autoreactivity through some innate receptors is an evolutionarily acceptable compromise.

Unlike innate immunity, adaptive immune responses by their nature produce somatic novelty, which can be inherently dangerous if not tightly regulated. And sometimes such regulation too falls short. As noted briefly in *Searching for Molecular Solutions*, self / non-self regulation operates at the T cell level by central tolerance in the thymus (clonal deletion) and active regulatory mechanisms in the periphery. Self-reactive B cells are also controlled or fail to receive help from absent self-reactive T cells. Yet these findings implicitly carry the information that potential self-reactivity is ever-present in normal individuals, and indeed a wealth of experimental data has shown that in many circumstances it is possible to 'break tolerance' against self-antigens<sup>94-96</sup>. To go into much detail here is beyond the scope of this necessarily short diversion into

autoimmunity, but a brief review of a situation where tolerance-breaking is actually a *good* thing will illustrate a few relevant points.

In the context of defining self and non-self, it is interesting to consider tumor-associated antigens. While some such molecules are aberrantly produced and unique to tumors, many others are in fact encoded by the normal genome. In the latter case, in order to use such antigens for recognition of tumors it is consequently necessary to mount an immune response against what is at least nominally self. Some 'normal' tumor antigens are only expressed in a very limited cellular and/or developmental compartment, and by such 'sequestration' may have escaped the usual tolerance mechanisms. An example of these are the X-chromosome-linked 'cancer testis antigens', which are only normally expressed in testicular germ cells and placental tissue but are commonly reactivated in tumors<sup>97</sup>. In other cases, tumor-associated antigens correspond to differentiation antigens of the specific cell lineage involved, such as a number of characterized melanoma antigens. Peptide fragments of these proteins found in both normal melanocytes and melanomas are presented to cytotoxic T cells by Class I MHC molecules<sup>98</sup>, and responses can be demonstrated in a majority of late-stage melanomas despite disease progression<sup>99</sup>. Clearly, self-reactive T cells exist initially for this to occur, and can be expanded by presentation of antigen, but appropriate co-activation of the innate immune system is needed for strong immunity to occur<sup>100,101</sup>.

Attempts to induce anti-tumor immunity against tumor differentiation antigens thus represent an inverse therapeutic situation to autoimmunity, where the immune response itself is the problem. Indeed, with melanoma as the example again, an effectively engineered response against melanocytic antigens can

target normal melanocytes as well as melanoma cells<sup>▼ 102</sup>, producing a depigmenting autoimmune reaction called vitiligo. Conversely, T cells from natural cases of vitiligo have also been considered as useful reagents for anti-melanoma therapy<sup>103</sup>.

In *Searching for Molecular Solutions*, biosystems for recognizing and detoxifying potentially noxious environmental chemicals were collectively referred to as the xenobiotic 'immune system'. With receptors that are germline-derived and not somatically variable, xenobiotic recognition and processing indeed has certain parallels with innate immune systems of diverse organisms. The evolutionary 'logic' for the development of a system for detecting and clearing dangerous chemical input from the environment is compelling, and in turn provides an obvious fitness benefit. But another interesting parallel between xenobiotic and conventional immune systems arises from the observation that at times, the biological interface with xenobiotic challenges may actually be counterproductive, resulting in autotoxic effects which would not otherwise have occurred. A case in point is neurodegeneration caused by specific toxic chemicals, which reproduce the effects of Parkinson's disease. Xenobiotic enzyme action on certain pyridine substituents can convert them into N-methylated pyridines, known toxins for dopaminergic neurones whose ablation triggers Parkinsonian symptoms<sup>104</sup>. Once again, a 'good' system can have strongly negative effects under some circumstances, emphasising once more that all bioprotective systems are evolutionary products which act on a balance-sheet of factors towards the greatest selective fitness benefit. And a certain level of collateral damage is an inevitable by-product.

---

▼ If an autoimmune reaction accompanying anti-tumor immunotherapy damages vital tissues, the treatment approach is not useful. In the case of melanoma, a co-reaction against melanocytes leading to depigmentation may be disfiguring, but is not life threatening in the same way as the melanoma tumor itself. It is therefore considered the lesser of two evils if it occurs during immunotherapy. Immunotherapeutic approaches in general have not yet reached a level of success where they are used as standard treatments.

We should not take these kinds of comparisons too far, though. Autotoxic failures of the xenobiotic immune system appear to be much simpler in concept than the generation of 'horror autotoxicus' in the adaptive immune system. To fully understand the latter is to understand all facets of immune regulation itself, which we cannot yet claim. As an example, only within the last few years has an entire subclass of T helper cells (the T<sub>H</sub>17 subset) been recognized, which has profound significance in the generation of autoimmunity<sup>105,106</sup>. So there is clearly much still to be learned, and this huge topic is beyond our present scope.

At the end of this short tour of the dark side of the adaptive immune system, are there any implications here for molecular design? Perhaps not for the immediate future, but a thorough study of the cellular and molecular perturbations which culminate in autoimmunity may well have ramifications for the understanding and design of higher-level molecular systems in general. One finding (discussed above) which is of interest in this regard is actually most directly concerned with the innate immune system, although its relevance to autoimmunity is clear.

This refers to recent work indicating indirect regulation of signal discrimination of certain nucleic acid-sensing Toll-like receptors, involving differential processing and transport for foreign vs. self nucleic acids<sup>86</sup>. This control mechanism can be considered as a paradigm for a particular kind of supra-molecular two-tiered (or multi-tiered) recognition system. In this sort of arrangement, recognition is not solely the function of a single receptor but is relegated between an end-receptor (for signaling) and primary screening mechanisms. In the biological example of Toll-like receptor (specifically TLR9) nucleic acid recognition, the 'primary screening' is natural sequestration or differential nuclease action towards self-nucleic acids compared with non-self DNA or RNA. In principle, one could envisage a screening system (combined with suitable molecular compartmentalization) where one (or more) 'pre-receptors' could 'gate' and disqualify potential non-specific molecules which might interfere with proper

binding of the desired ligand to the end-receptor. Such a composite approach to recognition could avoid having to compromise the function of a specialized receptor which may be suited for some ligands but not ideal for others. It is also conceptually reminiscent of the evolutionary subfunctionalization of protein function following gene duplication, as discussed in the Chapter 2 of *Searching for Molecular Solutions*.

#### Section A4: ***Immune / Nervous System Comparisons***

Relevant to the Section 'GOD in the Brain and its Extensions', beginning on p. 91 of *Searching for Molecular Solutions*.

##### *A Tale of Two Systems*

We have already encountered pathological interaction between the immune and nervous systems, in the example of the autoimmune condition multiple sclerosis ([Section A3](#) above). In contrast to such aberrations, a large body of evidence exists for normal cross-communication between immunological and neurological signaling, as many neurological mediators also impact upon the immune system<sup>107-110</sup>), and physical interfacing between the immune system and the sympathetic nervous system has been defined<sup>111</sup>. Perhaps this is not so surprising when we consider that functional immune systems must integrate successfully into the functioning of organisms as a whole. Nevertheless, it is not the observation that neural and immune systems 'talk' to each other that is of primary interest for our purposes, but how much their higher-level functions have in common and where they diverge. In this context itself, respective mechanisms for somatic diversification are the main focus.

Some parallels that one could draw between the immune and higher nervous systems, such as their shared high-level complexities, are true but not very useful without further information. It might also be noted that (apart from the above-mentioned signaling molecules involved in cross-talk), within both of these physiological systems, many molecular mediators are shared. For example, the

calcium-dependent protein phosphatase ♡ calcineurin is (as its name implies) important in certain neurological capacities, including synaptic regulation and memory <sup>112,113</sup>. Calcineurin is, however, also required by the immune system for regulation of T cell-specific transcription factors <sup>114,115</sup>, which in turn affect thymic T cell repertoire control <sup>116</sup>. But such apparent similarities can be deceptive, at least in the absence of additional data. Merely showing that two biological systems (within the same organism) share molecular factors does not in itself necessarily point towards higher-level similarities of the systems in question, since the use of common components in diverse functional settings is a widespread phenomenon \*. This again reflects genomic parsimony (referred to in Chapter 9 of *Searching for Molecular Solutions*), and is one of the reasons why ~20,000 genes can direct the development and functioning of highly complex life forms. But note the qualification regarding ‘additional data’. Shared molecular components between major physiological systems *may* indicate meaningful common higher-level function, but this can only be ascertained from further careful biological studies.

With this in mind, it is interesting to note that is a profitable exercise to evaluate mice for neurological changes following artificial ‘knock-out’ of many genes <sup>112</sup>. Many such ‘knock-out’ mouse lines were established for the study of specific genes in non-neurological contexts, but co-expression of such genes in neural tissues may render them (and the corresponding knock-out mice) of neurological significance as well. Another interesting case in point in this regard (which is still not fully resolved) is the observation that one of the essential genes for immunoglobulin and T cell receptor gene rearrangements (RAG-1 ^) is expressed at low levels in mouse brain, but mice with the RAG-1 gene ‘knocked out’ artificially show profound immunological but no apparent neural changes <sup>117</sup>.

---

♡ Protein phosphatases specifically remove phosphate groups from proteins, thus reversing the effects of phosphate-transferring kinase enzymes.

\* This theme is discussed in more detail in the file SMS–Extras–Ch9/Section A15.

^ The RAG genes and proteins are considered in the above Additional Material section [A1](#).

Since RAG-2 expression was not found in parallel with RAG-1, it is possible that RAG-1 has a different function in the murine central nervous system.

<b>Table 3A4.1</b>			
<b>Similarities between immune and nervous systems</b>			
<b>Shared or Analogous Attribute</b>	<b>Immune System</b>		<b>Central / Peripheral Nervous System / Brain</b>
<b>1. Memory mechanism</b>	Generation of long-lived T and B memory cells		Neurological memory
<b>2. Learning mechanism</b>	'Education' of T cells in the thymus		Learning by synaptic potentiation
<b>3. Cell-cell information transfer</b>	Immunological Synapses		Neurological Synapses
<b>4. High levels of apoptosis during development</b>	During T cell development and immune peripheral selection		During central nervous system development
<b>5. Self / Non-self discrimination</b>	<b>1</b>	Thymic education of T cells *	<b>A</b> Olfactory discrimination of kin / non-kin MHC
	<b>2</b>	Active peripheral control by regulatory T cells	<b>B</b> Specificity of neuronal wiring
	<b>3</b>	B cell self-recognition control mechanisms	<b>C</b> Higher-level self-recognition

**Footnotes to Table 3A4.1:**

\* Thymic 'education' includes both tuning the T cell repertoire towards MHC recognition and removal of self-reactive clones.

---

Apart from these considerations, some direct analogies between the brain / nervous system and immune systems do exist, and have been noted for a considerable time <sup>118</sup>. Five such categories of cross-system parallels are listed in Table 3A4.1. Some of these, such as memory and immunological / neurological synapses were noted in Chapter 3 of *Searching for Molecular Solutions*. These effects in both systems can be considered as specific instances of cell-based information transfer. For immune systems, this is usually a paired cell-cell arrangement, but the neural system involves a network of cells interconnecting via synapses, with the additional involvement of electrochemical signaling. Neural signaling thus involves a relay system in pre-determined network patterns. At each synapse, molecular recognition events between neurotransmitters and receptors occur. It might be considered that a primordial nervous system is likely to have evolved from cell-cell signaling involving specific molecular recognition events, as occurs in the immune systems (with the electrochemical processes evolving later).

Both immunological and neural systems make use of learning-based processes. Neurological learning is obvious enough on the scale of a whole animal, but is believed to be associated at the cellular level with long-term potentiation of neuronal synaptic activity <sup>♥</sup>. The fourth attribute of Table 3A4.1 refers to the high level of programmed cell death (apoptosis) seen during the development of both immune and nervous systems. This is certainly the case in the context of T cell

---

<sup>♥</sup> Long Term Potentiation is a mechanism whereby repeated neuronal activation during learning results in strengthening of specific neuronal synaptic connections <sup>112</sup>.

development, but extensive cell death is also a known feature of neural development <sup>119</sup>.

The fundamental importance of self / non-self recognition was made repeatedly in Chapter 3, and we can draw some parallels here with the brain and central nervous system (attribute 5 of Table 3A4.1). These have been subdivided into three subcategories in each case, and the neural instances span different levels of organization. The first of the latter (5A of Table 3A4.1) is analogous with self / non-self recognition, though at the level of whole organisms. This refers to the ability of animals to use olfactory perception to distinguish closely related individuals from those of the same species but with no kinship, and to thereby direct mating preferences towards outbreeding. It is thus a special case of allorecognition (the ability to distinguish a genetically distinct member of one's own species), which was noted in an immunological sense in Chapter 3. Behavioral channeling of mate-selection has two ramifications which improve fitness: the avoidance of inbreeding <sup>▼</sup> and its amplification of the effects of deleterious genes, and the maximization of MHC diversity. In fact, it has been recently found that the basis for this kin / non-kin discrimination is olfactory detection of non-self MHC peptides <sup>120-122</sup>.

From the level of the whole organism, we drop down to at the level of neurons themselves with the second subcategory B (within attribute 5 of Table 3A4.1). During the development of neural circuitry, neurons extend processes (dendrites) which distinguish between those of a common cellular origin (self-dendrites at the cellular level) and others. This is required to avoid crossing of self-dendritic projections, and to thereby maintain the integrity of neural signaling. In turn, this necessitates a highly diverse process for recognition specificity. It has been found in the fruitfly *Drosophila* that expression of alternative forms of a cell adhesion molecule (DSCAM; capable of very high diversity by means of alternative splicing at the mRNA level <sup>123</sup>) mediates this 'self-avoidance'

---

<sup>▼</sup> Thus a natural 'incest repellent' of sorts.

phenomenon<sup>124,125</sup>. Vertebrates also are faced with this type of neural self-recognition dictate, but are believed to solve it through other forms of molecular diversity<sup>126</sup>.

A third level (subcategory C; Table 3A4.1) of 'self / non-self discrimination' can be noted at a higher level of neurological functioning. Organisms beyond a certain threshold of neural development can cognitively distinguish themselves from others (as with a mirror test). It is notable that self / non-self recognition is an important area within robotics and artificial intelligence development<sup>127</sup>.

As emphasized in *Searching for Molecular Solutions*, certain processes within the adaptive immune system have a distinctly Darwinian flavor, by exhibiting functional selection from a large pool of alternatives, and amplification of selected cellular clones. Parallels have also been drawn between Darwinian processes and the development of the brain, most notably by Gerald Edelman<sup>128,129</sup>. The theory of Neural Darwinism (or Neural Group Selection) postulates that specific populations of neural synapses are somatically positively selected (by synaptic potentiation mechanisms) in response to individual experience. In this view, 'working circuitry' is thus selected based on its beneficial adaptive properties for an organism's behavior. Principles of this theory have also been applied in the design of real-world automata (the 'Darwin series'; Darwin I-X<sup>130-132</sup>). The problem, in the opinion of some observers, has been the labeling of the theory as 'Darwinian' more so than the details of the proposal itself<sup>133,134</sup>, since the neural synaptic groups which undergo positive selection do not replicate. (Francis Crick wryly termed the theory 'Neural Edelmanism'<sup>133</sup>). So this theory of neural functional selection, irrespective of its virtues or deficiencies, does not appear to have a strongly analogous counterpart in the operation of the adaptive immune system (and was not included in Table 3A4.1 for this reason). Nevertheless, meaningful comparisons between immune and neural systems,

---

<sup>129</sup> Himself a Nobel prize winner for the structure of immunoglobulins (1972).

and the above-mentioned expression of RAG-1 in murine neurons, raise the question of diversification mechanisms in the brain and central nervous system. This topic was alluded to in Searching for Molecular Solutions, and extended in the file SMS-CitedNotes-Ch3/Section 5.

## References:

1. Schatz, D. G., Oettinger, M. A. & Baltimore, D. The V(D)J recombination activating gene, RAG-1. *Cell* **59**, 1035-48 (1989).
2. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517-23 (1990).
3. Shinkai, Y. et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855-67 (1992).
4. Mombaerts, P. et al. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869-77 (1992).
5. Rooney, S., Chaudhuri, J. & Alt, F. W. The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol Rev* **200**, 115-31 (2004).
6. Grawunder, U. & Harfst, E. How to make ends meet in V(D)J recombination. *Curr Opin Immunol* **13**, 186-94 (2001).
7. Jung, D. & Alt, F. W. Unraveling V(D)J recombination; insights into gene regulation. *Cell* **116**, 299-311 (2004).
8. Schatz, D. G. & Spanopoulou, E. Biochemistry of V(D)J recombination. *Curr Top Microbiol Immunol* **290**, 49-85 (2005).
9. Schatz, D. G. Antigen receptor genes and the evolution of a recombinase. *Semin Immunol* **16**, 245-56 (2004).
10. Pancer, Z. et al. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430**, 174-80 (2004).
11. Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* **124**, 815-22 (2006).
12. Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744-51 (1998).
13. Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463-70 (1998).
14. Rothenberg, E. V. & Pant, R. Origins of lymphocyte developmental programs: transcription factor evidence. *Semin Immunol* **16**, 227-38 (2004).
15. Miller, W. J., McDonald, J. F. & Pinsky, W. Molecular domestication of mobile elements. *Genetica* **100**, 261-70 (1997).
16. Hughes, A. L. Genomic catastrophism and the origin of vertebrate immunity. *Arch Immunol Ther Exp (Warsz)* **47**, 347-53 (1999).

17. Fugmann, S. D., Messier, C., Novack, L. A., Cameron, R. A. & Rast, J. P. An ancient evolutionary origin of the Rag1/2 gene locus. *Proc Natl Acad Sci U S A* **103**, 3728-33 (2006).
18. Jones, J. M. & Gellert, M. The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev* **200**, 233-48 (2004).
19. Boehm, T. Quality control in self/nonself discrimination. *Cell* **125**, 845-58 (2006).
20. Verkoczy, L. K., Martensson, A. S. & Nemazee, D. The scope of receptor editing and its association with autoimmunity. *Curr Opin Immunol* **16**, 808-14 (2004).
21. Youinou, P. et al. B lymphocytes on the front line of autoimmunity. *Autoimmun Rev* **5**, 215-21 (2006).
22. Zhang, Z., Burrows, P. D. & Cooper, M. D. The molecular basis and biological significance of VH replacement. *Immunol Rev* **197**, 231-42 (2004).
23. Cascalho, M., Ma, A., Lee, S., Masat, L. & Wabl, M. A quasi-monoclonal mouse. *Science* **272**, 1649-52 (1996).
24. Koralov, S. B., Novobrantseva, T. I., Konigsmann, J., Ehlich, A. & Rajewsky, K. Antibody repertoires generated by VH replacement and direct VH to JH joining. *Immunity* **25**, 43-53 (2006).
25. Lopez-Macias, C. et al. Secondary rearrangements and hypermutation generate sufficient B cell diversity to mount protective antiviral immunoglobulin responses. *J Exp Med* **189**, 1791-8 (1999).
26. Cascalho, M. et al. A mouse with a monoclonal primary immunoglobulin repertoire not further diversified by V-gene replacement. *Dev Immunol* **7**, 43-50 (1999).
27. Darlow, J. M., Farrell, A. M. & Stott, D. I. Non-functional immunoglobulin G transcripts in a case of hyper-immunoglobulin M syndrome similar to type 4. *Immunology* **111**, 212-22 (2004).
28. Darlow, J. M. & Stott, D. I. V(H) replacement in rearranged immunoglobulin genes. *Immunology* **114**, 155-65 (2005).
29. Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* **197**, 1291-6 (2003).
30. Pham, P., Bransteitter, R., Petruska, J. & Goodman, M. F. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* **424**, 103-7 (2003).
31. Honjo, T., Nagaoka, H., Shinkura, R. & Muramatsu, M. AID to overcome the limitations of genomic information. *Nat Immunol* **6**, 655-61 (2005).
32. Neuberger, M. S. & Rada, C. Somatic hypermutation: activation-induced deaminase for C/G followed by polymerase eta for A/T. *J Exp Med* **204**, 7-10 (2007).

33. Papavasiliou, F. N. & Schatz, D. G. Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. *Cell* **109 Suppl**, S35-44 (2002).
34. Yuan, D., Witte, P. L., Tan, J., Hawley, J. & Dang, T. Regulation of IgM and IgD heavy chain gene expression: effect of abrogation of intergenic transcriptional termination. *J Immunol* **157**, 2073-81 (1996).
35. Dudley, D. D., Chaudhuri, J., Bassing, C. H. & Alt, F. W. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv Immunol* **86**, 43-112 (2005).
36. Leach, D. R. Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* **16**, 893-900 (1994).
37. Courey, A. J. & Wang, J. C. Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. *Cell* **33**, 817-29 (1983).
38. Britten, R. J. Forbidden synonymous substitutions in coding regions. *Mol Biol Evol* **10**, 205-20 (1993).
39. Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D. & Scharpe, S. Proline motifs in peptides and their biological processing. *Faseb J* **9**, 736-44 (1995).
40. Tendulkar, A. V., Ogunnaike, B. & Wangikar, P. P. Gaussian mixture modeling of alpha-helix subclasses: structure and sequence variations. *Pac Symp Biocomput*, 291-302 (2006).
41. Serrano, L., Neira, J. L., Sancho, J. & Fersht, A. R. Effect of alanine versus glycine in alpha-helices on protein stability. *Nature* **356**, 453-5 (1992).
42. Sahr, K. E. et al. Spectrin cagliari. an Ala-->Gly substitution in helix 1 of beta spectrin repeat 17 that severely disrupts the structure and self-association of the erythrocyte spectrin heterodimer. *J Biol Chem* **268**, 22656-62 (1993).
43. Lopez-Llano, J., Campos, L. A. & Sancho, J. Alpha-helix stabilization by alanine relative to glycine: roles of polar and apolar solvent exposures and of backbone entropy. *Proteins* **64**, 769-78 (2006).
44. Green, H. & Wang, N. Codon reiteration and the evolution of proteins. *Proc Natl Acad Sci U S A* **91**, 4298-302 (1994).
45. Tuller, T., Chor, B. & Nelson, N. Forbidden penta-peptides. *Protein Sci* **16**, 2251-9 (2007).
46. Janeway, C. A., Jr., Travers, P., Walport, M. & Shlomchik, M. *Immunobiology. The Immune System in Health and Disease* (Garland Science Publishers, 2004).
47. Rodi, D. J., Soares, A. S. & Makowski, L. Quantitative assessment of peptide sequence diversity in M13 combinatorial peptide phage display libraries. *J Mol Biol* **322**, 1039-52 (2002).

48. Krumpe, L. R. et al. T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries. *Proteomics* **6**, 4210-22 (2006).
49. Donnelly, M. L. et al. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J Gen Virol* **82**, 1013-25 (2001).
50. de Felipe, P. et al. E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends Biotechnol* **24**, 68-75 (2006).
51. Ryan, M. D. & Drew, J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *Embo J* **13**, 928-33 (1994).
52. Donnelly, M. L., Gani, D., Flint, M., Monaghan, S. & Ryan, M. D. The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J Gen Virol* **78 ( Pt 1)**, 13-21 (1997).
53. Xu, M. Q. & Perler, F. B. The mechanism of protein splicing and its modulation by mutation. *Embo J* **15**, 5146-53 (1996).
54. Pietrokovski, S. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci* **7**, 64-71 (1998).
55. Anraku, Y., Mizutani, R. & Satow, Y. Protein splicing: its discovery and structural insight into novel chemical mechanisms. *IUBMB Life* **57**, 563-74 (2005).
56. Pietrokovski, S. Intein spread and extinction in evolution. *Trends Genet* **17**, 465-72 (2001).
57. Perler, F. B. InBase: the Intein Database. *Nucleic Acids Res* **30**, 383-4 (2002).
58. Gimble, F. S. & Thorner, J. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature* **357**, 301-6 (1992).
59. Liu, X. Q. Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu Rev Genet* **34**, 61-76 (2000).
60. Hanada, K., Yewdell, J. W. & Yang, J. C. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* **427**, 252-6 (2004).
61. Vigneron, N. et al. An antigenic peptide produced by peptide splicing in the proteasome. *Science* **304**, 587-90 (2004).
62. Telenti, A. et al. The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. *J Bacteriol* **179**, 6378-82 (1997).
63. Derbyshire, V. et al. Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proc Natl Acad Sci U S A* **94**, 11466-71 (1997).
64. Mills, K. V. & Paulus, H. Reversible inhibition of protein splicing by zinc ion. *J Biol Chem* **276**, 10832-8 (2001).
65. Hampikian, G. & Andersen, T. Absent sequences: nullomers and primes. *Pac Symp Biocomput*, 355-66 (2007).

66. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A* **97**, 10701-5 (2000).
67. Liu, C. C. et al. Protein evolution with an expanded genetic code. *Proc Natl Acad Sci U S A* **105**, 17688-93 (2008).
68. Silverstein, A. M. Autoimmunity versus horror autotoxicus: the struggle for recognition. *Nat Immunol* **2**, 279-81 (2001).
69. Sadd, B. M. & Siva-Jothy, M. T. Self-harm caused by an insect's innate immunity. *Proc Biol Sci* **273**, 2571-4 (2006).
70. Snelgrove, R., Williams, A., Thorpe, C. & Hussell, T. Manipulation of immunity to and pathology of respiratory infections. *Expert Rev Anti Infect Ther* **2**, 413-26 (2004).
71. Spencer, S. L., Gerety, R. A., Pienta, K. J. & Forrest, S. Modeling somatic evolution in tumorigenesis. *PLoS Comput Biol* **2**, e108 (2006).
72. Thorsby, E. & Lie, B. A. HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. *Transpl Immunol* **14**, 175-82 (2005).
73. Winchester, R. The genetics of autoimmune-mediated rheumatic diseases: clinical and biologic implications. *Rheum Dis Clin North Am* **30**, 213-27, viii (2004).
74. Ballestar, E., Esteller, M. & Richardson, B. C. The epigenetic face of systemic lupus erythematosus. *J Immunol* **176**, 7143-7 (2006).
75. Kantarci, O. & Wingerchuk, D. Epidemiology and natural history of multiple sclerosis: new insights. *Curr Opin Neurol* **19**, 248-54 (2006).
76. Giovannoni, G. & Ebers, G. Multiple sclerosis: the environment and causation. *Curr Opin Neurol* **20**, 261-8 (2007).
77. Wandstrat, A. & Wakeland, E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat Immunol* **2**, 802-9 (2001).
78. Mackay, I. R. The etiopathogenesis of autoimmunity. *Semin Liver Dis* **25**, 239-50 (2005).
79. Germain, R. N. The art of the probable: system control in the adaptive immune system. *Science* **293**, 240-5 (2001).
80. Dyment, D. A., Ebers, G. C. & Sadovnick, A. D. Genetics of multiple sclerosis. *Lancet Neurol* **3**, 104-10 (2004).
81. Medzhitov, R. & Janeway, C. A., Jr. How does the immune system distinguish self from nonself? *Semin Immunol* **12**, 185-8; discussion 257-344 (2000).
82. Krieg, A. M. CpG motifs in bacterial DNA and their immune effects. *Annu Rev Immunol* **20**, 709-60 (2002).
83. Akira, S. TLR signaling. *Curr Top Microbiol Immunol* **311**, 1-16 (2006).
84. Sioud, M. Innate sensing of self and non-self RNAs by Toll-like receptors. *Trends Mol Med* **12**, 167-76 (2006).

85. Janeway, C. A., Jr. & Medzhitov, R. Innate immune recognition. *Annu Rev Immunol* **20**, 197-216 (2002).
86. Yasuda, K. et al. Endosomal translocation of vertebrate DNA activates dendritic cells via TLR9-dependent and -independent pathways. *J Immunol* **174**, 6129-36 (2005).
87. Martin, D. A. & Elkon, K. B. Autoantibodies make a U-turn: the toll hypothesis for autoantibody specificity. *J Exp Med* **202**, 1465-9 (2005).
88. Hardin, J. A. & Mimori, T. Autoantibodies to ribonucleoproteins. *Clin Rheum Dis* **11**, 485-505 (1985).
89. Routsias, J. G., Vlachoyiannopoulos, P. G. & Tzioufas, A. G. Autoantibodies to intracellular autoantigens and their B-cell epitopes: molecular probes to study the autoimmune response. *Crit Rev Clin Lab Sci* **43**, 203-48 (2006).
90. Burlingame, R. W. & Cervera, R. Anti-chromatin (anti-nucleosome) autoantibodies. *Autoimmun Rev* **1**, 321-8 (2002).
91. Gomez-Puerta, J. A., Burlingame, R. W. & Cervera, R. Anti-chromatin (anti-nucleosome) antibodies. *Lupus* **15**, 408-11 (2006).
92. Leadbetter, E. A. et al. Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. *Nature* **416**, 603-7 (2002).
93. Lau, C. M. et al. RNA-associated autoantigens activate B cells by combined B cell antigen receptor/Toll-like receptor 7 engagement. *J Exp Med* **202**, 1171-7 (2005).
94. von Herrath, M. G. & Oldstone, M. B. Virus-induced autoimmune disease. *Curr Opin Immunol* **8**, 878-85 (1996).
95. Ohashi, P. S. & DeFranco, A. L. Making and breaking tolerance. *Curr Opin Immunol* **14**, 744-59 (2002).
96. Zehn, D. & Bevan, M. J. T cells with low avidity for a tissue-restricted antigen routinely evade central and peripheral tolerance and cause autoimmunity. *Immunity* **25**, 261-70 (2006).
97. Simpson, A. J., Caballero, O. L., Jungbluth, A., Chen, Y. T. & Old, L. J. Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* **5**, 615-25 (2005).
98. Engelhard, V. H., Bullock, T. N., Colella, T. A., Sheasley, S. L. & Mullins, D. W. Antigens derived from melanocyte differentiation proteins: self-tolerance, autoimmunity, and use for cancer immunotherapy. *Immunol Rev* **188**, 136-46 (2002).
99. Rivoltini, L. et al. Immunity to cancer: attack and escape in T lymphocyte-tumor cell interaction. *Immunol Rev* **188**, 97-113 (2002).
100. Sanchez-Perez, L. et al. Killing of normal melanocytes, combined with heat shock protein 70 and CD40L expression, cures large established melanomas. *J Immunol* **177**, 4168-77 (2006).

101. Tormo, D. et al. Therapeutic efficacy of antigen-specific vaccination and toll-like receptor stimulation against established transplanted and autochthonous melanoma in mice. *Cancer Res* **66**, 5427-35 (2006).
102. Yee, C. et al. Melanocyte destruction after antigen-specific immunotherapy of melanoma: direct evidence of t cell-mediated vitiligo. *J Exp Med* **192**, 1637-44 (2000).
103. Oyarbide-Valencia, K. et al. Therapeutic implications of autoimmune vitiligo T cells. *Autoimmun Rev* **5**, 486-92 (2006).
104. Williams, A. C., Cartwright, L. S. & Ramsden, D. B. Parkinson's disease: the first common neurological disease due to auto-intoxication? *Qjm* **98**, 215-26 (2005).
105. Chen, Z. & O'Shea, J. J. Th17 cells: a new fate for differentiating helper T cells. *Immunol Res* **41**, 87-102 (2008).
106. Korn, T., Bettelli, E., Oukka, M. & Kuchroo, V. K. IL-17 and Th17 Cells. *Annu Rev Immunol* **27**, 485-517 (2009).
107. Dunzendorfer, S. & Wiedermann, C. J. Neuropeptides and the immune system: focus on dendritic cells. *Crit Rev Immunol* **21**, 523-57 (2001).
108. Wrona, D. Neural-immune interactions: an integrative view of the bidirectional relationship between the brain and immune systems. *J Neuroimmunol* **172**, 38-58 (2006).
109. McCann, S. M., De Laurentiis, A. & Rettori, V. Chronology of advances in neuroendocrine immunomodulation. *Ann N Y Acad Sci* **1088**, 1-11 (2006).
110. Gonzalez-Rey, E., Varela, N., Chorny, A. & Delgado, M. Therapeutical approaches of vasoactive intestinal peptide as a pleiotropic immunomodulator. *Curr Pharm Des* **13**, 1113-39 (2007).
111. Elenkov, I. J., Wilder, R. L., Chrousos, G. P. & Vizi, E. S. The sympathetic nerve--an integrative interface between two supersystems: the brain and the immune system. *Pharmacol Rev* **52**, 595-638 (2000).
112. Matynia, A., Kushner, S. A. & Silva, A. J. Genetic approaches to molecular and cellular cognition: a focus on LTP and learning and memory. *Annu Rev Genet* **36**, 687-720 (2002).
113. Sharma, S. K., Bagnall, M. W., Sutton, M. A. & Carew, T. J. Inhibition of calcineurin facilitates the induction of memory for sensitization in *Aplysia*: requirement of mitogen-activated protein kinase. *Proc Natl Acad Sci U S A* **100**, 4861-6 (2003).
114. Macian, F. NFAT proteins: key regulators of T-cell development and function. *Nat Rev Immunol* **5**, 472-84 (2005).
115. Lee, M. & Park, J. Regulation of NFAT activation: a potential therapeutic target for immunosuppression. *Mol Cells* **22**, 1-7 (2006).
116. Starr, T. K., Jameson, S. C. & Hogquist, K. A. Positive and negative selection of T cells. *Annu Rev Immunol* **21**, 139-76 (2003).

117. Chun, J. J., Schatz, D. G., Oettinger, M. A., Jaenisch, R. & Baltimore, D. The recombination activating gene-1 (RAG-1) transcript is present in the murine central nervous system. *Cell* **64**, 189-200 (1991).
118. Jerne, N. K. Towards a network theory of the immune system. *Ann Immunol (Paris)* **125C**, 373-89 (1974).
119. Blaschke, A. J., Staley, K. & Chun, J. Widespread programmed cell death in proliferative and postmitotic regions of the fetal cerebral cortex. *Development* **122**, 1165-74 (1996).
120. Leinders-Zufall, T. et al. MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**, 1033-7 (2004).
121. Slev, P. R., Nelson, A. C. & Potts, W. K. Sensory neurons with MHC-like peptide binding properties: disease consequences. *Curr Opin Immunol* **18**, 608-16 (2006).
122. Boehm, T. & Zufall, F. MHC peptides and the sensory evaluation of genotype. *Trends Neurosci* **29**, 100-7 (2006).
123. Schmucker, D. et al. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-84 (2000).
124. Soba, P. et al. Drosophila sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. *Neuron* **54**, 403-16 (2007).
125. Matthews, B. J. et al. Dendrite self-avoidance is controlled by Dscam. *Cell* **129**, 593-604 (2007).
126. Zipursky, S. L., Wojtowicz, W. M. & Hattori, D. Got diversity? Wiring the fly brain with Dscam. *Trends Biochem Sci* **31**, 581-8 (2006).
127. Biever, C. The Robots with a Sense of Self. *New Scientist* **194**, 30-31 (2007).
128. Edelman, G. *Neural Darwinism. The Theory of Neuronal Group Selection* (Basic Books., 1987).
129. Edelman, G. M. Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron* **10**, 115-25 (1993).
130. Reeke, G., Jnr, Sporn, O. & Edelman, G. M. Synthetic neural modeling: The "Darwin" series of recognition automata. *Proc. IEEE* **78**, 1498-1530 (1990).
131. Edelman, G. M. et al. Synthetic neural modeling applied to a real-world artifact. *Proc Natl Acad Sci U S A* **89**, 7267-71 (1992).
132. Krichmar, J. L., Nitz, D. A., Gally, J. A. & Edelman, G. M. Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. *Proc Natl Acad Sci U S A* **102**, 2111-6 (2005).
133. Crick, F. Neural edelmanism. *Trends Neurosci* **12**, 240-8 (1989).
134. van Belle, T. Is neural Darwinism Darwinism? *Artif Life* **3**, 41-9 (1997).