The adaptive immune system is a distributed complex system composed of circulating random detectors

Rob J de Boer
Theoretical Biology, Utrecht University,
Santa Fe Institute, Utrecht Center Quantitative Immunology

Explain to you how the adaptive immune system works.

This will be a complexity story of random detectors

that are storing decisions and keeping life-long memories.

We will make a random virus detector based upon immune principles.

This virus scanner can detect unknown viruses (it knows the unknowns).

The adaptive immune system is made of lymphocytes

Lymphocytes are white blood cells. About 10^{12} cells (1 kg).

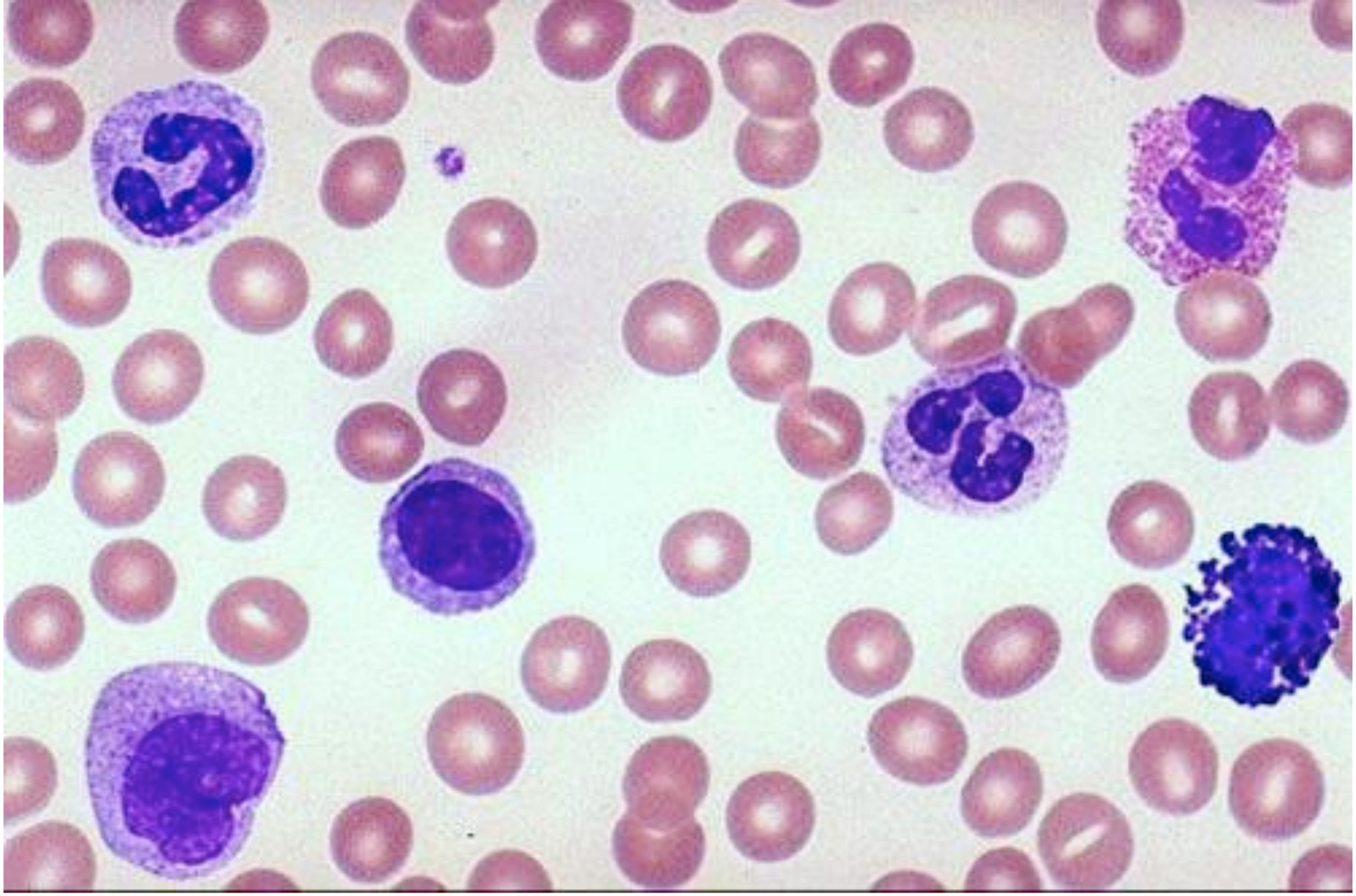
They reside in various tissues and circulate via blood and lymph.

Special tissues: lymph nodes, bone marrow, thymus, and spleen.

B lymphocytes are born in the bone marrow and produce antibodies.

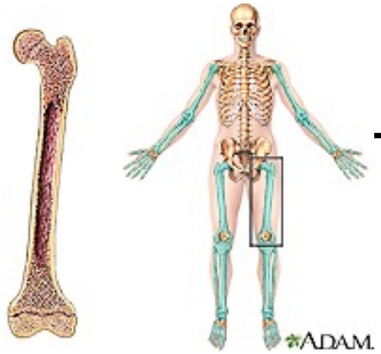
T lymphocytes are born in the thymus and kill aberrant cells.

Lymphocytes express a randomly generated protein
that randomly binds a very small fraction of all possible proteins.
This random protein is called the lymphocyte receptor.



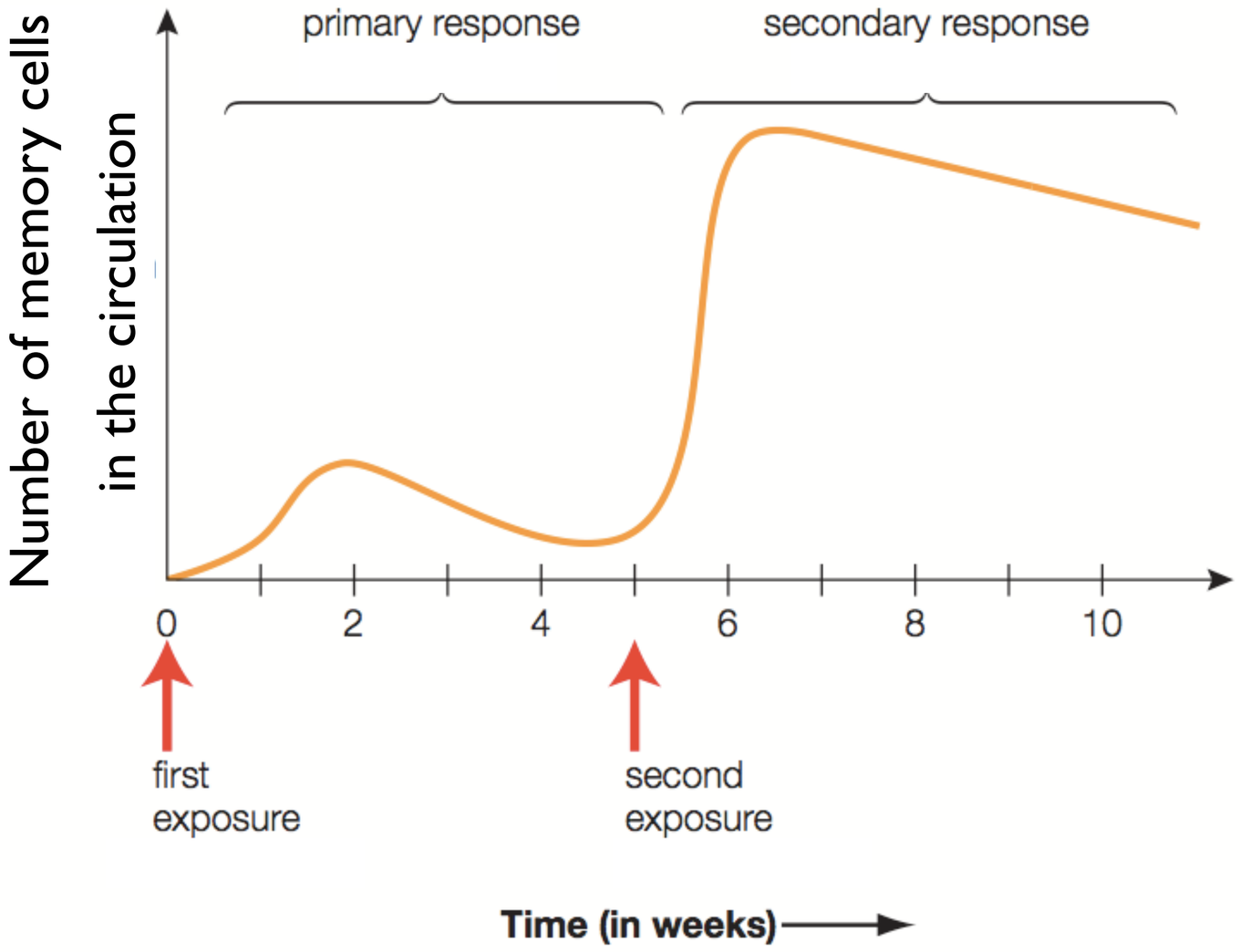
Monocyte Lymphocyte Neutrophil

Naive lymphocytes

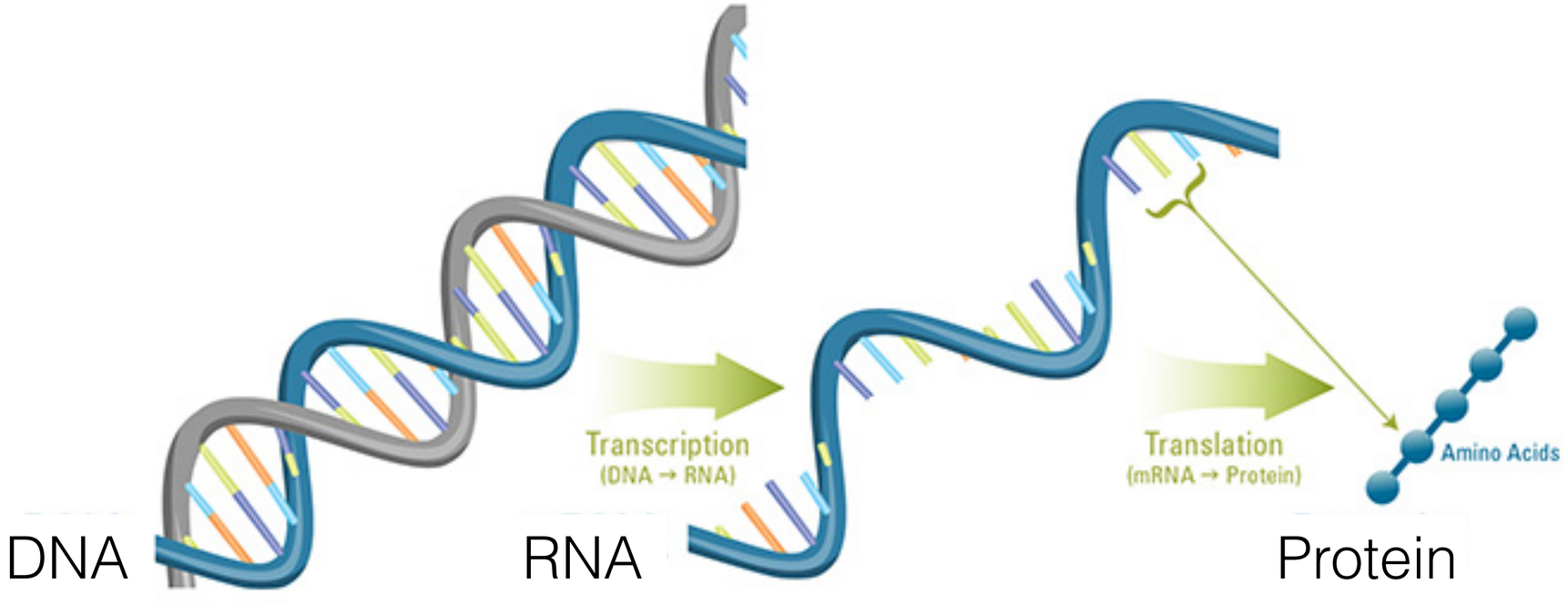


Clonal expansion

Memory lymphocytes


primary response
secondary response
Number of memory cells
in the circulation
0
2
4
6
8
10
first
exposure
second
exposure
Time (in weeks)

How can 10^4 genes make 10^8 proteins?



DNA makes RNA makes protein

Variable (V), diversity (D), joining (J) gene regions each containing many variants: diversity by combinatorics, and random insertions and deletions

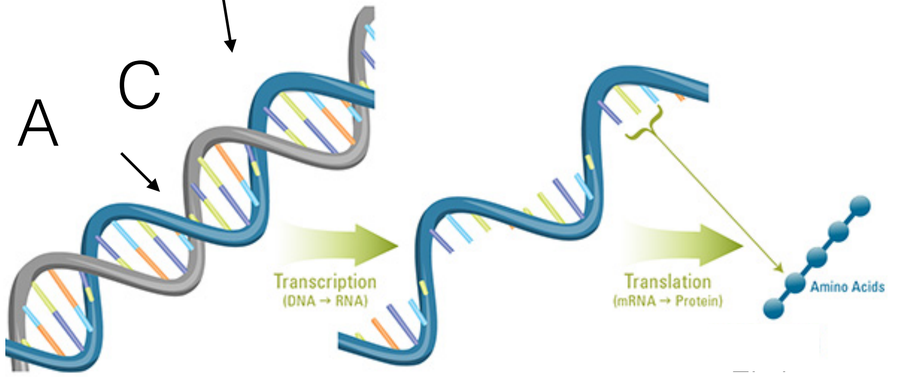
J_i

V_i

DNA fragments recombine into >10^8 different genes that can nowadays be sequenced from blood samples, and be error-corrected by bioinformatics

A

C



recombined DNA

RNA

Protein

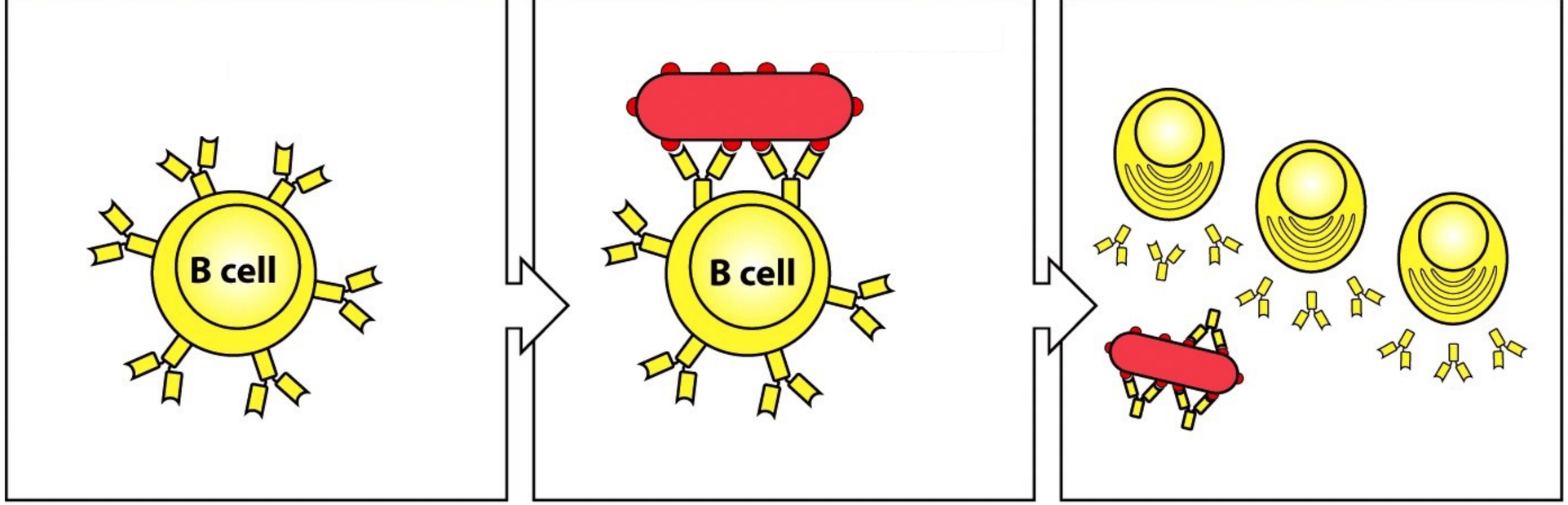
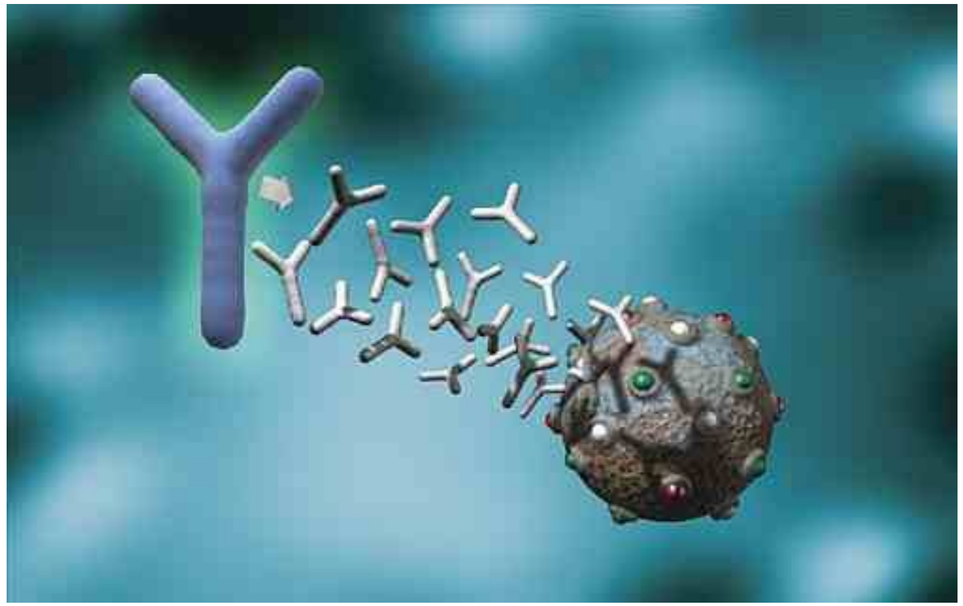
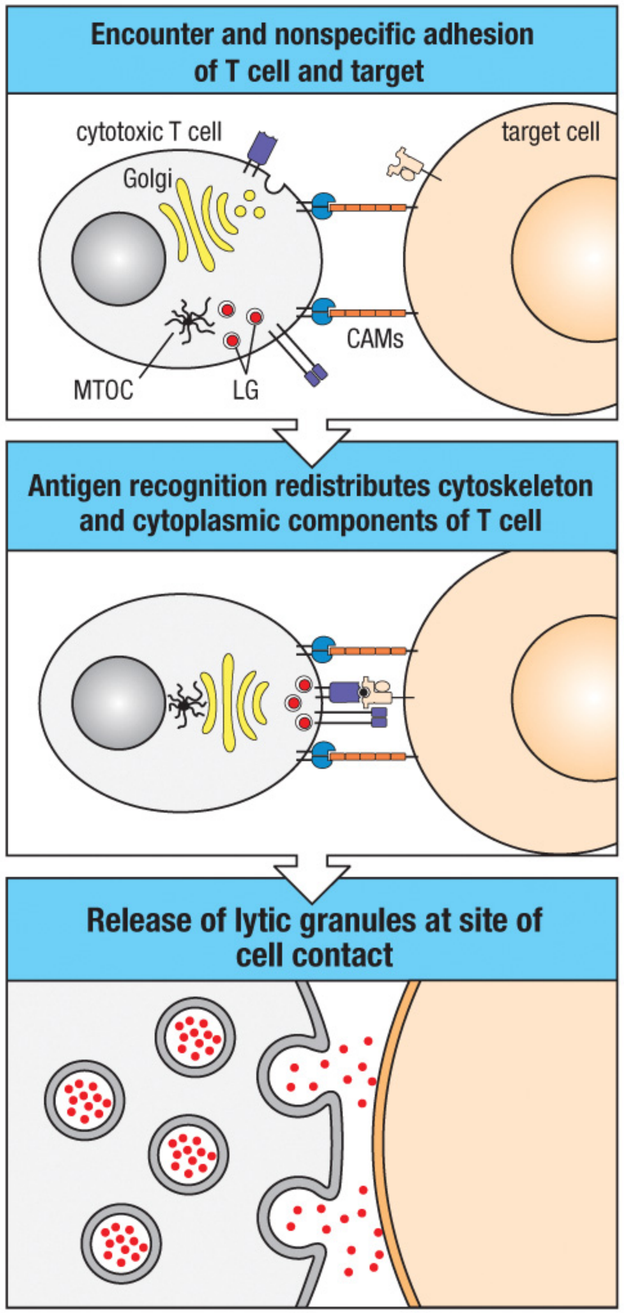


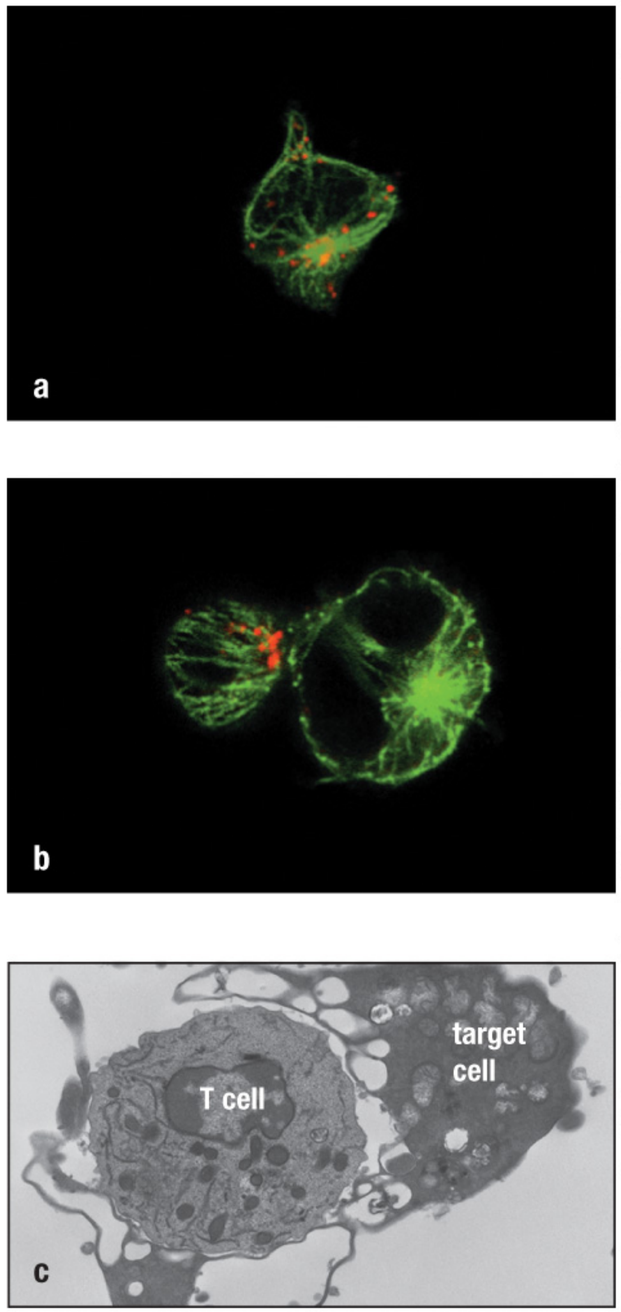
Figure 4.1 The Immune System, 3ed. (© Garland Science 2009)



B cells produce their receptor as antibody

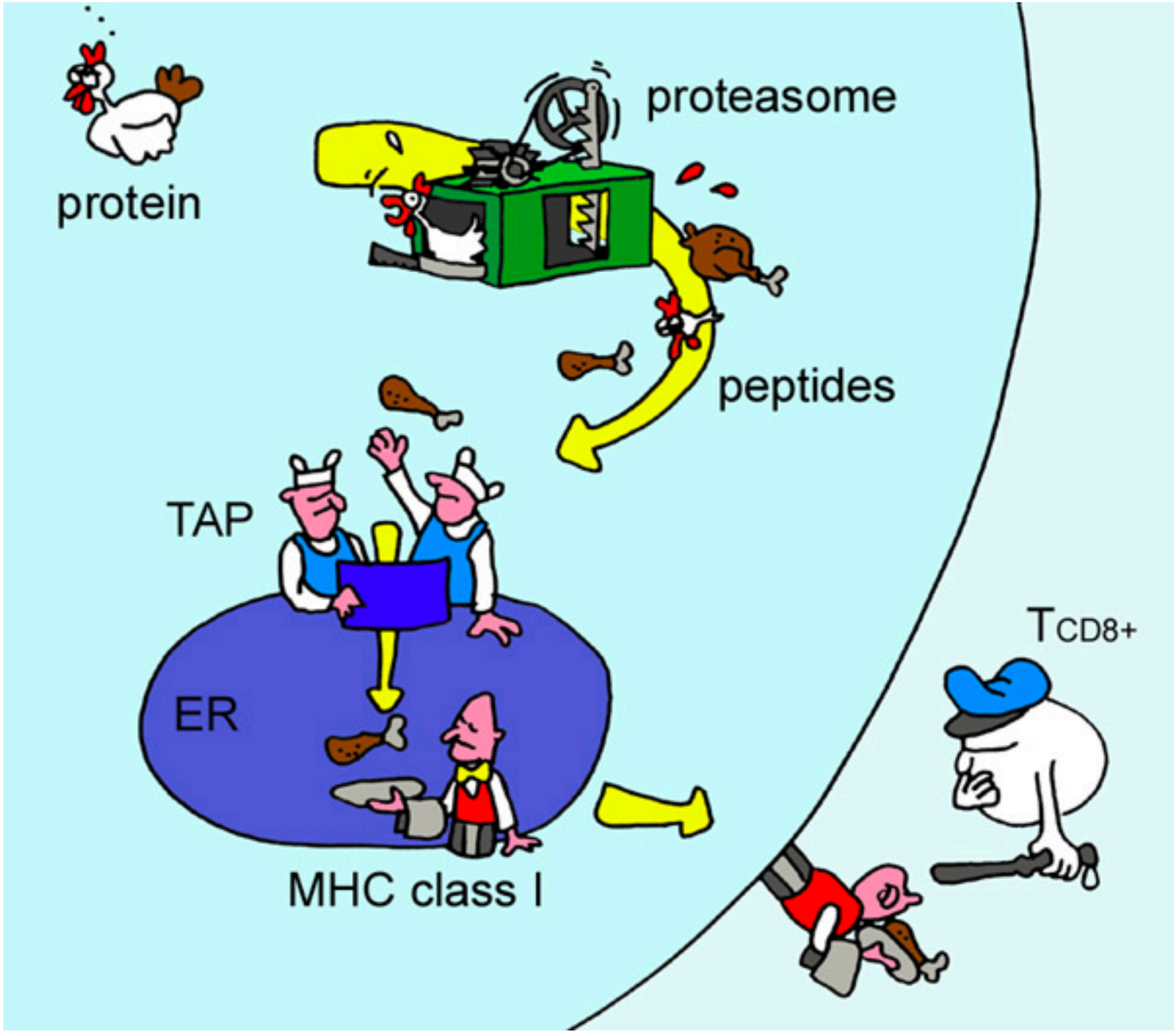
T cells kill by cell-to-cell contacts





Panels a and b: Photographs courtesy of G. Griffiths and J. Stinchcombe. Panel c from P.A. Henkart and E. Martz (eds), *Second International Workshop on Cell Mediated Cytotoxicity*. © 1985 Kluwer/Plenum Publishers. With permission from Springer Science and Business Media.

Special molecules (MHC) **sample** content of cell and present those on cell surface



Eric Reits

Nature Reviews Immunology 2003.

doi:10.1038/nri1250

This seems relatively simple

Distributed system of 10^{11} cells each expressing a random receptor.

Upon detecting a foreign protein about one in 10^5 receptors "fires".

These rare cells get amplified by cell division

and form a large clone of cells expressing the same receptor.

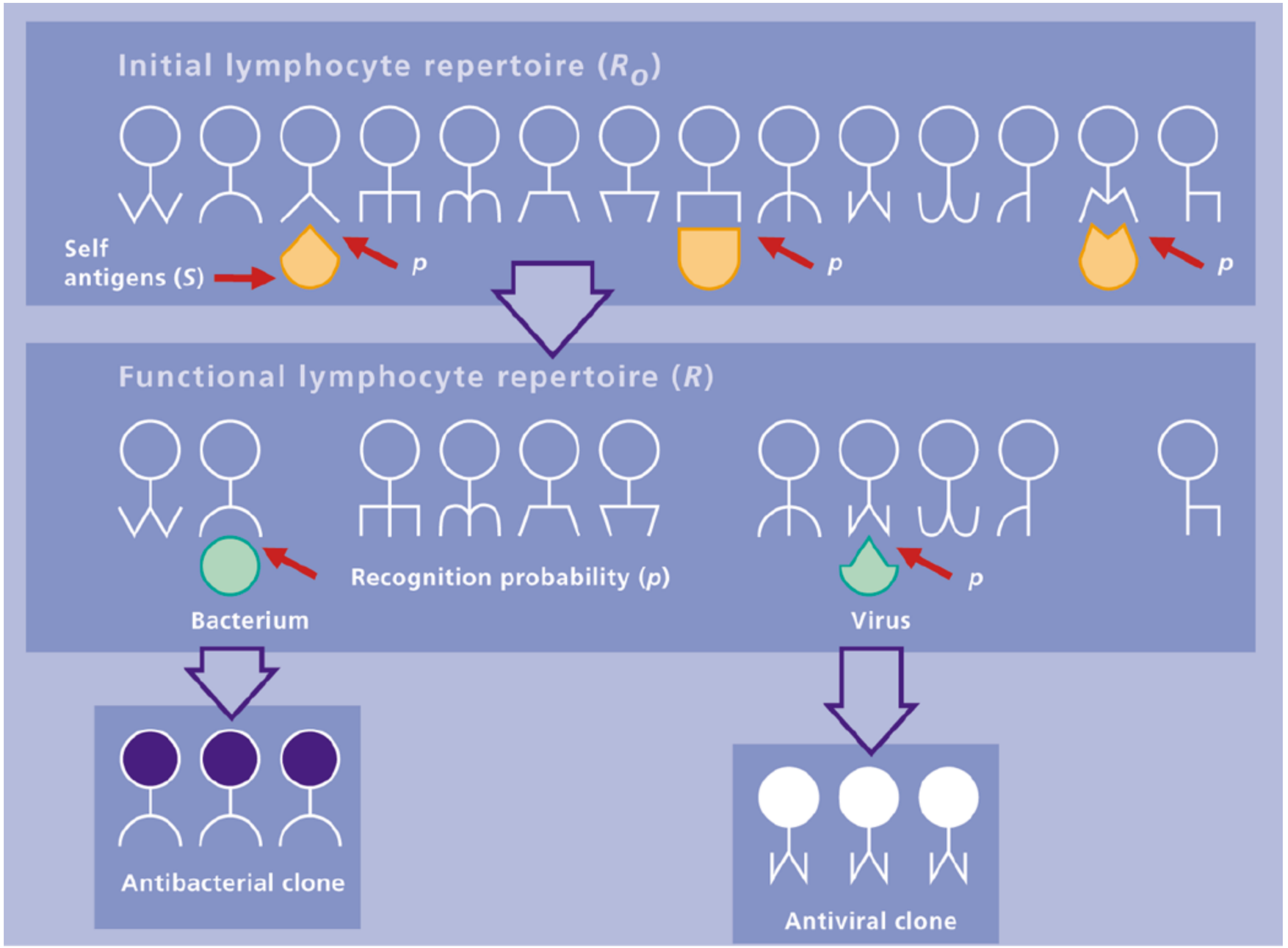
Effector cells of this large clone circulate and clear the invader.

This accounts for memory because a subsequent response to the same protein starts with much larger initial numbers.

(Memory cells are dynamically maintained for life)

Problem ?

A simple mathematical model



Central tolerance

S number of self epitopes evoking tolerance (10^5) [Burroughs.i04]
R_0 potential repertoire (before tolerance) (huge: 10^{12} cells)
R "functional" repertoire after tolerance (> 10^9) [Qi.pnas14]
p recognition probability (precursor freq. 10^{-5}) [Blattman.jem02, Su.i13]

Size of functional repertoire:

$$R = R_0(1-p)^S$$

$R = R_0P_S = R_0e^{-pS}$

Probability of responding to a foreign epitope:

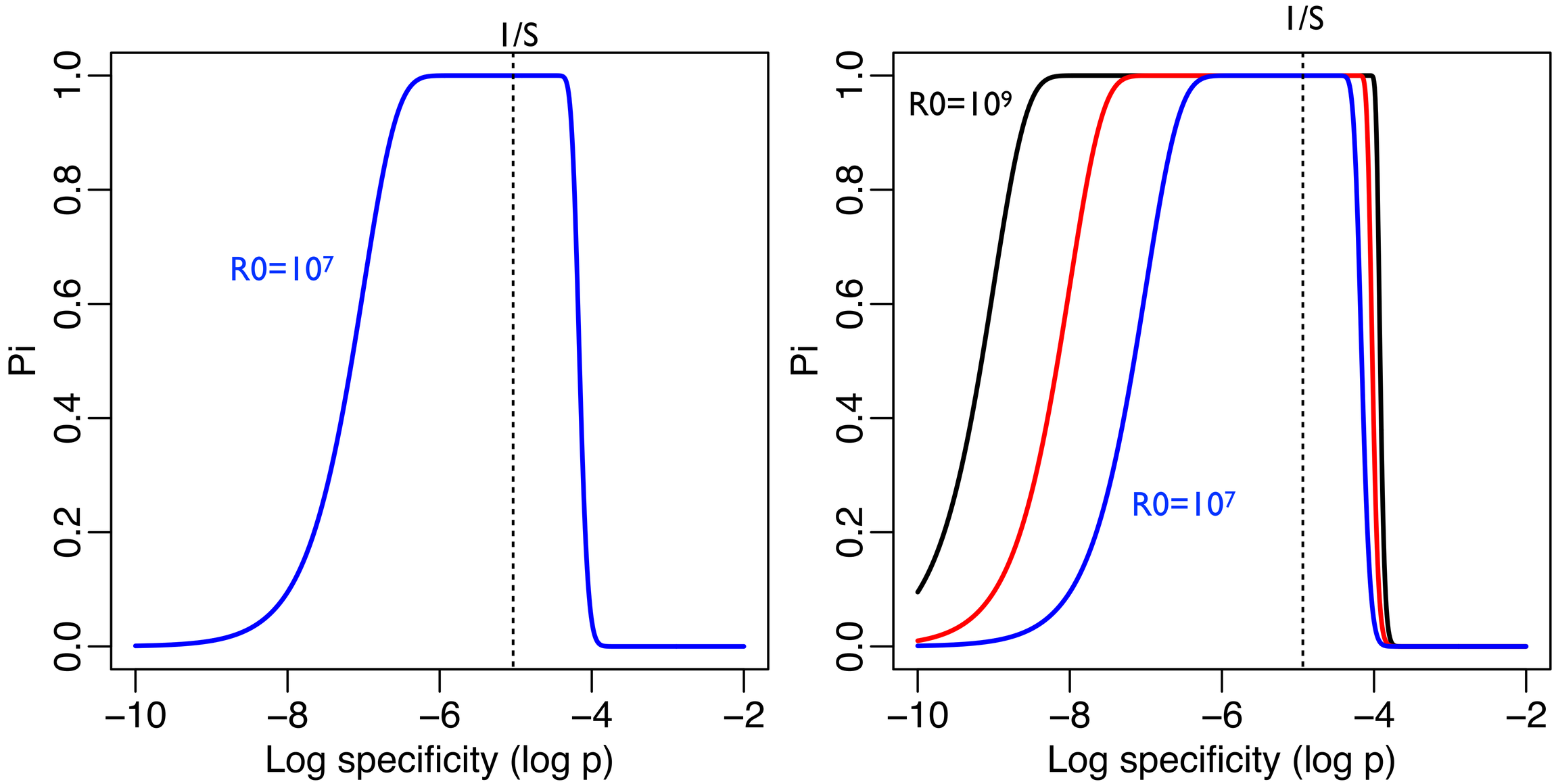
$$P_i = 1-(1-p)^R$$

Taking the derivative of P_i to p gives the optimum

$$p = \frac{1}{S} \simeq 10^{-5}$$

De Boer & Perelson Proc Royal Soc 1993

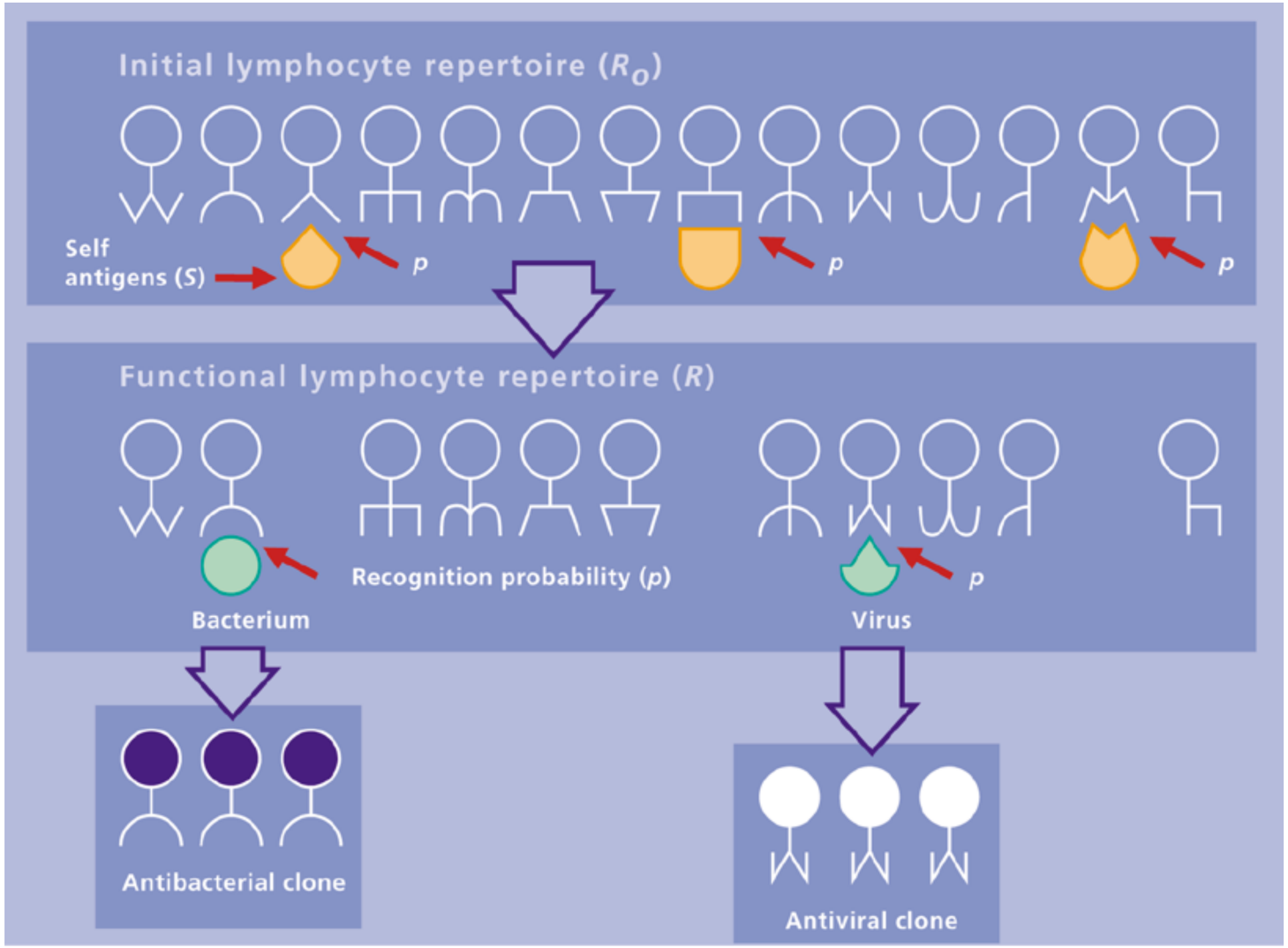
A very broad optimum



Too specific

Too degenerate

Immune response to self tissues are inappropriate



$$R_0 = 10^{12}$$
$$S = 10^5$$

$p = 10^{-5}$

$$R = R_0(1-p)^S$$

$$P_i = 1-(1-p)^R$$

Optimize (Pi'=0):

$$p = \frac{1}{S} \simeq 10^{-5}$$

Receptors have to specific to avoid massive deletion [De Boer, Perelson, Borghans, 1993, 1999]

Required repertoire investment

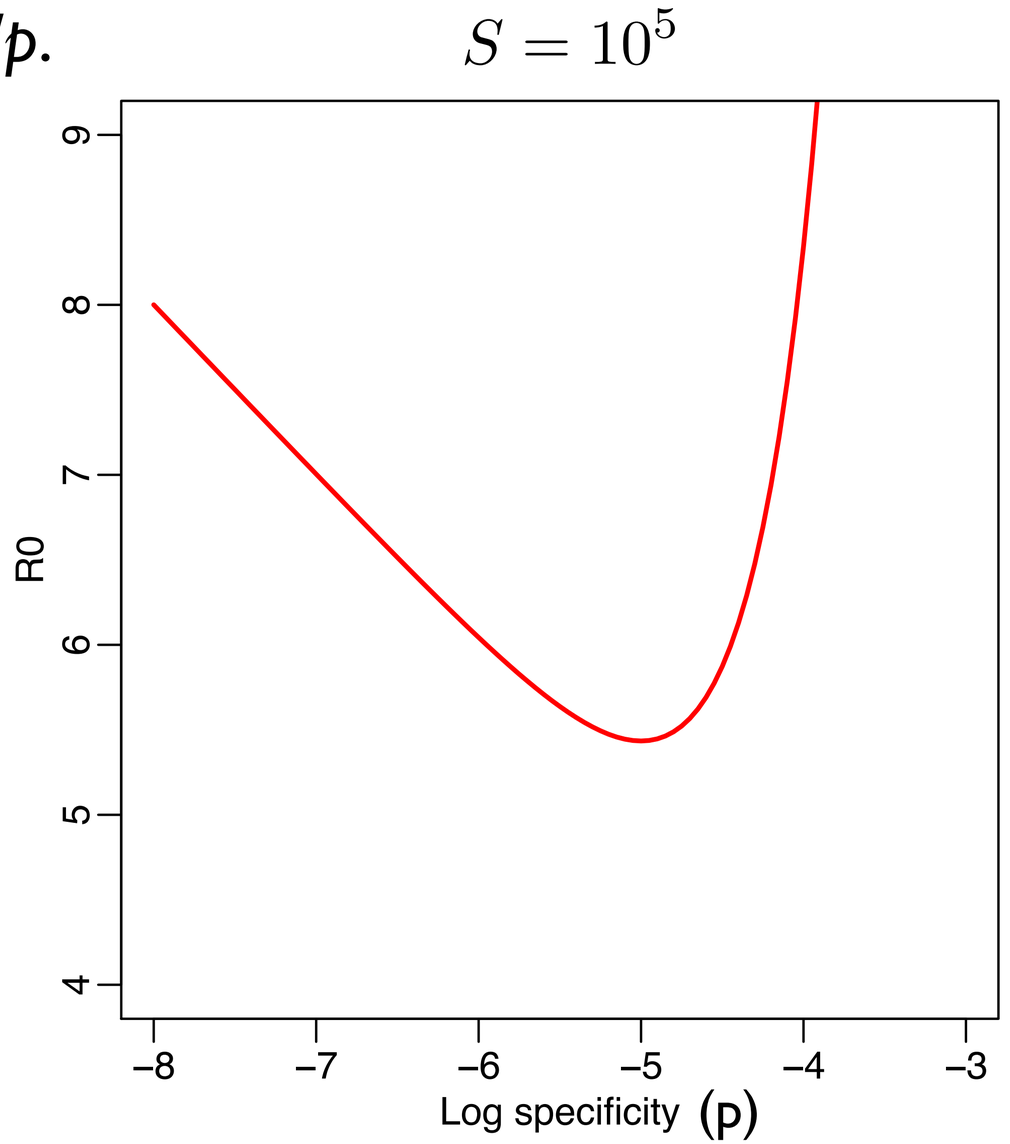
With a recognition probability of p per clone, one expects a response to every challenge when $R = 1/p$.

Hence:

$$R = R_0(1-p)^S$$

$$\frac{1}{p} = R_0(1-p)^S \simeq R_0 \mathrm{e}^{-pS}$$

$$\text{or} \quad R_0 = \frac{1}{p(1-p)^S} \simeq \frac{\mathrm{e}^{pS}}{p} \ .$$



A large fraction of the detectors is silenced

Lymphocytes are specific to have a chance to survive self tolerance.
Because they are specific we need a large random repertoire
to cover the world of all possible proteins.

Those that survive circulate and are allowed to respond to anything new

Problem 2: harmless environmental proteins (food)

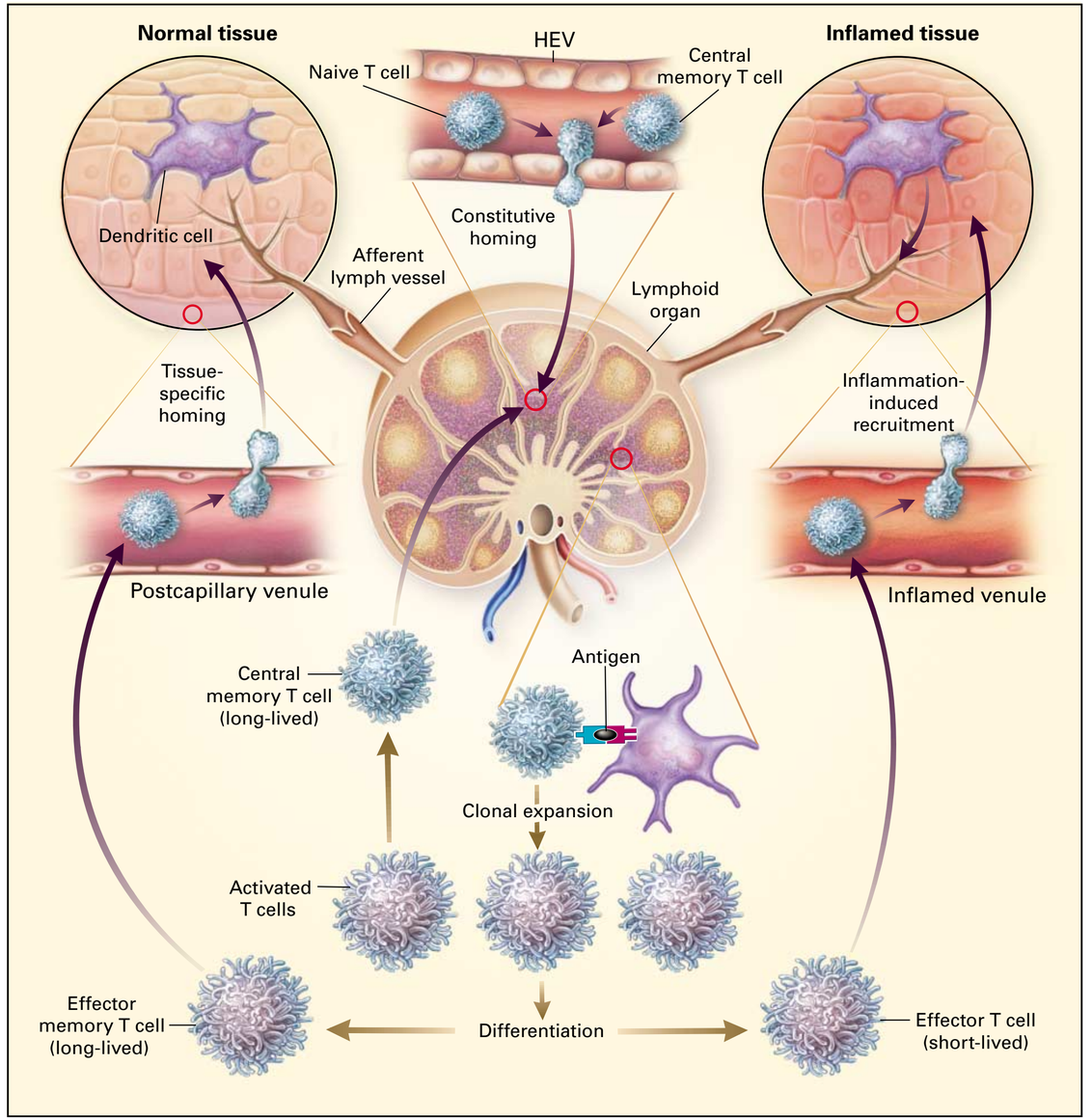
Immune responses develop in draining lymph nodes

Dendritic cells (DC) scan peripheral tissues, and migrate to draining lymph nodes to present their proteins.

Millions of different naive lymphocytes migrate through lymph nodes, and scan the DCs.

Only 1:100000 cells will become activated, expand, and emigrate as effector cells that move back to the inflamed tissues.

The first signal is the activation of DC that carry information to the draining lymph node.



Von Andrian & Mackay, NEJM, 2000

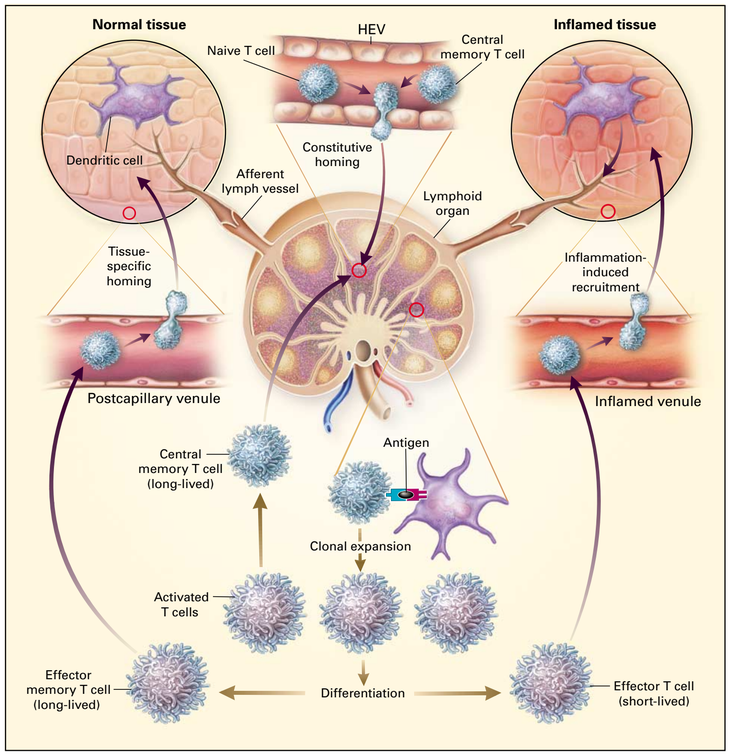
Supervised learning by innate immune system

Most circulating white blood cells don't express random receptors.

Macrophages, neutrophils, granulocytes are much more abundant.

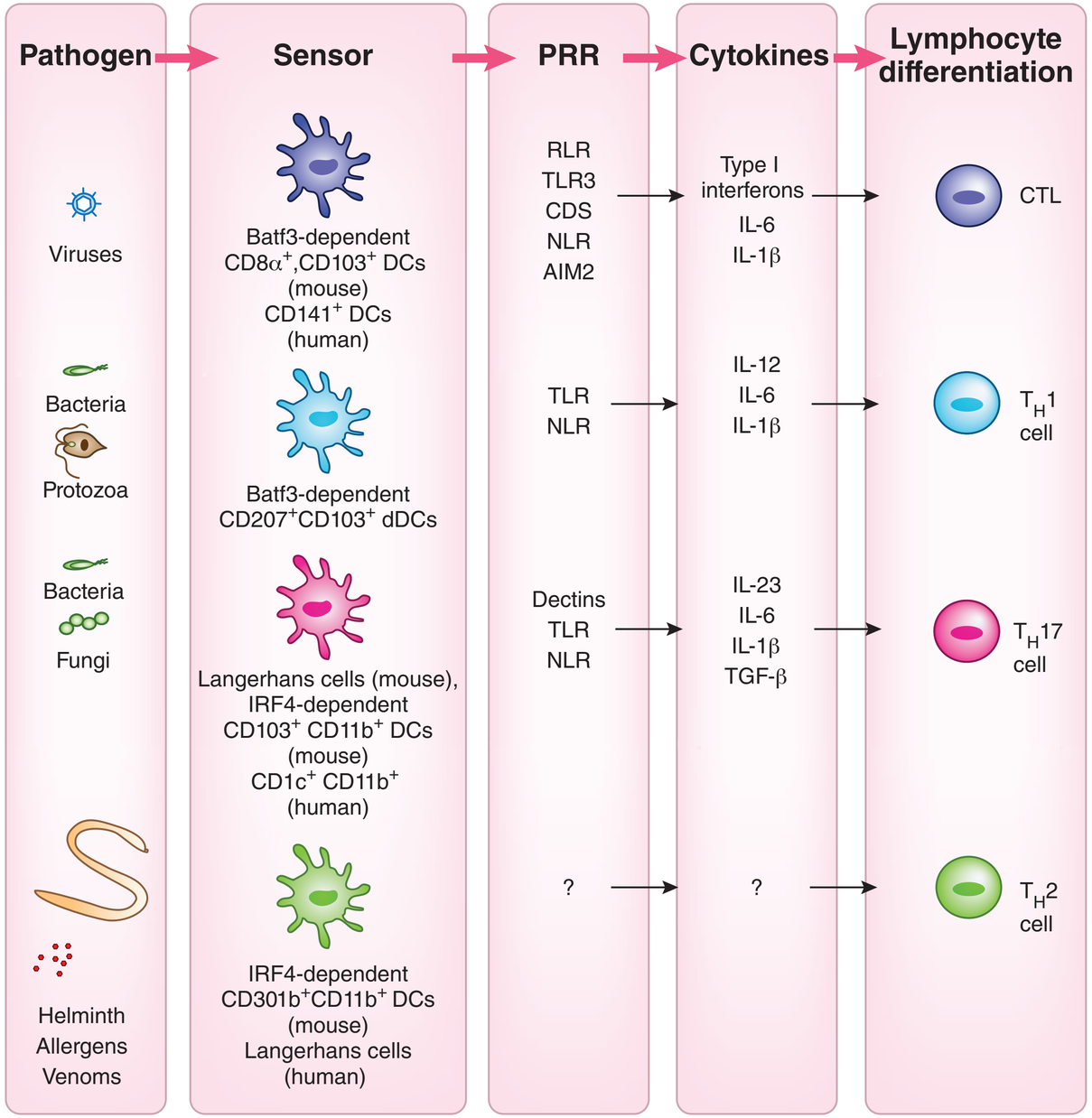
These "innate" immune cells express conserved receptors

evolved for a specific groups of pathogens (viruses, bacteria, worms).



Upon detecting a particular pathogen innate cells secrete molecules that "instruct" the activated naive lymphocytes in their environment.

Lymphocytes are activated within a particular context



Macrophages, dendritic cells, neutrophils, natural killer cells & infected cells have conserved sensors, induce inflammation and provide information in the form of cytokines.

Innate control of adaptive immunity

Iwasaki & Medzhitov

Nature Immunology 2015

Supervised learning by innate immune system

Upon detecting a particular pathogen innate cells secrete molecules that "instruct" the activated naive lymphocytes in their environment.

Instructed lymphocytes thus develop an appropriate phenotype,
change their gene expression and remember that for life.

The decision on what to do comes from the innate system,
and this gets "**uploaded**" in the adaptive system as life-long immunity.

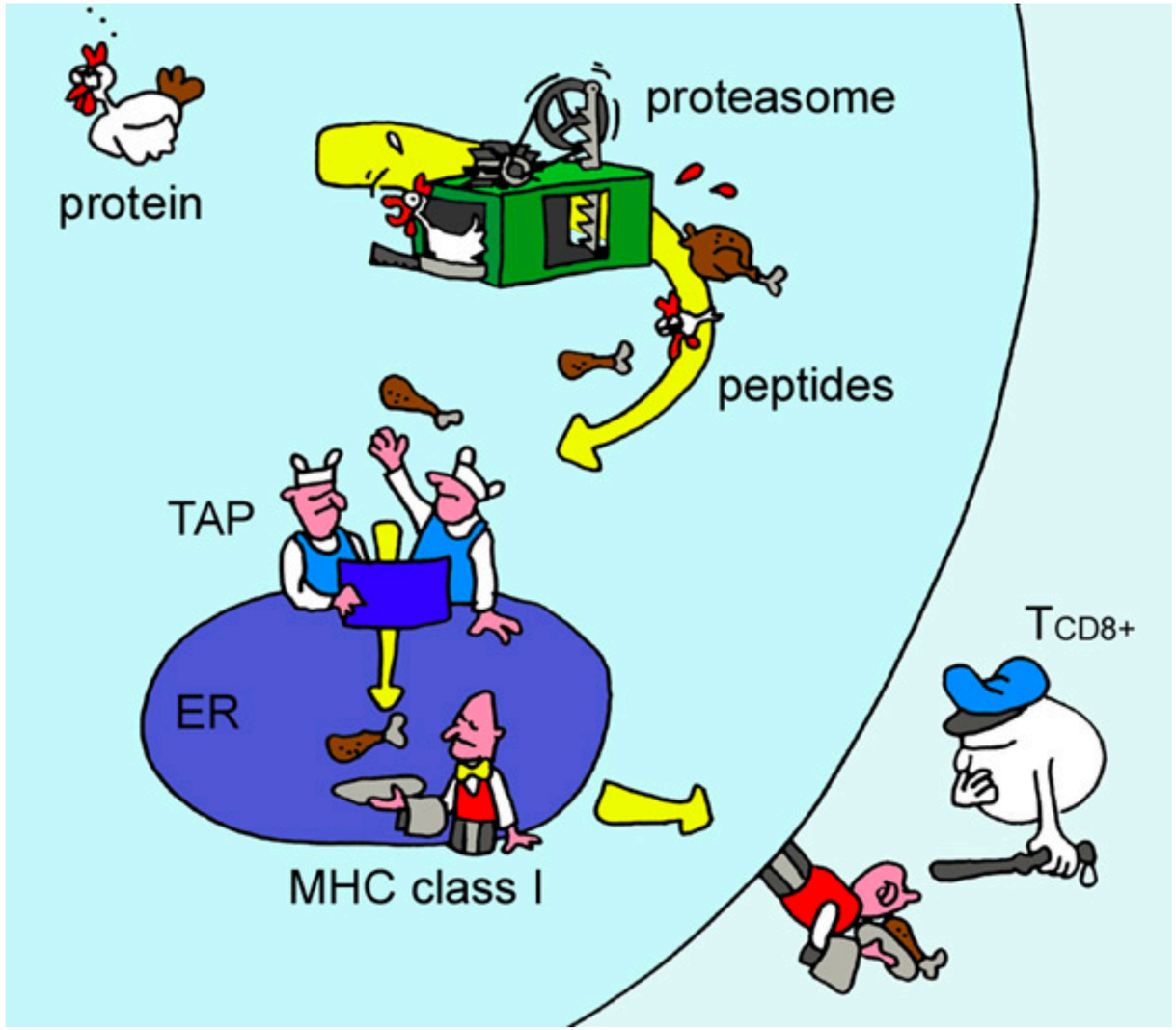
A few more details

T lymphocytes bind small samples (peptides) of proteins on specialized presentation molecules that differ from individual to individual.

MHC polymorphism: we all make unique responses to viral proteins,
viruses cannot predict which proteins will be targeted.
Immune escapes of a virus useless in the next host

There are also "regulatory" T cells broadcasting the message that they see self.

MHC molecules present content of cell in the form of small samples



We all have different MHC molecules by inheriting these genes from our parents.

MHC genes are polymorphic.

Why would that be?

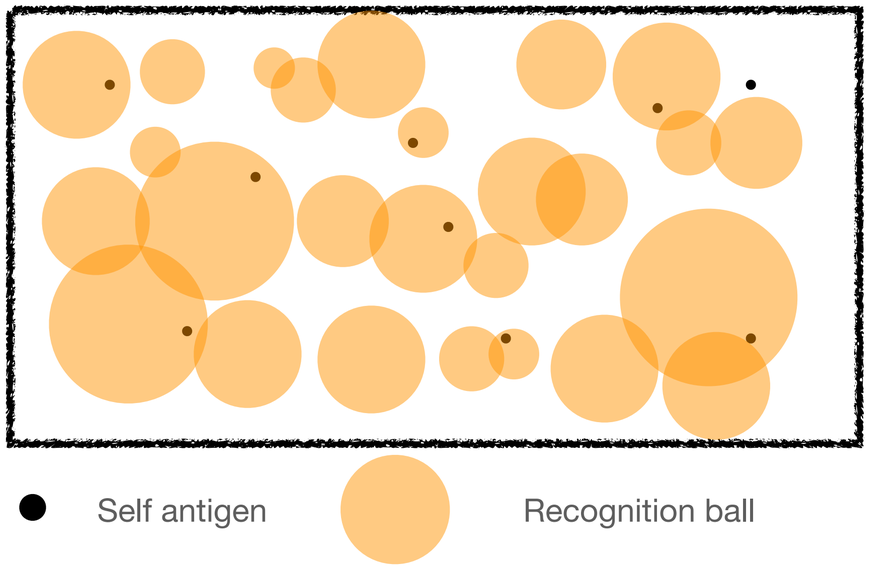
From: "Making sense of mass destruction: quantitating MHC class I antigen presentation"

Jonathan W. Yewdell, **Eric Reits** & Jacques Neefjes, Nature Reviews Immunology 2003

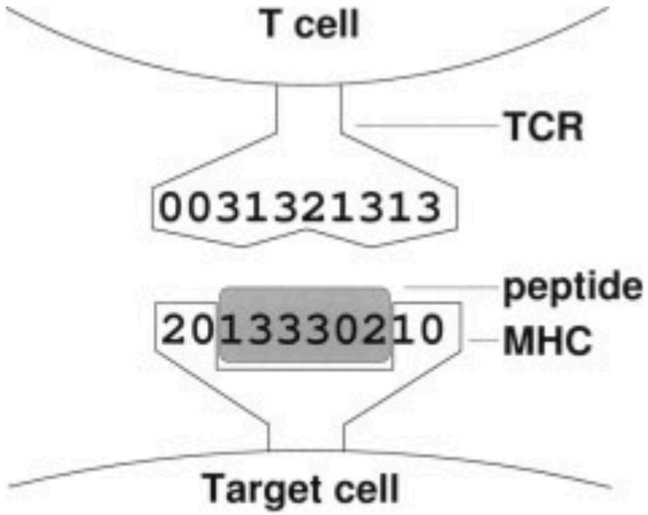
Modeling receptors as strings or as circles in shape space

This allows one to develop computer virus detectors (Forrest, 1997)
Allows one to study affinity maturation by mutating letters in agent based models.
Repertoire development: negative selection weeds out the most crossreactive receptors.

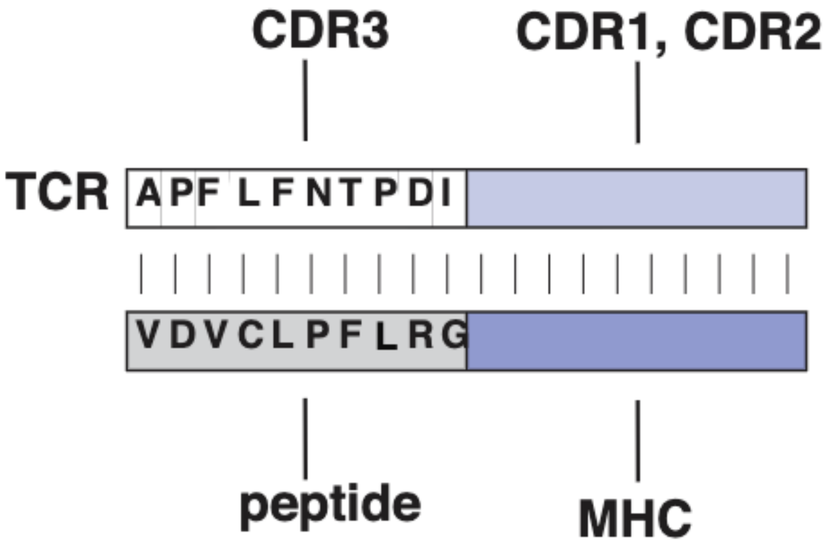
Perelson & Oster: shape space



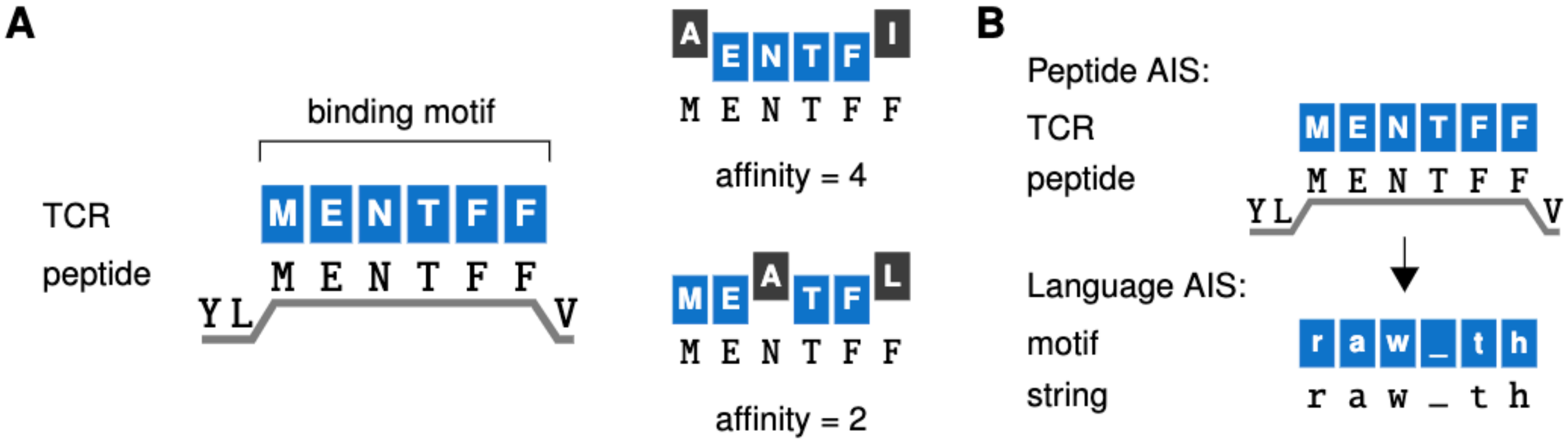
Chao et al Eur J Imm 2005



Kosmrlj et PNAS 2008



Distinguish languages by sampling words: generalization



Is T Cell Negative Selection a Learning Algorithm?

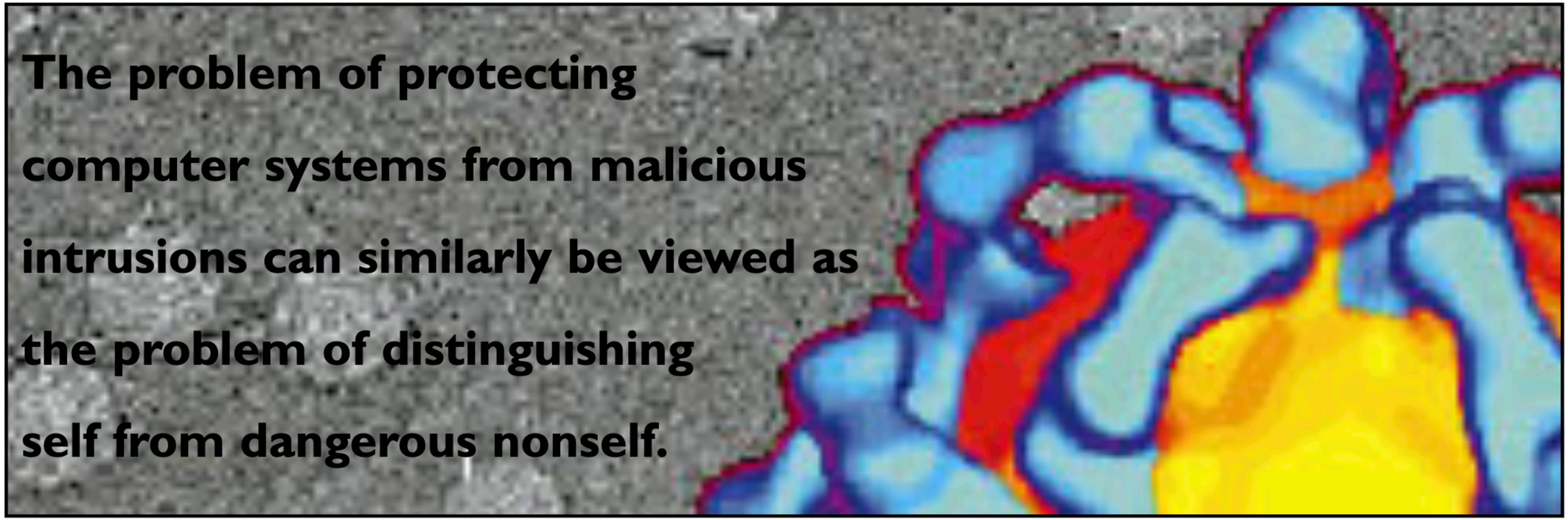
Inge M. N. Wortel [1,*], Can Keşmir [2], Rob J. de Boer [2], Judith N. Mandl [3] and Johannes Textor [1,2,*]

1 Department of Tumor Immunology, Radboud Institute for Molecular Life Sciences, Geert Grooteplein 26-28, 6525 GA Nijmegen, The Netherlands
2 Theoretical Biology, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands; c.kesmir@uu.nl (C.K.); r.j.deboer@uu.nl (R.J.d.B.)
3 Department of Physiology, McGill University, 3649 Promenade Sir William Osler, Montreal, QC H3G 0B1, Canada; Judith.mandl@mcgill.ca
* Correspondence: inge.wortel@radboudumc.nl (I.M.N.W.); johannes.textor@radboudumc.nl (J.T.)

Received: 21 January 2020; Accepted: 7 March 2020; Published: 11 March 2020

More difficult: only a fraction of self used for training.

Computer immunology



Stephanie Forrest (forrest@cs.unm.edu) is an associate professor in the Department of Computer Science at the University of New Mexico in Albuquerque and wrote this article while on sabbatical at the MIT Artificial Intelligence Laboratory.
Steven A. Hofmeyr (steveah@cs.unm.edu) is a Ph.D. student in the Department of Computer Science at the University of New Mexico in Albuquerque.
Anil Somayaji (soma@cs.unm.edu) is a Ph.D. student in the Department of Computer Science at the University of New Mexico in Albuquerque.

COMMUNICATIONS OF THE ACM October 1997/Vol. 40, No. 10

Coverage

CD8 T cells bind protein fragments of 9 amino acids, and there are about 10^7 unique 9-mers in the human self.

The whole world of 9-mers corresponds to $20^9=2^9 10^9=512 \times 10^9$ unique 9-mers.

Hence the self is expected to cover $c = 10^7/20^9 = 1/51200$ of the whole peptide space.

If a pathogen presents n=10 peptides (epitopes) these are expected to not overlap with self.

The TCR only binds 5 of the 9 amino acids.

Hence the peptide universe for CD8 T cells is 20^5 unique peptides.

Since there are about 2×10^6 unique self 5-mers in the human self the coverage would be

$$c = \frac{2 \times 10^6}{20^5} = \frac{1}{2^4 \times 10^{-1}} = \frac{5}{8}$$

About half of the peptides presented would therefore overlap with self.

A random virus detector: initialization

```
A = 8
K = 5
L = 7
S = 1e4
R = 1e4
letters = string.ascii_lowercase[:A]
nKmers = 1 + L - K
print("Total number of kmers:",A**K)
print("Number of kmers per detector:",nKmers)
pk = np.power(A, -float(K))
p = 1 - np.power((1 - pk), nKmers)
print("Specificity p = %.3e"%p)
c = S / A**K
print("Self covers a fraction c = %.3f of the sequence space"%c)
```

#letters in alphabet

length of k-mer

length of detector

number of self k-mers

size of repertoire

Make a dictionary with all k-mers

```
aaa: T [13, 23, 17]
aab: F [101, 213]
...: . []
zzz: F []
```

```
repertoire = [0, 0, 1, 0, 2, ...., 0]
```

```
a_matrix = [[i for i in letters] for j in range(K)]
all_kmers = ["".join(f) for f in list(itertools.product(*a_matrix))]
kmer_dict = {kmer:[False, []] for kmer in all_kmers}
```

```
self_kmers = nrandom_strings(K, S)
for kmer in self_kmers:
    kmer_dict[kmer][0] = True
```

S k-mers are set to be self

Build a random repertoire of R detectors

```
repertoire = []
detec_id = 0
n_naives = 0
while n_naives < int(R):
    Lmer = arandom_string(L) # make one new detector
    repertoire.append(0)
    for i in range(nKmers):
        kmer = Lmer[i:i+K]
        # add detector (if not already present for this k-mer):
        if not detec_id in kmer_dict[kmer][1]:
            kmer_dict[kmer][1] += [detec_id]
        # set state to tolerant if k-mer in self repertoire:
        if kmer_dict[kmer][0]:
            repertoire[detec_id] = 1
    # add to number of naive detectors if not tolerant:
    if repertoire[detec_id] == 0:
        n_naives += 1
    detec_id += 1
```

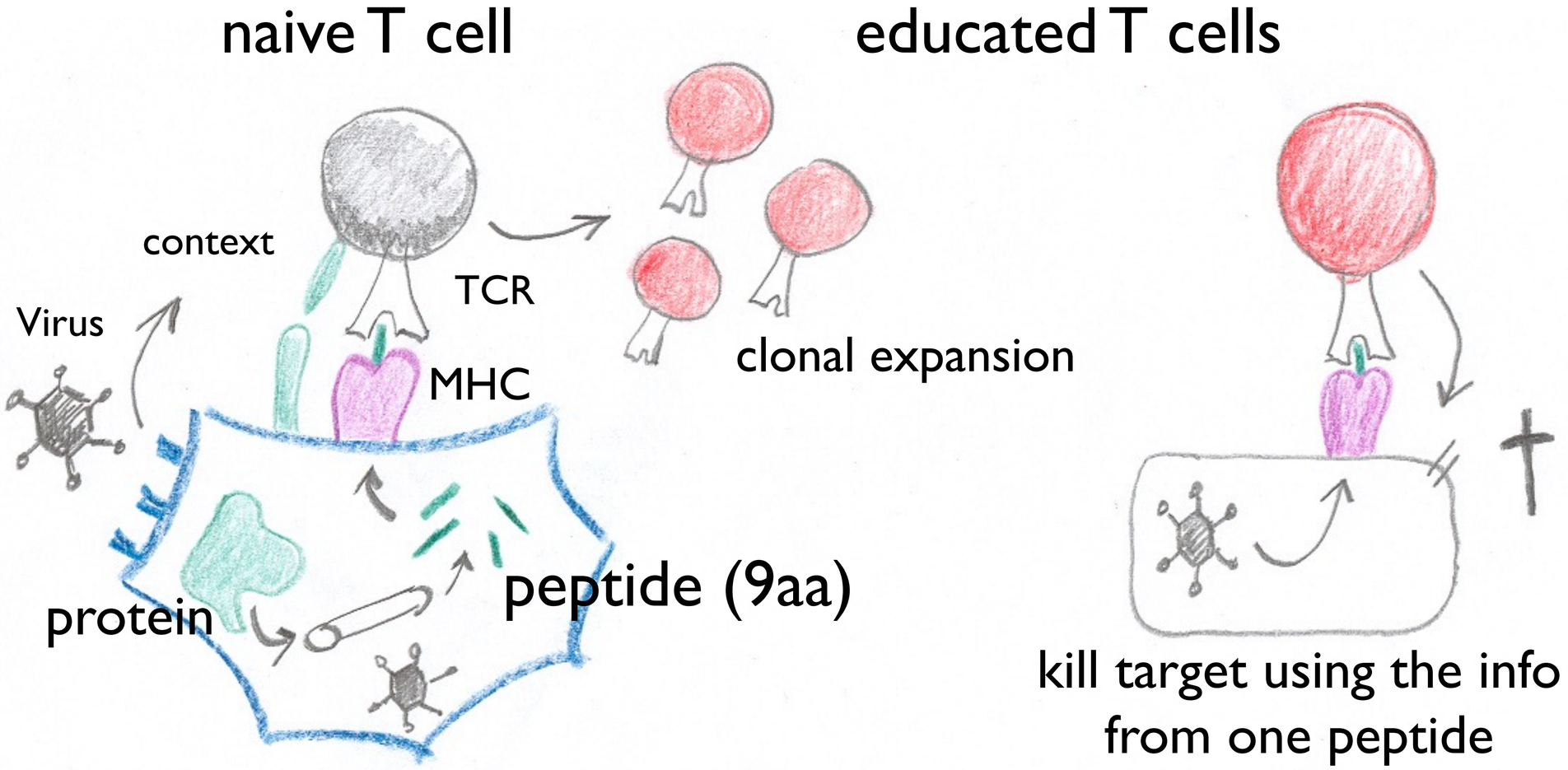
```
repertoire = [0, 0, 1, 0, 2, ...., 0]
```

Challenge with intruders

```
n_intruders = 0
alive = True
while alive:
    n_intruders += 1
    intruder = nrandom_strings(K, 10)
    responders = []
    for kmer in intruder:
        responders += kmer_dict[kmer][1]
    resp_states = [repertoire[i] for i in responders]
    distribution = np.bincount(resp_states, minlength = 3)

    if distribution[0] + distribution[2] > 0:  # respond
        response = True
        for i in responders:                   # update states
            if repertoire[i] == 0:
                repertoire[i] = 2
    else:                                      # no response
        alive = False
        print("No response to intruder %i"%n_intruders)
```

Immune systems samples a few peptides and stores contextual information in memory cells



MHC molecules presenting peptides to T cells are **polymorphic**

Due to a rare allele advantage: its is good to be different