ImmunoGrid

Large scale simulation of the immune system

Andrew Emerson, **Silvia Giuliani**, Elda Rossi

CINECA, Bologna Italy

The Mammalian Immune System

- A complex and adaptive learning system

- Evolved to defend an individual against foreign invaders

- Operates at multiple levels: from molecule to cell, organ and organism

A Complex system

- The human immune system has immense diversity:

  - $>10^{13}$ MHC class I haplotypes

  - 10^7-10^{15} different T-cell receptors

  - 10^{12} B-cell clonotypes in each individual

  - 10^{11} possible linear MHC-binding epitopes composed of nine amino acids

  - $>>10^{11}$ different conformational epitopes

  - $>10^9$ combinatorial antibodies

Immunology: Successes and Failures

- Vaccines have been instrumental in controlling many diseases

  - Eradication of smallpox

  - Near eradication of polio

- But many diseases are still poorly protected against

  - e.g. failure of the BCG vaccine against TB in some communities

Modelling the immune system

- Computer models are used to complement and not replace actual testing or experiments

- System-level models have a long history, and include models of T cell responses to viruses, analysis of MHC diversity under host-pathogen co-evolution, B cell maturation, etc.

Limits of models

The main problems that prevent the use of computational models in practical applications are:

- large combinatorial complexity of the human immune system,

- lack of understanding of specific molecular interactions

- correlation of model parameters to real-life measurements.

Grid computing can provide powerful computational infrastructure and capacity that can match the complexity of the real human immune system.

ImmunoGrid

- "…a 3 year project funded by the European Union which will establish an infrastructure for the simulation of the immune system that integrates processes at molecular, cellular and organ levels."

- To be designed for applications that support clinical outcomes such as design of vaccines and immunotherapies and optimization of immunization protocols."

Project Objectives

- Create computational models for the real-size human immune system.

- Standardize immune system concepts, bioinformatics tools and information resources to enhance the computational models for pre-clinical and clinical applications.

- Validate these models with experimental data (mice) and disseminate the tools developed to users such as vaccine and immunotherapy researchers and developers.

Immunogrid: Partners

- CINECA, Bologna, Italy (Project coordinator)

- University of Queensland, Australia (Scientific coordinator) - Vladimir Brusic

- CNR, Rome, Italy – Filippo Castiglione

- CNRS, Montpellier, France - Marie Paule Lefranc

- Technical University of Denmark – Soren Brunak

- Birkbeck College, University of London, UK - David Moss

- Department of Experimental Pathology, University of Bologna, Italy – PierLuigi Lollini

- University of Catania, Catania, Italy – Santo Motta
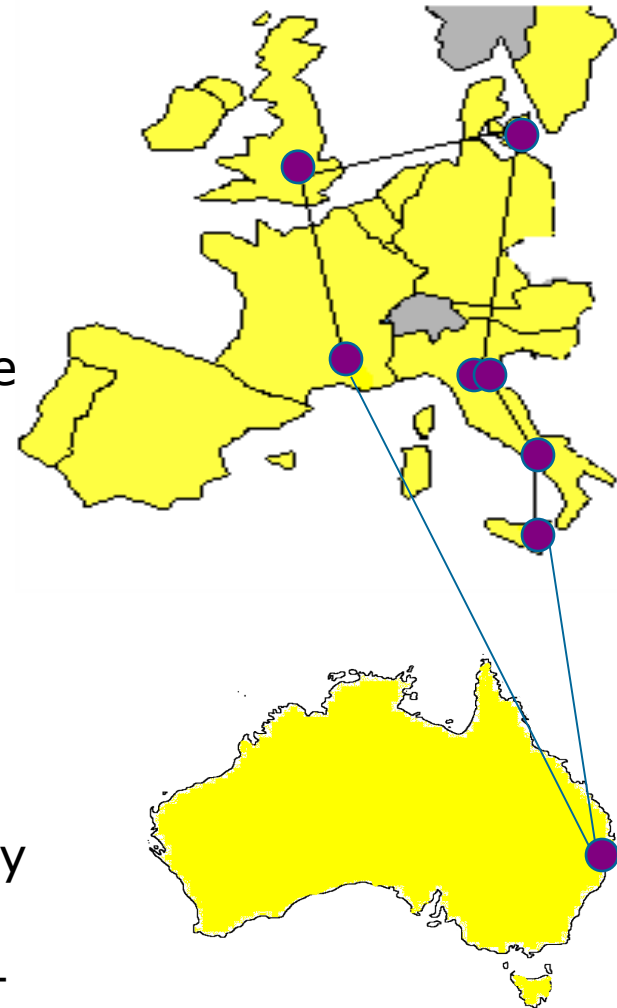
The starting point

- We have a starting point to construct the simulator

  – C-ImmSim (agent-based model)

- The "current" version of **C-ImmSim** (v.6.2) is available under the GNU General Public License

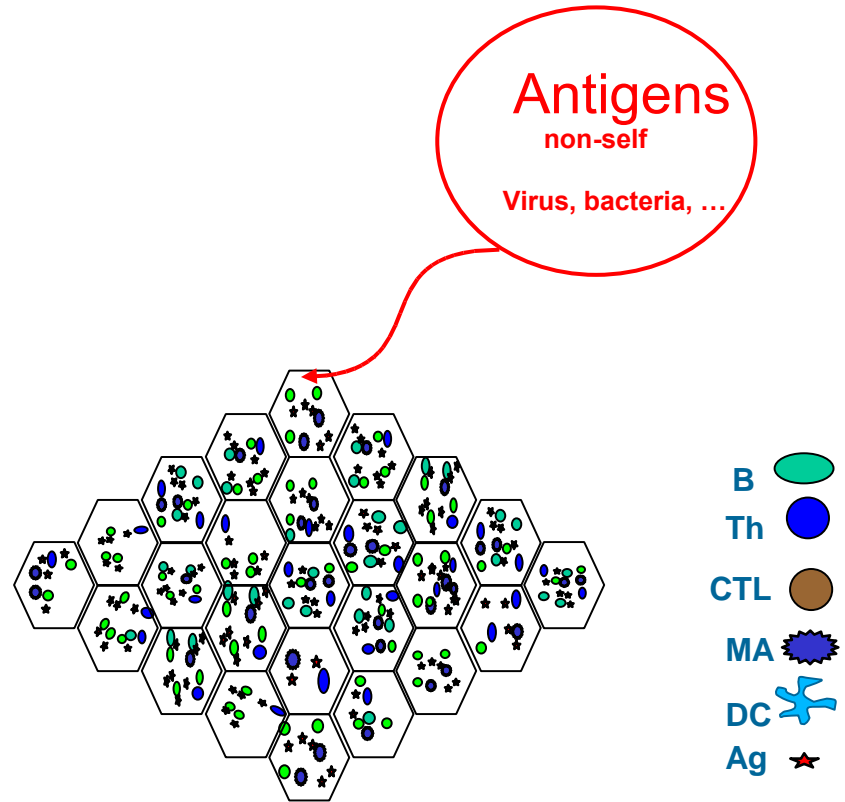 http://www.iac.cnr.it/~filippo/cimmsim.html

C-ImmSim

- Stochastic model able to simulate a wide range of immunological phenomena :
  - Viral infection
    - Generic virus
    - HIV-1
    - EBV
  - Generic bacterial infection
  - Hypersensitivity reactions
  - Cancer growth

2D, C-ImmSim

Antigens
non-self

Virus, bacteria, …

1mm³ tissue of
mouse immune organ

B
Th
CTL
MA
DC
Ag

Molecules

Molecules are represented by binary strings of bits:
- B-cell receptors (BCR), immunoglobulins
- T-cell receptors (TCR)
- Major Histocompatibility Complex (class I and II)

The antigen is represented, **at least**, by two binary strings:
- The epitope (*i.e.,* the BCR's binding site).
- The peptide (*i.e.,* the MHC class I and II's binding site).

Peptide

Epitope

receptor

lymphocyte

MHC

l =12 allows for a potential repertoire of.

2^{12} = 4096 distinct cell receptors and antibodies

the simplest antigen is identified by 2^{24} bits

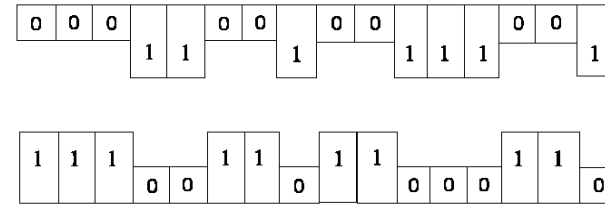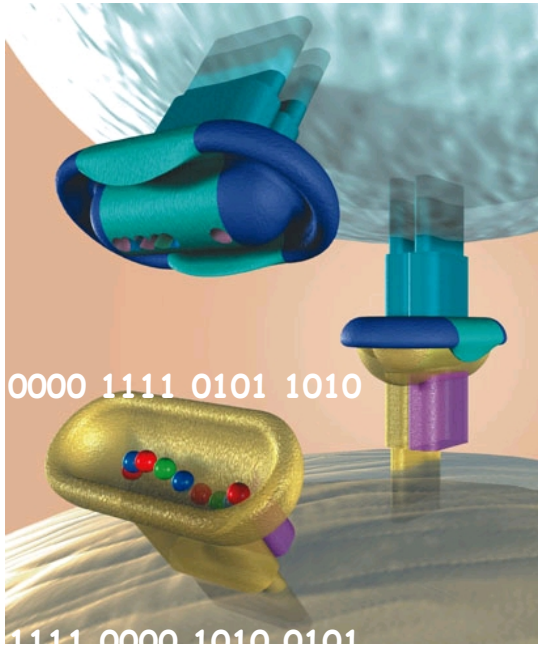2^{24} = 1.6 x 10^7, potential number of different antigens.

l =24 is the maximum bit-string length reached up to date.

True potential BRC repertoire is 10^{11} (~2^{33}).
10^{16} (~2^{50}) is the potential repertoire for TCR

Molecular affinity

0000 1111 0101 1010

1111 0000 1010 0101
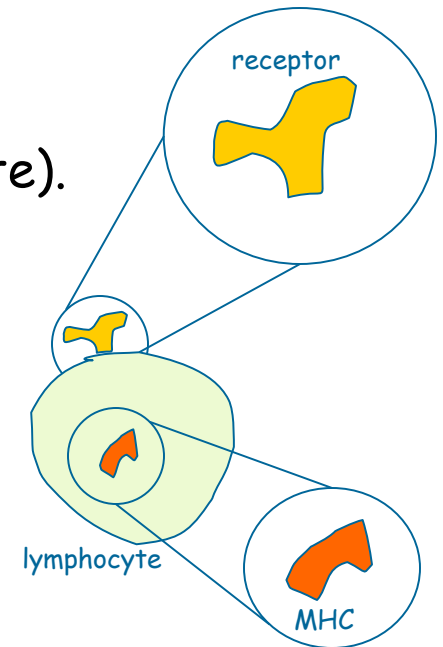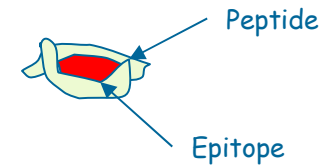
Introducing Molecular Detail

Introduce a pre-computed lookup table of affinities for each pair of peptides from a suitable set of peptides (basic components of cell receptors, antigens, MHC, etc)

F (sequence) = C-ImmSim parameters

Peptide set

(C-ImmSim)
IMMUNOGRID

database
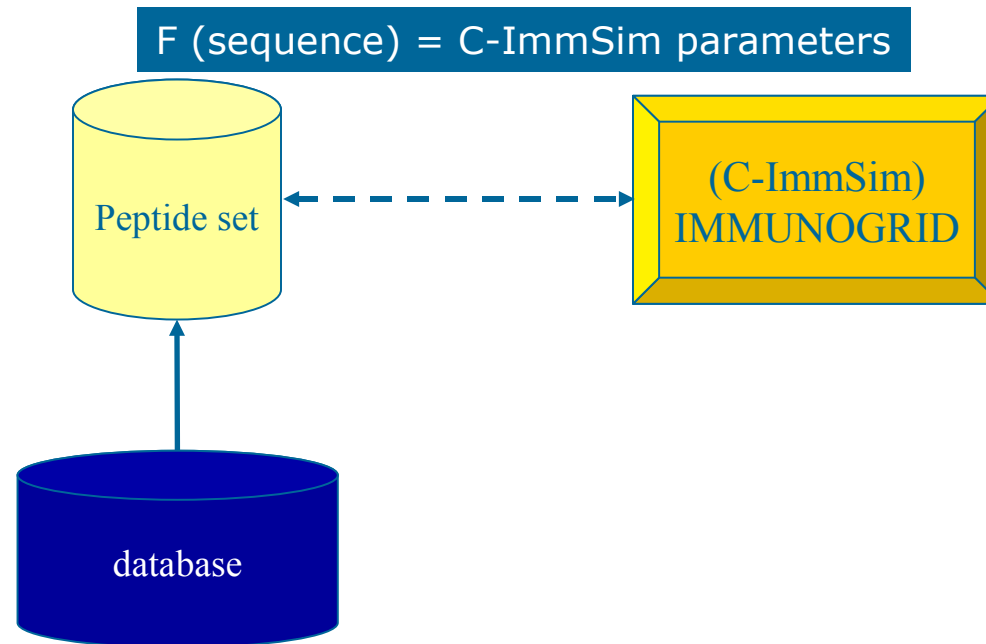
Grid implementation of simulator

- The grid part of the project has not started yet (it starts in year 2) because the implementation will depend on the progress made in other parts of the project (particularly on improvements in the model and extensions made to the simulator program).

- However, we expect the Grid to be used mainly for "number crunching", rather than data federation etc.

Immunogrid CPU requirements

- Depends on what you want to simulate. There are currently different versions of the simulator, with different CPU/memory requirements, e.g.

    - cancer immunoprevention model ("SimTriplex" simulator)

    - Virus infection (e.g. HIV-1 or EBV). Requires more CPU time (+memory) due to need to model virus mutation

    - Generic bacterial infection, hypersensitivity reactions, etc

Immunogrid CPU requirements

- Currently a Grid is not used to perform these ("simple") simulations.

- However, we need to increase the complexity of the models in order to match the complexity we see in the *real* human immune system

Model improvements

- **Dimensions and dimensionality**

  - Most of the models use 2D grids and represent small slices of real tissues (a few mm^3). Need to use 3D grids and scale up to real organ sizes, e.g. 10^3 mm^3 for a mouse.

- **Increase the number of molecule types**

  - CPU times scales as 2^l where l is the bit string length. Currently l=12 (4096 molecule types)but may need l>30 or more to match the immune system repertoire.

- **Introduce molecular complexity**

  - Peptide binding affinities currently modelled by comparing "bit strings" (Hamming distance). A future development could be to read more accurate values of the affinities from a database during the simulation.

Grid implementation

- First requirement of the Grid will be to simulate more complex and larger models of tissues or single organs.

- If the Grid nodes need to heavily communicate the most appropriate Grid available at the European level is DEISA.

- If the Grid nodes need to loosely communicate the most appropriate Grid available at the European level is EGEE.

DEISA is a consortium of leading national supercomputing centres that currently deploys and operates a persistent, production quality, distributed supercomputing environment with continental scope.

The purpose of the infrastructure is to enable scientific discovery across a broad spectrum of science and technology, by enhancing and reinforcing European capabilities in the area of high performance computing.

http://www.deisa.org
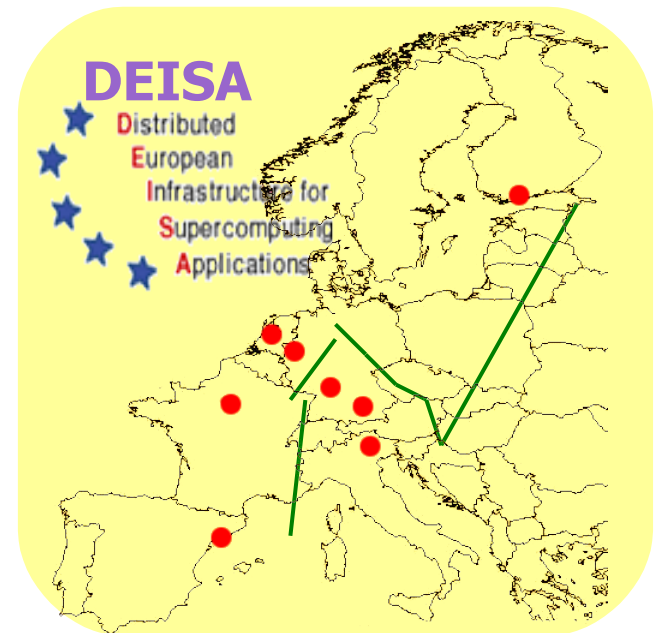
DEISA at-a-glance

- European network of supercomputer centres for high performance computing applications

- IBM-AIX "supercluster" based on IBM SP machines in Germany, France, Italy and Finland.

- Future addition of Linux clusters from Netherlands, UK and Spain (BSC) ("full heterogenous grid")

Using DEISA today

No need to grid-enable applications but currently must run on IBM SP-AIX.

Main integration feature of DEISA is the global distributed file system (gfs)

- a file on the gfs can be accessed with equal performance from any node on the system (data is NOT replicated)

Other possible Grid applications

With the opportunities offered by Grid technologies we can envisage other scenarios later in the project

- modelling many different immune system organs in the same individual

- or even different individuals and populations of individuals (e.g tens or hundreds of mice)

Population statistics are of course important for immunotherapy and vaccines
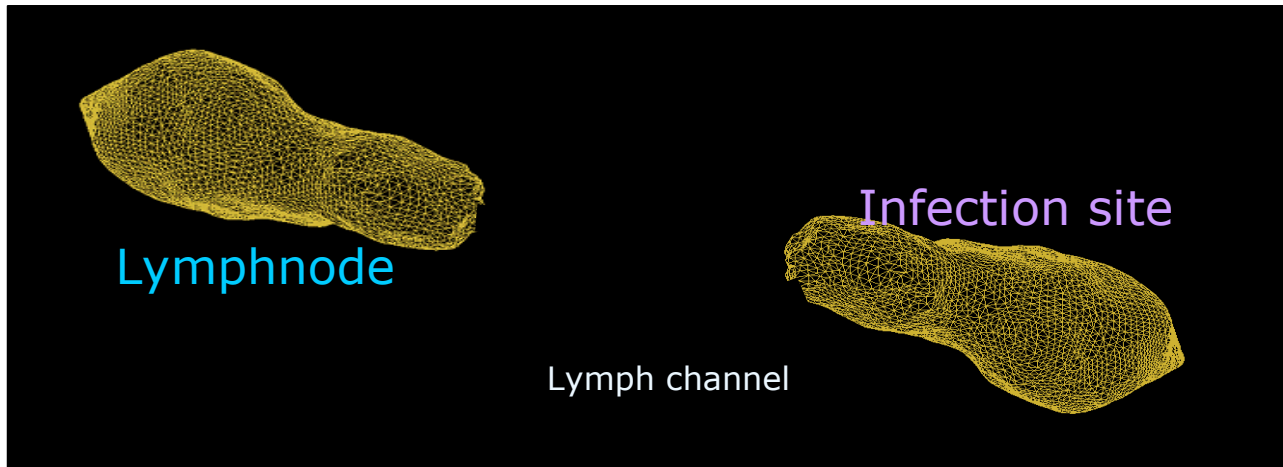
Immunogrid CPU requirements

Example: One of the latest simulations models two lymph nodes joined by a lymph channel:
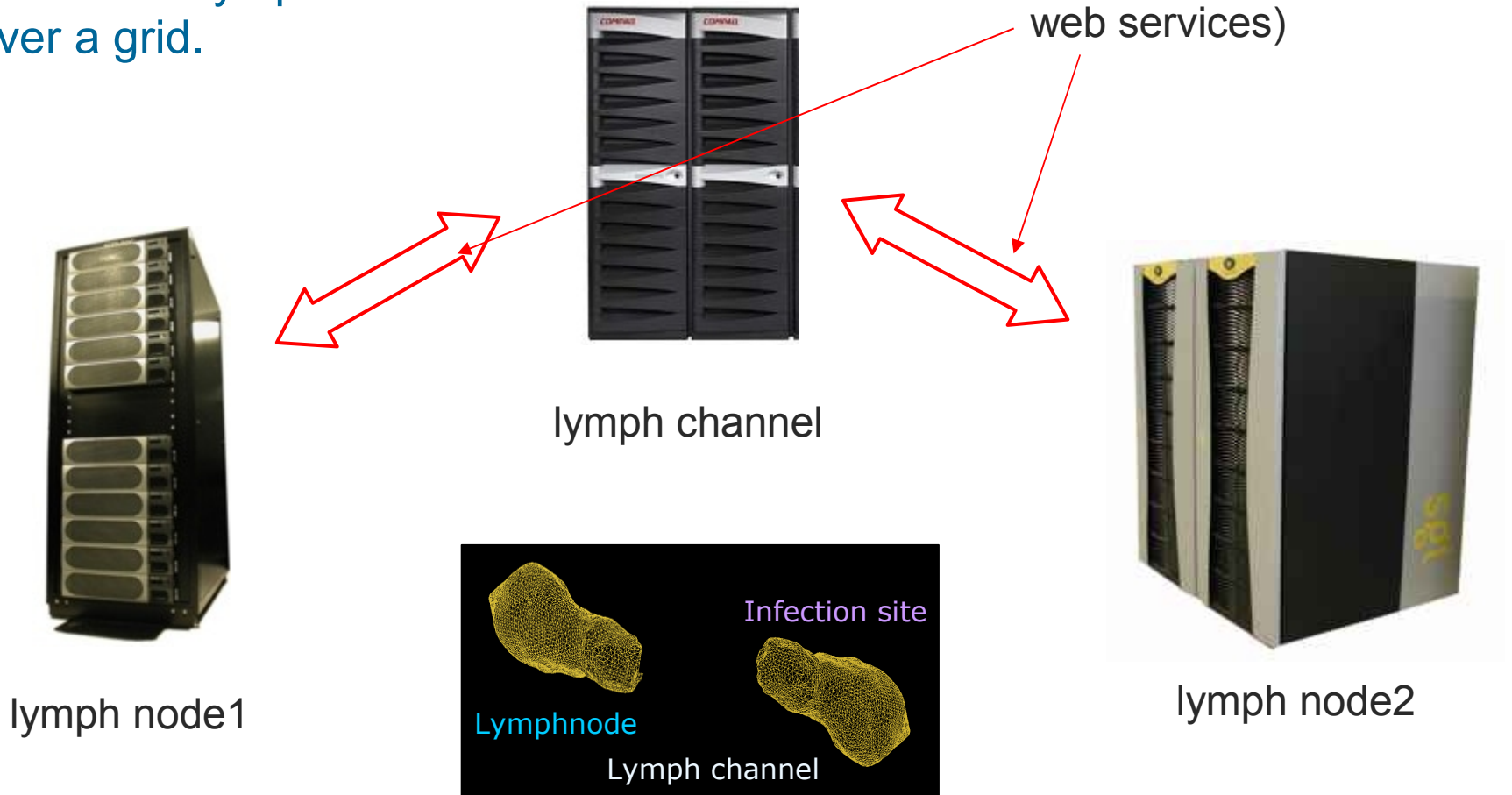
A few days on a shared memory machine with at least 5Gb memory.

(It doesn't include the chemotaxis, movement along a chemical gradient)

Lymphnode

Infection site

Lymph channel

Possible scenario:
simulating two lymph
nodes and lymph channel
over a grid.

communications (e.g.
web services)

lymph channel

lymph node1

Infection site

Lymphnode

Lymph channel

lymph node2

Grid implementation of immune system models

tissue level (a few mm3)

↓

whole organ level (~1000 mm3)

↓

two or more organs in the same individual (e.g. 2 lymph nodes joined by lymph channel)

↓

many individuals in to get population statistics

current situation: possible without grid technology

next step for immunogrid: large cluster or grid such as DEISA (communication between grid nodes required)

each organ on a separate grid node, communicating perhaps by web-services ?

non-interacting individuals, => any high throughput grid (e.g. EGEE, condor, seti@home)

Acknowledgments

- All the partners of the Immunogrid project

- The European Comission for funding (contract no. FP6-2004-IST-028069)

The Adaptive Immune Response

Immunoglobulin

peptide

T cell

B cell

T cell Receptor

MHC

Trimolecular complex