

ISMB/ECCB 2004 Tutorial Proposal

Immunoinformatics: Recent developments and applications

The immune system is one of the best examples of highly complex biological systems. The initiation, regulation and termination of an immune response involve a large number of cells of different types and several stimulatory/inhibitory signals delivered locally and systemically. It is widely accepted that bioinformatics in a broad sense can reveal some answers to the key questions in such a complex system. Therefore, recently the word “Immunoinformatics” has been introduced to define the field of bioinformatics research for better understanding of the immune system.

Often decisions made during an immune response, e.g. whether or not to respond to a microbial infection, or which type of response to make, are based on the information that exist in microbial proteins. These proteins might carry regions that are recognized by B lymphocytes. This recognition can initiate a cascade of processes in the host which results in antibody production against the microbial protein. Similarly, an infected cell can “present” peptides that are generated from the degradation of microbial proteins to immune cells. Indeed, the cellular arm of the immune system, e.g. cytotoxic T lymphocytes, constantly screens cells of the host for such peptides (“epitopes”) and destroys the cells that present them. In other words, the cellular arm of the immune system sees the world through these peptides. The presentation of the peptides to the immune cells is done by major histocompatibility complex (MHC) molecules, which has the largest degree of polymorphism among mammalian proteins. Large part of the immunoinformatics research involves predicting which peptides are most likely to be presented by individual MHC molecules, i.e. how different hosts perceive their world.

The proposed immunoinformatics tutorial will consist of three parts: i) a basic introduction to the immune system, ii) the nature of the immunological data and the details of bioinformatics tools to deal with it, and iii) the applications of immunoinformatics tools, which has a wide range starting from the vaccine design to the evolution of the immune system. We will spend majority of tutorial (3/4) on the techniques concerning the epitopes of the cellular immune response, as we (and the immunoinformatics field) have much more experience and information on generation and presentation of these epitopes.

During the introduction, we aim to make the rest of the tutorial understandable to bioinformaticians without a basic immunological knowledge. The emphasis will be only on the parts of the immunology that will be covered in this tutorial, as a more complete introduction to the immunology would easily occupy one day tutorial on its own.

The second part of the tutorial will be more technical. We will start with the cellular immune response, and go through the steps involved in the processing and presentation of proteins. For each step we will summarize the available data, the nature of the data, and how optimal bioinformatic prediction tools that can be generated from this data. Often the amount of data is very limited, and therefore we will also go through special techniques we developed to compensate the lack of data. Later we will switch to B cell recognition of proteins. B cells can recognize continuous regions (also called sequential or linear epitopes) on the proteins as well as discontinuous regions, which consists of amino acids that are not adjacent in the primary sequence but are brought into proximity by the folding of the polypeptide chain. Therefore, the structural information, e.g. which residues are solvent accessible, or which residues are in contact with each other, is very useful for prediction of B cell epitopes. We will present different methods for predicting continuous and discontinuous epitopes using sequence and structure information.

The final part of the tutorial will include a case story on vaccine design. One of the challenges for the immunoinformatics in the vaccine design field is the genetic description of the human population. As mentioned above, MHC molecules show a large degree of polymorphism, each molecule having a distinct specificity. That is, often the peptides presented by genetically different individuals do not overlap. Therefore, a widely used vaccine should contain multiple peptides, where each individual in the population can present at least one of the peptides. Recently, the immunoinformatics group in CBS has designed a SARS vaccine by combining advanced bioinformatics and high-throughput immunology. We have scanned the SARS genome for cytotoxic

T cell epitopes (cellular immune response) for all nine human MHC supertypes in total covering > 99% of all major human populations. For each supertype, we have selected the 15 top candidates for test in biochemical binding assays. These assays identified more than 85 potential vaccine candidates.

Brief description of the instructor(s): This tutorial will be given by the immunoinformatics team in the Center for Biological Sequence Analysis (CBS) at the Technical University of Denmark. The members of the team are Dr. Morten Nielsen, Dr. Claus Lundegaard, Dr. Peder Worning, Dr. Can Keşmir, and Dr. Ole Lund. Here we give only the descriptions for the ones that would be in charge of the tutorial.

Can Keşmir: Can Keşmir (CK) has 12 years experience as researcher in immunology and bioinformatics. She started her education in bioinformatics in 1991 in Dr. Søren Brunak's group (later named as Center for Biological Sequence Analysis, CBS, in Technical University of Denmark). During 1995-1998 she prepared her thesis in theoretical immunology under supervision of Dr. Rob de Boer. Dr. de Boer's group in Utrecht University (NL) is one of the very few theoretical immunology groups in the world with an internationally respected reputation (see www-binf.bio.uu.nl/rdb/immune for an overview of the work). CK is leading the immunoinformatics branch of this group. Moreover, she is part of the immunoinformatics group in CBS, Denmark. She has been involved in development of several immunoinformatics tools. The most significant one is NetChop (www.cbs.dtu.dk/services/NetChop/), which turned out to be the most reliable proteasome cleavage prediction tool available so far. Since 1999 she is teaching Bioinformatic Pattern Analysis course annually in Utrecht University (see www-binf.bio.uu.nl/BPA/). She is also part of the team giving the Immunological Bioinformatics PhD course in CBS (see www.cbs.dtu.dk/courses/27485.imm/index.php).

Ole Lund: Ole Lund has 12 years experience as researcher in immunology and bioinformatics. In 1995 Mr. Lund defended his PhD thesis "Non linear dynamics in HIV infection" at the Physics Institute at the Technical University of Denmark. From 1993 to 1998 he was employed at the Department of Infectious diseases at Hvidovre, Hospital, Copenhagen, where he established mathematical ADME models of gene therapy against HIV infection, and clarified some of the mechanisms behind antibody mediated enhancement of HIV infection by establishing mathematical models and making computer simulations. In this period he split his time between the Department of Infectious Diseases and the Center for Biological Sequence Analysis (CBS) at the Technical University of Denmark. From 1998 to 2001 he worked at Structural Bioinformatics (SBI) Advanced Technologies, Hørsholm, Denmark, where he participated in establishing a bioinformatics research group and developed improved methods for prediction of protein structure. From December 2001 to July 2002 he worked as a consultant for Plexus Vaccine Inc., where he was responsible for establishing a research group in Denmark. Since August 2002 he has been employed as an associate professor leading the four people strong Immunological Bioinformatics group at the CBS. Mr. Lund's group has developed methods for prediction of T cell epitopes, B cell epitopes, and databases of human pathogens and has developed the world's first PhD level Immunological Bioinformatics course.

Expected goals, objectives, and motivations: The goal/objective is to give an introduction to using bioinformatics in immunology. We want to demonstrate that the bioinformatics tools have a large range of applications starting from vaccine design and extending to more fundamental questions like the study of evolution of the immune system.

The motivation is to attract the attention of bioinformaticians in immunoinformatics, because the immunological area represents a large scientific field where we believe that the bioinformatics can make a large impact.

Tutorial level: Intermediate

Intended audience: The tutorial will be most interesting for bioinformaticians working in general biomedicine. Also "beginner" bioinformaticians with somewhat limited experience in bioinformatics, but have high affinity in immunological/medical questions are welcome, however we will assume that the audience knows the basis of data based learning (e.g. artificial neural networks, and hidden markov models) and phylogenetic methods.

Duration: half day (4hrs).

Detailed outline: The tutorial will consist of 5 lectures given by the team. The talk on MHC binding predictions and MHC polymorphism will last one hour, whereas other lectures are approximately 30 minutes. There will be a total of half an hour for break and half an hour for questions.

- **Introduction to the immune system:** Only the relevant points concerning the immunoinformatics research will be given.
- **Antigen processing in the cellular immune response:** the nature of the data that is available and how to handle it.
- **Prediction of MHC binding:** Data processing, short summary of different techniques with special emphasis on possible drawbacks. MHC polymorphism and supertype structure will be reviewed during this lecture
- **B cell epitope identification:** Use of sequence and structure information
- **A case study:** Integrating the Immunoinformatics tools to design a SARS vaccine.
- Discussion and further questions.

Publications relevant for this proposal

- [1] J. K. Christensen, K. Lamberth, M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemoller, S. Buus, S. Brunak, and O. Lund. Selecting informative data for developing peptide-MHC binding predictors using a query by committee approach. *Neural. Comput.*, 15:2931–2942, 2003.
- [2] M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemoller, K. Lamberth, S. Buus, S. Brunak, and O. Lund. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein. Sci.*, 12:1007–1017, 2003.
- [3] P. Saxova, S. Buus, S. Brunak, and C. Kesmir. Predicting proteasomal cleavage sites: a comparison of available methods. *Int. Immunol.*, 15:781–787, 2003.
- [4] O. Lund, M. Nielsen, C. Kesmir, J.K. Christensen, C. Lundegaard, P. Worning, and S. Brunak. Web-based tools for vaccine design. In B.T. Korber, C. Brander, B.F. Haynes, R. Koup, C. Kuiken, J.P. Moore, B.D. Walker, and D. Watkins, editors, *HIV Molecular Immunology*, pages 45–51, Los Alamos, NM, 2002. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.
- [5] S. Buus, S. L. Lauemoller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm, and S. Brunak. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue. Antigens.*, 62:378–384, 2003.
- [6] C. Kesmir, V. Van Noort, R. J. De Boer, and P. Hogeweg. Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome. *Immunogenetics.*, 55:437–449, 2003.
- [7] K. Yusim, C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detours, and B. T. Korber. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J. Virol.*, 76:8757–8768, 2002.
- [8] C. Kesmir, A. K. Nussbaum, H. Schild, V. Detours, and S. Brunak. Prediction of proteasome cleavage motifs by neural networks. *Protein. Eng.*, 15:287–296, 2002.
- [9] B. Korber, B. Gaschen, K. Yusim, R. Thakallapally, C. Kesmir, and V. Detours. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.*, 58:19–42, 2001.
- [10] S. L. Lauemoller, C. Kesmir, S. L. Corbet, A. Fomsgaard, A. Holm, M. H. Claesson, S. Brunak, and S. Buus. Identifying cytotoxic T cell epitopes from genomic and proteomic information: "The human MHC project.". *Rev. Immunogenet.*, 2:477–491, 2000.
- [11] T. N. Petersen, C. Lundegaard, M. Nielsen, H. Bohr, J. Bohr, S. Brunak, G. P. Gippert, and O. Lund. Prediction of protein secondary structure at 80 *Proteins.*, 41:17–20, 2000.
- [12] J. Gorodkin, O. Lund, C. A. Andersen, and S. Brunak. Using sequence motifs for enhanced neural network prediction of protein distance constraints. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 0:95–105, 1999.
- [13] C. Kesmir and R. J. De Boer. A mathematical model on germinal center kinetics and termination. *J. Immunol.*, 163:2463–2469, 1999.
- [14] J. E. Hansen, O. Lund, J. O. Nielsen, S. Brunak, and J. E. Hansen. Prediction of the secondary structure of HIV-1 gp120. *Proteins.*, 25:1–11, 1996.