

Organization LUND UNIVERSITY Department of Theoretical Physics Sölvegatan 14A SE-223 62 LUND Sweden	Document name DOCTORAL DISSERTATION	
	Date of issue November 2007	Subject designator
	Project name	
Author Liwen You	Sponsoring organization	
Document title Computational Prediction Models for Proteolytic Cleavage and Epitope Identification		
Abstract <p>The biological functions of proteins depend on their physical interactions with other molecules, such as proteins and peptides. Therefore, modeling the protein-ligand interactions is important for understanding protein functions in different biological processes. We have focused on the cleavage specificities of HIV-1 protease, HCV NS3 protease and caspases on short oligopeptides or in native proteins; the binding affinity of MHC molecules with short oligopeptides and identification of T cell epitopes. We expect that our findings on HIV-1 protease, HCV NS3 protease and caspases generalize to other proteases.</p> <p>In this thesis, we have performed analysis on these interactions from different perspectives — we have extended and collected new substrate data sets; used and compared different prediction methods (e.g. linear support vector machines, neural networks, OSRE method, rough set theory and Gaussian processes) to understand the underlying interaction problems; suggested new methods (i.e. a hierarchical method and Gaussian processes with test reject method) to improve predictions; and extracted cleavage rules for protease cleavage specificities.</p> <p>From our studies, we have extended oligopeptide substrate data sets and collected native protein substrates for HIV-1 protease, and a new oligopeptide substrate data set for HCV protease. We have shown that all current HIV-1 protease oligopeptide substrate data sets and our HCV data set are linearly separable; for HIV-1 protease, size and hydrophobicity are two important physicochemical properties in the recognition of short oligopeptide substrates to the protease; and linear support vector machine is the state-of-the-art for this protease cleavage prediction problem. Our hierarchical method combining protein secondary structure information and experimental short oligopeptide cleavage information can improve the prediction of HIV-1 protease cleavage sites in native proteins. Our rule extraction method provides simple and accurate cleavage rules with high fidelity for HIV-1 and HCV proteases. For MHC molecules, we showed that high binding affinities are not necessarily correlated to immunogenicity on HLA-restricted peptides. Our test reject method combined with Gaussian processes can simplify experimental design by reducing false positives for detecting potential epitopes in large pathogen genomes.</p>		
Key words binding affinity, caspase, cleavage prediction, cleavage specificity, epitope, false positive, Gaussian process, HCV, hierarchical method, HIV, immunology, MHC, OSRE, protease-peptide interaction, rule extraction, sequence analysis, SVM		
Classification system and/or index terms		
Supplementary bibliographical information		Language English
ISSN and key title		ISBN 978-91-628-7218-2
Recipient's notes	Number of pages 211	Price
	Security class	

 DOKUMENTTABLAD
 eni SIS 61 41 21

Distributor
 Liwen You
 Department of Theoretical Physics, Sölvegatan 14A, SE-223 62 Lund, Sweden

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources the permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Liwen You

Date 2007-10-18