ว.วิทย. มข. 40(3) 790-801 (2555) KKU Sci. J. 40(3) 790-801 (2012)

Immunoinformatics ศาสตร์ใหม่เพื่อการวิจัยวัคซีน

Immunoinformatics: A New Tool for Vaccine Research

อุมาพร ทาไธสง[1*] *และ ทรงศักดิ์ ทองชูศักดิ์*[2]

บทคัดย่อ

สารสนเทศภูมิคุ้มกันเป็นศาสตร์ใหม่ในวิทยาภูมิคุ้มกันที่ใช้เทคนิคทางชีวสารสนเทศสำหรับศึกษาปฏิสัมพันธ์ระหว่างแอนติเจนและระบบภูมิคุ้มกันของโฮสต์ วิธีดังกล่าวนี้สามารถใช้ในการเก็บรวบรวมข้อมูลทางภูมิคุ้มกัน วิเคราะห์การเกิดปฏิกิริยาข้ามกลุ่มของแอนติบอดี วิเคราะห์ความหลากหลายของแอนติเจน รวมทั้งการวิเคราะห์กระบวนการย่อยและตกแต่งแอนติเจน และนำเสนอแอนติเจน ในปัจจุบันสารสนเทศภูมิคุ้มกันกลายเป็นส่วนสำคัญของการวิจัยด้านภูมิคุ้มกันโดยเฉพาะอย่างยิ่งการศึกษาเกี่ยวกับภูมิคุ้มกันแบบพึ่งเซลล์ในโรคติดเชื้อ โรคภูมิต้านตนเองและมะเร็ง บทความเรื่องนี้จะเน้นเกี่ยวกับการใช้วิธีการทางคอมพิวเตอร์เพื่อทำนาย T-cell epitope จากลำดับกรดอะมิโนของแอนติเจนจากเชื้อก่อโรคและเนื้องอก

ABSTRACT

Immunoinformatics is a new filed in immunology which demonstrates the use of bioinformatics for studying the interaction between antigens and host immune system. The method can be used to organize immunological data, analysis of antibody cross reaction, analysis of antigenic diversity and simulate the antigen processing and presentation. Currently, immunoinformatics become an important part of immunological research especially study of cell mediated immunity for infectious disease, autoimmune and cancer. This article will be focused on the use of computational strategies to predict T-cell epitopes from amino acid sequence of antigens derived from pathogens or tumors.

[1]ภาควิชาจุลชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา จ.ชลบุรี 20131

[2]บริษัทสยามไบโอไซเอนซ์ จำกัด กรุงเทพฯ 10330

*Corresponding Author, E-mail: umaporn@buu.ac.th

คำสำคัญ: สารสนเทศภูมิคุ้มกัน ชีวสารสนเทศ T-cell epitope แอนติเจน
Keywords: Immunoinformatics, Bioinformatics, T-cell epitope, Antigen

บทนำ

Immunoinformatics เป็นศาสตร์ใหม่ของภูมิคุ้มกันวิทยาที่อาศัยเทคโนโลยีชีวสารสนเทศ (bioinformatics) มาใช้ในการวางแผนและวิเคราะห์งานวิจัยทางด้านภูมิคุ้มกันวิทยา (Flower, 2007) วิธีดังกล่าวนี้มีความสะดวก ราคาถูก และช่วยให้ได้ผลการทดลองที่รวดเร็วกว่าวิธีดั้งเดิม ด้วยเหตุนี้ immunoinformatics จึงเป็นศาสตร์ใหม่ที่นักวิจัยควรให้ความสำคัญ ซึ่ง immunoinformatics มีพัฒนาการมาจาก bioinformatics โดยมุ่งเน้นการวิเคราะห์ข้อมูลทางด้านดีเอ็นเอ และโปรตีนแอนติเจน หน้าที่เซลล์ในระบบภูมิคุ้มกันหรือเชื้อก่อโรค (Korber et al., 2006) เนื่องจากพัฒนาการทางด้านเครื่องมือวิจัยทางชีววิทยาการแพทย์ได้เจริญรุดหน้าไปอย่างรวดเร็ว ส่งผลให้ข้อมูลต่าง ๆ มีปริมาณมากขึ้นและยากต่อการวิเคราะห์ด้วยวิธีธรรมดาทั่วไป ดังนั้นนักวิทยาศาสตร์จึงได้นำเทคโนโลยีทางด้านคอมพิวเตอร์และระบบเครือข่ายอินเทอร์เน็ตเข้ามาใช้ประกอบการวิเคราะห์ข้อมูล ซึ่งวิธีดังกล่าวพบว่าสะดวก รวดเร็ว และแม่นยำ เนื่องจากศาสตร์นี้เป็นเรื่องใหม่สำหรับประเทศไทย บทความนี้จึงมีวัตถุประสงค์เพื่อเผยแพร่ความรู้ดังกล่าวให้กับนักวิจัยไทยและผู้สนใจทั่วไป เพื่อเป็นแนวทางในการศึกษาวิจัยภูมิคุ้มกันวิทยาแนวใหม่ต่อไป

เครื่องมือและโปรแกรมที่เกี่ยวข้องกับ Immunoinformatics

สำหรับเครื่องมือที่เกี่ยวข้องกับ Immunoinformatics นั้น โดยทั่วไปประกอบด้วยส่วนสำคัญ 2 ส่วนคือ ฐานข้อมูลชีวโมเลกุล (molecular databases) และโปรแกรมสำหรับวิเคราะห์หรือทำนายข้อมูล ในส่วนของฐานข้อมูลนั้นสามารถแบ่งย่อยออกเป็นสองส่วนคือ ฐานข้อมูลปฐมภูมิ (primary sequence database) และฐานข้อมูลทุติยภูมิ หรือที่เรียกว่าฐานข้อมูลเฉพาะทาง (secondary sequence หรือ specialized database) ในส่วนของโปรแกรมนั้นมีตั้งแต่โปรแกรมอย่างง่ายเพื่อวิเคราะห์ข้อมูลระดับดีเอ็นเอและโปรตีน ไปจนถึงโปรแกรมที่ใช้ในการวิเคราะห์คุณสมบัติบางประการของโปรตีน ตัวอย่างเช่น โปรแกรมที่ใช้ทำนายความเป็นแอนติเจนหรือปฏิสัมพันธ์ระหว่างการจับกันของโปรตีน ซึ่งจะกล่าวในลำดับต่อไป

1. ฐานข้อมูลชีวโมเลกุล (Molecular databases)

1.1 ฐานข้อมูลปฐมภูมิ (primary sequence database) ได้แก่ ฐานข้อมูลที่เก็บรวบรวมข้อมูลดีเอ็นเอและโปรตีนของสิ่งมีชีวิตทุกชนิด เช่น GenBank, SwissProt, Protein Data Bank (PDB) (Mizrachi, 2008) เป็นต้น ฐานข้อมูลดังกล่าวมีข้อเสียคือข้อมูลมักมีความซ้ำซ้อน เนื่องจากไม่มีการคลีน (clean) ข้อมูล (data cleaning) บางครั้งอาจพบข้อมูลที่มีความผิดปกติเนื่องจากความผิดพลาดในการส่งข้อมูลของลำดับดีเอ็นเอและลำดับกรดอะมิโน (submit sequence) เช่น unknown amino acid sequence XXXXX หรือมีแต่กรดอะมิโนเมทไธโอนีน (methionine) เพียงชนิดเดียว ข้อดีของฐานข้อมูลประเภทนี้คือมีข้อมูลครบถ้วนและอยู่บนระบบเครือข่ายอินเทอร์เน็ต หากต้องศึกษาเกี่ยวกับ

แหล่งที่มาหรือความแปรผันของโปรตีนแอนติเจนของเชื้อสายพันธุ์ที่มีลำดับพันธุกรรมเหมือนกัน

1.2 ฐานข้อมูลทุติยภูมิ หรือฐานข้อมูลเฉพาะทาง (secondary sequence หรือ specialized database) ได้แก่ฐานข้อมูลที่ผ่านการคลีนข้อมูลแล้ว ฐานข้อมูลชนิดนี้ประกอบด้วยข้อมูลเฉพาะทางเรื่องใดเรื่องหนึ่ง เช่น ฐานข้อมูลของปัจจัยก่อความรุนแรงของ *Candida albicans* (*Candida albicans* virulence factor database) (Tongchusak et al., 2005) และฐานข้อมูลของ human leukocyte antigen (HLA) (Robinson, 2003) เป็นต้น ข้อดีของข้อมูลประเภทนี้คือเป็นข้อมูลประเภทเดียว มักจะมีวิธีการสืบค้นที่ตรงกับความต้องการของผู้ใช้เฉพาะเรื่อง ไม่มีการซ้ำซ้อนของข้อมูลและอยู่บนระบบเครือข่ายอินเทอร์เน็ต ซึ่งสามารถแบ่งปันข้อมูลระหว่างนักวิจัยได้ อีกประการหนึ่งคือข้อมูลมีการปรับปรุงให้ทันสมัยอยู่ตลอดเวลา สำหรับข้อเสียคือ ใช้วิเคราะห์เฉพาะเรื่องใดเรื่องหนึ่งเท่านั้น ไม่สามารถเปรียบเทียบข้อมูลกับเรื่องอื่นได้โดยตรง ต้องอาศัยการเก็บข้อมูลแล้วนำมาวิเคราะห์ภายหลัง

2. โปรแกรมสำหรับวิเคราะห์หรือทำนายข้อมูล

2.1 โปรแกรมพื้นฐาน (basic program) ได้แก่ โปรแกรมสำหรับการค้นหาข้อมูล (data mining) โปรแกรมเปรียบเทียบลำดับดีเอ็นเอหรือโปรตีน เช่น pairwise alignment, basic local alignment search tool (BLAST) (Altschul et al., 1990) และ multiple sequence alignment programs (ClustalW, MAFFT, MUSCLE) (Subramanian et al., 2008) เป็นต้น โปรแกรมสำหรับดูโครงสร้างสามมิติของโปรตีน เช่น PyMol (DeLano, 2000), RasMol (Sayle and Milmer-White, 1995; Bernstein, 2000) และ Jmol (Hanson, 2010) เป็นต้น

2.2 โปรแกรมเฉพาะทาง (specialized program) ได้แก่ โปรแกรมการทำนาย B-cell epitope (Reitmaier, 2007) และโปรแกรมการทำนาย HLA binding peptide (Lin et al., 2008) เป็นต้น

การประยุกต์ใช้ Immunoinformatics ในงานวิจัยภูมิคุ้มกันวิทยา

1. Immunoinformatics กับโรคภูมิแพ้ (Allergy)

สารก่อภูมิแพ้คือกลุ่มของสารชีวโมเลกุลจากสิ่งมีชีวิต หรือโลหะจากสิ่งของเครื่องใช้ในชีวิตประจำวันที่มีคุณสมบัติกระตุ้นปฏิกิริยาภูมิแพ้ ในที่นี้จะกล่าวถึงเฉพาะสารก่อภูมิแพ้จากสิ่งมีชีวิตที่เป็นสาเหตุของภูมิแพ้ สารเหล่านี้เป็นโปรตีนเมื่อเข้าสู่ร่างกายครั้งแรกจะถูกจับกินด้วย antigen presenting cell (APC) เช่น langerhans cell ที่ผิวหนัง เป็นต้น (Kay, 2001) จากนั้น APC จะย่อยและตกแต่งแอนติเจน (antigen processing) รวมทั้งนำเสนอแอนติเจน (antigen presentation) ให้กับ T-cells เมื่อ T-cells รับรู้แอนติเจน (antigen recognition) จะถูกกระตุ้น มีผลทำให้ T-cells เพิ่มจำนวนและเปลี่ยนแปลง (proliferation and differentiation) ไปเป็น T helper 2 cell ซึ่งหลั่งไซโตไคน์ (cytokine) ชนิด interleukin (IL)-4, IL-5, IL-9 และ IL-13 (Kay, 2001) โดยไซโตไคน์เหล่านี้จะไปส่งเสริม B-cells ให้เกิด antibody class switching เป็นแอนติบอดีชนิด IgE ในขณะเดียวกัน IL-4, IL-5 และ IL-9 ยังไปกระตุ้นพัฒนาการของเซลล์ต้นกำเนิดให้เป็น Eosinophil และ Mast cell (Kay, 2001) เมื่อสารก่อภูมิแพ้เข้ามาใน

ร่างกายครั้งที่สองจะถูกจับด้วยแอนติบอดีชนิด IgE เกิดเป็น immune complex จากนั้น complex จะจับกับ IgE receptor บนผิวของ mast cells และ basophil แบบแน่นมาก (high affinity) (Kay, 2001) ส่งผลให้เซลล์ดังกล่าวมีการหลั่งสาร allergic mediator ออกมาในปริมาณมากทำให้มีอาการต่างๆ เช่น คัดจมูก จาม น้ำมูกไหล คันคอ เจ็บคอ ไอ มีเสมหะ เสียงแหบแห้ง และลงไปยังหลอดลม ทำให้หลอดลมตีบตัน เป็นหอบหืด หรือเป็นผื่นคัน ลมพิษ ตาบวม ท้องเสีย แสบตา คันตา หนังตาบวม น้ำตาไหล ทั้งนี้อาการดังกล่าวขึ้นอยู่กับระบบของร่างกายที่เกิดปฏิกิริยาภูมิแพ้ (Kay, 2001) จากการศึกษาพบว่าสารก่อภูมิแพ้จากสิ่งมีชีวิตหลายชนิดมีความคล้ายคลึงกันในเชิงโครงสร้างของโปรตีน และสามารถทำให้เกิดภูมิแพ้ได้โดยไม่ต้องได้รับสารเดิมนั้นอีก เช่น สารก่อภูมิแพ้ในเกสรตัวผู้ของหญ้ามีความคล้ายกับแอปเปิ้ล และมะเขือเทศ เป็นต้น เรียกปฏิกิริยาที่เกิดจากสารก่อภูมิแพ้เหล่านี้ว่า ปฏิกิริยาข้ามกลุ่มของสารก่อภูมิแพ้ (allergen cross-reactivity) (Brusic and Petrovsky, 2003)

immunoinformatics ที่เกี่ยวข้องกับโรคภูมิแพ้ ประกอบด้วย ฐานข้อมูลและโปรแกรมที่ใช้วิเคราะห์ข้อมูล ซึ่งรายชื่อเว็บไซต์ที่เกี่ยวข้องกับสารก่อภูมิแพ้ แสดงดังตารางที่ 1 และสำหรับการทำนายปฏิกิริยาข้ามกลุ่มของสารก่อภูมิแพ้นั้น สามารถทำได้ทางอ้อมด้วยวิธีการเปรียบเทียบลำดับกรดอะมิโน (sequence comparison) โดยอาศัยโปรแกรม BLAST หรือ multiple sequence alignment เพื่อหาแนวโน้มของปฏิกิริยาข้ามกลุ่ม (potential cross reactivity) และแนวโน้นของโปรตีนที่เป็นสารก่อภูมิแพ้ (potential allergenic protein) (Brusic and Petrovsky, 2003) จากรูปที่ 1 แสดงการทำนายปฏิกิริยาข้ามกลุ่มของ B-cell epitope ของสารก่อภูมิแพ้ชนิดเอนไซม์ serine protease (serine protease allergens B-cell epitope) ของเชื้อรา *Aspergillus niger* โดยใช้โปรแกรม BLAST ของ NCBI จากผลการเปรียบเทียบลำดับกรดอะมิโนจะเห็นได้ว่านอกจากเชื้อราชนิดนี้แล้วยังมีเชื้อราชนิดอื่นที่มีการสร้างสารก่อภูมิแพ้ชนิดดังกล่าว อย่างไรก็ตามวิธีการดังกล่าวนี้เป็นวิธีการวิเคราะห์ลำดับกรดอะมิโนแบบต่อเนื่อง (continuous หรือ linear epitope) แต่ในความเป็นจริงแล้ว ธรรมชาติของ B-cell epitope ส่วนใหญ่เป็นแบบไม่ต่อเนื่อง (discontinuous หรือ comformational epitope) (Abbas et al., 2000) ดังนั้นการทำนาย allergenicity อีกวิธีหนึ่งคือการวิเคราะห์ที่โครงสร้างสามมิติ (3-dimensional structure) ซึ่งโครงสร้างดังกล่าวได้มาจากการทำการทดลอง protein crystallography หรือการทำแบบจำลองโครงสร้างสามมิติของโปรตีนที่สนใจโดยวิธี homology modeling (หาข้อมูลเพิ่มเติมได้ที่ Swiss-model; http://swissmodel.expasy.org/SWISS-MODEL.html) วิธีดังกล่าวนี้เรียกว่า structural comparison

2. immunoinformatics กับมะเร็งและโรคติดเชื้อ

การประยุกต์ใช้ bioinformatics กับมะเร็งและโรคติดเชื้อนั้นส่วนใหญ่เป็นการทำนาย HLA-binding peptide ในกระบวนการตอบสนองทางภูมิคุ้มกันนั้น เมื่อแอนติเจนถูกจับกินโดย APC แล้วจะถูกย่อยเป็นเปปไทด์ (peptide) สายสั้น ๆ และนำเสนอบน HLA ให้กับ T-cells เพื่อเหนี่ยวนำให้เกิดการตอบสนองทางภูมิคุ้มกันในลำดับต่อไป การนำเสนอแอนติเจนบนโมเลกุล HLA นั้นมีสองระบบคือ

2.1 การนำเสนอแอนติเจนบน Class I HLA (Class I HLA antigen presentation)

เป็นกระบวนการนำเสนอแอนติเจนที่อยู่ภายในเซลล์ (endogenous antigens) เช่น แอนติเจนของตัวเอง (self-antigen) แอนติเจนของมะเร็ง (tumor antigen) และแอนติเจนของไวรัส (viral antigen) เป็นต้น โปรตีนเหล่านี้จะถูกติดฉลากด้วยโปรตีนยูบิไควติน (ubiquitin) และย่อยผ่านโปรติ-เอโซม (proteasome) (หรือ immunoproteasome ของ dendritic cells) ได้เป็นเปปไทด์สายสั้น ๆ ซึ่งจะถูกส่งผ่านเข้าไปในเอนโดพลาสมิกเรติคูลัม (endoplasmic reticulum) ผ่านทางโปรตีน TAP1/TAP2 (transporter associated with antigen processing) และฝังลงในร่อง (peptide binding groove) ของ Class I HLA และนำเสนอให้กับ CD8+ T-cell ซึ่งนำไปสู่การตอบสนองทางภูมิคุ้มกันในที่สุด โดยเปปไทด์ที่จับกับ Class I HLA นั้นส่วนใหญ่มีความยาวเท่ากับ 8-11 กรดอะมิโน แต่มีเพียง 8-9 กรดอะมิโนเท่านั้นที่จับกับ HLA โดยตรง (Abbas et al., 2000)

2.2 การนำเสนอแอนติเจนบน Class II HLA (Class II HLA antigen presentation)

เป็นกระบวนการนำเสนอแอนติเจนที่มาจากภายนอกเซลล์ (Exogenous antigens) ได้แก่ แบคทีเรีย รา สารก่อภูมิแพ้ เป็นต้น ขั้นตอนการนำเสนอแอนติเจนเริ่มต้นจากแอนติเจนถูกจับกินด้วยฟาโกไซต์โดยกระบวนการฟาโกไซโทซิส (phagocytosis) หรือนำเข้าสู่เซลล์โดย receptor mediated endocytosis จากนั้นแอนติเจนจะถูกย่อยด้วยเอนไซม์ในฟาโกไลโซโซม (phagolysosome) และรวมกับ MHC II rich compartment ที่อยู่ในไซโท-พลาซึม ซึ่งชิ้นส่วนของเปปไทด์จะถูกฝังลงในร่องของ Class II HLA และนำเสนอให้กับ CD4+ T-cell นำไปสู่การตอบสนองทางภูมิคุ้มกันตามลำดับ โดยเปปไทด์ที่จับกับ Class II HLA นั้นส่วนใหญ่มีความยาวเท่ากับ 12-20 กรดอะมิโน แต่มีเพียง 9 กรดอะมิโนเท่านั้นที่จับกับ HLA โดยตรง (Abbas et al., 2000)

ตารางที่ 1 แสดงรายชื่อเว็บไซต์ที่เกี่ยวข้องกับชีววิทยาของสารก่อภูมิแพ้ (allergen biology)

Website name	URL
Nomenclature and general information databases:	
All Allergy	http://allallergy.net/
IUIS (International Union of Immunological Societies)	http://www.allergen.org
Allergome	http://www.allergome.org
CSL (Central Science Laboratory, UK)	http://www.csl.gov.uk/allergen/index.htm
National Center for Food Safety and Technology	http://www.iit.edu/~sgendel/fa.htm
InformAll	http://foodallergens.ifr.ac.uk/
Cross-referenced databases with tools to compare sequences:	
ADFS – Allergen Database for Food Safety	http://allergen.nihs.go.jp/ADFS/
Cross-referenced databases with tools for prediction of allergenicity:	
SDAP (Structural database of Allergenic proteins) and SDAP-FOOD	http://fermi.utmb.edu/SDAP
Web servers for allergenicity prediction:	
Allermatch	http://www.allermatch.org/
AlgPred	http://www.imtech.res.in/raghava/algpred/

รูปที่ 1 แสดงวิธีการทำนายปฏิกิริยาข้ามกลุ่มของ B-cell epitope ของสารก่อภูมิแพ้ชนิดเอนไซม์ serine protease (serine protease allergens B-cell epitope) ของเชื้อรา *Aspergillus niger* โดยวิธีการเปรียบเทียบลำดับกรดอะมิโน ขั้นตอนการทำนายเริ่มต้นโดยการส่งข้อมูลของลำดับกรดอะมิโนของสารก่อภูมิแพ้ที่ทราบชนิดแล้ว (known allergen) ลงในโปรแกรม BLAST ของ NCBI จากนั้นดาวน์โหลดลำดับกรดอะมิโนทั้งหมดมาทำการเปรียบเทียบแบบ multiple sequence alignment โดยใช้โปรแกรม Bioedit ในการ alignment

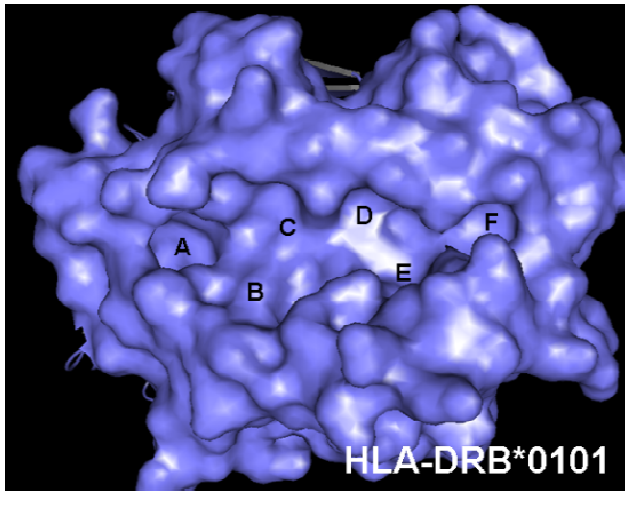
จากหลักการจับกันอย่างจำเพาะของเปปไทด์แอนติเจน (antigenic peptide) กับ HLA นี้เอง ทำให้เกิดการพัฒนาเครื่องมือทางชีวสารสนเทศ (bioinformatics) สำหรับการทำนาย HLA binding peptide ขึ้น โดยธรรมชาติของโมเลกุล HLA นั้น ส่วนที่ใช้ในการนำเสนอเปปไทด์แอนติเจนนั้นมีลักษณะคล้ายพาน ประกอบด้วยร่อง (peptide binding groove) ที่มีคุณสมบัติเป็น hydrophobic สำหรับให้เปปไทด์จับ จำนวน 6 ร่องด้วยกันได้แก่ A-F (รูปที่ 2) สมมุติว่าเปปไทด์แอนติเจนมีความยาวเท่ากับ 9 กรดอะมิโน ดังนั้นกรดอะมิโนตำแหน่งที่ 1 2 และ 3 (a1, a2 และ a3) ของเปปไทด์แอนติเจนจะฝังลงในร่อง A, B และ C ส่วนกรดอะมิโนตำแหน่งที่ 6 7 และ 9 (a6, a7 และ a9) ของเปปไทด์แอนติเจนจะฝังลงในร่อง D, E และ F โดยกรดอะมิโนตำแหน่งที่ 2 และ 9 จะมีอิทธิพลต่อการจับกันของ HLA และเปปไทด์ (Matsumura et al., 1992)

ปัจจุบันพบว่า HLA ของมนุษย์มีทั้งหมด 7,527 อัลลีล (allele) แบ่งออกเป็น Class I HLA จำนวน 5,880 อัลลีล และ Class II HLA จำนวน 1,647 อัลลีล (Robinson, 2003) โดยอัลลีลต่าง ๆ นี้สามารถนำมาจัดกลุ่มตามรูปแบบ motif ของร่องที่ใช้จับกับเปปไทด์ได้เป็นกลุ่มต่าง ๆ เรียกว่า supertype เช่น supertype A1 A2 A3 และ B44 เป็นต้น เป็นที่ทราบแน่ชัดว่าลักษณะของเปปไทด์แบบใดสามารถจับกับ HLA อัลลีล (allele) ไหน เปปไทด์หนึ่ง ๆ อาจก่อหรือไม่ก่อให้เกิดการตอบสนองทางภูมิคุ้มกัน ทั้งนี้ขึ้นอยู่กับชนิดของ HLA แต่โดยส่วนมากจะตอบสนองทางภูมิคุ้มกันใน HLA supertype เดียวกัน นอกจากนี้ยังมีเปปไทด์บางประเภทที่สามารถจับกับ HLA ได้มากกว่า 1 supertype (multiple HLA supertype) เรียกว่า promiscuous peptide ปัจจุบันมีฐานข้อมูลที่รวบรวมลำดับกรดอะมิโนของ T-cell epitope และ HLA-ligand ที่ได้จากการทดลองไว้ด้วยกันมีชื่อว่า SYFPEITHI (http://www.syfpeithi.de/home.htm) ข้อมูลเหล่านี้เองทำให้สามารถทำนาย T-cell epitope ได้ ตามคุณสมบัติทางเคมีฟิสิกส์ (physicochemical property) ของกรดอะมิโนแต่ละตำแหน่ง ขั้นตอนการพัฒนา algorithm สำหรับทำนายนั้นมีอยู่หลายวิธีด้วยกัน ซึ่งอาศัยหลักการทางคณิตศาสตร์และสถิติ ได้แก่ matrix-based method และ machine learning เป็นต้น โดยความแม่นยำของแต่ละวิธีขึ้นอยู่กับจำนวนเปปไทด์ที่ได้จากการทดลองจริงที่ใช้เป็น train dataset ของ HLA แต่ละอัลลีล ถ้าหากมีเปปไทด์จำนวนมากจะยิ่งทำให้ผลการทำนายมีความแม่นยำสูงขึ้น ซึ่งในที่นี้ไม่กล่าวถึงรายละเอียดของวิธีการทำนาย

สำหรับการประยุกต์ใช้ immunoinformatics กับมะเร็งและโรคติดเชื้อนั้นอาจทำได้โดยอาศัยโปรแกรมที่แสดงดังตารางที่ 2 และ 3 ซึ่งเป็นโปรแกรมสำหรับทำนาย T-cell epitope ที่เปิดให้ใช้บริการเป็นสาธารณะ ขั้นตอนการทำนายนั้นอันดับแรกต้องมีโมเลกุลเป้าหมายที่เราสนใจ เพื่อทำการศึกษาภูมิคุ้มกันแบบจำเพาะที่เกิดขึ้นภายหลังจากได้รับสิ่งแปลกปลอม (adaptive immunity) เช่น แอนติเจนของมะเร็ง (tumor antigen) หรือปัจจัยก่อความรุนแรง (virulence factors) ของเชื้อก่อโรค เป็นต้น คัดเลือกลำดับกรดอะมิโนของโปรตีนแอนติเจนที่ต้องการและใส่ลงในโปรแกรมที่จะทำนาย ซึ่งโปรแกรมจะทำการอ่านลำดับกรดอะมิโนทีละ 9 กรดอะมิโน โดยเลื่อนไปทางขวาทีละ 1 กรดอะมิโน แล้วคำนวณหาค่าความเป็นไปได้ที่เปปไทด์จะจับกับ HLA ซึ่งรายงานผลเป็นตัวเลข ข้อควรระวังในการอ่านผลการทำนายคือการแปลความหมายของตัวเลข ซึ่งแต่ละโปรแกรมให้ค่า

การทำนายที่แตกต่างกัน ตัวอย่างเช่น IEDB (ANN) รายงานค่าตัวเลขน้อยไปหามาก แสดงว่าค่าน้อยหมายถึงเปปไทด์มีโอกาสจับได้ดีกว่าค่ามาก (the lower the high binder) ในทางตรงกันข้ามนั้น Multipred (ANN) รายงานค่าตัวเลขมากไปหาน้อย แสดงว่าค่ามากหมายถึงเปปไทด์มีโอกาสจับได้ดีกว่าค่าน้อย (the higher the better)

ในการทำนายนั้นจะได้ผลการทำนายเป็นเปปไทด์จำนวนมากตามความยาวของโปรตีน ปัญหาสำคัญคือ ต้องใช้เปปไทด์ทั้งหมดนี้หรือไม่ ซึ่งจริง ๆ แล้วไม่ต้องใช้ โดยแต่ละโปรแกรมได้กำหนดค่า threshold ของการเลือกเปปไทด์ไปใช้อยู่แล้ว แต่บางครั้งอาจมีจำนวนมากเกินงบประมาณการวิจัย ดังนั้นในการคัดกรองเปปไทด์เพื่อนำไปใช้จึงควรมีเปปไทด์ควบคุม (control peptides) ในที่นี้คือเปปไทด์ที่ได้จากการทดลองจริง (experimentally characterized T-cell epitopes) ประโยชน์ของเปปไทด์ควบคุม คือช่วยทำให้ค่า threshold น้อยลง วิธีการหาค่า threshold ทำได้โดยทำนาย binding score ของเปปไทด์ควบคุม จากนั้นทำการจัดเรียง (sort) จากน้อยไปหามาก หากต้องการ epitope จำนวน 100% ให้ใช้ค่าสูงสุดเป็น threshold และหากต้องการ 50% หรือน้อยกว่านั้นจะต้องใช้เท่าไหร่ ตารางที่ 4 เป็นตัวอย่างวิธีการหาค่า threshold ของ T-cell epitope ของมะเร็งที่จำเพาะกับ HLA-A*0201 จากตารางพบว่า เปปไทด์ควบคุมมีจำนวนทั้งหมด 84 เปปไทด์ แต่ละเปปไทด์มีค่าการจับกันที่ได้จากการทำนายของโปรแกรมจาก immune epitope database (IEDB) เปปไทด์ทั้งหมดคิดเป็น 100% หากต้องการให้ได้ positive peptide ครอบคลุมประมาณ 10% ต้องใช้ binding score ที่ threshold เท่ากับ 9 nM ดังนั้นทุกครั้งที่ทำการทำนายแอนติเจนของมะเร็งจะใช้ค่า binding score ของเปปไทด์ควบคุมซึ่งเป็นตัวบ่งชี้ธรรมชาติของเปปไทด์จริง เป็นแนวทางในการเลือกจำนวนเปปไทด์ที่จะนำไปใช้ในการทดลอง ข้อสำคัญคือควรหมั่นตรวจสอบจำนวนเปปไทด์ควบคุมจากวรรณกรรม (literature) บ่อย ๆ เพื่อให้ได้ค่าการทำนายที่แม่นยำมากที่สุด สำหรับโปรแกรมที่ใช้เป็นมาตรฐานในการทำนายนั้นสามารถตรวจสอบได้จากงานวิจัยของ Lin et al. (2008)



รูปที่ 2 แสดง HLA-binding groove ของ HLA-DRB1*0101 ซึ่งประกอบด้วยร่อง A-F (ดัดแปลงจาก: protein data bank)

ตารางที่ 2 รายชื่อ prediction servers สำหรับทำนาย HLA-I binding peptides (URLs as of April 2007)

Servers	URLs
BIMAS	http://www-bimas.cit.nih.gov/molbio/hla_bind/
HLA Ligand	http://hlaligand.ouhsc.edu/prediction.htm
IEDB (ANN)	http://tools.immuneepitope.org/analyze/html/mhc_binding.html
IEDB (ARB)	http://tools.immuneepitope.org/analyze/html/mhc_binding.html
IEDB (SMM)	http://tools.immuneepitope.org/analyze/html/mhc_binding.html
MAPPP (Bimas)	http://www.mpiib-berlin.mpg.de/MAPPP/binding.html
MAPPP (SYFPEITHI)	http://www.mpiib-berlin.mpg.de/MAPPP/binding.html
MHC Binder Prediction	http://www.vaccinedesign.com/
MHC-BPS	http://bidd.cz3.nus.edu.sg/mhc/
MHC-I (Multiple matrix)	http://atom.research.microsoft.com/hlabinding/hlabinding.aspx
MHC-I (Single matrix)	http://atom.research.microsoft.com/hlabinding/hlabinding.aspx
MHCPred (Interactions)	http://www.jenner.ac.uk/MHCPred/
MHCPred (Amino Acids)	http://www.jenner.ac.uk/MHCPred/
MULTIPRED (ANN)	http://sdmc.lit.org.sg:8080/multipred1/
MULTIPRED (HMM)	http://sdmc.lit.org.sg:8080/multipred1/
MULTIPRED (SVM)	http://sdmc.lit.org.sg:8080/multipred1/
NetMHC (ANN)	http://www.cbs.dtu.dk/services/NetMHC/
NetMHC (Weight Matrix)	http://www.cbs.dtu.dk/services/NetMHC/
nHLAPred (ANNPred)	http://www.imtech.res.in/raghava/nhlapred/neural.html
nHLAPred (ComPred)	http://www.imtech.res.in/raghava/nhlapred/comp.html
PepDist	http://www.pepdist.cs.huji.ac.il/
PeptideCheck (Matrix)	http://www.peptidecheck.org/
Predep	http://margalit.huji.ac.il/Teppred/mhc-bind/index.html
ProPred1	http://www.imtech.res.in/raghava/propred1
Rankpep	http://bio.dfci.harvard.edu/Tools/rankpep.html
SMM	http://zlab.bu.edu/SMM/
SMMPMBEC	http://www.mhc-pathway.net/smmpmbec
SVMHC (MHCPEP)	http://www.sbc.su.se/~pierre/svmhc/new.cgi
SVMHC (SYFPEITHI)	http://www.sbc.su.se/~pierre/svmhc/new.cgi
SVRMHC	http://SVRMHC.umn.edu/SVRMHCdb
SYFPEITHI	http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.htm

ตารางที่ 3 รายชื่อ prediction servers สำหรับทำนาย HLA class II binding peptides (URLs as of December 2007)

Servers	URLs
HLA-DR4Pred (ANN)	http://www.imtech.res.in/raghava/hladr4pred/index.html
HLA-DR4Pred (SVM)	http://www.imtech.res.in/raghava/hladr4pred/index.html
IEDB (ARB)	http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html
IEDB (SMM)	http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html
IEDB (Saturniolo)	http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html
IEDB (Consensus)	http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html
MHC Binder Prediction	http://www.vaccinedesign.com
MHC2Pred	http://www.imtech.res.in/raghava/mhc2pred
MHC-BPS	http://bidd.cz3.nus.edu.sg/mhc
MHCPred	http://www.jenner.ac.uk/MHCPred
Multipred1 (ANN)	http://antigen.i2r.a-star.edu.sg/multipred1
Multipred1 (HMM)	http://antigen.i2r.a-star.edu.sg/multipred1
Multipred1 (SVM)	http://antigen.i2r.a-star.edu.sg/multipred1
NetMHCII	http://www.cbs.dtu.dk/services/NetMHCII
NetMHCIIpan	http://www.cbs.dtu.dk/services/NetMHCIIpan
PeptideCheck (Matrix)	http://www.peptidecheck.org
ProPred	http://www.imtech.res.in/raghava/propred
Rankpep	http://bio.dfci.harvard.edu/Tools/rankpep.html
SVMHC	http://www-bs.informatik.uni-tuebingen.de/SVMHC/index_html
SVRMHC	http://SVRMHC.umn.edu/SVRMHCdb
SYFPEITHI	http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.htm

ประโยชน์ของการทำนาย HLA binding peptide นั้นสามารถนำไปประยุกต์ใช้ในการหาโมเลกุลเป้าหมายอันจะนำไปสู่การพัฒนาวัคซีนในอนาคต หรือใช้ประโยชน์ในการศึกษาเพื่อการค้นพบ T-cell epitope ใหม่ ๆ เนื่องจากวิธีการทำ epitope mapping แบบดั้งเดิมนั้นต้องอาศัยการสังเคราะห์เปปไทด์หลาย ๆ เส้น และมีราคาแพงจนไม่สามารถสังเคราะห์แบบเลื่อนทีละหนึ่งกรดอะมิโนได้เหมือนโปรแกรม ข้อดีของการใช้โปรแกรมในการทำนายนั้นจะช่วยประหยัดเวลาและค่าใช้จ่าย รวมทั้งช่วยคัดกรองเปปไทด์ที่ไม่สามารถจับกับ HLA ได้ให้ด้วย

สรุป

การประยุกต์ใช้วิธีทางสารสนเทศภูมิคุ้มกันในงานวิจัยเกี่ยวกับการตอบสนองทางภูมิคุ้มกันแบบพึ่งเซลล์ (cell mediated immunity) เป็นวิธีการที่เป็นประโยชน์ต่อการหาโมเลกุลเป้าหมายเพื่อการทำ immunotherapy ในโรคภูมิแพ้ โรคติดเชื้อ และมะเร็ง

วิทยา ทั้งนี้เพราะวิธีดังกล่าวเป็นวิธีที่รวดเร็วกว่า ดีกว่าและมีราคาถูกกว่าการทำวิจัยแบบดั้งเดิมซึ่งปกติจะใช้เวลานานและสิ้นเปลืองงบประมาณในการทำการทดลอง นอกจากนี้วิธีการทางสารสนเทศยังช่วยลดจำนวนชุดของการทดลองซึ่งปกติมีปริมาณมากลงไปได้มากกว่า 50% ทำให้นักวิจัยได้โมเลกุลเป้าหมายเร็วขึ้น อันจะนำไปสู่การกำหนดทิศทางในการควบคุม รักษาโรคต่าง ๆ ได้ทันกับอุบัติการณ์ของผู้ป่วยที่เพิ่มขึ้นเรื่อย ๆ ในแต่ละปี

ตารางที่ 4 แสดงผลการคำนวณหาค่า binding score ของเปปไทด์ควบคุมจากแอนติเจนของมะเร็งซึ่งมีความจำเพาะกับ HLA-A*0201 โดยการทำนายด้วยคอมพิวเตอร์ด้วยโปรแกรมของ IEDB เพื่อนำไปใช้เป็น threshold ในการคัดเลือกเปปไทด์สำหรับทำการทดลองจริง

No.	Sequence	ANN IC50 (nM)	No.	Sequence	ANN IC50 (nM)	No.	Sequence	ANN IC50 (nM)	No.	Sequence	ANN IC50 (nM)
1	LLLEAVPAV	6.71	22	VLAGVGFFI	17.77	43	RLQGISPKI	75.24	64	ALCNTDSPL	899.07
2	SLDDYNHLV	6.94	23	WLQYFPNPV	18.02	44	TLAPATEPA	105.09	65	PLTSIISAV	1450.94
3	FMNKFIYEI	7.05	24	KIFGSLAFL	18.95	45	WLPKILGEV	111.47	66	STAPPAHGV	1639.37
4	FLYTLLREV	7.44	25	ILLVVVLGV	19.96	46	CLTSTVQLV	112.08	67	TLHCDCEIL	2143.24
5	FVWLHYYSV	7.79	26	IMNDMPIYM	22.43	47	CQLAKTCPV	121.86	68	IISAVVGIL	2817.52
6	LLLAGLFSL	7.97	27	ALSPVPPVV	22.61	48	RLGPTLMCL	161.69	69	SILLRDAGL	2849.43
7	FLDPRPLTV	8.85	28	GLIQLVEGV	24.48	49	ALCRWGLLL	172.99	70	PLQPEQLQV	5104.16
8	GLMKYIGEV	9	29	QLMPYGCLL	24.52	50	SLEENIVIL	183.3	71	HACWPAFTV	5703.06
9	ILLWQPIPV	9.04	30	RLLQETELV	25.05	51	HLYQGCQVV	191.95	72	MKWVESIFL	8283.3
10	SLVDVMPWL	9.57	31	AILALLPAL	26.64	52	VLSVNVPDV	209.64	73	NLVKQKPQI	9302.3
11	FLWGPRALV	9.99	32	LLGATCMFV	26.67	53	CMHLLLEAV	220.13	74	STPPPGTRV	12744.3
12	FLAKLNNTV	10.52	33	YVDPVITSI	26.69	54	STAPPVHNV	236.63	75	SSKALQRPV	15021.1
13	QLMAFNHLI	10.77	34	LLLLTVLTV	29.17	55	ILSLELMKL	265.87	76	ELVSEFSRM	17968.7
14	MLWGWREHV	11.7	35	KIWEELSVL	29.83	56	LLNQLQVNL	269.7	77	AYACNTSTL	18725.1
15	FLTGNQLAV	13.01	36	ALQPGTALL	30.86	57	PLFQVPEPV	272.46	78	LVQENYLEY	21481.7
16	KASEKIFYV	13.32	37	KASEYLQLV	33	58	ATVGIMIGV	297.62	79	LVHFLLLKY	22519.3
17	VLQELNVTV	14.8	38	FLIIWQNTM	38.16	59	TLEEITGYL	366.12	80	AGYLMELCM	23098.8
18	YMNGTMSQV	15	39	VVLGVVFGI	38.8	60	ILLRDAGLV	413.76	81	AKYLMELTM	23579.6
19	KVAELVHFL	15.23	40	ILHNGAYSL	47.09	61	VLVKSPNHV	443.08	82	PVNPGVGQC	24002.1
20	YLSGANLNL	15.43	41	RLTRFLSRV	54.99	62	RLARLALVL	632.21	83	AGYLMELCC	24283.2
21	YAIDLPVSV	17.73	42	ALIHHNTHL	56.15	63	GVLVGVALI	794.47	84	AGYLMELCF	24348.1

เอกสารอ้างอิง

Abbas, A.K., Lichtman, A.H., and Pillai, S. (2000). Cellular and Molecular Immunology. (4th edition). Philadelphia: Saunders/Elsevier. pp. 79-102.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215 (3): 403-410.

Bermstein, H.J. (2000). Recent changes to RasMol, recombining the variants. Trands Biochem. Sci. 25 (9): 453-455.

Brusic, V., and Petrovsky, N. (2003). Bioinformatics for characterisation of allergens, allergenicity and allergic cross reactivity. Trends Immunol. 24 (5): 225-228.

DeLano, W.L. (2000). The PyMOL molecular graphics system. DeLano Scientieic, San Carlos.

Flower, D.R. (2007) Immunoinformatics and the in silico prediction of immunegenicity. An introduction. Meth Mol Biol. 409: 1-15.

Hanson, R.M. (2010). Jmol – a paradigm shift in crystallographic visualization. J. Appl. Crystallogr. 43 (5): 1250-1260.

Kay, A.B. (2001). Allergy and allergic diseases. First of two parts. N Engl J Med. 344 (1): 30-37.

Korber, B., LaBute, M., and Yusim, K. (2006). Immunoinformatics comes of age. PLoS Comput Biol. 2 (6): e71.

Lin, H.H., Ray, S., Tongchusak, S., Reinherz, E.L., Brusic, V. (2008). Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunol. 9: 8.

Matsumura, M., Fremont, D.H., Peterson, P.A., and Wilson, I.A. (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. Science 257: 927-934.

Mizrachi, I.K. (2008). Managing sequence data. Methods Mol Biol. 452: 3-27.

Reitmaier, R. (2007). Review of Immunoinformatic approaches to in-silico B-cell epitope prediction. Nature Precedings. Posted 5 Jul, 10.1038/npre.2007.353.1.

Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P., Marsh, S.G.E. (2003). IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Research 31: 311-314.

Sayle, R.A., Milner-White, E.J. (1995). RasMol: Biomolecular graphics for all. Trands Biochem. Sci. 20 (9): 374.

Subramanian, A.R, Kaufmann, M., and Morgenstern, B. (2008). DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. Algorithms Mol Biol. 3: 6.

Tongchusak, S., Chaiyaroj, S.C., Veeramani, A., Koh, J.L.Y., and Brusic, V. (2005). CandiVF–*Candida albicans* Virulence Factor Database. Int J Pep Res and Thera. 11 (4): 271-277.

❑❑❑❑❑