Self/Nonself 1:3, 255-258; July/August/September 2010 © 2010 Landes Bioscience

The self/nonself issue
A confrontation between proteomes

Darja Kanduc

Department of Biochemistry and Molecular Biology; University of Bari; Bari, Italy

Key words: epitopic peptides, immune recognition, proteomic analysis, self/nonself

Defining self and nonself is the most compelling challenge in science today, at the basis of the numerous questions that remain unanswered in the immunology-pathology-therapy debate. The generation of the antibody repertoire, the complicated scenario offered by tolerance and autoimmunity, natural auto-antibodies and their relationship to autoimmune diseases, and positive and negative selection are only a few examples of the unresolved immunological questions. In this context, we proposed that sequence similarity to the host proteome modulates antigen peptide recognition and immunogenicity. Using the available proteome assemblies of viruses, bacteria and higher vertebrates, and applying the low-similarity criterion, we are systematically defining the proteomic similarity of B-cell epitopes already validated experimentally. Here, we report further data documenting that a low similarity to the host proteome is the common property that defines the immunological "nonself" nature of antigenic sequences in cancer, autoimmunity, infectious diseases and allergy.

Identifying Epitopes

Antibodies are immunoglobulin proteins that interact with specific areas on the surface of antigen proteins. These areas are referred to as B-cell epitopes. Identifying B-cell epitopes in order to induce a specific antibody response constitutes the unresolved core of immunology. Practically, the identification of epitopes in proteins is the fundamental, preliminary step in designing effective immunotherapy for cancer and infectious diseases as well as autoimmune pathologies.[1-3] As a logical consequence, recent decades have seen determined efforts aimed at identifying and defining antigen epitopes. To understand the molecular determinants characterizing epitopic structures, immunology has used the self/nonself concept.[4-6] For decades, immunologists have based their studies, reasoning, experiments and clinical treatments on the idea that the immune system works by distinguishing between self and nonself.[4,5] However, theoretical and experimental considerations lead to the recognition that there are no known molecular mechanisms that can explain how peptides of self-origin can be discriminated qualitatively from peptides of nonself-origin.[6] So, for example, how can the hexapeptide VLDVGG, which occurs in two different human proteins, be discriminated from the hexapeptide VLDVGG, which occurs in 380 bacterial, viral, protozoan and other organism proteins? How can the heptapeptide PPPPPPP, which occurs in 625 human proteins, be catalogued as self or nonself and distinguished from the heptapeptide PPPPPPP, which occurs in 12,883 bacterial, viral, protozoan and other organism proteins?

Epitopes: A Problem of Numbers

To answer questions on immunology helpful to immunotherapy, numerous predictive programs and algorithms have been developed for exact epitope identification. Different epitope qualities have been proposed and investigated, including protein hydrophobicity[7-9] protein hydrophilicity,[10] the protrusion index,[11] protein flexibility,[12] and protein secondary structure and conformational parameters.[13,14] However, the results of these studies are inconsistent and very little progress in B-cell epitope prediction has been made.[15,16]

Generally, the lack of success in B-cell epitope definition is due, mainly, to the difficulties associated with the task. The optimal amino acid (aa) length of a B-cell epitope is five aa,[17] but longer epitopes have been described. In general, linear B-cell epitopes have been described as varying in length, with up to 16 residues reported.[18,19] Considering the 20 naturally occurring amino acids, the potential epitope repertoire ranges from 3,200,000 different linear 5-mers to 655,360,000,000,000 different linear 16-mers. In addition, epitopic amino acid sequences may originate from conformational folding: antigen proteins are long amino acid chains folded into an enormous number of shapes. Searching through all possible foldings to evaluate the working structure-function of an epitopic sequence is a practically impossible task. Finally, epitopic sequences can be derived from post-translational modifications; they can also be site-switched by the epitope spreading phenomena, and can exist as cryptic epitopes. Again, this makes the possible determinant configurations numerically infinite.

Correspondence to: Darja Kanduc; Email: d.kanduc@biologia.uniba.it; dkanduc@gmail.com
Submitted: 01/11/10; Accepted: 01/29/10
Previously published online: www.landesbioscience.com/journals/selfnonself/article/11897

The Proteome-Guided Definition of the Immune Object

The proteomic era has provided comprehensive proteome databases for numerous (micro)organism types that can now be analyzed in detail using large-scale proteomics scanning. Using the available proteome repertoire, we advanced and explored the hypothesis that the immunogenicity of peptide sequences is modulated by low similarity to the host's proteome. During the past decade, a series of experimental models involving different disease-associated proteins[20-31] have substantiated and supported the low-similarity hypothesis.

Experimental data from our laboratory have been confirmed by numerous reports in the epitope mapping literature. A systematic analysis of epitopic peptide sequences in cancer, autoimmunity, allergy and infectious diseases indicates that the epitopic boundaries of an amino acid sequence are dictated by the low-similarity hypothesis.[28-30] Only peptide motifs with no/low similarity to the host proteome are epitopic targets in the humoral immune response. Additionally, only amino acid sequences that are scarce (or absent) in host proteins appear to participate in antigen-antibody interactions. In this way, we have validated hundreds of epitopes as low-similarity amino acid sequences.[28-30]

This structured survey adds to the tabulation of low-similarity epitopes obtained through proteomic scanning. As a pragmatic example of proteome-guided epitope definition, **Table 1** provides a sequence similarity analysis of the most recent data in epitope mapping of disease-associated antigens. The Table shows that amino acid fragment(s) endowed with a low level of similarity to the host proteome represent the common signature of experimentally validated epitopes, regardless of whether they are linear, discontinuous or mimic, independently of their being from microbial antigens, tumor-associated proteins or allergens.

Perspectives: Scientific and Clinical Applications

Here, we analyze a robust set of recent experimental reports suggesting that a low level of sequence similarity to the host proteome defines the B-cell epitope pool in the humoral immune response. The data add to and support experimental results from our laboratory[21-27,31] and reported meta-analyses.[28-30]

Theoretically, proteomic similarity analyses might elucidate the regulatory mechanisms/factors that dictate peptide immunogenicity assessment. From a clinical perspective, low-similarity peptides may have strong repercussions affecting the rational development of peptide-based treatments in cancer, infection and autoimmunity. De facto, the most attractive feature of the similarity concept is that it appears to guarantee the highest specificity and lowest cross-reactivity when designing effective, safe and theoretically infallible immunotherapeutic tools. The Ehrlichian idea of therapeutic agents equipped with high affinity to the causative agent, and efficacy at concentrations harmless to the patient, appears feasible.

Scientifically, the solution to what has been called the top-ranking mystery in science appears to be at hand, i.e., the mechanism of self versus nonself recognition in the immune system.[53] In fact, the low-similarity hypothesis might represent a quantum leap forward in our understanding of the rules dictating the immunological antigen-antibody epitope-paratope interaction at the peptide level.

References

1. Van Regenmortel MH. What is a B-cell epitope? Methods Mol Biol 2009; 524:3-20.
2. Kanduc D. Peptimmunology: immunogenic peptides and sequence redundancy. Curr Drug Discov Technol 2005; 2:239-44.
3. Lim PL, Zouali M. Pathogenic autoantibodies: emerging insights into tissue injury. Immunol Lett 2006; 103:17-26.
4. Tauber AI. The elusive immune self: a case of category errors. Perspect Biol Med 1999; 42:459-74.
5. Cohn M. On the opposing views of the self-nonself discrimination by the immune system. Immunol Cell Biol 2009; 87:113-9.
6. Pradeu T, Carosella ED. On the definition of a criterion of immunogenicity. Proc Natl Acad Sci USA 2006; 103:17858-61.
7. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157:105-32.
8. Manavalan P, Ponnuswamy PK. Hydrophobic character of amino acid residues in globular proteins. Nature 1978; 275:673-4.
9. Sweet RM, Eisenberg D. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. J Mol Biol 1983; 171:479-88.
10. Parker JMR, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography retention data: correlation of predicted surface residues with antigenicity and X-ray derived accessible sites. Biochemistry 1986; 25:5425-32.
11. Thornton JM, Edwards MS, Tayler WR, Barlow DJ. Location of 'continuous' antigenic determinants in the protruding regions of proteins. EMBO J 1986; 5:409-13.
12. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins: a tool for the selection of peptide antigens. Naturwissenschaften 1985; 72:212-3.
13. Garnier J, Robson B. The GOR method for predicting secondary structures in proteins. In: Fasman GD, editor. Prediction of Protein Structure and the Principles of Protein Conformation. Vol. 11. Plenum Press 1989; 417-65.
14. Lewis PN, Momany FA, Scheraga HA. Folding of polypeptide chains in proteins: a proposed mechanism for folding. Proc Natl Acad Sci USA 1971; 68:2293-7.
15. Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. Protein Sci 2005; 14:246-8.
16. Gowthaman U, Agrewala JN. In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. J Proteome Res 2008; 7:154-63.
17. Lucchese G, Stufano A, Trost B, Kusalik A, Kanduc D. Peptidology: short amino acid modules in cell biology and immunology. Amino Acids 2007; 33:703-7.
18. El-Manzalawy Y, Dobbs D, Honavar V. Predicting flexible length linear B-cell epitopes. Comput Syst Bioinformatics Conf 2008; 7:121-32.
19. Sollner J, Grohmann R, Rapberger R, Perco P, Lukas A, Mayer B. Analysis and prediction of protective continuous B-cell epitopes on pathogen proteins. Immunome Res 2008; 4:1.
20. Willers J, Lucchese A, Kanduc D, Ferrone S. Molecular mimicry of phage displayed peptides mimicking GD3 ganglioside. Peptides 1999; 20:1021-6.
21. Natale C, Giannini T, Lucchese A, Kanduc D. Computer-assisted analysis of molecular mimicry between human papillomavirus 16 E7 oncoprotein and human protein sequences. Immunol Cell Biol 2000; 78:580-5.
22. Kanduc D, Fanizzi FP, Lucchese G, Stevanovic S, Sinha AA, Mittelman A. NMR probing of in silico identification of anti-HPV16 E7 mAb linear peptide epitope. Peptides 2004; 25:243-50.
23. Mittelman A, Tiwari R, Lucchese G, Willers J, Dummer R, Kanduc D. Identification of monoclonal anti-HMW-MAA antibody linear peptide epitope by proteomic database mining. J Invest Dermat 2004; 123:670-5.
24. Dummer R, Mittelman A, Fanizzi FP, Lucchese G, Willers J, Kanduc D. Nonself discrimination as a driving concept in the identification of an immunodominant HMW-MAA epitopic peptide sequence by autoantibodies from melanoma cancer patients. Int J Cancer 2004; 111:720-6.
25. Willers J, Lucchese A, Mittelman A, Dummer R, Kanduc D. Definition of anti-tyrosinase MAb T311 linear determinant by proteome-based similarity analysis. Exp Dermatol 2005; 14:543-50.
26. Angelini G, Bonamonte D, Lucchese A, Favia G, Serpico R, Mittelman A, et al. Preliminary data on *Pemphigus vulgaris* treatment by a proteomics-defined peptide: a case report. J Transl Med 2006; 4:43.
27. Kanduc D, Tessitore L, Lucchese G, Kusalik A, Farber E, Marincola FM. Sequence uniqueness and sequence variability as modulating factors of human anti-HCV humoral immune response. Cancer Immunol Immunother 2008; 57:1215-23.
28. Kanduc D. Correlating low-similarity peptide sequences and allergenic epitopes. Curr Pharm Des 2008; 14:289-95.

Table 1. B-cell epitopes are (or contain) sequences with zero/low-similarity to the host proteome

Antigen	Aa position	Epitope sequence[a]	Matches[b]	Proteome[c]	Ref.
Prostate specific antigen	27–31	GGWEC	0	M	31
Melanin-concentrating hormone receptor 1	254–265	lYWFTLy	0	H	32
Osteopontin	31–37	qlYNKYP	0	M	33
Coagulation factor VIII	Mimotope	npveNMMDRdsq	1	H	34
	Mimotope	qspWQTWFtral	1	H	34
Adenosine A2a receptor	172–178	lfEDVVP	4	M	35
Collagen type IV, α3chain	24–38	FTRHSqttANPSCpe	0	R	36
Receptor tyr kinase-like	518–525	reEFRHEa	1, 2	M	37
CXCR1 chemokine	Mimotope	SFlWDf	3	M	38
	Mimotope	SAMWDf	0	M	38
	Mimotope	TNMWDf	0	M	38
	Mimotope	iTMWDF	1	M	38
	Mimotope	SDWWDf	0	M	38
CXCR2 chemokine	Mimotope	FWDDFw	1	M	38
	Mimotope	LWDDFw	2	M	38
	Mimotope	mWNDFW	0	M	38
	Mimotope	FWLDFw	0	M	38
DNA	Mimotope	DWEYSvwlsn	0	M	39
Bla g 4 allergen	118–152	cpaaanGHVIYvqlrltWRRFHpklgdkeMIQHYT	2, 2, 0	H	40
Blo t 12 isoallergen	30–49	htepddHHEKPtTQCTHeet	0, 0	H	41
	73–92	teeTHHSDdlivHEGGKtyh	0, 2	H	41
	111–130	iicsksgslWYITVmPCSIG	0, 1	H	41
HIV-1 envelope gp41	633–639	iWNNMTw	0	H	42
			0	M	42
HIV-1 envelope gp41	633–642	iwnnMTWMQw	0	H	42
			0	M	42
HIV-1 gp41	675–681	elDKWAS	1	M	43
HIV envelope gp41	670–677	wNWFDltn	0	M	44
Porcine reproductive-respi ratory syndrome virus gp5	152–156	rLYRWR	0	M	45
	169–178	eGHLIDlkrv	4	M	45
	196–200	QWGRL	1	M	45
	196–200	QWGRP	0	M	45
Rabies virus glycoprotein	Mimotope	kRDSTW	2	H	46
	Mimotope	kYLWSK	0	H	46
	Mimotope	kYWLSR	2	H	46
	Mimotope	kYWWSK	0	H	46
	Mimotope	kYAWSR	0	H	46
	Mimotope	kYSMSK	0	H	46
Japanese encephalitis virus NS1 protein	146–150	EHARW	1	M	47
Influenza A H5N1 HA	193–199	qNPTTYi	1	M	48
Influenza A H5N1 HA	121/164[d]	sWS/YNN	0	H	49
Influenza A fusion peptide HA2 glycopolypeptide		glfgAIAGF	1	M	50
K. pneumoniae adhesin	Mimotope	qktlakSTYMSa	0	M	51
Meningococcal factor H-binding protein	25/57[d]	DHK/YGn	1	M	52

The analysis was restricted to reports published in 2009 in PubMed. [a]Low similarity 5-mer given in capital letters. Based on the minimum length of an immune unit of 5–6 amino acids,[17] each epitope has been dissected into pentapeptide sequences. The pentapeptides overlapped by four residues, i.e., were shifted by one amino acid. Then, each pentapeptide was analyzed for the number of occurrences in the host proteome using described methodologies.[20-31] Any occurrence is called a match. A pentapeptide with up to five perfect matches to the host proteome was considered a low-similarity sequence. [b]The number of matches refers to the low-similarity 5-mer in capital letters. [c]Host proteome: H, human; M, murine; R, rat. [d]Discontinuous.

29. Kanduc D. "Self-Nonself" peptides in the vaccine design. Curr Pharm Des 2009; 15:3283-9.
30. Kanduc D, Serpico R, Lucchese A, Shoenfeld Y. Correlating low-similarity peptide sequences and HIV B-cell epitopes. Autoimmun Rev 2008; 7:291-6.
31. Stufano A, Kanduc D. Proteome-based epitopic peptide scanning along PSA. Exp Mol Pathol 2009; 86:36-40.
32. Gavalas NG, Gottumukkala RV, Gawkrodger DJ, Watson PF, Weetman AP, Kemp H. Mapping of melanin-concentrating hormone receptor 1 B cell epitopes predicts two major binding sites for vitiligo patient autoantibodies. Exp Dermatol 2009; 18:454-63.
33. Dai J, Cao Z, Kang Y, Fan K, Ji G, Yang H, et al. A functional motif QLYxxYP is essential for osteopontin induced T lymphocyte activation and migration. Biochem Biophys Res Commun 2009; 380:715-20.
34. Kessel C, Kreuz W, Klich K, Becker-Peters K, Vorpahl F, Dietrich U, et al. Multimerization of peptide mimotopes for blocking of Factor VIII neutralizing antibodies. ChemMedChem 2009; 4:1364-70.
35. By Y, Durand-Gorde JM, Condo J, Lejeune PJ, Mallet B, Carayon P, et al. Production of an agonist-like monoclonal antibody to the human A2A receptor of adenosine for clinical use. Mol Immunol 2009; 46:400-5.
36. Reynolds J, Abbott DS, Karegli J, Evans DJ, Pusey CD. Mucosal tolerance induced by an immunodominant peptide from rat {alpha}3(IV)NC1 in established experimental autoimmune glomerulonephritis. Am J Pathol 2009; 174:2202-10.
37. Gardberg AS, Dice L, Pridgen K, Ko J, Patterson P, Ou S, et al. The x-ray structures of amyloid beta-related peptides complexed to antibodies. Biochemistry 2009; 48:5210-7.
38. Houimel M, Mazzucchelli L. Identification of biologically active peptides that inhibit binding of human CXCL8 to its receptors from a random phage-epitope library. J Leukoc Biol 2009; 85:728-38.
39. Mihaylova N, Voynova E, Tchorbanov A, Dolashka-Angelova P, Bayry J, Devreese B, et al. Simultaneous engagement of FcgammaIIb and CD22 inhibitory receptors silences targeted B cells and suppresses autoimmune disease activity. Mol Immunol 2009; 47:123-30.
40. Shin KH, Jeong KY, Hong CS, Yong TS. IgE binding reactivity of peptide fragments of *Bla g* 4, a major German cockroach allergen. Korean J Parasitol 2009; 47:31-6.
41. Zakzuk J, Jiménez S, Cheong N, Puerta L, Lee BW, Chua KY, et al. Immunological characterization of a *Blo t* 12 isoallergen: identification of immunoglobulin E epitopes. Clin Exp Allergy 2009; 39:608-16.
42. Benferhat R, Krust B, Rey-Cuillé MA, Hovanessian AG. The caveolin-1 binding domain of HIV-1 glycoprotein gp41 (CBD1) contains several overlapping neutralizing epitopes. Vaccine 2009; 27:3620-30.
43. Fiebig U, Schmolke M, Eschricht M, Kurth R, Denner J. Mode of interaction between the HIV-1-neutralizing monoclonal antibody 2F5 and its epitope. AIDS 2009; 23:887-95.
44. Pejchal R, Gach JS, Brunel FM, Cardoso RM, Stanfield RL, Dawson PE, et al. A conformational switch in HIV gp41 revealed by the structures of overlapping epitopes recognized by neutralizing antibodies. J Virol 2009; 83:8451-62.
45. Zhou YJ, Yu H, Tian ZJ, Liu JX, An TQ, Peng JM, et al. Monoclonal antibodies and conserved antigenic epitopes in the C terminus of GP5 protein of the North American type porcine reproductive and respiratory syndrome virus. Vet Microbiol 2009; 138:1-10.
46. Houimel M, Dellagi K. Peptide mimotopes of rabies virus glycoprotein with immunogenic activity. Vaccine 2009; 27:4648-55.
47. Wang B, Hua RH, Tian ZJ, Chen NS, Zhao FR, Liu TQ, et al. Identification of a virus-specific and conserved B-cell epitope on NS1 protein of *Japanese encephalitis* virus. Virus Res 2009; 141:90-5.
48. Wang SF, Chen KH, Thitithanyanont A, Yao L, Lee YM, Chan YJ, et al. Generating and characterizing monoclonal and polyclonal antibodies against avian H5N1 hemagglutinin protein. Biochem Biophys Res Commun 2009; 382:691-6.
49. Khurana S, Suguitan AL Jr, Rivera Y, Simmons CP, Lanzavecchia A, Sallusto F, et al. Antigenic fingerprinting of H5N1 avian influenza using convalescent sera and monoclonal antibodies reveals potential vaccine and diagnostic targets. PLoS Med 2009; 6:1000049.
50. Prabhu N, Prabakaran M, Ho HT, Velumani S, Qiang J, Goutama M, et al. Monoclonal antibodies against the fusion peptide of hemagglutinin protect mice from lethal influenza A virus H5N1 infection. J Virol 2009; 83:2553-62.
51. Li Y, Han WY, Li ZJ, Lei LC. Klebsiella pneumoniae MrkD adhesin-mediated immunity to respiratory infection and mapping the antigenic epitope by phage display library. Microb Pathog 2009; 46:144-9.
52. Beernink PT, LoPasso C, Angiolillo A, Felici F, Granoff D. A region of the N-terminal domain of meningococcal factor H-binding protein that elicits bactericidal antibody across antigenic variant groups. Mol Immunol 2009; 46:1647-53.
53. Koshland DJ Jr. Recognizing self from nonself. Science 1990; 248:4961.