Amino Acids (2008) 35: 209–216
DOI 10.1007/s00726-007-0563-7
Printed in The Netherlands

Amino Acids

Mapping the human proteome for non-redundant peptide islands

G. Capone[1], **A. De Marinis**[1], **S. Simone**[1], **A. Kusalik**[2], and **D. Kanduc**[1,3]

[1] Department of Biochemistry and Molecular Biology "Ernesto Quagliariello", University of Bari, Bari, Italy
[2] Department of Computer Science, University of Saskatchewan, Saskatoon, Canada
[3] Clinical Center, Immunogenetics Laboratory, Department of Transfusion Medicine, NIH, Bethesda, MD, USA

Received April 20, 2007
Accepted May 8, 2007
Published online August 15, 2007;

Summary. We describe immune-proteome structures using libraries of protein fragments that define a structural immunological alphabet. We propose and validate such an alphabet as i) composed of letters of five consecutive amino acids, pentapeptide units being sufficient minimal antigenic determinants in a protein, and ii) characterized by low-similarity to human proteins, so representing structures unknown to the host and potentially able to evoke an immune response. In this context, we have thoroughly sifted through the entire human proteome searching for non-redundant protein motifs. Here, for the first time, a complete sequence redundancy dissection of the human proteome has been conducted. The non-redundant peptide islands in the human proteome have been quantified and catalogued according to the amino acid length. The library of uniquely occurring *n*-peptide sequences that was obtained is characterized by a logarithmic decrease of the number of non-redundant peptides as a function of the peptide length. This library represents a highly specific catalogue of molecular protein signatures, the possible use of which in cancer/autoimmunity research is discussed, with a major focus on non-redundant dodecamer sequences.

Keywords: Human proteome – Redundant peptide sequences – Quantitative proteomic redundancy – Qualitative proteomic redundancy

Introduction

Alignment of protein sequences provides information on residue conservation, variability, dispensability. This information allows reconstruction of the structural-functional profile of proteins through the definition of common structural motifs, interacting domains, and regulatory regions. The distribution of amino acid residues, dipeptides and, in general, short amino acid sequences is fundamental in the comparative analysis of proteins from different kingdoms, and of related proteins forming families. The final results are of interest not only in comparative genomics and proteomics studies, but also in experimental pathology and therapeutics.

De facto, research is progressively shifting the focus from analysis of disease-associated proteins as whole entities to identification of the specific peptide sequences at the root of the protein derangements leading to the altered pathways involved in disease pathogenesis (Bielekova et al., 2000; Yan et al., 2006). In parallel, identifying short amino acid motifs endowed with antigenic/immunogenic properties might offer the potential to provoke/neutralize (auto)immune responses and, consequently, to design immunotherapeutic treatments in cancer diseases as well as in infectious or autoimmune pathologies (Hartemann-Heurtier et al., 2004; Fontoura et al., 2005; Kanduc, 2005; Mahdavi and Monk, 2005; Wang and Walfield, 2005). These considerations are strengthened by an ample literature documenting that short peptides are critically involved in modulating/interfering with essential cell reactions (Lucchese et al., 2006, and refs. therein).

Our labs are exploring the possible immunological significance of peptide sequence redundancy in the human proteome. The rationale is the following. If it is true that normal auto-antigens are tolerated through the purging of all antigen receptors that could possibly recognize and attack auto-antigens (Tsubata and Honjo, 2000; Kamradt and Mitchison, 2001; Ohashi, 2003), i.e. through the clonal deletion process, then it is logical to hypothesize that peptide sequences uniquely expressed in a proteome (or expressed late in the temporal cell ontology or rarely expressed or present at subluminal concentration level or hidden in immune privileged sites) have more chances to escape the tolerance process. In other words, the molecular sequences with no/low similarity to the host proteome

have more chances to induce an immune response. Consequently, identification and localization of unique amino acids stretches in the human proteome is of remarkable importance since these non-redundant peptide sequences might represent i) ''pathogenicity islands'', by being potentially able to induce specific destructive autoimmune reactions or, alternatively, ii) ''ideal targets'', if immune reactions are desired as in cancer immunotherapeutic approaches.

In this study we searched the human proteome for such sole occurrence strings. To this end, we systematically analyzed the entire set of human proteins that comprehensively form the human proteome for peptide sequences that are present only and exclusively in one human protein. We report such an analysis of non-redundant peptide islands and discuss its possible use in studying/interpreting basic phenomena such as antigenicity/immunogenicity, as well as in designing/developing new therapeutical peptide-based approaches.

Materials and methods

Sequence similarity analysis for each human protein was conducted against the human proteome. The human proteome was obtained from EBI's Integr8 site (ebi.ac.uk/integr8); the analyzed proteome consisted of 37991 proteins when downloaded on May 30, 2006. In determining the intra-human similarity pattern, the human proteins to be analyzed were manipulated and analysed at the 5-mer level as follows.

Each entire protein was decomposed *in silico* to a set of 5-mers sequentially overlapped by four residues. Source protein and position were recorded for each 5-mer. The resultant dataset was then analyzed based on frequency of occurrence of individual 5-mers. Pentamers which only occurred once in the set of human proteins were identified and collected into a library. In silico digestion was performed by a custom program written in C for efficiency. All other analysis was performed using LINUX/UNIX shell scripts and standard LINUX/UNIX utilities. The zero-redundant peptide sequences were further analyzed through PIR, Expasy, NCBI, and PubMed databanks for potential disease relationships.

Results

Intra-proteomic peptide non redundancy

We explored the human proteome for the degree of internal non-redundancy using the pentapeptide as a length unit, i.e. a peptide unit effective from the biological as well immunological point of view (Niman et al., 1983; Reddehase et al., 1989; Liu et al., 1992; Lucchese et al., 2006; Opdenakker et al., 2006). We asked two questions: 1) How many pentapeptides are uniquely expressed in the human proteins? 2) Is peptide sequence non-redundancy influenced by the peptide length?

The quantitative data we found are reported in Table 1 listing the number of unique occurrences of *n*-peptides (with *n* from 5 to 16 amino acids) and the relative *n*-peptide overlaps disseminated through the more than 36,000 human proteins that comprehensively form the human proteome in the integr8 database (www.ebi.ac.uk/integr8). In other words, Table 1 provides numerical

Table 1. *n*-Peptide redundancy in the human proteome

Sequence length (aa)	Total occurrences			
	Theoretical (10^6)	Actual		
		Unique[a]	Including overlaps[b]	Formed by 5-mers occurring only once[c]
5	3.2	2 388 563	16 249 364	464 485
6	64	8 247 275	16 210 640	123 854
7	1 280	10 431 975	16 171 995	41 060
8	25 600	10 797 988	16 133 421	14 437
9	512 000	10 912 559	16 094 868	4 860
10	10 240 000	10 982 062	16 056 379	783
11	204 800 000	11 033 022	16 017 956	201
12	4 096 000 000	11 071 755	15 979 581	58
13	81 920 000 000	11 101 910	15 941 254	16
14	1 638 400 000 000	11 125 333	15 902 957	6
15	32 768 000 000 000	11 143 351	15 864 691	2
16	655 360 000 000 000	11 156 519	15 826 455	1

[a] Number of unique *n*-peptides in human proteome

[b] Occurrences of *n*-peptides in human proteome (including duplicate instances of same *n*-peptide)

[c] Number of *n*-peptides formed by 5-mers expressed only once in the human proteome, i.e. with zero redundancy

Table 2. Human proteins hosting long non-redundant peptide motifs and relevant to autoimmune diseases

Protein description-name	ID	Aa pos	Peptide sequence	Autoimmune pathology	References
Ubiquitin-activating enzyme E1	P22314	705	DCVTWACHHWHTQYSN	Pemphigus foliaceus	Liu et al. (1992)
Matrix metalloproteinase-9	P14780	114	LKWHHHNITYWIQN	Rheumatoid arthritis	Opdenakker et al. (2006)
				Multiple sclerosis	Descamps et al. (2003)
				Lupus erythematosus	Dubois et al. (1999)
				Experimental autoimmune encephalomyelitis	Matache et al. (2003)
ADAMTS-15,metalloproteinase	Q8TE58	689	FTKPMHGYNFVVA	Asthma	Paulissen et al. (2006)
Cytochrome P450 4F3	Q08477	84	FGDMCCWWVGPWH	Type 1 diabete	Park et al. (2004)
				Autoimmune hepatitis	Vergani and Mieli-Vergani (2004)
				Autoimmune polyglandular syndrome type 1	Liiv et al. (2002)
				Autoimmune endocrinopathies	Hrda et al. (2004)
Coagulation factor VIII	P00451	1832	KTYFWKVQHHMAP	Hemophilia A	Lacroix-Desmazes et al. (2002); Di Giambattista et al. (2001); Vianello et al. (1999) Jacquemin and Saint-Remy (1998) Cabane and Bossi (1996)
Vesicular inhibitory amino acid transporter GABA and glycine transporter	Q9H598	334	SEFHCMMNWTHIA	Mouse model of amyotrophic lateral sclerosis	Raiteri et al. (2005)
Neuronal acetylcholine receptor protein subunit α-9	Q14CY7	456	DRFFMWIFFIMV	Myasthenia	Ricny et al. (2002)
Deoxycytidine kinase	P27707	156	QDWHDWMNNQFG	Graves' and Hashimoto's diseases	Karbownik et al. (2003)
Tryptophan pyrrolase	P14902	32	LPDFYNDWMFIA	Tryptophan chronic immune activation in neurologic/psychiatric disorders	Schrocksnadel et al. (2006)
Psoriasis susceptibility 1	Q5ST21	49	MEPANHFWHAGD	Psoriasis	Szczerkowska-Dobosz (2005); Ozawa and Aiba (2004)
Laeverin CHL2 antigen	Q6Q4G3	425	VTMNWWNNIWLN	Rheumatoid arthritis	Haas et al. (2006)
Toll-like receptor-9	Q9NR96	170	RFLFMDGNCYYK	Systemic lupus erythematosus	Yu et al. (2006); Prinz et al. (2006)
				Multiple sclerosis	
Protein Wnt-5a	P4122	68	KHLYQDHMQYIG	Rheumatoid arthritis	Sen et al. (2001)

ID Primary accession number. Only non-redundant peptide motifs $\geq$12 aa long have been considered

values for uniquely expressed pentapeptides as well as for the intra-proteomic non-redundancy of peptides of different amino acid length and formed only and exclusively by 5-mers expressed only once in the human proteome, i.e. with a content of redundancy equal to zero. That is, Table 1 lists the number of: i) pentapeptides expressed only once in the human proteome, ii) esapeptides formed by two sequential 5-mers, overlapping by four residues and each of which is expressed only once in the human proteome, iii) eptapeptides formed by three 5-mers, overlapping by four residues and each of which is expressed only once in the human proteome, and so on. As an example, the non-redundant DRFFMWIFFIMV sequence we find in the neuronal acetylcholine receptor protein subunit alpha-9 (Acc. Q14CY7, also see Table 2 for description) is a dodecapeptide formed by eight consecutive pentamers each overlapping the next by 4 amino acids, i.e. DRFFM, RFFMW, FFMWI, FMWIF, and so on, each pentameric subunit being present only and exclusively in the neuronal acetylcholine receptor protein subunit alpha-9.

On the whole, the numbers reported in Table 1 clearly document that the human proteome presents a high number of pentapeptides occurring only once (464485 unique pentapeptides equal to 19.4% of the actual pentapeptides found in the human proteome). Then, at a deeper analysis, it is interesting that, although the number of pentapeptides expressed only once in the human proteome is high, the

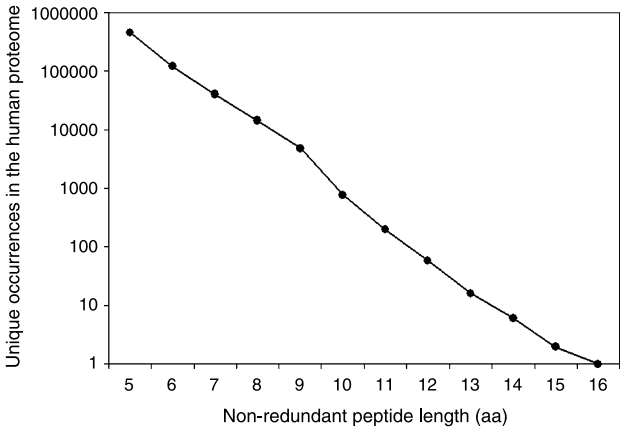

Fig. 1. Quantitation of peptide sequence non-redundancy in the human proteome as a function of peptide length

peptide sequences with more non-redundant pentapeptides in a row drastically decrease in count. Plotting peptide sequence non-redundancy in the human proteome as a function of peptide length reveals the behavior illustrated by Fig. 1.

Three observations are worth noting in Fig. 1. First, the number of n-mer peptides (N) formed by sole occurrence 5-mers and the length of the non-rendundant peptides (n) are inversely related by a logarithmic relationship. Indeed, plotting the number of n-mer peptides (N) formed by sole occurrence 5-mers on a log scale against n shows a linear behavior. That is, Fig. 1 indicates that the logarithm of the number of n-peptides formed by sole occurrence pentamers (N) is a linear function of the peptide length n. Second, the figure indicates that for each additional amino acid in the length, the number of non-redundant peptides is approximately 29% of the previous one, thus suggesting powerful constraint(s) that limit the presence of non-redundant peptide sequences. Third, Fig. 1 defines 16 amino acids as the length constraint to peptide non-redundancy in the human proteome. The longest human peptide entirely formed by non-redundant pentamers is a 16 amino acid long fragment. Biologically, on the whole Fig. 1 suggests that level of peptide sequence redundancy is not randomly distributed in the human proteome.

The mathematical treatment of the data listed in Table 1 and graphically reported in Fig. 1 is detailed in Box 1.

Location of the non-redundant sequences in human proteins

Next, we analyzed the location of the non-redundant sequences in the human proteins, searching for possible relationship(s) between non-redundant fragments and cancer/autoimmunity auto-antigens. The analysis was restricted

Box 1. N is exponentially related to the peptide length n

The relationship illustrated in Fig. 1 can be described by the equation

$$\log(N) = a + bn \tag{1}$$

or alternatively

$$N = 10^{(a+bn)} \tag{2}$$

where N = number of peptides of length n formed by sole occurrence 5-mers and n = the peptide length (with $5 \leq n \leq 16$). The formula

$$f(n) = a + bn$$

is the general form for a straight line where a is the "y-intercept" (the value of $\log(N)$ for a hypothetical value of 0 for n) and b is the slope of the line (i.e. the slope on the plot of $\log(N)$ versus n). We can obtain the values for a and b by performing a linear regression using a least-squares technique and the data reported in Table 1. We obtain: $a = 8.338747086$; $b = -0.536531468$, with $r = -0.997853259$. The value of r, the Pearson correlation coefficient, indicates how closely the data fits a linear relationship, with $r = -1$ or $r = 1$ for a perfect straight line and $r = 0$ for completely unrelated data (e.g. random data). A value of $r < -0.5$ or $r > 0.5$ is usually taken as indicating the presence of a linear relationship. In this case $r = -0.997853259$ which is highly supportive of the conclusion that a linear equation can describe the relationship between n and $\log(N)$. The fact that r is negative indicates that the two variables (n and $\log(N)$ in this case) are inversely related; i.e. as n gets larger, $\log(N)$ gets smaller.

From Eq. (2) above and the values for a and b we can give a formula for N, the number of n-mer peptides (N) formed by sole occurrence 5-mers. This formula is

$$N = 10^{(8.338747086 - 0.536531468n)} \tag{3}$$

Rearranging the exponent gives

$$\begin{aligned} &8 + 0.338747086 - 0.536531468n \\ &\quad = 0.338747086 + 8 - 0.536531468n \end{aligned}$$

so

$$N = 10^{(0.338747086 + 8 - 0.536531468n)} \tag{4}$$

Using the mathematical identity $10^{x+y} = 10^x \times 10^y$, Eq. (4) can be rewritten as

$$N = 10^{0.338747086} \times 10^{(8 - 0.536531468n)} \tag{5}$$

The first term on the right-hand side of this equation is a constant and can be simplified. Therefore, a final

expression for N is

$$N = 2.18145915 \times 10^{(8-0.536531468n)} \quad (6)$$

Solving the equation for values of n from 5 to 16, it turns out that for each additional amino acid in the length, the number of non-redundant peptides is approximately 29% of the previous one.

to non-redundant fragments having a number of amino acids $\geq$12. The data we obtained are reported in Tables 2 and 3. It can be seen that the longest human peptide entirely formed by non-redundant pentamers, i.e. the 16 amino acid long fragment DCVTWACHHWHTQYSN, is present in the ubiquitin-activating enzyme E1, Acc. P22314 (Table 2). That is, the DCVTWACHHWHTQYSN peptide is formed by 12 overlapping 5-mers, none of which recurs in the approximately 36 thousand human proteins we investigated. Importantly, Table 2 illustrates that the DCVTWACHHWHTQYSN peptide as well as a number of peptide fragments endowed with zero redundancy at pentameric level are located in proteins which, to some extent, are involved in autoimmune diseases, from multi-

Table 3. Human proteins hosting long non-redundant peptide motifs and relevant to cancer diseases

Protein description – name	ID	Aa pos	Sequence	Tumor pathology	References
Ubiquitin-activating enzyme E1	P22314	705	DCVTWACHHWHTQYSN	Human cancers	Sun (2006)
Matrix metalloproteinase-9	P14780	114	LKWHHHNITYWIQN	Prostate, endometrial, rectal cancer	Van Themsche et al. (2007); Unsal Kilic et al. (2007); Nabha et al. (2006)
TBC1 domain family member 8	O95759	582	KREHMMSCYWEQPR	Neuroblastoma	Roberts et al. (1998)
Ubiquitin-protein ligase	Q96EP1	632	RPDCYWGRNCRTQ	Human cancers	Pray et al. (2002)
Guanylate cyclase subunit β-2	O75343	279	QMIWMESMWCMVY	Gastric carcinoma	Behrends and Vehse (2000)
Coagulation factor VIII	P00451	1832	KTYFWKVQHHMAP	Prostatic adenoma; colorectal tumors	Barrier et al. (2006); Moon et al. (2006)
Huntingtin-interacting protein 14	Q8IUH5	366	TKFWMYVTWFFWF	Human tumors	Ducker et al. (2004)
Breast cancer anti-estrogen resistance protein 3	Q5TEW3	635	GTDMWEKNDQSCE	Breast cancer	Liang et al. (2006)
Ribosome biogenesis block of proliferation 1 protein	Q14137	462	VIVCHGMVYNDL	Colorectal tumor	Killian et al. (2006)
Ameloblastin	Q3B861	434	PEMMHDAWHFQE	Odontogenic tumors	Perdigao et al. (2004)
Breast epithelial mucin-associated antigen	Q3KP51	142	MVIWGANA	Prostate cancer	Teh et al. (2004)
Deoxycytidine kinase	P27707	156	QDWHDWMNNQFG	Pancreatic cancer, hairy cell and lymphocytic leukaemia. Human tumors	Vernejoul et al. (2006); Juliusson and Liliemark (2006); Pauwels et al. (2006); Lotfi et al. (2006)
Dermatopontin	Q07507	125	DREWQFYCCRYS	Prostate cancer; leiomyomas	Takeuchi et al. (2006); Catherino et al. (2004)
Dihydropyrimidinase,	Q14117	282	THYWNKEWHHAA	Oral squamous cell; gastric cancer	Garnis et al. (2004); Abe et al. (2003)
Axonemal dynein intermediate chain 2	Q9GZS0	399	SIMWTKYHMAYL	Ovarian cancer; prostate cancer	Ding et al. (2005); Bull et al. (2001)
Protocadherin fat 2	Q5W9G8	3844	QRHVNDHEWHSI	Neurofibromatosis-2	Willecke et al. (2006)
Tryptophan pyrrolase	P14902	32	LPDFYNDWMFIA	Hepatocarcinoma and other neoplasms; renal tumors	Vazquez et al. (1999); Cozzolino et al. (1969); Allegri et al. (1965)
Transmembrane protease	Q17RH4	153	KGCEHGWWEINE	Prostate cancer; ovarian cancer	Saleem et al. (2006); Tanimoto et al. (2005)
Retinoblastoma-associated factor 600	Q5T4S7	1671	FMNQHWYHCHTC	Retinoblastoma	Huh et al. (2005)
Oncostatin-M specific receptor β subunit	Q99650	288	CTHKNWCNWQIT	Melanoma; malignant mammary epithelial cells	Lacreusette et al. (2007); Liu et al. (1998)
Zinc in-between-RING-finger ubiquitin-associated domain protein	Q96EP0	411	QVWYCIHCTFCN	Breast cancer	Thompson et al. (2004)
Protein Wnt-5a	P41221	68	KHLYQDHMQYIG	Neuroblastoma; ewing sarcoma; breast cancer; colorectal cancer	Blanc et al. (2005); Uren et al. (2004); Jonsson et al. (2002); Smith et al. (2002)

ID = primary accession number. Only non-redundant peptide motifs $\geq$12 aa long have been considered

ple sclerosis to lupus erythematosus. In parallel, it was found that a number of non-redundant peptides are present in a number of auto-antigens studied in main tumoral pathologies (Table 3).

Discussion

We use computational biology to select peptide sequences having the immunogenic potential to evoke humoral response. The selection criterion is based on the analysis of the sequence redundancy level to the human proteome. The eventual goals are 1) to elucidate the regulatory mechanisms/factors that dictate peptide immunodominance assessment, and 2) to use peptide selection as a basis of innovative technologies and biological applications in disease prevention/cure. Indeed, the characterization of peptide sequences uniquely expressed in tumor auto-antigens might be therapeutically used in order to produce specific antibodies which selectively hit tumor cells. Of not secondary importance, such an approach would offer effective, safe and apparently ''infallible'' tool to specifically destroy tumors by hitting peptide motifs uniquely expressed in the tumoral antigens, without cross reactions with other cell components. The current antigen-specific immunotherapeutical protocols target not only antigen-positive tumor cells, but also normal tissues expressing the shared tumor antigen by molecular mimicry and/or epitope spreading (Miller et al., 2001; Tian et al., 2002). In this context, the data reported in Table 3 might be a concrete step along new, safe and practically ''infallible'' cancer therapies.

In addition, our data are of interest in furthering our understanding of the molecular mechanisms involved in peptide antigenicity and/or immunogenicity (Van Regenmortel, 2004). If immune reactions are related to a molecular definition of amino acid sequences never/sporadically encountered by the immune system, it is logical that auto-antigen peptide epitopes able to evoke humoral autoimmune response are defined by low sequence similarity to the proteome. The data reported in Table 2 may thus constitute a first ''sequence block'' to be investigated in the context of autoimmunity and related phenomena (i.e. epitope spreading, cripticity, glycosylation) and possibly might lead to the exact definition of peptides to be used in order to block and neutralize circulating autoantibodies in autoimmune diseases, with the concomitant gain of a complete absence of cross-reactions.

Acknowledgements

We are deeply grateful to: Cecilia Arighi, Georgetown University Medical Center, for making always available PIR facility, and Mick Bickis, Guglielmo Lucchese and Angela Stufano for precious discussion, help in the data analysis and manuscript reading.

References

Abe S, Kojima M, Tamura H, Kurihara H, Kitago M, Kobayashi T, Nakamura T, Ogihara T (2003) A clinical results of TS-1 in advanced and recurrent gastric cancer in our hospital. Gan To Kagaku Ryoho 30: 963–970

Allegri G, Benassi CA, Boccu E, De Nadai A, Persissinotto B (1965) Tryptophan pyrrolase, kynureninase and kynurenine transaminase activities of human renal tumours. Br J Cancer 19: 754–760

Barrier F, de La Guerenne A, Semjen E, Wierre L, Horrent S, Decaudin B, Resibois JP, Weillaert MP, Azar R, Odou P (2006) Spontaneously acquired haemophilia: report of a patient with prostatic adenoma. Ann Biol Clin (Paris) 64: 347–352

Behrends S, Vehse K (2000) The beta(2) subunit of soluble guanylyl cyclase contains a human-specific frameshift and is expressed in gastric carcinoma. Biochem Biophys Res Commun 271: 64–69

Bielekova B, Goodwin B, Richert N, Cortese I, Kondo T, Afshar G, Gran B, Eaton J, Antel J, Frank JA, McFarland HF, Martin R (2000) Encephalitogenic potential of the myelin basic protein peptide (amino acids 83–99) in multiple sclerosis: results of a phase II clinical trial with an altered peptide ligand. Nat Med 6: 1167–1175

Blanc E, Roux GL, Benard J, Raguenez G (2005) Low expression of Wnt-5a gene is associated with high-risk neuroblastoma. Oncogene 24: 1277–1283

Bull JH, Ellison G, Patel A, Muir G, Walker M, Underwood M, Khan F, Paskins L (2001) Identification of potential diagnostic markers of prostate cancer and prostatic intraepithelial neoplasia using cDNA microarray. Br J Cancer 84: 1512–1519

Cabane J, Bossi P (1996) Acquired hemophilia caused by anti-factor VIII autoantibodies: a model of autoimmunity. Rev Med Interne 17: 115–116

Catherino WH, Leppert PC, Stenmark MH, Payson M, Potlog-Nahari C, Nieman LK, Segars JH (2004) Reduced dermatopontin expression is a molecular link between uterine leiomyomas and keloids. Genes Chromosomes Cancer 40: 204–217

Cozzolino G, Scianaro L, Dominici G, Milia U (1969) Tryptophan-pyrrolase activity in patients with neoplasms of various types. Boll Soc Ital Biol Sper 45: 539–542

Descamps FJ, Van den Steen PE, Nelissen I, Van Damme J, Opdenakker G (2003) Remnant epitopes generate autoimmunity: from rheumatoid arthritis and multiple sclerosis to diabetes. Adv Exp Med Biol 535: 69–77

Di Giambattista M, Branckaert T, Laub R (2001) Mapping of natural anti-factor VIII antibodies in plasma pools from healthy donors: use of rationally designed synthetic peptides. Biologicals 29: 229–232

Ding W, Tang Q, Espina V, Liotta LA, Mauger DT, Mulder KM (2005) A transforming growth factor-beta receptor-interacting protein frequently mutated in human ovarian cancer. Cancer Res 65: 6526–6533

Dubois B, Opdenakker G, Carton H (1999) Gelatinase B in multiple sclerosis and experimental autoimmune encephalomyelitis. Acta Neurol Belg 99: 53–56

Ducker CE, Stettler EM, French KJ, Upson JJ, Smith CD (2004) Huntingtin interacting protein 14 is an oncogenic human protein: palmitoyl acyltransferase. Oncogene 23: 9230–9237

Fontoura P, Garren H, Steinman L (2005) Antigen-specific therapies in multiple sclerosis: going beyond proteins and peptides. Int Rev Immunol 24: 415–446

Garnis C, Coe BP, Zhang L, Rosin MP, Lam WL (2004) Overexpression of LRP12, a gene contained within an 8q22 amplicon identified by high-resolution array CGH analysis of oral squamous cell carcinomas. Oncogene 23: 2582–2586

Gautam AM, Pearson CI, Smilek DE, Steinman L, McDevitt HO (1992) A polyalanine peptide with only five native myelin basic protein residues induces autoimmune encephalomyelitis. J Exp Med 176: 605–609

Gusfield D (1997) Algorithms on strings, trees, and sequences: computer science and computational biology. Cambridge University Press, Cambridge

Haas CS, Creighton CJ, Pi X, Maine I, Koch AE, Haines GK, Ling S, Chinnaiyan AM, Holoshitz J (2006) Identification of genes modulated in rheumatoid arthritis using complementary DNA microarray analysis of lymphoblastoid B cell lines from disease-discordant monozygotic twins. Arthritis Rheum 54: 2047–2060

Hartemann-Heurtier A, Mars LT, Bercovici N, Desbois S, Cambouris C, Piaggio E, Zappulla J, Saoudi A, Liblau RS (2004) An altered self-peptide with superagonist activity blocks a CD8-mediated mouse model of type 1 diabetes. J Immunol 172: 915–922

Hemmer B, Kondo T, Gran B, Pinilla C, Cortese I, Pascal J, Tzou A, McFarland HF, Houghten R, Martin R (2000) Minimal peptide length requirements for CD4(+) T cell clones-implications for molecular mimicry and T cell survival. Int Immunol 12: 375–383

Hrda P, Sterzl I, Matucha P, Korioth F, Kromminga A (2004) HLA antigen expression in autoimmune endocrinopathies. Physiol Res 53: 191–197

Huh KW, DeMasi J, Ogawa H, Nakatani Y, Howley PM, Munger K (2005) Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600. Proc Natl Acad Sci USA 102: 11492–11497

Jacquemin MG, Saint-Remy JM (1998) Factor VIII immunogenicity. Haemophilia 4: 552–557

Jonsson M, Dejmek J, Bendahl PO, Andersson T (2002) Loss of Wnt-5a protein is associated with early relapse in invasive ductal breast carcinomas. Cancer Res 62: 409–416

Juliusson G, Liliemark J (2006) Purine analogues: rationale for development, mechanisms of action, and pharmacokinetics in hairy cell leukaemia. Hematol Oncol Clin North Am 20: 1087–1097

Kamradt T, Mitchison NA (2001) Tolerance and autoimmunity. N Engl J Med 344: 655–664

Kanduc D (2005) Peptimmunology: immunogenic peptides and sequence redundancy. Curr Drug Discov Technol 2: 239–244

Karbownik M, Brzezianska E, Zasada K, Lewinski A (2003) Expression of genes for certain enzymes of pyrimidine and purine salvage pathway in peripheral blood leukocytes collected from patients with Graves' or Hashimoto's disease. J Cell Biochem 89: 550–555

Killian A, Sarafan-Vasseur N, Sesboue R, Le Pessot F, Blanchard F, Lamy A, Laurent M, Flaman JM, Frebourg T (2006) Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis. Genes Chromosomes Cancer 45: 874–881

Lacreusette A, Nguyen JM, Pandolfino MC, Khammari A, Dreno B, Jacques Y, Godard A, Blanchard F (2007) Loss of oncostatin M receptor beta in metastatic melanoma cells. Oncogene 26: 881–892

Lacroix-Desmazes S, Misra N, Bayry J, Mohanty D, Kaveri SV, Kazatchkine MD (2002) Autoantibodies to factor VIII. Autoimmun Rev 1: 105–110

Liang Y, Brekken RA, Hyder SM (2006) Vascular endothelial growth factor induces proliferation of breast cancer cells and inhibits the anti-proliferative activity of anti-hormones. Endocr Relat Cancer 13: 905–919

Liiv I, Teesalu K, Peterson P, Clemente MG, Perheentupa J, Uibo R (2002) Epitope mapping of cytochrome P450 cholesterol side-chain cleavage enzyme by sera from patients with autoimmune polyglandular syndrome type 1. Eur J Endocrinol 146: 113–119

Liu J, Hadjokas N, Mosley B, Estrov Z, Spence MJ, Vestal RE (1998) Oncostatin M-specific receptor expression and function in regulating cell proliferation of normal and malignant mammary epithelial cells. Cytokine 10: 295–302

Liu Z, Diaz LA, Haas AL, Giudice GJ (1992) cDNA cloning of a novel human ubiquitin carrier protein. An antigenic domain specifically recognized by endemic pemphigus foliaceus autoantibodies is encoded in a secondary reading frame of this human epidermal transcript. J Biol Chem 267: 15829–15835

Lotfi K, Karlsson K, Fyrberg A, Juliusson G, Jonsson V, Peterson C, Eriksson S, Albertioni F (2006) The pattern of deoxycytidine- and deoxyguanosine kinase activity in relation to messenger RNA expression in blood cells from untreated patients with B-cell chronic lymphocytic leukaemia. Biochem Pharmacol 71: 882–890

Lucchese G, Stufano A, Trost B, Kusalik A, Kanduc D (2006) Peptidology: short amino acid modules in cell biology and immunology. Amino Acids DOI 10.1007/s00726-006-0458-z

Mahdavi A, Monk BJ (2005) Vaccines against human papillomavirus and cervical cancer: promises and challenges. Oncologist 10: 528–538

Matache C, Stefanescu M, Dragomir C, Tanaseanu S, Onu A, Ofiteru A, Szegli G (2003) Matrix metalloproteinase-9 and its natural inhibitor TIMP-1 expressed or secreted by peripheral blood mononuclear cells from patients with systemic lupus erythematosus. J Autoimmun 20: 323–331

Miller SD, Katz-Levy Y, Neville KL, Vanderlugt CL (2001) Virus-induced autoimmunity: epitope spreading to myelin autoepitopes in Theiler's virus infection of the central nervous system. Adv Virus Res 56: 199–217

Moon WS, Park HS, Yu KH, Jang KY, Kang MJ, Park H, Tarnawski AS (2006) Expression of angiopoietin 1, 2 and their common receptor Tie2 in human gastric carcinoma: implication for angiogenesis. J Korean Med Sci 21: 272–278

Nabha SM, Bonfil RD, Yamamoto HA, Belizi A, Wiesner C, Dong Z, Cher ML (2006) Host matrix metalloproteinase-9 contributes to tumor vascularization without affecting tumor growth in a model of prostate cancer bone metastasis. Clin Exp Metastasis 23: 335–344

Niman HL, Houghten RA, Walker LE, Reisfeld RA, Wilson IA, Hogle JM, Lerner RA (1983) Generation of protein-reactive antibodies by short peptides is an event of high frequency: implications for the structural basis of immune recognition. Proc Natl Acad Sci USA 80: 4949–4953

Ohashi PS (2003) Negative selection and autoimmunity. Curr Opin Immunol 15: 668–676

Opdenakker G, Dillen C, Fiten P, Martens E, Van Aelst I, Van den Steen PE, Nelissen I, Starckx S, Descamps FJ, Hu J, Piccard H, Van Damme J, Wormald MR, Rudd PM, Dwek RA (2006) Remnant epitopes, autoimmunity and glycosylation. Biochim Biophys Acta 1760: 610–615

Ozawa M, Aiba S (2004) Immunopathogenesis of psoriasis. Curr Drug Targets Inflamm Allergy 3: 137–144

Park Y, Kim S, Park S, Kim T, Yu L, Rewers M, Eisenbarth GS (2004) Differential expression of nonislet autoimmunity: comparison of Korean and U.S. patients with type 1 diabetes. Ann NY Acad Sci 1037: 69–73

Paulissen G, Rocks N, Quesada-Calvo F, Gosset P, Foidart JM, Noel A, Louis R, Cataldo DD (2006) Expression of ADAMs and their inhibitors in sputum from patients with asthma. Mol Med 12: 171–179

Pauwels B, Korst AE, Pattyn GG, Lambrechts HA, Kamphuis JA, De Pooter CM, Peters GJ, Lardon F, Vermorken JB (2006) The relation between deoxycytidine kinase activity and the radiosensitising effect of gemcitabine in eight different human tumour cell lines. BMC Cancer 6: 142

Perdigao PF, Gomez RS, Pimenta FJ, De Marco L (2004) Ameloblastin gene (AMBN) mutations associated with epithelial odontogenic tumors. Oral Oncol 40: 841–846

Pray TR, Parlati F, Huang J, Wong BR, Payan DG, Bennett MK, Issakani SD, Molineaux S, Demo SD (2002) Cell cycle regulatory E3 ubiquitin ligases as anticancer targets. Drug Resist Updat 5: 249–258

Prinz M, Garbe F, Schmidt H, Mildner A, Gutcher I, Wolter K, Piesche M, Schroers R, Weiss E, Kirschning CJ, Rochford CD, Bruck W, Becher B (2006) Innate immunity mediated by TLR9 modulates pathogenicity in an animal model of multiple sclerosis. J Clin Invest 116: 456–464

Raiteri L, Zappettini S, Stigliani S, Paluzzi S, Raiteri M, Bonanno G (2005) Glutamate release induced by activation of glycine and GABA

transporters in spinal cord is enhanced in a mouse model of amyotrophic lateral sclerosis. Neurotoxicology 26: 883–892

Reddehase MJ, Rothbard JB, Koszinowski UH (1989) A pentapeptide as minimal antigenic determinant for MHC class I-restricted T lymphocytes. Nature 337: 651–653

Ricny J, Simkova L, Vincent A (2002) Determination of anti-acetylcholine receptor antibodies in myasthenic patients by use of time-resolved fluorescence. Clin Chem 48: 549–554

Roberts T, Chernova O, Cowell JK (1998) NB4S, a member of the TBC1 domain family of genes, is truncated as a result of a constitutional t(1;10)(p22;q21) chromosome translocation in a patient with stage 4S neuroblastoma. Hum Mol Genet 7: 1169–1178

Saleem M, Adhami VM, Zhong W, Longley BJ, Lin CY, Dickson RB, Reagan-Shaw S, Jarrard DF, Mukhtar H (2006) A novel biomarker for staging human prostate adenocarcinoma: overexpression of matriptase with concomitant loss of its inhibitor, hepatocyte growth factor activator inhibitor-1. Cancer Epidemiol Biomarkers Prev 15: 217–227

Schrocksnadel K, Wirleitner B, Winkler C, Fuchs D (2006) Monitoring tryptophan metabolism in chronic immune activation. Clin Chim Acta 364: 82–90

Sen M, Chamorro M, Reifert J, Corr M, Carson DA (2001) Blockade of Wnt-5A/frizzled 5 signaling inhibits rheumatoid synoviocyte activation. Arthritis Rheum 44: 772–781

Smith K, Bui TD, Poulsom R, Kaklamanis L, Williams G, Harris AL (1999) Up-regulation of macrophage wnt gene expression in adenoma-carcinoma progression of human colorectal cancer. Br J Cancer 81: 496–502

Sun Y (2006) E3 ubiquitin ligases as cancer targets and biomarkers. Neoplasia 8: 645–654

Szczerkowska-Dobosz A (2005) Human leukocyte antigens as psoriasis inheritance and susceptibility markers. Arch Immunol Ther Exp (Warsz) 53: 428–433

Takeuchi T, Suzuki M, Kumagai J, Kamijo T, Sakai M, Kitamura T (2006) Extracellular matrix dermatopontin modulates prostate cell growth in vivo. J Endocrinol 190: 351–361

Tanimoto H, Shigemasa K, Tian X, Gu L, Beard JB, Sawasaki T, O'Brien TJ (2005) Transmembrane serine protease TADG-15 (ST14/Matriptase/MT-SP1): expression and prognostic value in ovarian cancer. Br J Cancer 92: 278–283

Teh BS, Ayala G, Aguilar L, Mai WY, Timme TL, Vlachaki MT, Miles B, Kadmon D, Wheeler T, Caillouet J, Davis M, Carpenter LS, Lu HH, Chiu JK, Woo SY, Thompson T, Aguilar-Cordova E, Butler EB (2004) Phase I-II trial evaluating combined intensity-modulated radiotherapy and in situ gene therapy with or without hormonal therapy in treatment of prostate cancer-interim report on PSA response and biopsy data. Int J Radiat Oncol Biol Phys 58: 1520–1529

Thompson HG, Harris JW, Lin L, Brody JP (2004) Identification of the protein Zibra, its genomic organization, regulation, and expression in breast cancer cells. Exp Cell Res 295: 448–459

Tian J, Olcott AP, Kaufman DL (2002) Antigen-based immunotherapy drives the precocious development of autoimmunity. J Immunol 169: 6564–6569

Tsubata T, Honjo T (2000) B cell tolerance and autoimmunity. Rev Immunogenet 2: 18–25

Unsal Kilic D, Uner A, Akyurek N, Erpolat P, Dursun A, Pak Y (2007) Matrix metalloproteinase-9 expression correlated with tumor response in patients with locally advanced rectal cancer undergoing preoperative chemoradiotherapy. Int J Radiat Oncol Biol Phys 67: 196–203

Uren A, Wolf V, Sun YF, Azari A, Rubin JS, Toretsky JA (2004) Wnt/Frizzled signaling in Ewing sarcoma. Pediatr Blood Cancer 43: 243–249

Van Regenmortel MH (2004) Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism. EMBO Rep 5: 1016–1020

Van Themsche C, Mathieu I, Parent S, Asselin E (2007) Transforming growth factor-beta3 increases the invasiveness of endometrial carcinoma cells through phosphatidylinositol 3-kinase-dependent upregulation of X-linked inhibitor of apoptosis and protein kinase C-dependent induction of matrix metalloproteinase-9. J Biol Chem 282: 4794–4802

Vazquez E, Gerez E, Caballero F, Polo C, Batlle A (1999) Drug metabolizing enzyme system and heme pathway in hepatocarcinogenesis. Cancer Biochem Biophys 17: 25–34

Vergani D, Mieli-Vergani G (2004) Mechanisms of autoimmune hepatitis. Pediatr Transplant 8: 589–593

Vernejoul F, Ghenassia L, Souque A, Lulka H, Drocourt D, Cordelier P, Pradayrol L, Pyronnet S, Buscail L, Tiraby G (2006) Gene therapy based on gemcitabine chemosensitization suppresses pancreatic tumor growth. Mol Ther 14: 758–767

Vianello F, Tison T, Tagariello G, Zerbinati P, Zanon E, Scarano L, Girolami A (1999) Serological markers of autoimmunity in patients with hemophilia A: the role of hepatitis C virus infection, alpha-interferon and factor VIII treatment in skewing the immune system toward autoreactivity. Blood Coagul Fibrinolysis 10: 393–397

Wang CY, Walfield AM (2005) Site-specific peptide vaccines for immunotherapy and immunization against chronic diseases, cancer, infectious diseases, and for veterinary applications. Vaccine 23: 2049–2056

Willecke M, Hamaratoglu F, Kango-Singh M, Udan R, Chen CL, Tao C, Zhang X, Halder G (2006) The fat cadherin acts through the hippo tumor-suppressor pathway to regulate tissue size. Curr Biol 16: 2090–2100

Yan SD, Xiong WC, Stern DM (2006) Mitochondrial amyloid-beta peptide: pathogenesis or late-phase development? J Alzheimers Dis 9: 127–137

Yu P, Wellmann U, Kunder S, Quintanilla-Martinez L, Jennen L, Dear N, Amann K, Bauer S, Winkler TH, Wagner H (2006) Toll-like receptor 9-independent aggravation of glomerulonephritis in a novel model of SLE. Int Immunol 18: 1211–1219

Authors' address: D. Kanduc, Clinical Center, Immunogenetics Laboratory, Department of Transfusion Medicine, NIH, 10 Center Dr. Bldg 10, Room 1N224, Bethesda, MD, USA,
Fax: +1-301-402-1360, E-mail: kanducd@cc.nih.gov