Self/Nonself 1:4, 328-334; October/November/December 2010 © 2010 Landes Bioscience

No human protein is exempt from bacterial motifs, not even one

Brett Trost,[1] Guglielmo Lucchese,[2] Angela Stufano,[2] Mik Bickis,[3] Anthony Kusalik[1] and Darja Kanduc[2,*]

[1]Department of Computer Science; and [3]Department of Mathematics and Statistics; University of Saskatchewan; Saskatoon, Canada; [2]Department of Biochemistry and Molecular Biology; University of Bari; Bari, Italy

Key words: bacterial versus human peptide overlap, molecular mimicry, microbial immune escape, autoimmunity, vaccines

The hypothesis that mimicry between a self and a microbial peptide antigen is strictly related to autoimmune pathology remains a debated concept in autoimmunity research. Clear evidence for a causal link between molecular mimicry and autoimmunity is still lacking. In recent studies we have demonstrated that viruses and bacteria share amino acid sequences with the human proteome at such a high extent that the molecular mimicry hypothesis becomes questionable as a causal factor in autoimmunity. Expanding upon our analysis, here we detail the bacterial peptide overlapping to the human proteome at the penta-, hexa-, hepta- and octapeptide levels by exact peptide matching analysis and demonstrate that there does not exist a single human protein that does not harbor a bacterial pentapeptide or hexapeptide motif. This finding suggests that molecular mimicry between a self and a microbial peptide antigen cannot be assumed as a basis for autoimmune pathologies. Moreover, the data are discussed in relation to the microbial immune escape phenomenon and the possible vaccine-related autoimmune effects.

Introduction

The sustained increase in the incidence of autoimmune diseases in the population[1-4] and the continuously expanding list of autoimmune pathologies and autoantigens[5] necessitate investigations into the role of molecular mimicry in the triggering of autoimmunity.[6] Molecular mimicry, i.e., the sharing of a linear amino acid sequence or a conformational fit between a microbe and a host self determinant, has been and still is the predominant field of investigation in autoimmunity research.[7-19] Recently, we reported that viral proteins overlap extensively with the human proteome,[20,21] with only a limited number of viral pentamers not found in the human proteome. In conflict to the dominant tendency to causally associate viral infections and autoimmune diseases, these findings support the view that molecular mimicry is over-emphasized as a critical mechanism during autoimmune disease pathogenesis. We reasoned that, if there is a link between viral infections and autoimmune reactions, then the documented extent of viral peptide overlapping in the human proteome would suggest that the entire world human population would suffer from autoimmunity. In addition, the analysis of a number of bacterial proteomes for amino acid sequence similarity to the human proteome demonstrated the sharing of hundreds of nonamer sequences between bacterial and human proteomes.[22] Again, the implications of these data appear of importance to define the current molecular mimicry model and, in general, to understand basic mechanisms in pathology and address research towards new directions. Here, as a further step in our studies on autoimmunity mechanisms, we detail the bacterial versus human peptide overlapping at the 5-, 6-, 7- and 8-mer levels, and demonstrate that no human proteins are exempt from the presence of bacterial motifs.

Results

Analyzing forty bacterial proteomes versus the *Homo sapiens* proteome: an overlap snapshot. Forty bacterial proteomes, 20 pathogenic and 20 non-pathogenic, were analyzed for peptide sharing with the human proteome at the penta-, hexa-, hepta- and octapeptide level to examine bacterial-versus-human similarity. Peptide similarity analysis of bacterial proteomes versus the human proteome was conducted as already described in detail[20-22] and produced the data illustrated in **Table 1**.

Table 1, last line, shows that combining all bacterial proteomes into one protein set and then computing the overlap of this set with the human proteome gives 15,260,383 perfect pentapeptide matches distributed through 36,014 human proteins; 9,133,718 perfect hexapeptide matches distributed through 36,014 human proteins; 1,643,139 perfect heptapeptide matches distributed throughout 35,906 human proteins; and 200,708 perfect octapeptide matches distributed throughout 31,170 human proteins. That is, the bacterial-versus-human overlap at the penta- and hexapeptide levels spans the entire human proteome: there does

*Correspondence to: D. Kanduc; Email: d.kanduc@biologia.uniba.it and dkanduc@gmail.com
Submitted: 07/12/10; Revised: 08/10/10; Accepted: 08/11/10
Previously published online: www.landesbioscience.com/journals/selfnonself/article/13315
DOI: 10.4161/self.1.4.13315

Table 1. Peptide sharing between bacterial proteomes and the human proteome at the 5-, 6-, 7- and 8-mer level

Taxa Id	Bacterium name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
299768	*Streptococcus thermophilus*	450402	3863966	35961	92.8	448812	345526	33679	31.8	447222	29009	13533	3.5	445632	5821	2258	0.5
367928	*Bifidobacterium adolescentis*	593219	4720571	35984	91.6	591592	464558	34565	31.8	589965	41915	16817	3.6	588338	7312	3025	0.4
206672	*Bifidobacterium longum*	631639	4930744	35983	91.7	629916	507033	34673	32.5	628193	46282	17784	3.7	626470	7796	3266	0.4
257314	*Lactobacillus johnsonii*	580247	4373499	35966	91.6	578438	394509	34122	29.9	576629	32622	14862	3.2	574820	5814	2354	0.4
272621	*Lactobacillus acidophilus*	576525	4285917	35956	91.4	574666	378992	33885	29.0	572807	31372	14409	3.0	570948	5704	2190	0.4
203120	*Leuconostoc mesenteroides*	598355	4551589	35963	92.0	596353	419771	34303	30.4	594351	34970	15491	3.3	592349	6662	2432	0.5
416870	*Lactococcus lactis*	672683	4980061	35981	92.1	670299	495503	34595	31.5	667915	41383	17246	3.5	665531	6919	2831	0.4
393595	*Alcanivorax borkumensis*	896972	6304479	35999	91.5	894220	736390	35320	33.5	891468	62704	21456	4.1	888716	9487	4176	0.5
220668	*Lactobacillus plantarum*	904305	5835444	35989	90.8	901306	623037	35071	30.0	898307	52498	19685	3.3	895308	8195	3575	0.4
226185	*Enterococcus faecalis*	940332	6145880	35988	91.6	937095	668944	35158	31.1	933858	57666	20168	3.3	930621	9470	3796	0.4
420662	*Methylibium petroleiphilum*	1363510	7268505	36004	91.3	1359155	1153424	35558	35.9	1354800	123368	26348	5.1	1350445	19643	7555	0.7
251221	*Gloeobacter violaceus*	1359892	7559021	36004	91.7	1355488	1155772	35593	35.8	1351084	114279	26125	4.7	1346680	18514	7090	0.5
369723	*Salinispora tropica*	1499556	7020220	35995	91.6	1495034	1268110	35537	37.5	1490512	142761	27415	5.5	1485990	20474	8513	0.6
78245	*Xanthobacter autotrophicus*	1598054	7692783	36005	91.6	1593085	1285165	35609	35.8	1588116	136687	27146	4.9	1583147	21199	7980	0.6
138119	*Desulfitobacterium hafniense*	1580893	8351354	36000	90.5	1575878	1125244	35636	31.4	1570863	97706	25570	3.5	1565848	14124	6018	0.4
318586	*Paracoccus denitrificans*	1541153	7483126	35998	90.8	1536137	1192734	35649	34.5	1531121	122240	26532	4.7	1526105	17463	7217	0.6
351746	*Pseudomonas putida*	1734619	8415989	36006	90.5	1729375	1327148	35661	33.9	1724131	126059	27425	4.3	1718887	16807	7352	0.5
222523	*Bacillus cereus*	1505863	7776742	36005	90.0	1499864	981068	35501	29.7	1493866	82158	23724	3.2	1487873	10991	4881	0.4
366394	*Sinorhizobium medicae*	1896855	8634389	36006	90.6	1890710	1395567	35700	33.6	1884565	133621	27663	4.2	1878420	20378	7549	0.5
224911	*Bradyrhizobium japonicum*	2582736	9728471	36007	89.8	2574488	1816906	35804	32.9	2566240	183683	29933	4.1	2557992	27201	9643	0.5
471472	***Chlamydia trachomatis***	**306365**	**3176212**	**35959**	93.3	**305482**	**276387**	**33051**	34.4	**304599**	**22712**	**11834**	4.1	**303716**	**3440**	**1592**	0.5
455434	***Treponema pallidum***	**345928**	**3356569**	**35952**	93.0	**344863**	**299569**	**33510**	33.8	**343798**	**26743**	**12789**	3.9	**342733**	**5561**	**2064**	0.5
392021	***Rickettsia rickettsii***	**313106**	**2762374**	**35941**	92.6	**311761**	**226202**	**31809**	30.7	**310416**	**21699**	**10218**	3.6	**309071**	**5559**	**1605**	0.7
458234	***Francisella tularensis***	**450826**	**3551140**	**35966**	91.8	**449318**	**298668**	**33231**	29.6	**447810**	**24243**	**12135**	3.2	**446302**	**4463**	**1687**	0.5
85962	***Helicobacter pylori***	**485536**	**3780814**	**35963**	92.2	**483971**	**359383**	**33668**	31.3	**482406**	**29876**	**13919**	3.5	**480841**	**4281**	**1925**	0.4
224326	***Borrelia burgdorferi***	**412046**	**2867951**	**35918**	93.1	**410464**	**257661**	**32030**	30.8	**408882**	**20885**	**11291**	3.3	**407300**	**3026**	**1426**	0.4
195099	***Campylobacter jejuni***	**531256**	**3866438**	**35955**	92.2	**529420**	**374675**	**33686**	30.7	**527584**	**30421**	**14358**	3.4	**525748**	**5188**	**2141**	0.4
374833	***Neisseria meningitidis***	**559567**	**4473458**	**35977**	92.2	**557569**	**433384**	**34379**	32.4	**555571**	**38343**	**16154**	3.8	**553573**	**6201**	**2582**	0.6
516950	***Streptococcus pneumoniae***	**624663**	**4783135**	**35976**	92.2	**622474**	**467024**	**34432**	31.7	**620285**	**38722**	**16550**	3.5	**618096**	**6560**	**2702**	0.4
257309	***Corynebacterium diphtheriae***	**716248**	**5461966**	**35986**	92.2	**713983**	**601188**	**34952**	34.0	**711718**	**51983**	**19542**	4.0	**709453**	**7482**	**3527**	0.5
212717	***Clostridium tetani***	**799625**	**4942017**	**35972**	91.3	**797211**	**519578**	**34461**	29.5	**794797**	**42607**	**17049**	3.1	**792383**	**6520**	**2876**	0.4
273036	***Staphylococcus aureus***	**725020**	**4995196**	**35982**	91.2	**722512**	**465501**	**34460**	28.8	**720004**	**36783**	**16275**	3.0	**717496**	**5780**	**2508**	0.4
262698	***Brucella abortus***	**874969**	**6056419**	**35991**	92.0	**871895**	**714382**	**35165**	33.8	**868821**	**65769**	**21087**	4.2	**865747**	**12203**	**4314**	0.7
400673	***Legionella pneumophila***	**1021398**	**6572966**	**35999**	90.7	**1018194**	**703289**	**35287**	30.2	**1014990**	**57003**	**20586**	3.3	**1011786**	**9586**	**3736**	0.4
520	***Bordetella pertussis***	**1051997**	**6271766**	**35998**	91.6	**1048737**	**884648**	**35324**	35.3	**1045477**	**91024**	**23660**	4.9	**1042217**	**14756**	**5916**	0.7
243277	***Vibrio cholerae***	**1138797**	**7077244**	**36000**	90.7	**1134997**	**833801**	**35408**	31.2	**1131197**	**70647**	**22617**	3.5	**1127398**	**12337**	**4680**	0.4
349746	***Yersinia pestis***	**1123279**	**7065400**	**36001**	90.8	**1119462**	**846102**	**35433**	31.9	**1115645**	**73246**	**22780**	3.7	**1111828**	**11446**	**4709**	0.5
83331	***Mycobacterium tuberculosis***	**1307195**	**6877511**	**35993**	91.5	**1302949**	**1082521**	**35477**	35.8	**1298704**	**115970**	**25679**	4.9	**1294460**	**19203**	**7141**	0.6
99287	***Salmonella typhimurium***	**1401436**	**7838576**	**36001**	90.0	**1396908**	**1015124**	**35570**	31.4	**1392380**	**89038**	**24671**	3.6	**1387852**	**12184**	**5316**	0.4
261594	***Bacillus anthracis***	**1439159**	**7615447**	**36005**	90.1	**1433570**	**944793**	**35483**	29.7	**1427981**	**76954**	**23306**	3.2	**1422392**	**11356**	**4851**	0.4
	Total bacterial overlap*:		**15260383**	**36014**			**9133718**	**36014**			**1643139**	**35906**			**200708**	**31170**	

The level of peptide overlap between 40 bacterial proteomes and the human proteome is shown. The filtered bacterial proteomes consisted of 128,248 unique proteins, while the human proteome contained 36,014 proteins at the time of the analysis. Bacteria that are pathogenic to humans are shown in bold. Information for each bacterium can be found at ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi. Column details are as follows: (1) Number of 5-mer occurrences in the bacterial proteome (including duplicate instances of same unique 5-mer); (2) Observed bacterial 5-mer occurrences in the human proteome (including multiple occurrences); (3) Number of human proteins involved in the pentapeptide overlap; (4) Percent of unique bacterial 5-mers which occur in the human proteome; (5) Number of 6-mer occurrences in the bacterial proteome (including duplicate instances of same unique 6-mer); (6) Observed bacterial 6-mer occurrences in the human proteome (including multiple occurrences); (7) Number of human proteins involved in the hexapeptide overlap; (8) Percent of unique bacterial 6-mers which occur in the human proteome; (9) Number of 7-mer occurrences in the bacterial proteome (including duplicate instances of same unique7-mer); (10) Observed bacterial 7-mer occurrences in the human proteome (including multiple occurrences); (11) Number of human proteins involved in the heptapeptide overlap; (12) Percent of unique bacterial 7-mers which occur in the human proteome; (13) Number of 8-mer occurrences in the bacterial proteome (including duplicate instances of same unique 8-mer); (14) Observed bacterial 8-mer occurrences in the human proteome (including multiple occurrences); (15) Number of human proteins involved in the octapeptide overlap; (16) Percent of unique bacterial 8-mers which occur in the human proteome. *Obtained by combining all bacterial proteomes into one protein set, and computing the overlap of this set with the human proteome. The human proteome was downloaded from UniProtKB[23] and analyzed by custom programs written in C[24] (see under Methods).

not exist one human protein that does not host a bacterial hexapeptide. Only 104 of the 36,104 human proteins (i.e., about 0.3% of the human proteome) are exempt from bacterial heptapeptide motifs. The human proteins that do not harbor bacterial motifs at the heptapeptide level are listed in **Box 1**.

Actually, the heptapeptide sharing between bacteria and human proteome is extensive and massive as schematized in the circle graph of **Figure 1**, illustrating the percentage distribution of bacterial heptapeptides throughout the human proteome.

The bacterial motif distribution through the human proteome decreases at the octamer level, but is still impressive: only 4,844 human proteins (just 13.44% of the human proteome) are exempt from bacterial 8-mers.

In addition, **Table 1** shows that the microbe's pathogenicity does not affect the level of bacterial overlaps through the human proteins (see the percent of unique bacterial *n*-mers which occur in the human proteome, columns 4, 8, 12 and 16 and in **Table 1**). As a further confirmation, log-log plotting the bacterial 5-mer occurrences in the human proteome as a function of the bacterial proteome length produces the graph illustrated in **Figure 2**. It can be seen that the bacterial versus human overlap is independent of the microbe's pathogenicity and expectedly depends almost exclusively on the size of the bacterial proteome.

Analyzing bacterial proteins versus the *Homo sapiens* proteome for heptapeptide sharing: the *Klebsiella pneumoniae* and *Proteus mirabilis* paradigmatic examples. **Table 1** suggests

that molecular mimicry between a self and a microbial peptide antigen cannot be assumed as a single or exclusive basis for autoimmune pathologies. Also, it has to be underlined that the data reported above analyze only 40 bacterial proteomes. Therefore, taking into account that the human organism hosts hundreds of bacterial organisms amounting to trillions of bacterial cells,[25,26] this study greatly understates the level of overlap between bacterial and human proteomes. But even considering one bacterial protein only, we are presented with a marked bacterial-versus-human peptide commonality. In this regard, scientific relevant models are offered by *Klebsiella pneumoniae* and *Proteus mirabilis*. In the past decades, there has been an intensive scientific debate because of a consecutive sequence of six amino acids, Gln-Thr-Asp-Arg-Glu-Asp (QTDRED) shared between HLA B27.1 and the nitrogenase reductase enzyme of *K. pneumoniae*.[27] This sequence commonality was invoked as a possible structural basis for cross-reactivity to occur and cause of ankylosing spondylitis. Following a profusion of inconclusive papers on the issue,[27-33] the attention successively shifted on the molecular mimicry between human motifs (EQRRAA and LRREI) and *P. mirabilis* peptide sequences (ESRRAL and IRRET) as a possible aetiological basis for autoimmune rheumatoid arthritis.[34]

Peptide overlap analysis shows that *K. pneumoniae* and *P. mirabilis* proteomes have a peptide platform in common with the human proteome. As examples, **Table 2** shows the heptapeptide sharing between the human proteome and *K. pneumoniae* ATP-binding protein and *P. mirabilis* ATP-dependent RNA helicase protein. The *K. pneumoniae* ATP-binding protein (UniProt accession number: B5XMS5_KLEP3, aa 1–233) is formed by 227 heptamers, 36 of which are present in the human proteome. Analogously, *P. mirabilis* RNA helicase protein (UniProtKB: B4ESW0_PROMH, aa 1–465, formed by a total of 459 heptamers) has 190 perfect heptapeptide sequences in common with the human proteome (**Table 2**). In conclusion, data from **Table 2** further demonstrate that molecular mimicry cannot be considered as a single or exclusive causal factor in the genesis of autoimmune phenomena.

BOX 1: List of the human proteins that do no host a bacterial heptapeptide.

CATR1; COAS3; CT187; CU094; FA27L; KR124; KR192; KR211; KR410; KR412; KR413; KR414; KRA42; KRA44; KRA47; KRA81; LCE2B; MT1F; MT1G; MT1M; MT2; MT4; RL41; SPHAR; SPR2A; SPR2B; SPR2D; SPR2E; SPR2F; TRG11; Q2XP30; Q9MY73; Q05CR9; Q05CT7; Q0QVY9; Q6EHZ1; Q6PK85; Q7Z4Q0; Q86YX3; Q9HCX8; Q9P1F9; Q5FC06; Q86XP7; Q8IVI0; Q9NY32; Q9UI53; Q6JTU6; Q6ZVA9; Q86TX6; Q8IWU1; Q8NI73; Q96EQ2; Q9BXV1; Q9H325; Q5HYP9; Q6GZ88; Q7Z5A1; Q9H3A8; Q9P145; Q9P1I0; Q4VFV5; Q8IVH9; Q9NZ11; Q9P1F8; Q13254; Q6AWA8; Q7Z425; Q96S45; A0A4R1; A0MA52; Q07603; Q8WYR5; Q96Q13; Q9BZU2; Q9NYD4; Q9P1E0; Q9P1E9; Q9UI79; Q147W9; Q6JV79; Q6JV82; Q9HAZ7; A2RUG3; Q96IP2; Q31629; Q495H9; Q5JT78; Q5JVP1; Q5T7W9; Q5TAP0; Q68K28; Q6ZQP6; Q71M31; Q7LCP5; Q7Z4E0; Q86SX0; Q86YX6; Q8NG36; Q8WV73; MORN4; Q96IR5; Q96JR7; Q9BZU0; Q9UI80

Human proteins reported as accession numbers.

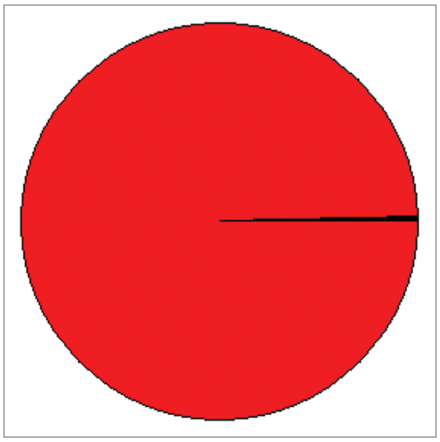

Figure 1. Distribution of bacterial heptapeptides throughout the human proteome: only 0.3% of the human proteome is exempt from bacterial heptapeptide motifs. The area corresponding to the human proteins containing bacterial heptapeptide(s) is reported in red. The area corresponding to the human proteins with no bacterial heptapeptide(s) is in black. The pie chart is a schematic representation of the peptide sharing between the forty bacteria under analysis in **Table 1** and *Homo sapiens*.

Discussion

Using a set of forty bacterial proteomes, this study shows that there does not exist a human protein exempt from a bacterial peptide overlap at the penta- and hexapeptide level and that only 104 proteins out of 36,104 do not host a microbial heptapeptide overlap. This finding is remarkable from a biological point of view and further supports a non-stochastic nature of the peptide overlapping between microbial and human proteomes. Our considerations are the following. Taking heptamer motifs as an example, we calculate that there are 1,280,000,000 possible heptapeptides that theoretically are available and might be used to build a human proteome exempt of bacterial heptapeptides. On the other hand, we know that the human proteome is formed by 36,103 proteins and 15,697,964 occurrences of 10,431,975 unique 7-mers[21] and, in the present study we find that only 104 human proteins out of 36,104 do not host a microbial heptapeptide overlap. That is, in face of the enormously high number of potential heptapeptides (1,280,000,000), the human proteome not only presents a high degree of repetitiveness in its 7-mer composition, but also utilizes heptapeptides common to bacteria so that almost no human protein is exempt from bacterial heptapeptide motifs. This peptide commonality has no mathematical justification. There is no shortage of possible heptapeptides, rather an incredibly huge number of heptapeptides are potentially available. Thus, we are forced to conclude that the redundancy present in the protein world is not stochastic (i.e., is not pure random chance), but reflects strong peptide usage bias.[21,35,36]

In addition, in light of the intensive research dedicated to understanding the function/effect of the presence of a single

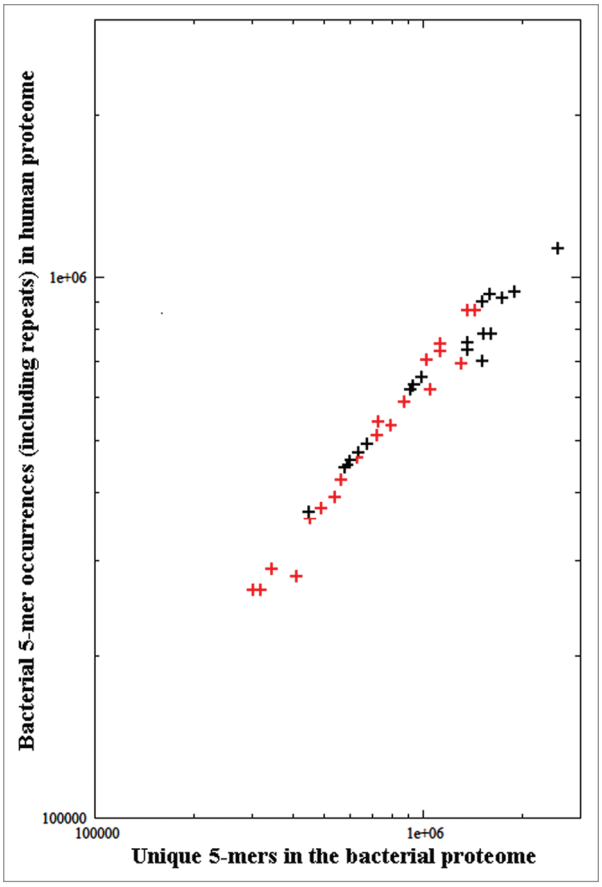

Figure 2. Bacterial 5-mer occurrences in the human proteome as a function of the bacterial proteome length. Symbol plus in black: non-pathogenic bacteria. Symbol plus in red: pathogenic bacteria. Information for bacteria can be found in **Table 1**.

bacterial match in a human protein looking for pathological correlates,[27-34] the present data are striking and seem to overturn our conceptualization of the relationship(s) between microbes and *Homo sapiens*. Actually, the data reported in this study are logical when analyzed in the light of the phylogenetic background linking bacteria and eukaryotes. Cells are of only two kinds: bacteria (or prokaryotes) and eukaryotes, which evolved from bacteria, possibly as recently as 800–850 My ago. As described in detail by Cavalier-Smith,[37] eukaryogenesis involved radical changes in almost every metabolic and structural aspect of the bacterial cell with a reorganization of the membrane and cytoskeleton apparatus and new chromosomal relationships to originate the eukaryotic nucleus and mitotic cycle. In this cell re-organization, new eukaryotic proteins evolved from old bacterial ones. Therefore, we can conclude that the data from **Tables 1 and 2** find a proper explanation in the evolutionary history of eukaryotes.

When analyzed from a pathological-clinical point of view, this report is of crucial importance in the study of autoimmune diseases for three reasons. As already discussed above, the data are of special relevance as regards the molecular mimicry hypothesis. Indeed, the molecular mimicry hypothesis suggests that, when bacterial/viral agents share epitopes with a host's protein, an immune response against the infectious agent may result in the formation of cross-reacting antibodies that bind the shared epitopes on the normal cell and result in the auto-destruction of the cell.[6] However, the extensive sequence similarity between bacteria and human proteins documented in this study suggests that molecular mimicry between a self and a microbial peptide antigen is inadequate to explain autoimmune pathologies.

Second, this study might contribute to explaining the microbial immune escape phenomenon. Scientific and clinical literature have been and are intensively debating the escape of microbes from immune control. A number of hypotheses and possible mechanisms have been proposed,[38-40] albeit with scarce results. Here, the quantitative analysis of *n*-peptide overlapping of bacterial versus human proteomes reported in **Tables 1 and 2** offers a logical and rational explanation to the *vexata quaestio* of microbial escape from immune surveillance. Indeed, the present data and our past studies[20,21] document that microbes are "a portion" of our human self and, consequently, presumably are subject to the same tolerance mechanisms that characterize human antigens and tissues. As a matter of fact, most chronic diseases, including pertussis,[41] tuberculosis,[42] leishmaniasis,[43] periodontitis,[44] gastritis,[45] to cite only a few of them, occur because an appropriate immune response required for pathogen clearance is not established. This causes a long-term pathogen colonization favored and progressively auto-sustained by pathogen-encoded molecules that enable the suppression of host immune response. The progressively increasing bacterial burden then causes a vicious cycle of bacterial proliferation and host tissue inflammation that translates into tissue damage, impaired function and eventual disease.

The third and most crucial consequence of this study is related to current anti-infectious vaccine preparations. Possibly as a consequence of immunotolerance mechanisms towards repeatedly shared peptide sequences, in general active vaccines produce a weak immune response; also, autoimmune cross-reactions are extremely rare events.[46-50] Under normal, non-stimulated conditions, the immune system fails to make immune responses to the infectious antigens present in the vaccines unless adjuvants are added.[17,48,49] As a rule, the current active vaccine formulations contain adjuvants to enhance immunogenicity.[50-52] The adjuvants serve to activate the immune system against microbial antigens that by themselves do not evoke immune responses, but rather are immunotolerated. However, as demonstrated in this and other studies of ours,[20-22] microbial antigens contain a high number of motifs shared with human proteins. Therefore, using viral or bacterial antigens in adjuvanted active vaccines will possibly trigger the immune system to react against the shared motifs (i.e., not only against the microbial antigen(s), but also against human self-molecules) with the concrete risk of developing adverse events and autoimmune pathologies in the human population.[53-56]

Methods

The human proteome was downloaded from UniProtKB (www.ebi.ac.uk),[23] and duplicated sequences and fragments were filtered out. After filtering, we were left with a human proteome consisting of 36,014 unique proteins, for a total of 15,806,702 amino acids. Bacterial proteomes were

Table 2. Heptapeptides are described by their amino acid position in the bacterial protein, amino acid sequence, and number of exact matches to the human proteome. The human proteins hosting bacterial heptapeptides are indicated by accession number (www.uniprot.org)

Aa Pos	Sequence	Matches	Human proteins hosting bacterial heptapeptides
***K. pneumoniae* ATP-binding protein:**			
41	VGTSGSG	1	Q8WXI7
44	SGSGKST	4	ABCB7; ABCBA; Q5T6J7; Q5T6J8
45	GSGKSTL	4	CAR15; CFTR; CN37; PEX1
46	SGKSTLL	4	CAR15; CFTR; Q9BX10; Q9P1K2
47	GKSTLLH	1	DIRA3
51	LLHLLGG	1	Q6IF12
103	SALENVA	1	CNR2
112	LLIGKKK	1	SEM4D
113	LIGKKKP	1	SEM4D
114	IGKKKPA	1	Q8N6H7
130	LQAVGLE	1	Q6JQN1
152	QRVAIAR	3	ABCB7; ABCBB; Q9HAQ7
153	RVAIARA	2	ABCB7; ABCBB
154	VAIARAL	2	ABCBB; CCR10
155	AIARALV	5	MDR1; MDR3; Q4G0Q4; Q6KG50; TAP2
164	PRLVLAD	1	TIE1
193	VAQRTAF	1	TJAP1
222	RLTADLT	1	RNF39
223	LTADLTL	1	RNF39
***P. mirabilis* ATP-dependent RNA helicase protein:**			
4	FTSLGLS	1	SYQ
5	TSLGLSE	1	SYQ
7	LGLSEAL	1	Q6ZMC8
8	GLSEALL	1	SNX7
9	LSEALLR	1	Q6W0C5
11	EALLRAI	1	ITIH2
25	PTPIQQQ	1	NGAP
32	AIEPILA	1	ANC5
46	AQTGTGK	2	DDX43; DDX53
47	QTGTGKT	7	DDX43; DDX53; KIF11; KIF3A; KIF3B; KIF3C; KIFC2
48	TGTGKTA	1	DDX27
49	GTGKTAA	4	DD19A; DD19B; DDX25; DDX27
50	TGKTAAF	4	DD19A; DD19B; DDX25; DDX27
59	PILEKLA	1	Q6VMQ6
79	ALILTPT	1	Q5T1V6
80	LILTPTR	1	Q5T1V6
81	ILTPTRE	1	Q5T1V6
82	LTPTREL	6	DDX24; DDX43; DDX47; DDX49; DDX53; Q5T1V6
83	TPTRELA	9	DDX24; DDX43; DDX46; DDX47; DDX49; DDX53; Q5T1V6; Q8NHQ9; Q8TEC9
86	RELAAQI	1	MA1C1
102	YLPIRSL	1	CN103
126	GVDVLIA	1	Q6ZND7
127	VDVLIAT	1	Q6ZND7
128	DVLIATP	1	Q6ZND7
129	VLIATPG	1	Q6ZND7
130	LIATPGR	2	DDX27; Q6ZND7
131	IATPGRL	11	DDX17; DDX23; DDX27; DDX43; DDX47; DDX49; DDX53; DDX54; DDX5; Q5T1V6; Q9UI98
132	ATPGRLL	2	DDX18; Q5T1V6
133	TPGRLLD	2	DDX18; Q5T1V6
153	VLVLDEA	3	Q8NAM8; Q8NHQ9; Q8TEC9
154	LVLDEAD	11	DDX10; DDX17; DDX1; DDX28; DDX3Y; DDX43; DDX48; DDX4; DDX5; Q8NHQ9; Q8TEC9
155	VLDEADR	9	DDX10; DDX17; DDX23; DDX3Y; DDX46; DDX4; DDX5; Q8NHQ9 Q8TEC9
156	LDEADRM	8	DDX17; DDX23; DDX27; DDX3Y; DDX41; DDX46; DDX4; DDX5
157	DEADRML	5	DDX17; DDX27; DDX3Y; DDX4; DDX5
158	EADRMLD	5	DDX17; DDX27; DDX3Y; DDX4; DDX5
159	ADRMLDM	4	DDX17; DDX3Y; DDX4; DDX5
160	DRMLDMG	4	DDX17; DDX3Y DDX4; DDX5
161	RMLDMGF	4	DDX17; DDX3Y; DDX4; DDX5
185	LLFSATF	5	DD19A; DD19B; DDX25; IRK10; Q86XP3
213	PKNSAAE	1	ARFP1
232	RKTELLS	1	CASC5
292	FKDGKLK	1	PGH1
302	ATDIAAR	1	DDX10
303	TDIAARG	1	DDX10
304	DIAARGL	1	DDX10
305	IAARGLD	1	DDX10
306	AARGLDI	10	DDX18; DDX21; DDX24; DDX27; DDX3Y; DDX4; DDX50; DDX54; Q59FR7; Q86XP3
307	ARGLDID	1	Q6ZUM4
328	EDYVHRI	1	DDX17
329	DYVHRIG	1	DDX17
330	YVHRIGR	6	DDX17; DDX1 DDX3Y; DDX41; DDX43; DDX4
331	VHRIGRT	6	DDX17; DDX3Y; DDX41; DDX43; DDX4; Q5VZQ4
332	HRIGRTG	10	DD19A; DD19B; DDX23; DDX25; DDX3Y; DDX41; DDX43; DDX4; DDX52; Q86XP3
332	HRIGRTG	10	DD19A; DD19B; DDX23; DDX25; DDX3Y; DDX41; DDX43; DDX4; DDX52; Q86XP3

Table 2. Heptapeptides are described by their amino acid position in the bacterial protein, amino acid sequence, and number of exact matches to the human proteome. The human proteins hosting bacterial heptapeptides are indicated by accession number (www.uniprot.org) (continued)

Aa Pos	Sequence	Matches	Human proteins hosting bacterial heptapeptides
333	RIGRTGR	10	DD19A; DD19B; DDX23; DDX25; DDX3Y; DDX41; DDX43; DDX4; DDX52; Q86XP3
334	IGRTGRA	4	DDX23; DDX43; DDX52; Q86XP3
337	TGRAAAT	1	ABCBB
341	AATGKAI	1	SMC3
391	KPKNKAR	1	KI21A
402	GGHGRAD	1	Q92827
438	KSKPARR	1	OR9Q2
446	RKHDDDR	3	Q2QGD7; ZXDA; ZXDB

downloaded from Integr8,[23] and each bacterial proteome was filtered in the same manner as the human proteome. The bacteria were chosen based on the following criteria: (1) known to be non-pathogenic or pathogenic; (2) phylogenetically different; (3) have proteomes established to a significant degree of completeness. In addition, the bacterial proteomes were chosen to span a range of proteome sizes, with the smallest bacterial proteome being 450,406 and 306,369 amino acids (for non-pathogenic and pathogenic bacteria, respectively), the largest being 2,582,740 and 1,439,163 amino acids (for non-pathogenic and pathogenic bacteria, respectively).

Sequence similarity analysis of each bacterial proteome to the human proteome was carried out using bacterial *n*-mers (with *n* from 5 to 8) sequentially overlapped by 4, 5, 6 and 7 residues, respectively. The scans were performed by custom programs written in C, which utilized suffix trees for efficiency.[24] The bacterial proteomes were manipulated and analyzed as follows. Each bacterial proteome was decomposed in silico to a set of penta-, hexa-, hepta- or octamers (including all duplicates). A library of unique penta-, hexa-, hepta- or octamers for each microbial proteome was then created by removing duplicates. Next, for each *n*-mer in the library, the entire human proteome was searched for instances of the same *n*-mer. Any such occurrence was termed an overlap or match. Cursory analysis (e.g., identification of unique overlapping *n*-mers, counts of unique overlapping *n*-mers, counts of duplications) were performed using LINUX/UNIX shell scripts and standard LINUX/UNIX utilities.

Authors' Contributions

B.T., G.L. and A.S. performed the computational analysis. M.B. and A.K. performed the mathematical analysis and supervised the computational analysis. D.K. proposed the original idea, interpreted the data, developed the research project and wrote the manuscript. All authors discussed the results and revised and commented on the manuscript with a particular contribution from A.K.

References

1. Redelings MD, McCoy L, Sorvillo F. Multiple sclerosis mortality and patterns of comorbidity in the United States from 1990 to 2001. Neuroepidemiology 2006; 26:102-7.
2. Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G. EURODIAB Study Group. Incidence trends for childhood type 1 diabetes in Europe during 1989–2003 and predicted new cases 2005–20: a multicentre prospective registration study. Lancet 2009; 373:2027-33.
3. Stark W, Huppke P, Gärtner J. Paediatric multiple sclerosis: the experience of the German Centre for Multiple Sclerosis in childhood and adolescence. J Neurol 2008; 255:119-22.
4. Ramírez-Zamora M, Burgos-Ganuza CR, Alas-Valle DA, Vergara-Galán PE, Ortez-González CI. Guillain-Barre syndrome in the paediatric age: epidemiological, clinical and therapeutic profile in a hospital in El Salvador. Rev Neurol 2009; 48:292-6.
5. Lernmark Å. Autoimmune diseases: are markers ready for prediction? J Clin Invest 2001; 108:1091-6.
6. Oldstone MBA. Molecular mimicry as a mechanism for the cause and as a probe uncovering etiologic agent(s) of autoimmune disease. Curr Top Microbiol Immunol 1989; 145:127-35.
7. Li de la Sierra I, Pernot L, Prangé T, Saludjian P, Schiltz M, Fourme R, et al. Molecular structure of the lipoamide dehydrogenase domain of a surface antigen from *Neisseria meningitidis*. J Mol Biol 1997; 269:129-41.
8. Karges WJ, Ilonen J, Robinson BH, Dosch HM. Self and nonself antigen in diabetic autoimmunity: Molecules and mechanisms. Mol Aspects Med 1995; 16:179-213.
9. Karopoulos C, Rowley MJ, Handley CJ, Strugnell RA. Antibody reactivity to mycobacterial 65 kDa heat shock protein: relevance to autoimmunity. J Autoimmun 1995; 8:235-48.
10. Steinman L, Oldstone MB. More mayhem from molecular mimics. Nat Med 1997; 3:1321-2.
11. O'Donohue J, McFarlane B, Bomford A, Yates M, Williams R. Antibodies to atypical mycobacteria in primary biliary cirrhosis. J Hepatol 1994; 21:887-9.
12. Oomes PG, Jacobs BC, Hazenberg MP, Bänffer JR, van der Meché FG. Anti-GM1 IgG antibodies and Campylobacter bacteria in Guillain-Barre' syndrome: evidence of molecular mimicry. Ann Neurol 1995; 38:170-5.
13. Maclaren NK, Alkinson MA. Insulin-dependent diabetes mellitus: The hypothesis of molecular mimicry between islet cell antigens and microorganisms. Mol Med Today 1997; 3:76-83.
14. Markesich DC, Sawai ET, Butel JS, Graham DY. Investigations on etiology of Crohn's disease. Humoral immune response to stress (heat shock) proteins. Dig Dis Sci 1991; 36:454-60.
15. Cunningham MW. Autoimmunity and molecular mimicry in the pathogenesis of post-streptococcal heart disease. Front Biosci 2003; 8:533-43.
16. Lamb DJ, El-Sankary W, Ferns GA. Molecular mimicry in atherosclerosis: a role for heat shock proteins inimmunisation. Atherosclerosis 2003; 167:177-85.
17. Waisbren BA Sr. Acquired autoimmunity after viral vaccination is caused by molecular mimicry and antigen complimentarity in the presence of an immunologic ajuvant and specific HLA patterns. Med Hypotheses 2008; 70:346-8.
18. Swanborg RH, Boros DL, Whittum-Hudson JA, Hudson AP. Molecular mimicry and *horror autotoxicus*: do chlamydial infections elicit autoimmunity? Expert Rev Mol Med 2006; 8:1-23.
19. Cunha-Neto E, Bilate AM, Hyland KV, Fonseca SG, Kalil J, Engman DM. Induction of cardiac autoimmunity in Chagas heart disease: a case for molecular mimicry. Autoimmunity 2006; 39:41-54.
20. Kusalik A, Bickis M, Lewis C, Li Y, Lucchese G, Marincola FM, et al. Widespread and ample peptide overlapping between HCV and *Homo sapiens* proteomes. Peptides 2007; 28:1260-7.
21. Kanduc D, Stufano A, Lucchese G, Kusalik A. Massive peptide sharing between viral and human proteomes. Peptides 2008; 29:1755-66.
22. Trost B, Kusalik A, Lucchese G, Kanduc D. Bacterial peptides are intensively present throughout the human proteome. Self/Nonself 2010; 1:1-4.
23. Kersey P, Bower L, Morris L, Horne A, Petryszak R, Kanz C, et al. Integr8 and Genome Reviews: integrated views of complete genomes and proteomes. Nucleic Acids Res 2005; 33:297-302.
24. Gusfield D. Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology. Cambridge University Press 1997.
25. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science 2005; 308:1635-8.
26. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science 2005; 307:1915-20.
27. Ewing C, Ebringer R, Tribbick G, Geysen HM. Antibody activity in ankylosing spondylitis sera to two sites on HLA B27.1 at the MHC groove region (within sequence 65–85) and to a *Klebsiella pneumoniae* nitrogenase reductase peptide (within sequence 181–199). J Exp Med 1990; 171:1635-47.
28. Schwimmbeck PL, Wu DTY, Oldstone MBA. Autoantibodies to HLA B27 in the sera of HLA B27 patients with ankylosing spondylitis and Reiter's syndrome. Molecular mimicry with *Klebsiella pneumoniae* as potential mechanism of autoimmune disease. J Exp Med 1987; 166:173.
29. Ebringer A, Cowling P, Ngwa-Suh N, James DCO, Ebringer R. Crossreactivity between *Klebsiella aerogenes* species and B27 lymphocyte antigens as an aetiological factor in ankylosing spondylitis. In: HLA and Disease J. Dausset and A. Svejgaard, editors. INSERM, Paris 1976; 58:27.

30. Schwimmbeck PL, Oldstone MBA. Autoimmune pathogenesis for ankylosing spondylitis (AS) and Reiter's syndrome (RS): autoantibodies against an epitope shared by HLA B27 and *Klebsiella pneumoniae* nitrogenase in sera of patients with AS and RS. Trans Assoc Am Phys 1987; 100:28-39.
31. Schwimmbeck PL, Oldstone MBA. Molecular mimicry between human leukocyte antigen B27 and Klebsiella. Consequences for spondyloarthropathies. Am J Med 1988; 85:51-3.
32. Husby G, Tsuchiya N, Schwimmbeck PL, Keat A, Pahle JA, Oldstone MBA. Williams RC Jr. Crossreactive epitope with *Klebsiella pneumoniae* nitrogenase in articular tissue of HLA-B27 positive patients with ankylosing spondylitis. Arthr Rheum 1989; 32:437-45.
33. Ogawasara M, Kono DH, Yu DTY. Mimicry of human histocompatibility HLA B27 antigens by *Klebsiella pneumoniae*. Infect Immun 1986; 51:901-8.
34. Wilson C, Tiwana H, Ebringer A. Molecular mimicry between HLA-DR alleles associated with rheumatoid arthritis and *Proteus mirabilis* as the aetiological basis for autoimmunity. Microbes Infect 2000; 2:1489-96.
35. Kusalik A, Trost B, Bickis M, Fasano C, Capone G, Kanduc D. Codon number shapes peptide redundancy in the universal proteome composition. Peptides 2009; 30:1940-4.
36. Capone G, Novello G, Fasano C, Trost B, Bickis M, Kusalik A, et al. The oligodeoxynucleotide sequences corresponding to never-expressed peptide motifs are mainly located in the non-coding strand. BMC Bioinformatics 2010; 11:383.
37. Cavalier-Smith T. Predation and eukaryote cell origins: a coevolutionary perspective. Int J Biochem Cell Biol 2009; 41:307-22.
38. Datta S, Panigrahi R, Biswas A, Chandra PK, Banerjee A, Mahapatra PK, et al. Genetic characterization of Hepatitis B virus in peripheral blood leukocytes: Evidence for selection and compartmentalization of viral variants with immune escape G145R mutation. J Virol 2009; 83:9983-92.
39. Middeldorp JM, Pegtel DM. Multiple roles of LMP1 in Epstein-Barr virus induced immune escape. Semin Cancer Biol 2008; 18:388-96.
40. van Kooyk Y, Appelmelk B, Geijtenbeek TB. A fatal attraction: *Mycobacterium tuberculosis* and HIV-1 target DC-SIGN to escape immune surveillance. Trends Mol Med 2003; 9:153-9.
41. McGuirk P, McCann C, Mills KH. Pathogen-specific T regulatory 1 cells induced in the respiratory tract by a bacterial molecule that stimulates interleukin 10 production by dendritic cells: a novel strategy for evasion of protective T helper type 1 responses by *Bordetella pertussis*. J Exp Med 2002; 195:221-31.
42. Boussiotis VA, Tsai EY, Yunis EJ, Thim S, Delgado JC, Dascher CC, et al. IL-10-producing T cells suppress immune responses in anergic tuberculosis patients. J Clin Invest 2000; 105:1317-25.
43. Belkaid Y, Piccirillo CA, Mendez S, Shevach EM, Sacks DL. $CD4^+CD25^+$ regulatory T cells control *Leishmania major* persistence and immunity. Nature 2002; 420:502-7.
44. Kinane DF, Marshall GJ. Periodontal manifestations of systemic disease. Aust Dent J 2001; 46:2-12.
45. Marshall BJ, Windsor HM. The relation of *Helicobacter pylori* to gastric adenocarcinoma and lymphoma: pathophysiology, epidemiology, screening, clinical presentation, treatment and prevention. Med Clin North Am 2005; 89:313-44.
46. Janeway CA Jr. Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb Symp Quant Biol 1989; 54:1-13.
47. Havarinasab S, Pollard KM, Hultman P. Gold- and silver-induced murine autoimmunity requirement for cytokines and CD28 in murine heavy metal-induced autoimmunity. Clin Exp Immunol 2009; 155:567-76.
48. Mizrahi M, Lalazar G, Ben Ya'acov A, Livovsky DM, Horowitz Y, Zolotarov L, et al. Beta-glycoglycosphingolipid-induced augmentation of the anti-HBV immune response is associated with altered CD8 and NKT lymphocyte distribution: a novel adjuvant for HBV vaccination. Vaccine 2008; 26:2589-95.
49. Fraser CK, Diener KR, Brown MP, Hayball JD. Improving vaccines by incorporating immunological coadjuvants. Expert Rev Vaccines 2007; 6:559-78.
50. Schmidt CS, Morrow WJ, Sheikh NA. Smart adjuvants. Expert Rev Vaccines 2007; 6:391-400.
51. Bryan JT. Developing an HPV vaccine to prevent cervical cancer and genital warts. Vaccine 2007; 25:3001-6.
52. Halperin SA, Dobson S, McNeil S, Langley JM, Smith B, McCall-Sani R, et al. Comparison of the safety and immunogenicity of hepatitis B virus surface antigen coadministered with an immunostimulatory phosphorothioate oligonucleotide and a licensed hepatitis B vaccine in healthy young adults. Vaccine 2006; 24:20-6.
53. Mandavilli A. When the vaccine causes disease Nat Med 2007; 13:274.
54. Caspi RR. Immunotherapy of autoimmunity and cancer: the penalty for success. Nat Rev Immunol 2008; 8:970-6.
55. Kanduc D. Quantifying the possible cross-reactivity risk of an HPV16 vaccine. J Exp Ther Oncol 2009; 8:65-76.
56. Ricco R, Kanduc D. Hepatitis B virus and *Homo sapiens* proteome-wide analysis: a profusion of viral peptide overlaps in neuron-specific human proteins. Biologics 2010; 4:75-81.