M. AVIUM BINDING TO HLA-DR EXPRESSED ALLELES IN SILICO: A MODEL OF PHENOTYPIC SUSCEPTIBILITY TO SARCOIDOSIS

C. Saltini[1], M. Pallante[1], E. Puxeddu[1], S. Contini[1], C.E. Voorter[2,3], M. Drent[3,4], M. Amicosante[1]

[1] Department of Internal Medicine, University of Rome "Tor Vergata", Rome, Italy; [2] Scuola di Dottorato in Scienze mediche e chirurgiche "Scienze pneumo-cardio-toraciche di interesse medico e chirurgico", University of Bologna, Italy; [3] Transplantation Immunology, Tissue Typing Laboratory, University Hospital Maastricht, Maastricht, The Netherlands; [4] Maastricht University Medical Centre, The Netherlands

ABSTRACT. Sarcoidosis is a systemic granulomatosis disease of unknown origin where a number of microbes, in particular *M. tuberculosis* and non-tuberculous mycobacteria, have been hypothesized to play a role in disease pathogenesis, possibly through bacterial antigen-driven hypersensitivity. To test this concept, we used bioinformatic tools allowing the identification of antigenic peptides in whole microbial genomes to analyze the interaction between the expressed HLA-DR gene allelic variants and the HLA-DR immunome of all pathogenic bacteria in a population of 149 sarcoidosis affected subjects and 447 controls, all HLA-typed at high resolution. We show here that patients with the Löfgren's syndrome, express HLA-DR alleles that recognize *in silico* a significantly higher number of bacterial antigen epitopes compared to the control population (18,496+9,114 vs 17,954+8,742; p<0.00001), and the chronic sarcoidosis affected population (17,954+8,742; p<0.00001 vs Löfgren's and controls). Further, the analysis of the ability of the HLA-DR allele combinations expressed by the Löfgren's and the chronic sarcoidosis affected subjects to recognize *M. avium* epitopes demonstrates that a significantly larger number of Löfgren's are capable of top affinity recognition, compared to chronic sarcoidosis (45% vs 17%, p<0.0037). Finally, both Löfgren's and chronic sarcoidosis subjects expressed HLA-DR allele combinations capable of *M. tuberculosis* and *M. avium* epitope recognition at higher affinity than tuberculosis affected subjects (p<0.01 all comparisons). In conclusion, we propose that - at least in a subgroup of affected subjects - sarcoidosis might be part of a spectrum of granulomatous responses to several agents where the Löfgren's syndrome represents the hyper-reactive end of the spectrum while pulmonary tuberculosis and atypical mycobacterial infections might represent the opposite end. *(Sarcoidosis Vasc Diffuse Lung Dis 2008; 25: 100-116)*

KEY WORDS: Löfgren's syndrome, chronic sarcoidosis, tuberculosis, HLA-DR, susceptibility, epitope prediction, T-cell response

INTRODUCTION

Sarcoidosis is a chronic disorder characterized by the accumulation of increased numbers of T-lym-

phocytes and activated mononuclear phagocytes in the lower respiratory tract and other sites of disease, leading to the formation of noncaseating granulomas in involved tissues such as the lung, skin and eyes. Several lines of evidence suggest that the accumulation of activated T-cells and macrophages in the lower respiratory tract is the result of chronic stimulation by some (yet unknown) antigen(s) (1). Firstly, in sarcoidosis activated alveolar macrophages possess increased ability of presenting antigens to T-cells (2) and express HLA-DR molecules at higher level at the surface of the cell (3). Secondly, T-cells accumulate in the lower respiratory tract where they release

interleukin 2 (4, 5) and interferon-γ at exaggerated rates (2). Thirdly, sarcoid lung T-lymphocytes display activated memory T-cell phenotypes (6, 7) and express oligoclonal Vβ T-cell receptor genes as they were locally expanded in response to a classical antigenic stimulus rather than a more broad-based stimulus such as from a superantigen (8, 9).

A number of potential causative agents have been implicated in the pathogenesis of sarcoidosis (10) and some sarcoidosis-like diseases and reactions could be identified as hypersensitivities to beryllium (11), titanium (12), zirconium (13) and aluminum (14). Recently, it has been shown that, in addition to chemicals, microbial agents otherwise capable of causing infection and disease may be implicated in the pathogenesis of granulomatous hypersensitivity pneumonias such as the hot tub lung, where *M. avium* has been identified as a causative agent and (15) and metal working fluid pneumonitis, where a variety of bacteria have been implicated (16). Furthermore, in a chronic granulomatous disease such as Chron's disease *M. avium* paratuberculosis has been implicated (17).

In sarcoidosis, since the application of the the polymerase chain reaction to the detection of mycobacterial genomes in clinical samples (18), the presence of nucleic acids of *M. tuberculosis,* non tuberculous mycobacteria and of *Propionibacterium acnes* has been observed both in bronchoalveolar lavage and tissue samples in different sarcoidosis patient populations (19-23). Collectively, these studies indicate that mycobacteria can be found in tissue samples in 23 to 30% of sarcoid patients, with a ten-fold higher probability compared to controls and suggest an association between mycobacteria and sarcoidosis (24-26).

In this context and with the knowledge that certain microbial agents could cause hypersensitivity reactions leading to granuloma formation in the absence of clinical infection, the observations that antibody and T-cell mediated immune responses against mycobacterial antigens, such as catalase-peroxidase (mKatG) (27), heat shock protein (Mtb-hsp) (28), and Antigen 85A (29) could be demonstrated in sarcoidosis and in the Kweim reaction, led to hypothesize that exposure to certain microbes might be causative of sarcoidosis in hyperreactive individuals (30-32).

Antigen presentation to CD4 T-cells is mediated by a class of cell surface receptors, the HLA class II proteins (HLA-DP, -DQ and -DR), which are capable of binding peptides derived from the digestion of microbes and other agents in the phago-lysosome of antigen presenting cells and to carry them to the cell surface and to present them to the CD4 T-cell expressing the appropriate T-cell antigen receptor. Importantly, the selection of antigenic peptides presented to, hence recognized by, T-cells is driven by the chemico-physical interaction between (i) the amino acid side chains which are lining the receptor-like pockets on the floor of the HLA antigen binding groove and (ii) the aminoacid side chains of the antigenic peptides (33, 34).

The HLA class II genes have been implicated in susceptibility to sarcoidosis (35, 36) and it has been shown that the expression of different HLA allelic variants is associated with different clinical presentations and disease outcomes. The Löfgren's syndrome, an acute form of sarcoidosis characterized by fever, erythema nodosum, bilateral hilar lymphoadenopaty and/or ankle arthritis is associated with the expression of the HLA-DRB1*0301 allele and has a favourable clinical outcome (37, 38). On the other hand, chronic forms of sarcoidosis with less favourable outcomes have been associated with the expression of the DRB1*1501 allele (37).

Key to the understanding of susceptibility to immune disorders is the knowledge that HLA alleles can select peptide antigens for antigen presentation with widely varying affinities, thus leading to different immune responses against the same agent. In this regard, it is well known that since the HLA polymorphisms characterizing allelic variants code for aminoacid changes in the peptide binding groove's, each HLA allele binds a unique set of aminoacid side chains. Each HLA class II allelic variant is therefore capable of selecting a discrete set of antigenic peptides for antigen presentation (39, 40) and to allow for the formation of unique HLA molecule/T-cell receptor/antigenic peptide complexes activating specific cytokine producing CD4 T-cells (41). Thus, since each individual coexpresses at least two HLA-DR molecules on the cell surface of the antigen presenting cell, it is reasonable to think that individual ability to generate immune responses to a given antigen shall be determined by the capability of each subject's two HLA-DR molecules combined to recognize antigen epitopes (42-44).

By using bioinformatics tools allowing the identification of antigenic peptides in whole microbial genomes (45), we have recently demonstrated that the expression of HLA-DR allele combinations endowed with a lower affinity for the *M. tubercuosis* genome and for its derived peptide antigens is associated with susceptibility to active tuberculosis (46). Interestingly, the HLA-DR alleles implicated in this reduced ability to bind and select for mycobacterial antigenic peptides are the same which have also been associated with chronic sarcoidosis. Even more interestingly, the alleles endowed with higher affinity for *M. tuberculosis* antigenic peptides, and with "protection" against tuberculosis, are the same which have been previously associated with the acute forms of sarcoidosis (46).

With the background that the affinity for peptide antigens of the HLA-DR expressed allelic variants plays a key role in determining the ability of the individual to recognize and react to bacterial antigens, hence in generating the ineffective immune response leading to bacterial invasion and active disease, we hypothesized that it may also play a role in determining susceptibility to hypersensitivity reactions to the same bacteria. To test this hypothesis, we set out to to analyze the interaction between the expressed HLA-DR gene allelic variants and the genomes of all pathogenic bacterial in the Löfgren's syndrome and in chronic sarcoidosis, two immunologically contrasting forms of sarcoidosis, in order to determine whether hypersensitivity to mycobacterial, propionibacterial or other bacterial antigens may play a potential role in the pathogenesis of sarcoidosis.

METHODS

Patients' characteristics

The study population was composed by the patients with sarcoidosis and matched controls of an already described population evaluated for genetic susceptibility factors in sarcoidosis. HLA-DR high resolution typing was available for all study subjects (35).

They were 149 patients with sarcoidosis, of which 39 with Löfgren's syndrome and 110 with chronic sarcoidosis, and 447 matched controls (Tab. 1). The quantitative implemented peptide binding motifs i.e., the algorithms with which to assess the

ability of a given HLA-DR allelic variant, are available only for 49 out of the more than 300 HLA-DRB1 alleles (47, 48) and cover, with at least one allele, about 90% of the HLA-DR variability of different human populations (49). For the purpose of this study, only the subjects carrying both HLA-DR alleles with an available HLA-DR binding motif were included (Tab. 1). Specifically, 349 out of 447 (78.1%) control subjects, 31 out of 39 (79.5%) patients with Löfgren's syndrome and 93 out of 110 (84.5%) patients with chronic sarcoidosis, could be subjected to the analysis.

The selected subgroups did not differ for demographical characteristics from the subgroups of subjects excluded from the study for having one or both HLA-DR alleles without a known binding motif (data not shown).

Finally, a group of patients with active tuberculosis and matched controls previously evaluated (46) were used as comparison. They include: (i) 303 patients with active tuberculosis of which 92 carrying both HLA-DR alleles with an available HLA-DR binding motif and (ii) 345 controls of which 111 carrying both HLA-DR alleles with an available HLA-DR binding motif (Tab. 1).

Genomes

One hundred and twenty four, non redundant, bacterial complete genomes, classified as "relevant human pathogens" in the Genome OnLine Database v2.0 (GOLD database update October 30th 2008, http://www.genomesonline.org/), were used to analyze the number of potential epitopes in each genome which were recognized by the study population. Bacterial genome characteristics (accession code, species, family, human disease and number of open reading frames) are listed in table 2.

Table 1. Study population

Study population	Cases (N)	Subjects with both HLA-DR alleles with known peptide binding motif[1]
Controls	447	349 (78.1%)
Löfgren	39	31 (79.5%)
Chronic Sarcoidosis	110	93 (84.5)
Tuberculosis[2]	303	92 (30.4%)

[1] Numbers of subjects for each study population used for the epitope prediction analysis
[2] Patient population already described (46)

Table 2. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations

Organism[1]	Disease	Gold Database Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Löfgren	Chronic Sarc vs CTR	Löfgren vs CTR	Löfgren vs Chronic Sarc
Acinetobacter baumannii	Meningitis, Pneumonia, Septicemia	Gc00522	3976	3352	17871±3963	17460±3881	18156±4008	0.1868	0.3504	0.1964
Acinetobacter calcoaceticus	Nosocomial infection	Gc00201	3598	3325	21074±4645	20626±4573	21373±4622	0.2036	0.3657	0.2168
Actinobacillus pleuropneumoniae	Necrotizing pleuropneumonia	Gc00708	2242	2036	13015±2854	12798±2808	13221±2833	0.2565	0.3499	0.2346
Aeromonas hydrophila hydrophila	Gastroenteritis, Septicemia, Food poisoning	Gc00456	4744	4122	24703±5644	24118±5315	25373±5553	0.1843	0.2633	0.1312
Aeromonas salmonicida salmonicida	Furunculosis	Gc00536	4702	4086	23482±5284	22944±4991	24103±5246	0.1890	0.2653	0.1355
Anaplasma marginale	Anaplasmosis	Gc00239	1197	949	6919±1536	6723±1466	7152±1639	0.1350	0.2108	0.0867
Anaplasma phagocytophilum	Anaplasmosis	Gc00351	1471	1264	7080±1571	6916±1544	7282±1617	0.1842	0.2470	0.1301
Arcobacter butzleri	Bacteremia, Gastroenteritis	Gc00657	2341	2259	16597±3902	16502±4006	16797±3699	0.4185	0.3915	0.3591
Bacillus anthracis	Anthrax	Gc00189	5227	5309	31589±7108	31303±7157	31971±6753	0.3655	0.3868	0.3245
Bacillus cereus	Food poisoning	Gc00173	5224	5603	33351±7428	33086±7480	33715±7080	0.3804	0.3964	0.3410
Bacillus licheniformis	Food poisoning	Gc00213	4222	4152	23972±5311	23715±5265	24447±5125	0.3388	0.3165	0.2507
Bacteroides fragilis	Diarrhea, Abscesses	Gc00260	5205	4184	30522±6601	30276±6557	31142±6555	0.3746	0.3083	0.2628
Bacteroides thetaiotaomicron	Peritonitis	Gc00126	6260	4778	35884±7660	35563±7597	36516±7751	0.3674	0.3154	0.2658
Bacteroides vulgatus	Opportunistic peritoneal disease	Gc00584	5163	4065	28893±6119	28651±6052	29445±6250	0.3597	0.3302	0.2742
Bartonella henselae	Bacillary angiomatosis	Gc00192	1931	1488	9206±1997	9048±1958	9427±2013	0.2483	0.2778	0.1780
Bartonella quintana	Bacillary angiomatosis, Endocarditis, Trench fever	Gc00191	1581	1142	7710±1699	7544±1666	7896±1705	0.1998	0.2801	0.1564
Bdellovibrio bacteriovorus	Cell lysis	Gc00168	3782	3587	21121±4611	20600±4493	21603±4929	0.1654	0.2899	0.1478
Bordetella parapertussis	infection Respiratory	Gc00147	4773	4185	22589±5550	21833±5067	23661±5670	0.1177	0.1520	**0.0469**
Bordetella pertussis	Respiratory infection	Gc00146	4086	3436	18390±4527	17731±4121	19260±4681	0.1024	0.1534	**0.0432**
Borrelia afzelii	Lyme disease, Acrodermatitis chronica atrophicans	Gc00408	905	856	7545±1708	7616±1778	7590±1624	0.3624	0.4436	0.4718
Borrelia burgdorferi	Lyme disease	Gc00012	910	851	7635±1733	7699±1803	7690±1652	0.3767	0.4331	0.4896

(continued)

Table 2 *(continued)*. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations

Organism[1]	Disease	Gold Database Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Löfgren	Chronic Sarc vs CTR	Löfgren vs CTR	Löfgren vs Chronic Sarc
Borrelia garinii	Lyme disease	Gc00211	904	832	7395±1673	7446±1734	7447±1603	0.3963	0.4338	0.4996
Borrelia hermsii	Tick-borne relapsing fever	Gc00792	922	819	7605±1724	7668±1791	7695±1634	0.3786	0.3905	0.4707
Borrelia turicatae	Tick-borne relapsing fever	Gc00791	9173	818	7763±1784	7828±1859	7860±1684	0.3778	0.3851	0.4663
Brucella abortus	Brucellosis	Gc00263	3287	3085	15773±3805	15314±3590	16508±3809	0.1484	0.1516	0.0584
Brucella melitensis	Brucellosis	Gc00074	3294	3198	16505±3949	16052±3726	17265±3953	0.1605	0.1524	0.0623
Brucella suis	Brucellosis, Fever, Infectious abortions	Gc00699	3323	3241	16270±3886	15819±3661	17000±3886	0.1574	0.1585	0.0641
Burkholderia cenocepacia	Chronic infection, Necrotizing Pneumonia	Gc00744	7900	7008	34542±8237	33395±7622	36080±8521	0.1131	0.1604	0.0508
Burkholderia mallei	Glanders, Pneumonia	Gc00518	5900	5852	24519±5794	23776±5247	25689±6100	0.1316	0.1420	**0.0471**
Burkholderia multivorans	Cepacia syndrome	Gc00689	6820	6121	32560±7692	31521±7102	33960±8004	0.1202	0.1668	0.0557
Burkholderia pseudomallei	Pneumonia, Bacteremia, Melioidosis	Gc00520	7100	7183	32157±7571	31144±6874	33613±7981	0.1216	0.1537	**0.0495**
Burkholderia vietnamiensis	Pneumonia	Gc00531	7430	7617	33383±7916	32235±7273	34852±8335	0.1036	0.1624	**0.0486**
Burkholderia xenovorans (fungorum)	Opportunistic infection	Gc00365	9279	8702	46606±10914	45002±10138	48476±11425	0.1010	0.1815	0.0561
Campylobacter concisus	Gingivitis, Periodontitis, Periodontosis, Gastroenteritis	Gc00644	2052	1929	11849±2753	11758±2783	12204±2635	0.3881	0.2453	0.2174
Campylobacter curvus	Gastroenteritis, Periodontal infection	Gc00628	1971	1931	12086±2771	11987±2783	12425±2642	0.3799	0.2564	0.2219
Campylobacter fetus fetus	Bacteremia, Infertility, Septicemia, Meningitis	Gc00466	1773	1719	11959±2849	11858±2909	12283±2689	0.3809	0.2716	0.2374
Campylobacter jejuni jejuni	Diarrhea, Gastroenteritis, Food poisoning	Gc00491	1616	1653	11426±2562	11457±2621	11577±2383	0.4583	0.3761	0.4113
Campylobacter lari	Bacteremia, Diarrhea, Food poisoning, Gastroenteritis	Gc00851	1562	1599	10676±2320	10703±2362	10803±2214	0.4603	0.3854	0.4187
Candidatus	Pneumonia Protochlamydia amoebophila	Gc00185	2414	2031	13565±2812	13476±2807	13470±2836	0.3928	0.4286	0.4961
Chlamydia muridarum	Pharyngitis, Bronchitis, Pneumonia, Respiratory infection	Gc00028	1072	904	6170±1313	6049±1293	6241±1360	0.2159	0.3859	0.2407

(continued)

Table 2 *(continued)*. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations

Organism[1]	Disease	Gold Database Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Löfgren	Chronic Sarc vs CTR	Löfgren vs CTR	Löfgren vs Chronic Sarc
Chlamydia trachomatis	Bronchitis, Heart disease, Pharyngitis, Pneumonia, Respiratory infection, Trachoma, Venereal disease	Gc00017	1042	895	6007±1291	5883±1267	6084±1331	0.2044	0.3743	0.2248
Chlamydophila abortus	Pharyngitis, Bronchitis, Pneumonia, Respiratory infection	Gc00262	1144	932	6630±1421	6521±1416	6687±1430	0.2550	0.4157	0.2869
Chlamydophila pneumoniae	Pharyngitis, Bronchitis, Pneumonia, Heart disease, Respiratory infection	Gc00143	1225	1113	7387±1560	7268±1543	7438±1582	0.2572	0.4305	0.2995
Chromobacterium violaceum	Diarrhea, Septicemia	Gc00157	4751	4407	22494±5283	21847±4846	23367±5441	0.1431	0.1898	0.0725
Citrobacter koseri	Bacteremia, Brain abscesses, Meningoencephalitis, Neonatal meningitis	Gc00642	4720	5003	28306±6188	27612±5922	28928±6208	0.1664	0.2961	0.1460
Clostridium botulinum	Botulism	Gc00741	3992	3655	23971±5592	23730±5681	24467±5397	0.3563	0.3178	0.2639
Clostridium difficile	Diarrhea, Colitis, Peritonitis	Gc00420	4290	3742	26564±6412	26124±6506	27255±6245	0.2791	0.2826	0.1996
Clostridium perfringens	Dysenteria, Enterocolitis, Enterotoxemia, Food poisoning, Gas gangrene	Gc00406	3256	2876	18878±4506	18730±4584	19400±4242	0.3898	0.2672	0.2370
Clostridium tetani	Tetanus	Gc00120	2799	2373	18240±4199	18081±4260	18631±4105	0.3730	0.3096	0.2656
Corynebacterium diphtheriae gravis	Diphtheria, Respiratory infection	Gc00163	2488	2272	12322±3052	11767±2926	12830±3196	0.0584	0.1882	**0.0447**
Corynebacterium jeikeium	Endocarditis, Septicemia, Meningitis, Nosocomial infection	Gc00271	2462	2104	11183±2838	10672±2686	11739±3007	0.0595	0.1494	**0.0327**
Corynebacterium urealyticum	Cystitis, Pyelitis, Urinary tract infection	Gc00755	2369	2024	10616±2640	10185±2486	11118±2738	0.0790	0.1560	**0.0401**
Coxiella burnetii	Food poisoning	Gc00692	2016	1930	12824±2643	12776±2602	12901±2601	0.4375	0.4386	0.4086
Ehrlichia chaffeensis	Ehrlichiosis, Human monocytic ehrlichiosis	Gc00353	1176	1105	7907±1729	7804±1763	7993±1720	0.3058	0.3958	0.3023
Enterobacter sakazakii	Meningitis, Septicemia, Necrotizing enterocolitis	Gc00638	4368	4277	24007±5218	23480±4948	24594±5261	0.1912	0.2746	0.1438

Table 2 *(continued)*. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations

Organism[1]	Disease	Gold Database Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Löfgren	Chronic Sarc vs CTR	Löfgren vs CTR	Löfgren vs Chronic Sarc
Enterococcus faecalis	Bacteremia, Endocarditis, Urinary infection	Gc00831	2739	2701	18533±4090	18323±4088	18760±3981	0.3306	0.3833	0.3026
Escherichia coli	Diarrhea	Gc00716	4686	4126	29392±6442	28654±6199	30050±6555	0.1618	0.2931	0.1433
Francisella philomiragia	Bacteremia, Pneumonia, Septicemia	Gc00722	2045	1911	14775±3443	14622±3505	14931±3225	0.3523	0.4040	0.3326
Francisella tularensis	Tularemia	Gc00241	1892	1603	11937±2715	11810±2749	12039±2605	0.3443	0.4206	0.3424
Fusobacterium nucleatum nucleatum	Periodontal infection	Gc00085	2170	2067	14465±3377	14476±3456	14784±3182	0.4891	0.3064	0.3309
Granulibacter bethesdensis	Chronic granulomatous	Gc00423	2708	2437	13697±3305	13213±3077	14332±3446	0.1017	0.1540	**0.0457**
Haemophilus ducreyi	Genital ulcer, Chancroid	Gc00142	1698	1717	9470±2049	9291±2014	9569±2051	0.2260	0.3986	0.2543
Haemophilus influenzae	Otitis, Meningitis, Septicemia, Sinusitis, Bronchitis	Gc00001	1830	1657	10640±2330	10462±2291	10852±2324	0.2560	0.3138	0.2077
Haemophilus somnus	Arthritis, Myocarditis, Pneumonia	Gc00743	2263	1980	11524±2533	11307±2499	11710±2538	0.2313	0.3476	0.2201
Helicobacter pylori	Gastric inflammation, Ulcer	Gc00879	1673	1567	10030±2103	9998±2111	10098±2064	0.4475	0.4316	0.4091
Klebsiella pneumoniae	Bacteremia, Pneumonia, Urinary tract infection	Gc00841	5641	5425	29171±6561	28491±6262	29909±6474	0.1855	0.2740	0.1406
Legionella pneumophila	Legionellosis	Gc00226	3345	2878	21074±4478	20803±4455	21141±4445	0.3014	0.4681	0.3571
Leptospira borgpetersenii hardjobovis	Leptospirosis	Gc00433	3920	2945	19500±4144	19288±4116	19667±4101	0.3306	0.4150	0.3290
Listeria monocytogenes	Food poisoning, Listeriosis	Gc00186	2905	2821	17937±4140	17600±4133	18305±3994	0.2431	0.3173	0.2044
Mycobacterium abscessus	Broncho-pulmonary infection, Respiratory infection, Wound infection	Gc00729	5067	4920	24543±6186	23410±5792	25751±6546	0.0562	0.1502	**0.0309**
Mycobacterium avium	Respiratory infection, Tuberculosis type pulmonary infection	Gc00462	5475	5120	23353±6182	22194±5736	24817±6661	0.0519	0.1049	**0.0182**
Mycobacterium avium paratuberculosis	Johne's disease, Paratuberculosis, Enteritis	Gc00169	4829	4350	21675±5713	20621±5284	23029±6117	0.0545	0.1047	**0.0184**
Mycobacterium bovis	Tuberculosis	Gc00138	4345	3920	19755±5050	18796±4688	20854±5432	**0.0497**	0.1247	**0.0221**
Mycobacterium bovis BCG	Bovine tuberculosis	Gc00489	4374	3952	19934±5100	18969±4732	21050±5486	0.0503	0.1234	**0.0220**

(continued)

Table 2 *(continued)*. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations

Organism[1]	Disease	Gold Database Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Löfgren	Chronic Sarc vs CTR	Löfgren vs CTR	Löfgren vs Chronic Sarc
Mycobacterium leprae	Leprosy, Hanson's disease	Gc00045	3268	1605	9120±2335	8658±2202	9609±2524	**0.0434**	0.1341	**0.0235**
Mycobacterium marinum	Tuberculosis	Gc00784	6636	5423	28943±7417	27506±6949	30423±7938	**0.0466**	0.1452	**0.0266**
Mycobacterium smegmatis	Soft tissue lesions	Gc00461	6988	6716	32660±8582	31102±8048	34619±9073	0.0579	0.1131	**0.0217**
Mycobacterium tuberculosis	Tuberculosis	Gc00578	4424	3950	20040±5081	19073±4713	21124±5483	**0.0492**	0.1294	**0.0232**
Mycobacterium ulcerans	Buruli ulcer	Gc00469	5631	4160	20218±5218	19169±4838	21314±5711	**0.0405**	0.1334	**0.0217**
Mycoplasma arthritidis	Arthritis	Gc00825	820	631	4728±998	4682±1013	4715±1041	0.3462	0.4724	0.4377
Mycoplasma genitalium	Urogenital infection, Respiratory infection, Non-gonococcal urethritis	Gc00002	580	477	4034±836	4017±849	4015±860	0.4314	0.4542	0.4974
Mycoplasma penetrans	Urogenital infection, Respiratory infection	Gc00110	1358	1037	8755±1969	8662±1991	8690±1990	0.3445	0.4304	0.4736
Mycoplasma pneumoniae	Pneumonia, Tracheobronchitis, Respiratory infection	Gc00005	816	689	6227±1330	6293±1364	6172±1331	0.3367	0.4118	0.3333
Neisseria gonorrhoeae	Gonorrhea	Gc00258	2153	2002	10711±2295	10488±2158	10987±2423	0.2008	0.2617	0.1413
Neisseria meningitidis	Meningitis, Septicemia	Gc00026	2272	2063	10104±2187	9898±2081	10378±2282	0.2079	0.2529	0.1401
Neorickettsia sennetsu	Sennetsu fever	Gc00352	859	932	5766±1265	5712±1268	5895±1223	0.3561	0.2934	0.2419
Nocardia farcinica	Nocardiosis, Mastitis	Gc00224	6021	5683	26116±6921	24917±6314	27781±7320	0.0656	0.1010	**0.0189**
Parabacteroides distasonis	Opportunistic peritoneal disease	Gc00583	4811	3850	28852±6264	28651±6238	29468±6162	0.3916	0.2999	0.2639
Pasteurella multocida	Pasteurellosis, Cholera, Septicemia	Gc00048	2257	2015	13526±2959	13315±2905	13708±2895	0.2700	0.3716	0.2578
Porphyromonas gingivalis	Dental plaque, Periodontal infection	Gc00809	2354	2090	13339±2871	13275±2830	13739±2836	0.4251	0.2282	0.2155
Propionibacterium acnes	Acne	Gc00204	2560	2297	12996±3347	12326±3179	13641±3565	**0.0418**	0.1536	**0.0277**
Proteus mirabilis	Encephalitis, Pneumonia, Pyelonephritis, Septicemia, Surgical wound infection, Ulcer, Urinary tract infection, Urolithiasis	Gc00758	4063	3693	22802±5068	22364±4999	23152±4996	0.2293	0.3560	0.2243
Pseudomonas aeruginosa	Opportunistic infection, Nocosomial infection	Gc00432	6537	5892	32928±7633	32280±7041	34239±7495	0.2301	0.1797	0.0946

Table 2 *(continued)*. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations

Organism[1]	Disease	Gold Database Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Löfgren	Chronic Sarc vs CTR	Löfgren vs CTR	Löfgren vs Chronic Sarc
Pseudomonas mendocina	Endocarditis, Spondylodiscitis	Gc00544	5072	4594	25846±6022	25287±5593	26784±5865	0.2099	0.2026	0.1022
Rickettsia bellii	Epidemic typhus	Gc00661	1528	1476	9556±2129	9492±2186	9582±2043	0.3992	0.4735	0.4200
Rickettsia s canadensi	Epidemic typhus	Gc00660	1159	1093	7112±1594	7078±1638	7144±1513	0.4272	0.4576	0.4217
Rickettsia conorii	Rocky Mountain Spotted Fever	Gc00062	1268	1374	8045±1775	7988±1816	8088±1709	0.3922	0.4489	0.3944
Rickettsia prowazekii	Typhus, Rocky Mountain Spotted Fever	Gc00018	1111	835	7260±1621	7249±1677	7277±1536	0.4777	0.4784	0.4684
Rickettsia rickettsii	Rocky Mountain Spotted Fever	Gc00736	1257	1345	7436±1648	7374±1687	7479±1583	0.3747	0.4447	0.3810
Rickettsia typhi	Typhus	Gc00208	1111	838	7186±1602	7164±1653	7212±1528	0.4544	0.4645	0.4432
Salmonella enterica enterica	Food poisoning, Gastroenteritis, Salmonellosis	Gc00833	4685	4318	25874±5635	25253±5414	26418±5671	0.1723	0.2980	0.1502
Salmonella enterica sv Paratyphi A	Food poisoning, Gastroenteritis, Salmonellosis, Typhoid fever	Gc00853	4581	4078	24585±5427	23997±5208	25131±5404	0.1708	0.3036	0.1537
Salmonella enterica sv Paratyphi A	Food poisoning, Gastroenteritis, Salmonellosis, Typhoid fever	Gc00238	4585	4093	28451±6197	27834±5959	29034±6155	0.1746	0.2958	0.1501
Salmonella enterica sv Typhi	Typhoid fever, Salmonellosis, Food poisoning	Gc00066	4809	4395	25456±5596	24843±5356	26012±5603	0.1951	0.3081	0.1688
Serratia proteamaculans	Pneumonia	Gc00647	5448	4891	30454±6854	29710±6536	31154±6733	0.1740	0.2929	0.1462
Shigella boydii	Dysenteria, Food poisoning	Gc00779	4615	4246	22229±4780	21699±4539	22648±4871	0.1686	0.3205	0.1621
Shigella dysenteriae	Dysenteria, Food poisoning	Gc00324	4369	4274	21088±4422	20522±4135	21444±4595	0.1331	0.3344	0.1488
Shigella flexneri	Dysenteria, Gastroenteritis, Food poisoning	Gc00130	4599	4068	23543±5078	22989±4839	24010±5130	0.1729	0.3121	0.1592
Shigella sonnei	Dysenteria, Food poisoning	Gc00323	5039	4461	24441±5249	23862±4983	24912±5377	0.1703	0.3163	0.1606
Staphylococcus aureus	Mastitis, Nosocomial infection	Gc00327	2515	2665	18106±4153	17948±4241	18367±3890	0.3727	0.3684	0.3140
Staphylococcus haemolyticus	Opportunistic infection	Gc00274	2685	2676	17340±3945	17207±4023	17582±3713	0.3869	0.3711	0.3238
Streptococcus agalactiae	Meningitis	Gc00100	2211	2094	12858±2930	12638±2948	13092±2877	0.2601	0.3353	0.2285
Streptococcus gordonii	Dental plaque, Endocarditis, Periodontal infection	Gc00643	2196	2051	13014±2945	12895±2958	13286±2829	0.3647	0.3106	0.2602

(continued)

Table 2 *(continued)*. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations

Organism[1]	Disease	Gold Database Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Löfgren	Chronic Sarc vs CTR	Löfgren vs CTR	Löfgren vs Chronic Sarc
Streptococcus mutans	Dental caries	Gc00109	2030	1960	12205±2735	12118±2770	12415±2651	0.3939	0.3406	0.3015
Streptococcus pneumoniae	Meningitis, Pneumonia, Otitis media	Gc00437	2000	1914	12092±2702	11967±2722	12331±2619	0.3458	0.3182	0.2580
Streptococcus pyogenes	Bone infection, Endocarditis, Mastoiditis, Meningitis, Myositis, Necrotizing fasciitis, Otitis, Pharyngitis, Pneumonia, Sinusitis, Tonsillitis	Gc00285	1838	1865	10563±2399	10376±2396	10801±2403	0.2530	0.2985	0.1976
Streptococcus suis	Meningitis, Endocarditis, Septicemia, Arthritis	Gc00546	2096	2186	12022±2696	11867±2694	12293±2660	0.3108	0.2955	0.2225
Treponema pallidum pallidum	Syphilis	Gc00789	1139	1028	6849±1410	6676±1310	7002±1533	0.1435	0.2829	0.1264
Tropheryma whipplei	Whipple's disease	Gc00123	925	783	5486±1270	5331±1245	5654±1276	0.1467	0.2404	0.1079
Ureaplasma parvum	Respiratory infection, Urinary tract infection	Gc00742	751	609	5510±1196	5469±1221	5498±1212	0.3847	0.4797	0.4533
Vibrio cholerae	Cholera, Food poisoning	Gc00557	4148	3875	22495±4906	21872±4736	22891±5091	0.1368	0.3337	0.1552
Vibrio parahaemolyticus	Gastroenteritis	Gc00124	5165	4832	27525±6192	26701±6047	28124±6419	0.1263	0.3035	0.1330
Yersinia enterocolitica	Gastroenteritis, Food poisoning	Gc00481	4615	3979	25630±5744	24958±5547	26182±5722	0.1696	0.2958	0.1485
Yersinia pestis	Bubonic and Pneumonic plague	Gc00538	4517	3850	23982±5310	23393±5121	24516±5316	0.1566	0.3041	0.1466
Yersinia pseudotuberculosis	Gastroenteritis	Gc00776	4695	4150	25461±5702	24847±5530	26020±5664	0.1766	0.3005	0.1556

[1] Non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database v2.0 (GOLD database update October 30th 2008, http://www.genomesonline.org/) under analysis
[2] Genome size in Kilobases
[3] Number of Open Reading Frame in the genome
[4] Mean Number of potential epitopes of the genome under analysis, equivalent to the affinity of 1% of the best binder peptides, recognized by the different study populations
[5] p value for the comparison between groups indicated. Students' t-test. Statistically significant comparisons (p<0.05) are indicated in bold

Software for the identification and enumeration of epitopes in whole genomes

To enumerate the T-cell epitopes present in the protein data sets of each microbial genome examined, we used a software specifically designed for the identification and enumeration of peptide binding epitopes to HLA-DR molecules in large set of protein databases, as already described (46).

For the purpose of this study, the bacterial genomes were analyzed for enumerating the epitopes in silico recognized by the combination of the HLA-DR alleles carried by single subjects at an affinity level equivalent or higher than the 1% of the

best binder peptides for the allele in analysis (50). The threshold of affinity is a preselected numerical value used to differentiate between binders and non binders. Any peptide frame scoring higher than this value is predicted as binder or vice versa. As this threshold correlates with the peptide score (40, 50) and with HLA-ligand interaction, it is an indicator of the likelihood that a predicted peptide is capable of binding to a given HLA-molecule.

Statistical analysis

All the data are expressed as mean $\pm$ standard deviation of the mean (SD), and frequency expressed as percentage. Interquartile distribution of epitope number recognized was used to classify allele combination. Comparisons between means are made by Student's t test. Comparison between frequencies was made by Chi-square, with Yates and Bonferroni's correction, Chi-square for trend and Fisher's exact test when appropriate. GraphPad Prism version 4.0 was used for all statistical analyses and graphs.

Results

Differential capability of chronic sarcoidosis and Löfgren's syndrome affected subjects to recognize genomes of human pathogen

The capability of the HLA-DR expressed allelic variants of the study subjects to bind bacterial peptide epitopes was determined as the number of epitopes bound *in silico* at the top 1% affinity. At an affinity level equivalent to that of the top 1% of the binder peptides for the HLA-DR molecules of the study subjects, Löfgren's syndrome affected subjects recognized a higher number of epitopes (mean epitope number for all the genomes: 18,496±9,114) than chronic sarcoidosis affected subjects (17,515±8,448; p<0.00001) and controls (17,954±8,742; p<0.00001). In addition, the comparison between chronic sarcoidosis and healthy controls showed that the capability of affected subjects to recognize bacterial pathogen genomes was significantly reduced compared to controls (p<0.00001).

Interestingly though, the capability of the HLA-DR allelic variants expressed by Löfgren's syndrome affected subjects to recognize a higher

number of epitopes than chronic sarcoidosis affected subjects was not generalized to all the genomes examined, but it was more prominent for certain bacterial pathogens. In particular, statistically significant differences in the mean number of epitopes recognized by Löfgren's syndrome affected subjects with respect to chronic sarcoidosis affected subjects was observed for the genomes of bacteria belonging to the genders Bordetella, Burkholderia, Granulibacter, Mycobacterium, Nocardia and Propionibactierium (Tab. 2 and Fig. 1). The largest difference was observed for *Mycobacterium avium*, where there was 12% difference between the average HLA-DR immunomes, i.e. the number of putative epitopes deduced from the whole bacterial genome that were capable of being bound by the HLA-DR expressed allele of each individual subject, of the two disease groups (Fig. 1).

Contribution of the different HLA-DRB1 alleles combination in the hyper-recognition of genomes by patients with Löfgren's syndrome

In order to analyze the nature of the exaggerated epitope recognition of bacterial genomes seen in the subjects with Löfgren's syndrome compared to those with chronic sarcoidosis, the data obtained in
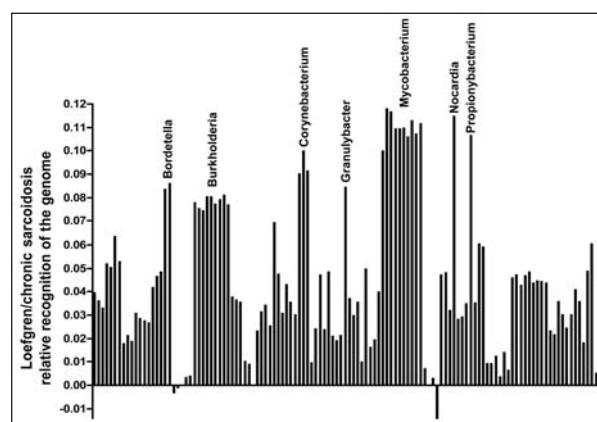

Fig. 1. Relative recognition of all the Löfgren's syndrome affected subjects with respect to chronic sarcoidosis patients of the HLA-DR immunomes deduced from the genomes of pathogenic bacteria. The ratio of the total number of the recognized epitopes (on the ordinate) in the two study groups is reported for 124 non redundant genomes of bacterial pathogens analysed (shown in alphabetic order on the abscissa). The bacterial groups showing statistically significant differences in HLA-DR immunome recognition are indicated on the graph

the study population for the HLA-DR immunome of *M. avium* i.e., the best binder of all tested genomes, were used.

The distribution of HLA-DR allelic variants in the study populations is shown in table 3. As previously reported (37), alleles HLA-DRB*0301 and *0407 were significantly more frequent in the Löfgren's syndrome affected while the allele *1501 was significantly over-represented in the chronic sarcoidosis affected group (Tab. 3).

Table 4 shows the distribution of *M. avium* genome recognition levels in the study groups, according to the percentile distribution from bottom to top recognition levels. Considering the affinity level equivalent to 1% of the higher affinity peptides, the 1,225 different combinations generated by the 49 HLA-DRB1 alleles for which the quantitative implemented peptide binding motifs is available, can recognize 6,149 to 43,781 epitopes in the genome of *M. avium*. Subjects with Löfgren's syndrome showed significantly higher levels of *M. avium* epitope recognition than subjects with chronic sarcoidosis as 45% of the Löfgren's affected were in the top 76-100° percentile compared to 17% of the chronic sarcoidosis affected (p=0.0037) and 26% of the controls (p=0.0386). Conversely, subjects with chronic sar-

Table 3. HLA-DRB1 allelic frequency analysis of the study population

HLA-DRB1 Allele	Controls		Chronic Sarcoidosis		Löfgren		CTR vs Chr Sarc χ-square p	CTR vs Loef χ-square p	Chr Sarc vs Loef χ-square p
	N	Freq	N	Freq	N	Freq			
0101	57	0.064	12	0.055	6	0.077	0.7250	0.8312	0.6627
0102	4	0.004	0	0.000	0	0.000	0.7152	0.7413	NA
0103	3	0.003	0	0.000	0	0.000	0.8932	0.5811	NA
0301	103	0.115	21	0.095	16	0.205	0.4746	0.0321*	0.0201*
0302	2	0.002	0	0.000	0	0.000	0.8519	0.3764	NA
0401	57	0.064	13	0.059	3	0.038	0.9199	0.5189	0.6876
0402	6	0.007	1	0.005	0	0.000	0.9108	0.9777	0.5872
0403	10	0.011	0	0.000	1	0.013	0.2393	0.6693	0.5872
0404	41	0.046	4	0.018	2	0.026	0.0936	0.5852	0.9473
0405	9	0.010	1	0.005	0	0.000	0.7048	0.7841	0.5872
0406	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
0407	9	0.010	1	0.005	4	0.051	0.7048	0.0116*	0.0246*
0408	8	0.009	0	0.000	0	0.000	0.3358	0.8528	NA
0701	74	0.083	15	0.068	6	0.077	0.5644	0.9725	0.9986
0801	18	0.020	8	0.036	1	0.013	0.2384	0.9832	0.5100
0803	6	0.007	1	0.005	0	0.000	0.9108	0.9777	0.5872
0804	2	0.002	2	0.009	0	0.000	0.3716	0.3764	0.9698
0812	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
0901	10	0.011	4	0.018	1	0.013	0.6194	0.6693	0.8444
1001	6	0.007	3	0.014	1	0.013	0.5435	0.9313	0.6039
1101	78	0.087	29	0.132	4	0.051	0.0598	0.3769	0.0823
1102	5	0.006	0	0.000	1	0.013	0.5831	0.9777	0.5872
1103	7	0.008	2	0.009	0	0.000	0.8156	0.9313	0.9698
1104	51	0.057	5	0.023	2	0.026	0.0555	0.3620	0.7725
1201	10	0.011	2	0.009	0	0.000	0.9244	0.7234	0.9698
1202	4	0.004	1	0.005	0	0.000	0.5831	0.7413	0.5872
1301	94	0.105	16	0.073	9	0.115	0.1876	0.9283	0.3524
1302	52	0.058	18	0.082	3	0.038	0.2543	0.6406	0.3039
1303	13	0.015	0	0.000	1	0.013	0.1474	0.7091	0.5872
1305	1	0.001	1	0.005	0	0.000	0.8519	0.1221	0.5872
1326	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
1401	33	0.037	13	0.059	3	0.038	0.1964	0.8079	0.6876
1404	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
1407	2	0.002	0	0.000	0	0.000	0.8519	0.3764	NA
1501	86	0.096	44	0.200	13	0.167	0.00002^	0.0754	0.6344
1502	9	0.010	0	0.000	0	0.000	0.2829	0.7841	NA
1601	19	0.021	2	0.009	1	0.013	0.3620	0.9305	0.7065
1602	1	0.001	1	0.005	0	0.000	0.8519	0.1221	0.5872

*Bonferroni corrected p>0.05
^Bonferroni corrected p=0.0013

Table 4. Recognition of *M. avium* genome by the HLA-DR allele combinations expressed by the study population subjects

Percentile group of the distribution of number of epitopes recognised in M. avium genome at affinity equivalent of 1% of best binder peptides (all combinations)	Study population			Chi-square p-value		
	Controls N (%)	Chronic Sarcoidosis N (%)	Löfgren N (%)	Controls vs Chronic Sarcoidosis	Controls vs Löfgren	Löfgren vs Chronic Sarcoidosis
A. 1-25° percentile (6149 - 17616 epitopes)	54 (15%)	14 (15%)	4 (13%)	0.9504	0.9039	1.0000
B. 26-50° percentile (17617 - 21464 epitopes)	83 (24%)	36 (39%)	5 (16%)	**0.0059**	0.4557	**0.0363**
C. 51-75° percentile (21465 - 26437 epitopes)	121 (35%)	27 (29%)	8 (26%)	**0.0363**	0.4232	0.9083
D. 76-100° percentile (26438 - 43781 epitopes)	91 (26%)	16 (17%)	14 (45%)	0.7328	**0.0386**	**0.0037**

coidosis showed a significantly higher level of *M. avium* epitope recognition in the lower 26°-50° percentiles distribution with 39% compared to the 16% of the Löfgren's affected (p=0.0363) and the 24% of the controls (p=0.0059; table 4).

Furthermore, in order to determine whether the exaggerated recognition of microbial genomes such as those of mycobacteria, shown by Löfgren's syndrome affected subjects, was associated with the expression of specific HLA-DR allelic variants rather than with the expression of combinations of HLA-DR alleles endowed with higher binding capabilities, we analyzed the distribution of HLA-DR alleles in the *M. avium* recognition percentile subgroups.

This analysis showed that there were no biases in the expression of HLA-DR variants in the Löfgren's syndrome affected population, neither in the 26-50° nor in the 76-100° percentile groups (Tab. 5). Only the association of the HLA-DRB1*1501 allele with chronic sarcoidosis remained statistically significant both in the higher and the lower affinity percentile groups (Tab. 5).

Assessment of HLA-DR M. avium and M. tuberculosis epitope affinity in sarcoidosis and tuberculosis

Altogether the data suggests that Löfgren's syndrome immune responsiveness might be characterized by the expression of HLA-DR genomes endowed with an exaggerated ability to recognize bacterial genomes while, in contrast, chronic sarcoidosis appears to be characterized by the expression of HLA-DR alleles with significantly lower ability to bind and present bacterial antigens. As already reported in tuberculosis affected patients, a biased expression of the same HLA-DR alleles expressed in sarcoidosis could be observed (46).

Strikingly, the comparison of epitope binding affinity of the HLA-DR allelic combinations expressed by sarcoidosis affected subjects with those of tuberculosis patients showed a statistically significant lower binding affinity, both toward the *M. avium* and the *M. tuberculosis* immunomes than that of sarcoidosis affected subjects. In addition, when the patient groups were ranked by their binding affinities, the Löfgren's syndrome affected ranked the highest, with the tuberculosis affected the lowest. Chronic sarcoidosis, although ranking lower than controls ranked higher than tuberculosis (Tab. 6), consistently with the concept that in tuberculosis the susceptibility to infection and active disease is likely associated with the expression of a defective epitope recognition repertoire (Fig. 2).

Discussion

We show here that patients affected by the Löfgren's syndrome express HLA-DR allelic combinations that recognize a significantly higher number of bacterial antigen epitopes compared to healthy individuals, as they do in comparison to chronic sarcoidosis.

With regard to sarcoidosis, the data are consistent with the observation of Grosser et al. (51) who recently reported the association of chronic sarcoidosis, the presence of mycobacterial DNA in affected tissue and the expression of DRB1*15, whereas acute sarcoidosis was associated with the absence of mycobacterial DNA and with the expression of HLA-DRB1*03. These data add further support to the concept that the manifestations and outcomes of sarcoidosis are driven by a complex interaction between antigen presenting molecules, antigens(s) and T-cell receptors leading to antigen clearance or persistence (52). In this hypothesis, HLA Class II mol-

Table 5. Analysis of the HLA-DRB1 alle frequency in the study population subgroups ranked by the percentile recognition of M. avium immunome

HLA-DRB1 Allele	Löfgren		Chronic Sarcoidosis		Controls		CTR vs Chr Sarc χ-square p	CTR vs Loef χ-square p	Chr Sarc vs Loef χ-square p
	N	Freq	N	Freq	N	Freq			
A. 76°-100° percentile of M. avium immunome									
0101	4	0,142857	2	0,0625	12	0,065934	0,752637	0,29569	0,545972
0102	0	0	0	0	3	0,016484	0,933209	0,86417	NA
0301	11	0,392857	12	0,375	71	0,39011	0,972137	0,856908	0,901165
0401	1	0,035714	1	0,03125	5	0,027473	0,644641	0,714702	0,532173
0404	2	0,071429	1	0,03125	6	0,032967	0,62522	0,645852	0,905486
0405	0	0	0	0	1	0,005495	0,324589	0,279601	NA
0408	0	0	0	0	2	0,010989	0,688938	0,625769	NA
0701	1	0,035714	3	0,09375	14	0,076923	0,976218	0,693497	0,703662
0801	1	0,035714	0	0	4	0,021978	0,889536	0,824374	0,94628
1101	2	0,071429	2	0,0625	11	0,06044	0,721656	0,844175	0,703662
1102	1	0,035714	0	0	0	0	NA	0,279601	0,94628
1104	1	0,035714	0	0	10	0,054945	0,366004	0,975771	0,94628
1301	3	0,107143	0	0	17	0,093407	0,147756	0,908241	0,191528
1305	0	0	0	0	1	0,005495	0,324589	0,279601	NA
1401	1	0,035714	4	0,125	12	0,065934	0,419583	0,844175	0,435253
1501	0	0	7	0,21875	13	0,071429	0,020827	0,298833	0,025735
B. 26°-50° percentile of M. avium immunome									
0401	1	0,1	6	0,083333	31	0,186747	0,067584	0,788224	0,669285
0402	0	0	1	0,013889	2	0,012048	0,60619	0,235259	0,245069
0404	0	0	1	0,013889	11	0,066265	0,169496	0,866468	0,245069
0405	0	0	1	0,013889	3	0,018072	0,750289	0,407113	0,245069
0408	0	0	0	0	3	0,018072	0,60619	0,407113	NA
0701	0	0	3	0,041667	18	0,108434	0,155789	0,574302	0,809451
0801	0	0	6	0,083333	6	0,036145	0,227863	0,775284	0,763971
0804	0	0	0	0	1	0,006024	0,666602	0,054871	NA
1101	2	0,2	13	0,180556	22	0,13253	0,446225	0,897053	0,773789
1102	0	0	0	0	3	0,018072	0,60619	0,407113	NA!
1104	0	0	1	0,013889	3	0,018072	0,750289	0,407113	0,245069
1301	3	0,3	11	0,152778	18	0,108434	0,456291	0,189295	0,477119
1302	0	0	6	0,083333	15	0,090361	0,941676	0,681216	0,763971
1501	4	0,4	23	0,319444	28	0,168675	0,01502	0,155656	0,881651
1502	0	0	0	0	2	0,012048	0,871005	0,235259	NA

Table 6. Comparison of the recognition of the HLA-DR immunomes of *M. avium* and *M. tuberculosis* in the study populations in comparison to tuberculosis patients.

Genome	Mean Number of potential epitopes (1% affinity)[1]				Student's t-test (p-value)[2]		
	Controls	Chronic Sarcoidosis	Löfgren	Tuberculosis[3]	TB vs Control	TB vs Löfgren	TB vs Chronic Sarcoidosis
M. avium	23353±6182	22194±5736	24817±6661	20709±6894	**0.0002**	**>0.00001**	**0.003**
M. tuberculosis	20040±5081	19073±4713	21124±5483	17925±5732	**0.0008**	**>0.00001**	**0.01**

[1] Mean Number of potential epitopes in the different study population of the *M. avium* and *M. tuberculosis*, equivalent or higher than the affinity of 1% of the best binder peptides for the subject's HLA-DR molecules
[2] p value for the comparison between groups indicated. Students' t-test. Statistically significant comparisons (p<0.05) are indicated in bold
[3] Matched controls of the tuberculosis patient populations did not differ from the control population of this study in terms of the number of potential epitopes recognized (*M. avium*: 23353±6182 vs 22591±7612; *M. tuberculosis*: 20040±5081 vs 19472±4972; p>0.05 all comparisons)

ecules with the higher affinity for mycobacterial antigenic epitopes might present to T cells wider antigen repertoires with more elevated efficiency. This would initiate a hypersensitivity response lead-
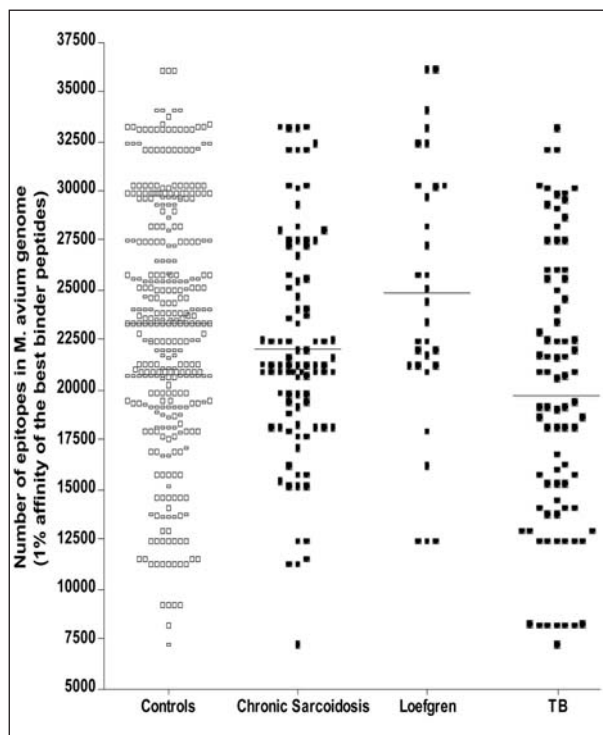
Fig. 2. Comparison of the recognition of the HLA-DR immunomes of *M. avium* in the study populations. Shown the numbers of *M. avium* epitopes recognized by each individual at an affinity equal or above that of the best 1% on binding peptides. Individual numbers of M. avium epitopes is shown on the ordinate, study populations are shown on the abscissa

ing to acute onset of disease, antigen clearance and spontaneous disease remission. To the contrary, HLA Class II molecules, such as DRB1*1501, due to their lower affinity for bacterial epitopes, would express a less efficient antigen presentation process, thereby dictating a sluggish immune response resulting in insidious clinical presentation, antigen persistence, continuing granuloma formation: in summary, to chronic sarcoidosis.

Following the same line of interpretation also for the *in silico* data presented here on tuberculosis patients, our observation would suggests that patients bearing even less recognizing HLA-DR allele combinations could be unable to mount an effective immune response, would fail to achieve bacterial containment and would, consequentially, develop active infectious disease.

Interestingly, it has been proposed that in tuberculosis a spectrum of immune reactions may be defined, where poor cellular mediated immunity, associated with exuberant antibody production, represents one end of the spectrum while good cellular mediated immunity, with little or no antibody formation, would be at the other end (53). Similarly, in leprosy the immune spectrum is represented by non-reactive disseminated lepromatous leprosy, associated with HLA-DR 15 alleles (54, 55), at one end, while tuberculoid leprosy − the disease presentation characterized by stronger T-cell mediated immunity and lesser invasiveness − is associated with HLA-DR3 alleles (56) and represents the opposite end. In this context, it is enticing to hypothesize that sarcoidosis may be part of a spectrum of granulomatous responses to one or more closely related bacteria where the Löfgren's syndrome represents hypersensitivity, similarly to the hot tub lung (15) and the metalworking fluid associated pneumonitis (16), while pulmonary infection, by atypical mycobacteria infection, *M. tuberculosis* or other bacteria might represent the other end of the spectrum.

In conclusion, even though these *in silico* immunogenetic data are representative of only about 80% of the study population evaluated and they still need to be confirmed in other sarcoid populations with different ethnical backgrounds and to be validated by *in vitro* and *ex vivo* confirmatory analysis, our findings support a pathogenetic role of specific groups of bacteria in the hypersensitivity granulomatous reaction(s) of sarcoidosis.

References

1. Crystal RG, Roberts WC, Hunninghake GW, Gadek JE, Fulmer JD, Line BR. Pulmonary sarcoidosis: a disease characterized and perpetuated by activated lung T-lymphocytes. Ann Intern Med 1981; 94: 73-94.
2. Venet A, Hance AJ, Saltini C, Robinson BWS, Crystal RJ. Enhanced Alveolar Macrophage-mediated Antigen-induced T Lymphocyte Proliferation in Sarcoidosis. J Clin Invest 1985; 75: 293-301.
3. Spurzem JR, Saltini C, Kirby M, Konishi K, Crystal RG. Expression of HLA class II genes in alveolar macrophages of patients with sarcoidosis. Am Rev Respir Dis 1989; 140: 89-94.
4. Pinkston P, Bitterman PB, Crystal RG. Spontaneous release of interleukin-2 by lung T-lymphocytes in active pulmonary sarcoidosis. N Engl J Med 1983; 308: 793-800.
5. Saltini C, Spurzem JR, Lee JJ, Pinkston P, Crystal RG. Spontaneous Release of Interleukin 2 by Lung T Lymphocytes in Active Pulmonary Sarcoidosis Is Primarily from the Leu3+DR+ T Cell Subset. J Clin Invest 1986; 77: 1962-70.
6. du Bois RM, Kirby M, Balbi B, Saltini C, Crystal RG: T-lymphocytes that accumulate in the lung in sarcoidosis have evidence of recent stimulation of the T-cell antigen receptor. Am Rev Respir Dis 1992; 154: 1205-11.
7. Saltini C, Kirby M, Trapnell BC, Tamura N, Crystal RG. Biased accumulation of T lymphocytes with "memory"-type CD45 leukocyte

common antigen gene expression on the epithelial surface of the human lung. J Exp Med 1990; 171: 1123-40.

8. Silver RF, Crystal RG, Moller DR. Limited heterogeneity of biased T-cell receptor Vβ gene usage in lung but not blood T cells in active pulmonary sarcoidosis. Immunology 1996; 88: 516-23.

9. Moller DR. T-cell receptor genes in sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis 1998; 15: 158-64.

10. Thomeer M, Demedts M, Wuyts W. Epidemiology of sarcoidosis. In: Dent M, Costabel U, eds. Sarcoidosis Eur Respir Mon 2005; 32: 13-22.

11. Saltini C, Winestock K, Kirby M, Pinkston P, Crystal RG. Maintenance of alveolitis in patients with chronic beryllium disease by beryllium-specific helper T-cells. N Engl J Med.1989; 320: 1103-9.

12. Redline S, Barna B, Tomashefski JF, Abraham JL. Granulomatous disease associated with pulmonary deposition of titanium. Br J Indust Med 1986; 43: 652-6.

13. Werfel U, Schneider J, Rödelsperger K, et al. Sarcoid granulomatosis after zirconium exposure with multiple organ involvement. Eur Respir J 1998; 12: 750.

14. De Vuyst P, Dumortier P, Schandené L, Estenne M, Verhest A, Yernault JC. Sarcoidlike lung granulomatosis induced by aluminum dusts. Am Rev Respir Dis 1987; 135: 493-7.

15. Hanak V, Kalra S, Timothy R, et al. Hot tub lung: Presenting features and clinical course of 21 patients. Respiratory Medicine 2006; 100: 610-5.

16. Hodgson MJ, Bracker A, Yang C, et al. Hypersensitivity Pneumonitis In a Metal-Working Environment. American J Industrial Medicine 2001; 39: 616-28.

17. Sechi LA, Scanu AM, Molicotti P, et al. Detection and Isolation of Mycobacterium avium subspecies paratuberculosis from intestinal mucosal biopsies of patients with and without Crohn's disease in Sardinia. Am J Gastroenterol 2005; 100: 1529-36.

18. Brisson-Noël A, Gicquel B, Lecossier D, Lévy-Frébault V, Nassif X, Hance AJ. Rapid diagnosis of tuberculosis by amplification of mycobacterial DNA in clinical samples. Lancet 1989; 2: 1069-71.

19. Bocart D, Lecossier D, De Lassence A, Valeyre D, Battesti JP, Hance AJ. A search for mycobacterial DNA in granulomatous tissues from patients with sarcoidosis using the polymerase chain reaction. Am Rev Respir Dis 1992; 145: 1142-8.

20. Saboor SA, Johnson NM, McFadden J. Detection of mycobacterial DNA in sarcoidosis and tuberculosis with polymerase chain reaction. Lancet 1992; 339: 1012-5.

21. Eishi Y, Suga M, Ishige I, et al. Quantitative analysis of mycobacterial and propionibacterial DNA in lymph nodes of Japanese and European patients with sarcoidosis. J Clin Microbiol 2002; 40: 198-204.

22. Ishige I, Eishi Y, Takemura T, et al. Propionibacterium acnes is the most common bacterium commensal in peripheral lung tissue and mediastinal lymph nodes from subjects without sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis 2005; 22: 33-42.

23. Ichikawa H, Kataoka M, Hiramatsu J, et al. Quantitative analysis of propionibacterial DNA in bronchoalveolar lavage cells from patients with sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis 2008; 25: 15-20.

24. Fidler HM. Mycobacteria and sarcoidosis: recent advances. Sarcoidosis 1994; 11: 66-8.

25. Mangiapan G, Hance AJ. Mycobacteria and sarcoidosis: an overview and summary of recent molecular biological data. Sarcoidosis 1995; 12: 20-37.

26. Gupta D, Agarwal R, Aggarwal AN, Jindal SK. Molecular evidence for the role of mycobacteria in sarcoidosis: a meta-analysis. Eur Respir J 2007; 30: 508-16.

27. Song Z, Marzilli L, Greenlee BM, et al. Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. J Exp Med 2005, 201: 755-67.

28. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M,

Dubaniewicz A, Singh M, Mysliwski A. Mycobacterial heat shock protein-induced blood T lymphocytes subsets and cytokine pattern: comparison of sarcoidosis with tuberculosis and healthy controls. Respirology 2007; 12: 346-54.

29. Hajizadeh R, Sato H, Carlisle J, et al. Mycobacterium tuberculosis Antigen 85A induces Th-1 immune responses in systemic sarcoidosis. J Clin Immunol 2007; 27: 445-54.

30. Drake WP. When a commensal becomes a pathogen. Sarcoidosis Vasc Diffuse Lung Dis 2008; 25: 10-11.

31. Moller DR. State of the Art. Potential Etiologic Agents in Sarcoidosis. The Proceedings of the American Thoracic Society 2007; 4: 465-8.

32. Tercelj M, Salobir B, Rylander R. Microbial antigen treatment in sarcoidosis a new paradigm? Med Hypotheses 2008; 70: 831-4.

33. Davenport MP, Quinn CL, Valsasnini P, Sinigaglia F, Hill AV, Bell JI. Analysis of peptide-binding motifs for two disease associated HLA-DR13 alleles using an M13 phage display library. Immunology 1996; 88: 482-6.

34. Mallios RR. Predicting class II MHC/peptide multi-level binding with an iterative stepwise discriminant analysis meta-algorithm. Bioinformatics 2001; 17: 942-8.

35. Voorter CE, Amicosante M, Berretta F, Groeneveld L, Drent M, van den Berg-Loonen EM. HLA class II amino acid epitopes as susceptibility markers of sarcoidosis. Tissue Antigens 2007; 70: 18-27.

36. Rossman MD, Thompson B, Frederick M, et al: HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. Am J Hum Genet 2003; 73: 720-35.

37. Berlin M, Fogdell-Hahn A, Olerup O, Eklund A, Grunewald J. HLA-DR predicts the prognosis in scandinavian patients with pulmonary sarcoidosis. Am J Respir Crit Care Med 1997; 156: 1601-5.

38. Grunewald J, Eklund A. Sex-specific manifestations of Lofgren's syndrome. Am J Respir Crit Care Med 2007; 175: 40-4.

39. Germain RN. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. Cell 1994; 76: 287-99.

40. Sturniolo T, Bono E, Ding J, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 1999; 17: 555-61.

41. Scherer A, Frater J, Oxenius A, et al. Quantifiable cytotoxic T lymphocyte responses and HLA-related risk of progression to AIDS. Proc Natl Acad Sci USA 2004; 101: 12266-70.

42. Stunz L, Karr RW, Anderson R. HLA-DRB1 and -DRB4 genes are differentially regulated at the transcriptional level. J Immunol 1989; 143: 3081-6.

43. Berdoz J, Gorski J,Termijtelen A, et al. Constitutive and induced expression of the individual HLA-DR h and a chain loci in different cell types. J Immunol 1987; 139: 1336-41.

44. Cotner T, Charbonneau H, Mellins E, Pious D. mRNA abundance, rather than differences in subunit assembly, determine expression of HLA-DRh1 and DRh3 molecules. J Biol Chem 1996; 264: 11107-11.

45. De Groot AS, Bosma A, Chinai N, et al. From genome to vaccine: in silico predictions, ex vivo verification. Vaccine 2001; 19: 4385-95.

46. Contini S, Pallante M, Vejbaesya S, et al. A model of phenotypic susceptibility to Tuberculosis: deficient in silico selection of Mycobacterium tuberculosis epitopes by HLA alleles. Sarcoidosis Vasc Diffuse Lung Dis 2008; 25: 21-8.

47. Marsh SGE. Nomenclature for factors of the HLA system, update January 1998. Tissue Antigens 1998; 51: 582-3.

48. Marsh SG. WHO Nomenclature Committee for Factors of the HLA System. Nomenclature for factors of the HLA system, update September 2006. Tissue Antigens 2006; 68: 540-2.

49. Kubo K, Yamazaki Y, Hanaoka M, et al. Analysis of HLA antigens in Mycobacterium avium-intracellulare pulmonary infection. Am J Respir Crit Care Med 2000; 161: 1368-71.

50. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. Bioinformatics 2001; 17: 1236-7.

51. Grosser M, Luther T, Fuessel M, Bickhardt J, Magdolen V, Baretton G. Clinical course of sarcoidosis in dependence on HLA-DRB1 allele frequencies, inflammatory markers, and the presence of M. tuberculosis DNA fragments. Sarcoidosis Vasc Diffuse Lung Dis 2005, 22: 66-74.

52. Moller DR, Chen ES. Genetic basis of remitting sarcoidosis. Triumph of the trimolecular complex? Am J Respir Cell Mol Biol 2002; 27: 391-5.

53. Lenzini L, Rottoli P, Rottoli L. The spectrum of human tuberculosis. Clin exp Immunol 1977; 27: 230-7.

54. Soebono H, Giphart MJ, Schreuder GM, Klatser PR, de Vries RR. Associations between HLA-DRB1 alleles and leprosy in an Indonesian population. Int J Lepr Other Mycobact Dis 1997; 65: 190-6.

55. Rani R, Fernandez-Vina MA, Zaheer SA, Beena KR, Stastny P. Study of HLA class II alleles by PCR oligotyping in leprosy patients from north India. Tissue Antigens 1993; 42: 133-7.

56. De Vries RPP. An immunogenetic view of delayed type hypersensitivity. Tubercle 1991; 72: 161-7.